Volume 6, Number 4 December 2010

IJCB INTERNATIONAL JOURNAL OF CENTRAL BANKING Volume 6, Number 4 December 2010

IJCB
INTERNATIONAL JOURNAL OF CENTRAL BANKING

Special Issue: The Theory and Practice of Macroprudential Regulation

IJCB International Journal of Central Banking

Volume 6, Number 4 December 2010

Special Issue:
The Theory and Practice of Macroprudential Regulation

The contents of this journal, together with additional materials provided by article authors, are available without charge at ***www.ijcb.org****.*

The *International Journal of Central Banking* is published quarterly (ISSN: 1815-4654). Online access to the publication is available free of charge at **www.ijcb.org**. Individual print subscriptions are available at an annual rate of $100 (USD).

Print subscription orders may be placed online at www.ijcb.org, by phone (+49 69 1344 7623), via fax (+49 69 1344 8553), or by e-mail (editor@ijcb.org).

Renewals, claims, address changes, and requests for permission to reprint material from this journal should be addressed to:

International Journal of Central Banking
DG Research
European Central Bank
Postfach 16 03 19
D-60066 Frankfurt
Germany

Phone: +49 69 1344 7623
Fax: +49 69 1344 6575
E-mail: editor@ijcb.org

ISSN: 1815-4654

The Theory and Practice of Macroprudential Regulation

Introduction to a Special Issue of the International Journal of Central Banking

Douglas Gale, Rafael Repullo, Til Schuermann, and Frank Smets

One of the main lessons from the financial crisis over the past three years has been the need for the establishment of a macroprudential supervisor that oversees the health and stability of the *overall* financial system. The principles, tools, and transmission channels of such a new macroprudential policy framework are, however, not yet fully understood. The second financial stability conference of the *International Journal of Central Banking* (IJCB), hosted by Banco de España in Madrid on June 17–18, 2010, dealt with the topic of "The Theory and Practice of Macroprudential Regulation." This issue of the IJCB contains selected papers and commentaries presented at that conference.

The first two papers deal with empirical work on the importance of bank capital and financial conditions of banks for their lending and the monetary transmission mechanism more generally. Jose Berrospide and Rochelle Edge (Federal Reserve Board) use a number of different methods for gauging the size of the effect of bank capital on the extension of bank credit. They find modest estimated effects and apply these estimates to investigate the impact of TARP capital injections. José-Luis Peydró (ECB) discusses the paper and points to some of the difficult identification problems associated with distinguishing between credit demand and supply effects. In the second paper, Ramona Jimborean and Jean-Stéphane Mésonnier (Banque de France) use a novel approach to show that common factors extracted from individual banks' liquidity and leverage ratios do predict macroeconomic developments in France. They also find that these bank factors are, however, largely irrelevant for the transmission of monetary policy once the development of credit aggregates is taken into account. In his discussion, Don Morgan (Federal Reserve Bank of New York) suggests that the framework may also be used to look at the macroeconomic effects of changes in bank capital and liquidity. In his commentary, Mark Gertler (New York University) provides a framework for thinking about the link between banking crises and real activity, and presents some suggestive evidence of the importance of bank credit risk in the current recession.

The third paper, by Francisco Covas (Federal Reserve Board) and Shigeru Fujita (Federal Reserve Bank of Philadelphia), shows that capital requirements significantly contribute to magnifying output fluctuations, using a general equilibrium model where the financing of capital goods production is subject to an agency problem. In his discussion, Javier Suarez (CEMFI) puts the framework used in the context of the wider literature and argues that the assessment of the likely effects of a significant rise in capital requirements may differ substantially in alternative models that pay explicit consideration to the frictions that affect the dynamics of bank capital accumulation. In his commentary, Douglas Gale (New York University) reviews the impact of capital requirements on risk taking and takes on the classical risk-shifting argument that underlies the claim that capital reduces risk. Gale points out that this partial equilibrium argument ignores the factors that determine the supply and cost of capital. He shows that, in a model in which managers have target rates of return which force them to "reach for yield," the conventional effects of greater capital on risk taking are turned upside down.

The fourth paper deals with empirical approaches to determine which financial institutions are systemically important. Chen Zhou (De Nederlandsche Bank) considers three measures of the systemic importance of a financial institution within an interconnected financial system and argues that size is not necessarily a good proxy of systemic importance. In his discussion, Stefan Straetmans (Maastricht University) expands on the benefits and shortcomings of multivariate extreme value analysis for measuring systemic risk. In his commentary, Jean-Charles Rochet (University of Zurich) presents a different perspective on regulations aimed at containing systemic risk. He proposes adopting a platform-based (instead of institution-based) regulatory perspective on systemic risk and encouraging a generalized move to central counterparty clearing.

Finally, under the heading of "It's Broke, Let's Fix It," the final paper, by Alan Blinder (Princeton University), presents a number of principles of sound regulation as well as a list of major recommendations and reviews the regulatory response in the United States in this area. At the conference, this paper introduced a panel discussion with Jean-Pierre Danthine (Swiss National Bank), Charles Goodhart (London School of Economics), and Jean-Pierre Landau (Banque de France). Jean-Pierre Danthine reviewed the Swiss experience of new capital and liquidity regulation and indicated that, although the balance sheets of the large Swiss banks have shrunk significantly, there has so far been little effect on lending. Charles Goodhart emphasized the importance of designing bank taxes well in order to prevent externalities and pointed to legal problems stemming from insufficient harmonization of national laws. Finally, Jean-Pierre Landau made a distinction between a buffer and an incentive approach to macroprudential regulation, noting that different instruments (such as capital requirements) may not necessarily serve both.

The Effects of Bank Capital on Lending: What Do We Know, and What Does It Mean?*

Jose M. Berrospide and Rochelle M. Edge
Federal Reserve Board

The effect of bank capital on lending is a critical determinant of the linkage between financial conditions and real activity, and has received especial attention in the recent financial crisis. We use panel regression techniques—following Bernanke and Lown (1991) and Hancock and Wilcox (1993, 1994)—to study the lending of large bank holding companies (BHCs) and find small effects of capital on lending. We then consider the effect of capital ratios on lending using a variant of Lown and Morgan's (2006) VAR model, and again find modest effects of bank capital ratio changes on lending. These results are in marked contrast to estimates obtained using simple empirical relations between aggregate commercial bank assets and leverage growth, which have recently been very influential in shaping forecasters' and policymakers' views regarding the effects of bank capital on loan growth. Our estimated models are then used to understand recent developments in bank lending and, in particular, to consider the role of TARP-related capital injections in affecting these developments.

JEL Codes: E51, G010, G21, G28, G32.

 We thank Brian Bucks, Mark Gertler, Diana Hancock, David Jones, Mike Kiley, Myron Kwast, Sylain Leduc, Nellie Liang, Alistair Milne, Dave Reifschneider, Jeremy Rudd, Klaus Schaeck, Skander Van den Heuvel, José-Luis Peydró, James Wilcox, and Egon Zakrajsek for comments on earlier versions of this paper and Peter Andrews, Matthew Osborne, Gianni De Nicolò, and Christina Wang for helpful conversations. The views expressed here are our own and do not necessarily reflect the views of the Board of Governors or the staff of the Federal Reserve System. Author e-mails: jose.m.berrospide@frb.gov and rochelle.m.edge@frb.gov. Mailing address: Mailstop 153, 20th and C Streets NW, Washington, DC 20551.

1. Introduction

The effect of changes in bank capital on the extension of bank credit is a key determinant of the linkage between financial conditions and real activity. Quantifying this relation has therefore been one of the most important research questions of the recent financial crisis. For example, early in the crisis—when the likelihood of a credit crunch was still under debate—the connection between bank capital and bank lending was central to any assessment of how losses in banks' mortgage-related portfolios would affect lending and thus real activity. Likewise, when the Troubled Asset Relief Program (TARP) moved to inject capital into banks through the Capital Purchase Program (CPP), the impact of the program on real activity largely focused on the effect of these injections on bank lending. More recently, this question has reemerged in light of proposals announced by the Basel Committee on Banking Supervision to raise banks' capital requirements and limit leverage ratios. In this case, policymakers and regulators need to weigh the effects of stricter capital requirements on economic activity against the benefits that would result from greater financial and macroeconomic stability.

In principle, there is a wide range of possible values for effect of a change in bank capital on bank assets and lending. On the one hand, a well-capitalized bank or a bank with access to additional sources of capital will be able to accommodate capital losses without reducing its assets (and hence its lending). By contrast, one could imagine a polar case in which banks very actively manage their assets in order to maintain a constant equity-capital-to-assets ratio (or "capital ratio")—for example, because they cannot raise equity to offset declines in their capital. In this case, a capital loss results in a reduction in assets, with the required reduction equal to the size of the bank's capital loss scaled up by the inverse of its capital ratio (that is, its leverage ratio). Since observed bank capital ratios roughly range from 0.08 to 0.125, leverage ratios range from $8(= 1/0.125)$ to $12.5(= 1/0.08)$. This means that a dollar's reduction in capital results in an \$8 to \$12.50 reduction in bank assets.[1] How

[1] If leverage is procyclical, the reduction in bank assets associated with a \$1 reduction in capital is even larger.

close we will be to either case in practice is, of course, an empirical question.

Despite its importance, there are relatively few recent estimates for the United States of the effect of changes in bank capital on lending. This is somewhat surprising given the existence of an active earlier literature on this topic. In particular, during the sluggish recovery that followed the 1990–91 recession, many observers debated whether the newly introduced capital regulations associated with Basel I were inhibiting lending. Although this debate did not yield a definitive conclusion, it did result in the development of empirical models that expressly sought to quantify the effect of bank capital on bank lending. For example, Hancock and Wilcox (1993, 1994) estimated models relating changes in individual banks' loan growth to measures of loan demand and bank capital, and Berger and Udell (1994) specified an equation relating the growth rate of various bank assets to different measures of bank capital ratios. Finally, Bernanke and Lown (1991) developed state-level equations linking bank loan growth to bank capital ratios and employment, along with bank-level equations for a single state (New Jersey).[2]

This paper draws on this earlier literature, as well as more recent approaches, to examine how bank capital affects bank lending in the United States. In all cases, we find relatively modest effects of bank capital on lending and more important roles for factors such as economic activity and increased perception of risk by banks. Using balance sheet data for large bank holding companies (BHCs), we consider variants of two of the approaches developed in the earlier literature: specifically, the Hancock and Wilcox (1993, 1994) and Bernanke and Lown (1990) models. For our purposes, the main difference between these two approaches is that the Bernanke-Lown approach considers the effect of *actual* BHC capital-to-asset ratios on BHC loan growth while the Hancock-Wilcox approach looks at

[2]Another important paper in this literature is Peek and Rosengren (1995), which used New England data to fit a model of bank deposits. Because it focuses on liabilities, this methodology is less suitable for studying the effect of bank capital on loan growth, however.

the effect of deviations of BHC capital levels relative to an estimated target. Both approaches have merits and shortcomings—which we will discuss at length—but both also agree in finding relatively small effects of BHC capital on loan growth.

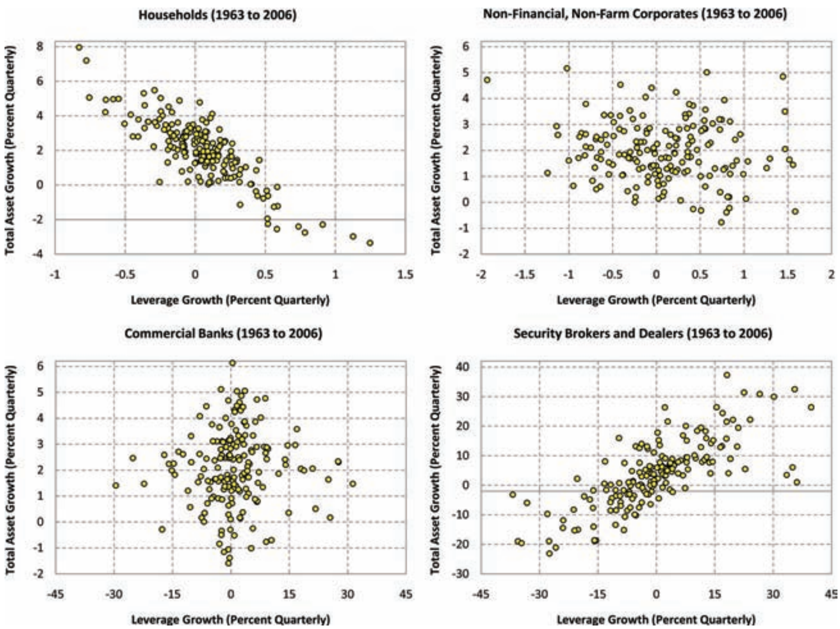
A potential concern surrounding both sets of results is survivor bias: Our panel only includes BHCs that were still in operation at the end of our sample period (2008:Q3). We address this by also considering a more aggregate approach—namely, a modified variant of one of Lown and Morgan's (2006) vector autoregression (VAR) models. This allows us to investigate the dynamic and general equilibrium effects of an exogenous change in bank capital ratios. Again we find modest effects of bank capital ratio changes on loan growth even using this very different approach.

Our estimates of the effect of bank capital on lending are considerably smaller than the effects suggested in recent statements by U.S. Treasury officials. For example, on the likely effects on lending of capital injections and anticipated capital raising following the stress tests, such statements suggest that a \$1 capital injection generates between \$8 and \$12 of lending capacity.[3] These magnitudes seem more consistent with the view described earlier that banks actively manage their assets to maintain constant bank capital ratios. This view has been quite prominent of late and is apparently based on a scatter plot for aggregate leverage and commercial bank asset growth reported by Adrian and Shin (2007).[4] This scatter plot for commercial banks is reproduced in the lower-left panel of figure 1 using our data; the figure also shows scatter plots for all of the other institutions that Adrian and Shin consider (specifically, households, non-financial and non-farm corporate firms, and security brokers and dealers). The sample period used in figure 1 is 1963 to 2006, the same as that employed by Adrian and Shin.[5] The roughly constant

[3]Our estimates are one-fifteenth to one-quarter as large. See Treasury Department Press Release (TG-95) "Treasury Secretary Timothy Geithner, Opening Remarks as Prepared for Delivery to the Congressional Oversight Panel," April 21, 2009, for the quoted lending estimates.

[4]Adrian and Shin (2007) do not focus on the effect of bank capital ratios on loan growth. Rather, their paper presents a pictorial representation of this relation as part of a set of motivating facts before moving on to consider the sources of procyclical leverage among securities brokers and dealers.

Figure 1. Assets and Leverage Growth of Various Institutions in the Flow of Funds Accounts



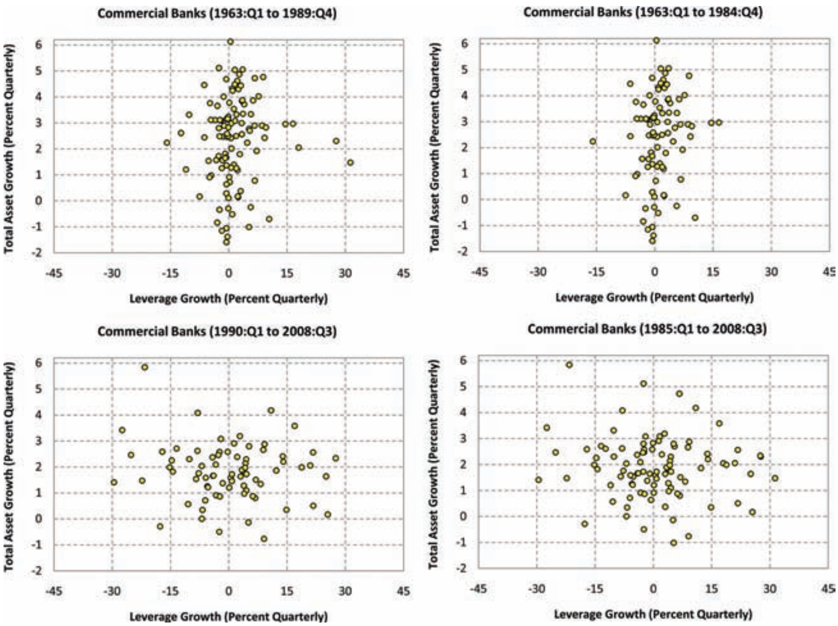
[5]The series used to generate the reported plots in figure 1 are as follows. The top-left panel gives the growth rate of the household sector's total assets (series FL152000005.Q in the Flow of Funds Accounts) and the growth rate of the household sector's leverage ratio (measured as the ratio of total assets—series FL152000005.Q—divided by household equity, i.e., total assets—series FL152000005.Q—less total liabilities—series FL154190005.Q). The top-right panel shows the growth rate of total assets for the non-financial and non-farm corporate sector (series FL102000005.Q) and the growth rate of the leverage ratio for this sector (total assets—series FL10200005.Q—divided by non-financial, non-farm corporate equity—series FL102000005.Q less series FL104190005.Q). The bottom-left panel plots the growth rate of total financial assets for the commercial banking sector (series FL764090005.Q) against the growth rate of the sector's overall leverage ratio (total financial assets—series FL764090005.Q—divided by commercial bank equity—series FL764090005.Q less FL764190005.Q). Finally, the bottom-right panel plots the growth rate of total financial assets for security brokers and dealers (series FL664090005.Q) against the growth rate of their overall leverage ratio (total financial assets—series FL664090005.Q—divided by broker-dealer equity—series FL664090005.Q less series FL664190005.Q).

leverage ratio apparent from the commercial bank panel—especially when compared with other classes of institutions—suggests a very active management of assets by commercial banks. This implies that a change in bank capital has a magnified effect—with the scaling factor equal to the leverage ratio—on asset levels and lending volumes.

Adrian and Shin's scatter plots have also been quite influential in shaping forecasters' views. For example, in a note written early in the crisis, Hatzius (2007) used these commercial bank scatter plots to estimate the reduction in loan volumes that would result from banks' mortgage-related portfolio losses. In a subsequent Brookings paper on the same issue, Hatzius (2008) again appealed to asset-leverage scatter plots to inform revised estimates of the effect of bank capital losses on loans.

Interestingly, reconciling our regression results with the Adrian-Shin scatter plots turns out to be very simple: Because of the major structural changes that occurred in the banking sector following the adoption of the Basel Accord, our empirical analysis begins in the early 1990s; in contrast, Adrian and Shin's scatter plots span a much longer period that starts in 1963. To see what effect the choice of sample period has on the analysis, consider the two left panels of figure 2, which give scatter plots for commercial bank asset and leverage growth over 1963:Q1 to 1989:Q4 (the top panel) and 1990:Q1 to 2008:Q3 (the bottom panel). As can be seen from comparing these plots, the feature of the data that has led to the view that commercial banks actively manage their assets to maintain constant leverage is much more of an artifact of the early part of the sample and is considerably less evident in the latter (post-Basel) part, which is naturally more relevant for considering current questions. Indeed, in the latter part of the sample the scatter plot is merely a cloud of points with no obvious correlation. The two right-hand panels of figure 2 provide a different split of the sample; specifically, one that breaks the sample in 1985:Q1 and therefore uses a more equal number of observations (eighty-eight and ninety-four observations, respectively) for each plot. Again, the feature of the data reported by Adrian and Shin (2007) and emphasized by Hatzius (2007, 2008) is very evident in the first part of the sample and not at all evident in the second

Figure 2. Assets and Leverage of Commercial Banks across Different Sample Periods



part.[6] Clearly, the scatter plots shown in the lower panels of figure 2 make the small effects of bank capital on loan growth found in our regressions less surprising.

A couple of other related recent studies are also worth noting. First, Francis and Osborne (2009) apply Hancock and Wilcox's approach to UK commercial bank data; like us, they find relatively modest effects of bank capital shortfalls on lending. Ciccarelli, Maddaloni, and Peydró-Alcalde (2010) build on Lown and Morgan's VAR model—albeit along a different dimension than we do—to study the credit channel of monetary policy in the euro area and the United States (in this case using the European Bank Lending Survey, or

[6]Another reason to split the sample in 1985 is that the Call Report data were less systematically collected and cleaned prior to 1985. Moreover, major changes were made to the reporting forms in early 1984 and this led to discontinuities in many of the existing time series, thereby making pre- and post-1985 comparisons difficult.

BLS, and the U.S. Senior Loan Officer Opinion Survey, or SLOOS). In contrast to our approach, however, Ciccarelli, Maddaloni, and Peydró-Alcalde (2010) do not explicitly consider the influence of bank capital by including a measure of it in their VAR model but rather rely on answers to "reasons for tightening/loosening lending standards" questions from the survey. Based on this approach, Ciccarelli, Maddaloni, and Peydró-Alcalde (2010) conclude that bank capital and liquidity have significant effects on GDP.[7]

Other methods for examining the effect of bank capital on lending have been pursued in a number of recent papers, albeit for countries other than the United States. For example, following Peek and Rosengren (1997), Puri, Rocholl, and Steffen (forthcoming) use loan applications from German Landesbanks to examine the effect of shocks to capital on the supply of credit by comparing the performance of affected and unaffected banks. Gianetti and Simonov (2010) use Japanese data to perform a similar exercise concerning bank bailouts. These papers do find a relevant role for capital in determining loan volumes, although they do not explicitly compare the magnitudes of the effects they find with those implied by the constant leverage view. Another group of papers use firm and bank loan-level data; these include Jiménez et al. (2010), who use Spanish data, and Albertazzi and Marchetti (2010), who use data on Italy. These papers find sizable effects of low bank capitalization and scarce liquidity on credit supply. (We contrast our methodology with the approaches that use firm and bank loan-level data later in the paper.)

We would note that although the above-mentioned loan-level studies for Spain and Italy suggest larger effects than what we find, other studies for the United States that use loan-level data (albeit for syndicated loans only) matched to bank and firm variables, such

[7]Results from our VAR analysis also show significant effects of lending standards on both bank lending and GDP growth. However, our reading of the "reasons for tightening or loosening lending standards" responses in the SLOOS differ notably from Ciccarelli, Maddaloni, and Peydró-Alcalde (2010). Indeed, we find that the two most important reasons for tightening lending standards in the SLOOS during the recent financial crisis were a more uncertain economic outlook and a reduced tolerance for risk. Deteriorations in banks' capital positions, in contrast, ranked as the least important reason. These responses are consistent with our empirical findings that bank loan growth in the United States appears to be determined more by factors like aggregate demand and bank risk.

as Santos and Winton (2010), obtain relatively small effects of bank capital on lending. In addition, Elliott (2010) uses simulation-based techniques and finds small effects of higher capital ratios on loan pricing and loan volumes for U.S. banks. Finally, De Nicolò and Lucchetta (2010) use aggregate data for the G-7 countries and conclude that credit demand shocks are the main drivers of bank lending cycles. Thus, our empirical findings of modest effects of capital on total loan growth appear consistent with other papers that employ U.S. data.

Our findings have several interpretations. First, in making their lending decisions, BHCs might not actively manage their assets on the basis of their capital positions.[8] If our results are correct, and other factors such as loan demand or increased bank risk play a more important role in determining BHC loan growth, then we have an explanation for the notable slowing of loan growth observed over the first three quarters of 2009: The impact of increases in the availability of loanable funds from various government actions was more than offset by weak economic conditions that reduced loan demand, as well as by increased risks (such as the reduced creditworthiness of borrowers).

Second, given the widespread consensus that banks entered the most recent crisis with too little capital relative to the riskiness of their portfolios (even though by most measures they were previously viewed as well capitalized), another interpretation of our results is that conventional measures of bank capital—such as those that we use—may not be relevant for assessing financial intermediaries' capital positions. If so, our results naturally call into question the prevailing risk-based measures employed by the regulatory capital framework and militate a strengthening in risk coverage, particularly in terms of incorporating capital requirements for off-balance-sheet operations (such as securitization activities) and counterparty exposures.

[8]This does not contradict the view that adequate levels of bank capital are crucial to prevent institutional failure, given the risks that attend financial intermediation: As shown by Berger and Bouwman (2009), increases in bank capital seem to be key to boosting bank profits and helping banks survive the effects of financial crises.

The balance of the paper is organized as follows. We describe our data in section 2. Section 3 contains empirical estimates of the relation between bank capital and bank loan growth at the institutional level, while section 4 presents our results using aggregate data. Sections 5 and 6 then use the models developed in sections 3 and 4 to address some questions of practical policy interest. Specifically, in section 5 we estimate the likely effect that TARP capital injections had on bank loans, and in section 6 we consider the factors underlying recent reductions in bank loan growth. Section 7 concludes.

2. Data

Because we use both panel and time-series methods, the paper employs a wide range of data sources. We discuss these data sources at length in this section.

2.1 Institution-Level Data

For our panel regression, which we conduct at the institution level, we use bank holding company data rather than data on the loans of individual commercial banks within a BHC. This reflects the observation that many decisions regarding a BHC's activities are taken for the institution as a whole rather than on a subsidiary-by-subsidiary basis.[9] Institution-level data are taken from the Consolidated Financial Statements for Bank Holding Companies (FR Y-9C); the data cover the period 1992:Q1 to 2009:Q3, with the exception of the data series that are based on regulatory capital (and that appear in some of our Bernanke-Lown type regressions), which begin only in 1996. Our sample consists of 165 large BHCs, all with total assets in excess

[9]For evidence supporting this view see Ashcraft (2004), who documents the practice of BHCs providing assistance—sometimes in the form of capital injections—to distressed subsidiaries, and Houston, James, and Marcus (1997), who report that BHC subsidiary loan growth is more sensitive to the cash-flow and capital positions of the holding company than to the subsidiary's position.

Table 1. Summary Statistics for BHC-Level Panel Data

Variable	Number of Observations	Median	Mean	Std. Dev.
Total Assets ($ Million)	11,107	3,595.30	27,302.91	145,408.60
Total Loan Quarterly Growth	11,107	2.15%	3.00%	5.03%
C&I Loan Quarterly Growth	11,107	2.09%	2.91%	8.82%
Equity/Assets	11,107	8.17%	8.39%	2.16%
Tier 1 Capital Ratio	7,320	10.72%	11.38%	3.29%
Total Capital Ratio	7,320	12.54%	13.24%	2.20%
TCE Ratio	9,139	6.83%	7.02%	2.20%
Return on Assets	11,107	0.1396	0.1296	0.8287
Securities/Assets	9,460	0.2257	0.2424	0.1131
Net Charge-offs/Assets	11,099	0.1687	0.2963	0.4288

Notes: This table presents summary statistics for the sample of 140 BHCs from 1992 to 2008 used in the panel estimation. The data are taken from the Consolidated Financial Statements for BHCs (FR Y-9C). The sample excludes institutions with missing observations in total assets, loans, and capital, and also BHCs that remain in the sample for less than thirty quarters. To minimize the influence of outliers, observations with more than 50 percent growth in total assets over a single quarter were also removed. The remaining variables were winsorized at the 1 percent and 99 percent levels.

of $3 billion as of 2008:Q3. This represents approximately 85 percent of the total assets in the banking sector.[10]

Table 1 reports summary statistics for the variables used in the panel regressions. Although the sample only includes large BHCs, there is a sizable variation in total assets. The median BHC in the sample has assets above $3 billion. However, the standard deviation of $145 billion is consistent with the fact that the four largest institutions in the sample have assets above $1 trillion, whereas total assets

[10]We exclude institutions with missing observations in total assets, loans, and capital, and also BHCs that remain in the sample for less than thirty quarters. In order to minimize the influence of extreme outliers, we remove observations with more than 50 percent growth in total assets over a single quarter and winsorize the remaining variables at the 1 percent and 99 percent levels. The final sample consists of about 11,107 bank-quarter observations for 140 institutions out of the 165 BHCs.

for the smallest ones are below $1 billion.[11] Quarterly growth in total and C&I (commercial and industrial) loans averages 3 percent across all BHCs and over time. The distribution of variables scaled by assets such as capital, securities, and net charge-offs exhibits smaller variances. Both the mean and the median values for the equity-to-assets (leverage) ratio are about 8 percent, while the mean and median of total and tier 1 risk-based capital ratios (measured as total and tier 1 capital over risk-weighted assets) are 12.5 percent and 13.25 percent (for total) and 10.75 percent and 11.25 percent (for tier 1). Finally, the tangible common equity (TCE) ratio (measured as tangible common equity over tangible assets) for the mean and median BHC is about 7 percent.

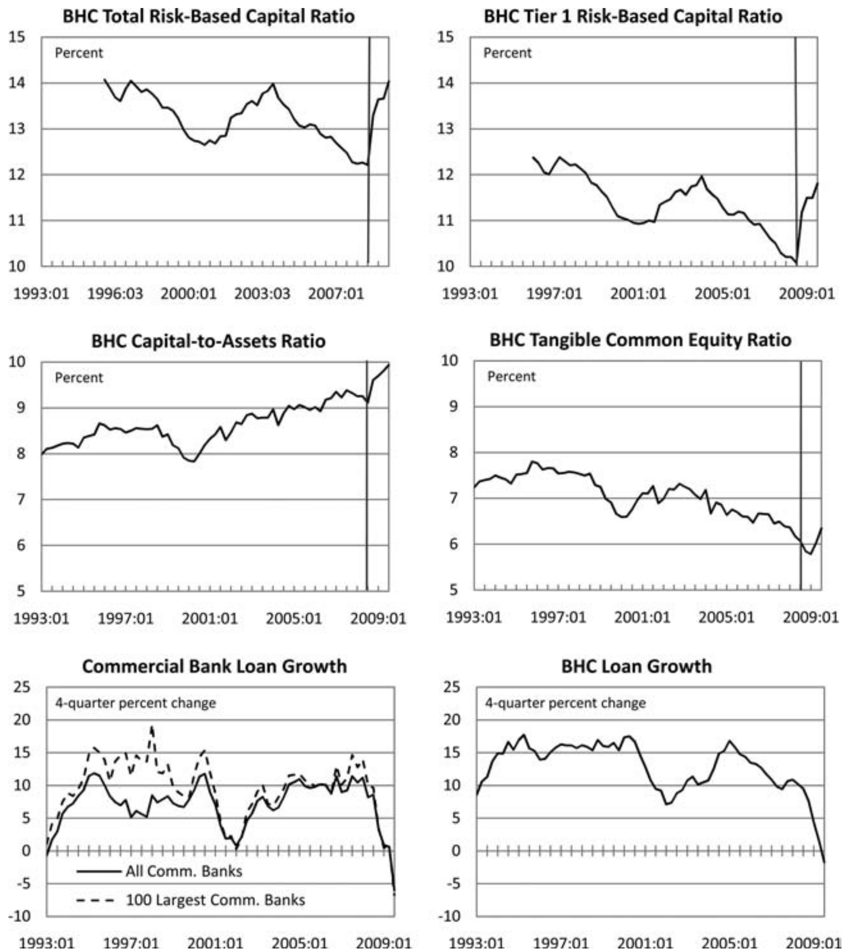
The first four panels of figure 3 depict the evolution of the average of different capital ratios over our sample. As is evident from the top two panels, throughout most of the sample period the average total and tier 1 risk-based capital ratios for large BHCs remained significantly above their well-capitalized regulatory minima.[12] Note also the divergent movements across different capital ratio measures: For example, while risk-based capital ratios have exhibited a downward trend since 2004, the capital-to-assets ratio—the middle-left panel—remained on an upward trend until mid-2007. This could partly reflect differences arising from the denominator of the capital ratio measures which, in turn, could be associated with increasing risk taking by banks during the last five years. In particular, excessive risk-taking behavior could have made risk-weighted assets increase faster than total assets.

As expected, the leverage ratio and the regulatory capital ratios all jump in 2008:Q4 after the implementation of the Treasury's capital injection program. (A vertical line is shown in the top four panels of figure 3 at 2008:Q3, just before the capital injections.) However, because the Treasury injected capital through purchases of preferred stock with warrants rather than common equity, the TCE ratio—the middle-right panel—is not affected by TARP capital infusions and seems to have maintained the downward trend that it has displayed

[11] As noted above, the BHCs in our sample have total assets in excess of $3 billion in 2008:Q3. In earlier periods, however, total assets can be smaller than the $3 billion cutoff.

[12] Among other criteria, a BHC is considered well capitalized if it has a total risk-based capital ratio of at least 10 percent and a tier 1 risk-based capital ratio of at least 6 percent.

Figure 3. BHC Capital Ratios and Commercial Bank and BHC Loan Growth



since 2003. Adverse market risk perceptions of the banking sector were more evident after the beginning of the financial crisis, and this explains a deeper contraction in the average TCE ratio after the second half of 2007. This market-focused measure of capitalization reached its lowest value in 2009:Q1 due to adverse market expectations in advance of the May 2009 release of the bank stress-test results (the average TCE ratio was below 6 percent, with ratios below 2 percent for some of the largest BHCs).

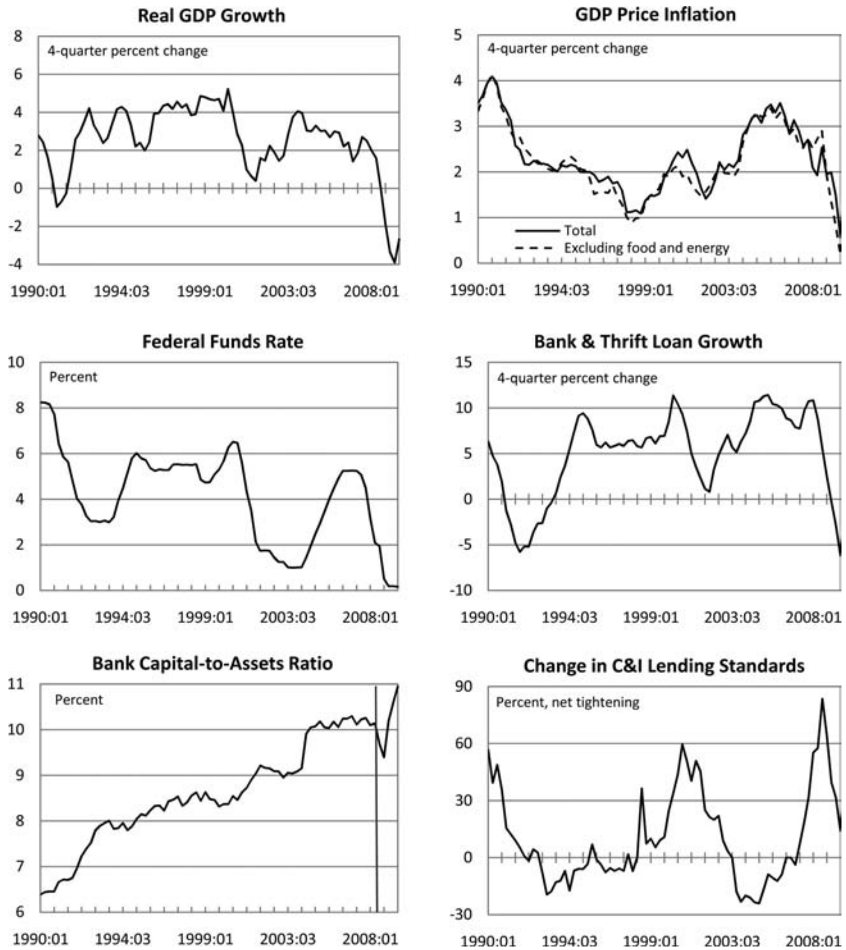
The bottom two panels of figure 3 plot the average four-quarter growth rate of total loans for the 165 BHCs in our sample, along with loan growth for all commercial banks and large banks only. As can be seen from these panels, commercial bank and BHC loan growth slowed notably in the 2001 recession but recovered rapidly thereafter, with the average four-quarter growth rate prior to 2007 fluctuating between 10 and 15 percent. Total growth in bank loans exhibits a severe contraction starting in early 2008, with a further slowing in growth following the ongoing deterioration in credit conditions during 2008. Negative four-quarter growth rates were evident in the first part of 2009.

2.2 Aggregate Data

Our panel regressions include a number of macroeconomic data series; these are plotted in figure 4. Specifically, our panel regression loan growth equations include real GDP growth (the top-left panel), as measured by the National Income and Product Accounts (NIPAs); inflation (the top-right panel, solid line), as measured by the overall GDP deflator; the federal funds rate (the middle-left panel); and (aggregate) net tightening of C&I lending standards (the bottom-right panel), as reported by the Federal Reserve Board's Senior Loan Officer Opinion Survey (SLOOS).

A large fraction of the aggregate variables that are included in the paper's panel regressions are also included in the VAR model that we consider in section 4. Indeed our panel equation for loan growth has a similar specification to our loan growth equation in the VAR, albeit without the institution-specific variables that the former equation has. In particular, our VAR model includes real GDP growth, along with inflation. However, the inflation measure used in the VAR is based on the GDP deflator excluding food and energy (see the dotted line in the top-right panel of figure 4). Our reason for using different inflation measures is attributable to some very large swings in overall GDP prices caused by swings in energy prices at the end of our data. In our panel regressions, the difference between the two inflation measures has a minimal impact on our results since our estimation period stops at the third quarter of 2008, before we observe significant movements in energy prices. In the VAR, these swings were not important for our impulse response

Figure 4. Aggregate Data Used in Panel and VAR Models



functions but did influence (in an implausible way) the composition of estimated non-financial structural shocks being derived from our model over the 2008 to 2009 period.[13]

---

[13]Specifically, when our model included the total GDP deflator, we found that the most important non-financial shock driving movements in bank and thrift loan growth over 2008 and 2009 was the price shock, which seemed somewhat counterintuitive. However, using the GDP deflator excluding food and energy eliminated this problem.

Like the panel regression, our VAR model also includes the federal funds rate and the net tightening in C&I lending standards from the SLOOS. Our measure of the aggregate capital asset ratio (the bottom-left panel) is for commercial banks and is taken from the Call Reports. Finally, for loan growth (the middle-right panel) we use bank and thrift depository loan growth as reported by the Flow of Funds Accounts. Our principal motivation for using commercial bank and thrift loans in the VAR rather than just commercial bank loans—which may seem more sensible given that all other banking-sector variables pertain only to commercial banks—arises from some very sharp spikes in the latter series caused by purchases of large thrifts by commercial banks, such as Bank of America's purchase of Countrywide and JP Morgan's purchase of Washington Mutual. Using the aggregate bank and thrift data eliminates these swings and does not greatly alter the VAR's impulse response functions, although clearly it leads to a more plausible estimated sequence of structural shocks.

3. Bank Capital and Loan Growth at the Institution Level

In this section we use institution-level BHC data and panel regression techniques to study the effects of BHC capital on loan growth. The idea here is to model the loan growth of our panel of BHCs as a function of supply and demand factors, with one of our supply factors being BHC capital. Specifically, the loan growth regressions that we estimate take the form

$$\begin{aligned}\Delta\% LOAN_{i,t} = {} & \sum_{s=1}^{4} \alpha_s \cdot \Delta\% LOAN_{i,t-s} + \sum_{s=1}^{4} \gamma_s \cdot \Delta\% GDP_{t-s} \\ & + \sum_{s=1}^{4} \delta_s \cdot INF_{t-s} + \sum_{s=1}^{4} \beta_s \cdot \Delta RFF_{t-s} + \sum_{s=1}^{4} \zeta_s \cdot STD_{t-s} \\ & + \phi \cdot LIQU_{i,t-1} + \chi \cdot CHG_{i,t-1} \\ & + \psi \cdot \begin{cases} CAPITALSURPLUS/SHORTFALL \text{ or} \\ CAPITAL\text{-}TO\text{-}ASSETS\ RATIO \end{cases} + \epsilon_{i,t}, \end{aligned} \tag{1}$$

where $\Delta\%LOAN_{i,t}$ denotes the growth rate of BHC loans, $\Delta\%GDP_t$ is real GDP growth, $\Delta\%INF_t$ is the inflation rate measured by the GDP deflator, ΔRFF_t is the change in the federal funds rate, and STD_t denotes lending standards. The bank-specific variable $LIQU_{i,t}$ denotes BHC liquidity and is measured by the ratio of securities to total assets, while the bank-specific variable $CHG_{i,t}$ is the ratio of net charge-offs to total assets.

Because our loan growth regression (1) controls for the effects of other supply and demand variables, we can interpret the coefficient ψ on the capital term as the effect of BHC capital conditions on loan growth. We consider two ways that capital can affect loan growth. In the first case—which corresponds to the approach of Hancock and Wilcox (1993, 1994)—it is the divergence between a BHC's actual and target capital (i.e., its capital surplus or shortfall) that is important for determining loan growth. In the second case—which corresponds to the approach of Bernanke and Lown (1991)—a BHC's capital-to-assets ratio influences loan growth.

The BHC capital-to-assets ratio measures included in the Bernanke-Lown regressions can be calculated from variables taken directly from the Consolidated Financial Statements for Bank Holding Companies (the FR Y-9C data). By contrast, the capital measures used in the Hancock-Wilcox regressions first require us to create time-varying desired/target capital levels for each BHC in our panel. From these time series, we can compute the percentage surplus or shortfall of actual capital relative to target. (The precise method and results for this stage of the Hancock-Wilcox estimation procedure are discussed in sub-section 3.1.) Note that the estimation period for these regressions is 1992:Q1 to 2008:Q3; that is, we stop the analysis around when the crisis intensified in late September 2008 and just before TARP capital injections began to occur. This prevents the TARP capital injections from distorting our regression coefficients.

3.1 Estimation of BHC Capital Surplus/Shortfall Measures

Following Hancock and Wilcox (1993, 1994) and Flannery and Rangan (2008), we use a partial adjustment model between actual and target capital holdings to estimate individual bank-specific capital targets. The target capital ratio, $k^*_{i,t}$, is modeled as a linear function of a vector of control variables $X_{i,t}$, which include bank-specific characteristics as well as institutional and aggregate determinants (so $k^*_{i,t} = \theta \cdot X_{i,t}$). Based on the rationale that costs of altering bank capital prevent banks from moving capital to its target level immediately, the actual bank capital ratio $k_{i,t}$ is assumed to follow a partial adjustment process of the form

$$k_{i,t} - k_{i,t-1} = \lambda \cdot \left(k^*_{i,t-1} - k_{i,t-1}\right) + \epsilon_{i,t}.$$

Substituting in our expression for $k^*_{i,t-1}$, rearranging terms, and adding a bank-specific constant yields our estimation equation:

$$k_{i,t} = \alpha_i + (1 - \lambda) \cdot k_{i,t-1} + \lambda \cdot \theta \cdot X_{i,t-1} + \epsilon_{i,t}. \qquad (2)$$

The bank-specific variables in $X_{i,t}$ include the log of total assets, the return on assets, and the net charge-off rate; these are intended to serve as proxies for size, earnings, and risk, respectively. Size captures the fact that larger banks are likely to face lower risks (due, for example, to greater diversification) and better access to sources of funding (thereby requiring less capital). Earnings are included because dividend payments are often slow to adjust; hence, an increase in earnings results in an accumulation of retained earnings and bank capital. Risk captures the fact that markets and regulators require more capital to be held against riskier assets. In the benchmark specification of our capital ratio equation we also include the loan/security composition of bank assets, which reflects the differential risk exposures implied by different types of assets, and the composition of bank loans, which for similar reasons could also affect target capital ratios. These terms are dropped in an alternative specification of equation (2) that we consider.[14] The institutional variable, called "regulatory pressure," is a dummy variable that equals unity if the BHC's equity capital ratio is less than 1.5 percentage

[14] All composition variables, such as the composition of security and loan holdings, are calculated as a share of bank assets.

points above the minimum of 5 percent.[15] We also include aggregate variables such as a measure of stock market volatility and the aggregate net charge-off rate.

Note that the only measure of capital that we consider in deriving our surplus/shortfall estimates is equity capital. Consequently, we do not estimate models of the form described by equation (2) for the three other capital ratios—i.e., the total and tier 1 risk-based capital ratios and the tangible common equity-to-assets ratio—that we reported in figure 3 and discussed in section 2.1. (The full set of capital ratios is examined when we consider the Bernanke-Lown version of equation (1).) We focus here on equity capital on the grounds that the Hancock-Wilcox loan growth regression implicitly seeks to capture a behavioral relation whereby BHCs alter the growth rate of their loan volumes—the dependent variable of equation (1)—so as to obtain a desired level of capital. If this is the case, however, it would then be most likely that such an adjustment would involve a fairly broad measure of capital (such as total equity capital) rather than a more narrow measure (such as tangible common equity). Similarly, we do not consider total and tier 1 risk-based capital ratios because these are essentially regulatory concepts for which the relevant target is the threshold that determines whether a BHC is adequately or well capitalized. Thus, it makes less sense to try to develop a model of the target for these capital ratios. In addition, even if BHCs did seek to target these ratios, they would be more likely to adjust their risk-based assets—that is, to adjust the composition of their loans as well as their volume—which would make this a less useful tool for analyzing the growth of aggregate loan volumes (the ultimate focus of our analysis).

Table 2 provides the estimation results for equation (2) under two different specifications of the $\theta \cdot X_{i,t-1}$ term. Other than the coefficients on bank size and aggregate stock market volatility, most of the coefficients reported in table 2 are close to what might be expected a priori. Size receives a positive and significant coefficient in both specifications, which implies that larger BHCs have higher equity capital ratios. This contradicts previous findings that larger BHCs have lower capital ratios, and suggests that there are some

[15] We take 5 percent to be the approximate level below which regulators in the United States would have concerns about a BHC's capital adequacy.

Table 2. Determinants of BHC Target Capital Ratios

	Model 1		Model 2	
	Estimate	***t*-stat**	**Estimate**	***t*-stat**
Lagged Capital Ratio	0.9108***	(87.60)	0.9189***	(87.91)
Size	0.0931***	(4.12)	0.0778***	(5.09)
ROA	0.0356	(1.48)	0.0373*	(1.76)
Aggregate Volatility	−0.0053***	(−4.67)	−0.0035***	(−4.09)
C&I Loan Share	−0.0097***	(−2.95)		
Real Estate Loan Share	−0.0065***	(−3.20)		
Consumer Loan Share	0.0001	(0.01)		
Securities Share	−0.0059***	(−2.92)		
Regulatory Pressure	−0.0260	(−0.58)	−0.0159	(−0.43)
Sector Charge-off/Loan	0.1074**	(2.45)	0.0781**	(2.58)
Charge-offs/Assets	0.0229	(0.60)	0.0094	(0.75)
Intercept	−0.1584	(−0.42)	−0.5334**	(−2.56)
Within R^2	0.852		0.870	
N	8,706		10,512	

Notes: This table presents the fixed-effect regression estimates of equation (2) for the determinants of BHCs' target capital ratios. Bank-specific variables include Size, measured as the log of total assets, the return on assets (ROA), the ratio of net charge-offs to total assets, the shares of C&I loans, real estate loans, consumer loans, and investment securities in total assets. The regression also includes an institutional variable, called regulatory pressure, defined as a dummy variable that equals one if the BHC's equity capital ratio is less than 1.5 percentage points above the minimum of 5 percent. Aggregate variables include the stock market volatility index provided by the Chicago Board Options Exchange (CBOE S&P 100 volatility index) and the ratio of net charge-offs to total loans for the commercial banking sector. Regressions include quarterly dummies to control for seasonal factors. Robust t-statistics are given in parentheses. *, **, and *** denote significance at the 10 percent, 5 percent, and 1 percent level, respectively.

economies of scale in holding capital at larger institutions.[16] We also find that larger shares of each broad category of loans and of

[16]Flannery and Rangan (2008) and Berger et al. (2008) find a negative relationship between capital ratios and size when they use either market capital or regulatory capital ratios. For the BHCs in our sample we also observe that larger BHCs operate with lower market capital and regulatory capital ratios. However, the divergence across different bank sizes is less evident when we use the equity capital ratio. Interestingly, Flannery and Rangan (2008) also consider the determinants of the book value equity ratio (defined as the ratio of the book value of common equity to the book value of total assets), which is a concept similar to the equity capital ratio that we use, and here they find an insignificant zero coefficient on bank size.

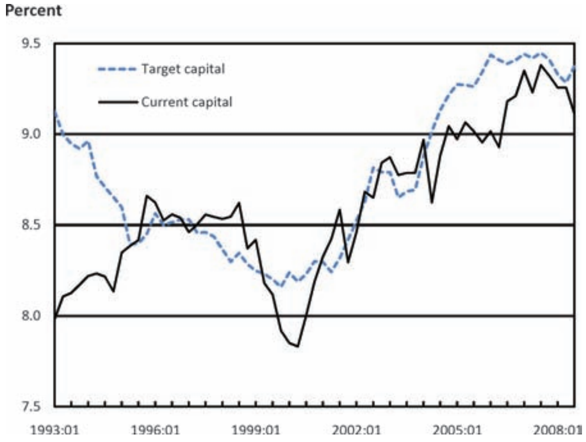
securities (relative to assets) are associated with lower capital ratios, which we interpret as indicating that BHCs with more portfolio diversification (loans categories and investment securities) hold less capital. The negative coefficient on stock market volatility appears to reflect the countercyclical nature of volatility and the procyclicality of bank capital. Both individual and aggregate net charge-off rates are associated with higher capital ratios, though only the aggregate measure seems statistically significant. These two variables control for bank risks and suggest that BHCs hold additional capital during bad times, when the credit quality of loans deteriorates. Our dummy control for regulatory pressure has a negative coefficient but is statistically insignificant, which suggests that regulatory restrictions do not seem to bind in a way that affects equity capital ratios. Finally, BHCs with higher profits tend to hold more capital, but this result also appears to be statistically weak.

In addition to the usual candidate determinants of the target capital ratio, the regression equation includes bank-specific fixed effects to control for possible omitted variables and to capture heterogeneous characteristics such as different risk preferences, governance structures, and managerial skills. The fixed effects explain a large fraction of the cross-sectional variation in target capital ratios without affecting the statistical significance of the other time-varying and firm-specific characteristics in the regression equation.[17] Interestingly, the coefficient on the lagged capital ratio implies a relatively quick adjustment of capital ratios to target: Our estimated average speed of adjustment of 36 percent per year is smaller than the 49 percent per year rate found by Flannery and Rangan (2008) but within the range of 28 percent to 41 percent reported by Berger et al. (2008).

As in Hancock and Wilcox, the estimates of the coefficients in equation (1) are used to calculate a time series for each BHC's target capital ratio $k^*_{i,t}$. These estimated target capital *ratios* are then used to construct target capital *levels* $K^*_{i,t}$. Capital surpluses, denoted by

[17]Although the presence of fixed effects in dynamic panel data estimation can lead to biased OLS estimates, simulations by Judson and Owen (1999) suggest that the the bias is minor in panels with more than thirty observations. Given that our panel uses fifty-nine periods of data and the minimum number of quarters for any BHC in our sample is thirty, we make no correction for potential short-panel bias.

Figure 5. Average Actual and Target BHC Capital-to-Assets Ratios



$Z_{i,t}$, are calculated as deviations of actual capital levels relative to target; that is, $Z_{i,t} = (K_{i,t} - K^*_{i,t})/K^*_{i,t}$.

Figure 5 plots the average actual and average target capital ratios—that is, $\bar{k}_t$ and $\bar{k}^*_t$—over our estimation period, where the averages are computed over the entire sample of BHCs. Differences between the average actual and average target capital ratios appear quite pronounced and quite persistent; nonetheless, they do appear to conform with certain aspects of the conventional wisdom concerning movements in these series. In particular, in line with the findings of the literature on bank capital management, we find that BHCs operated with significant capital surpluses over most of the late 1990s and early 2000s.[18] Furthermore, BHCs appeared to have operated with large capital shortfalls between 1993 and 1995, which may have reflected widespread capital pressures after the implementation of the Basel I capital guidelines around 1992.[19]

[18]For example, Flannery and Rangan (2008) and Berger et al. (2008) document a buildup of capital cushions during this period for large banks that was substantially above the "well-capitalized" regulatory minima (note that the 1990s and mid-2000s represented a period of unusual profitability for the banking sector).

[19]Hancock and Wilcox (1994) find that by the end of 1991 more than 40 percent of banks had capital shortfalls relative to their bank-specific targets.

3.2 Estimation of the Effect of Capital Surpluses or Shortfalls on Loan Growth

To gauge the effects of capital surpluses or shortfalls on bank loans, we include our estimates of bank capital surpluses—denoted by $Z_{i,t}$, above—in a regressions model of bank loan growth. Our model is similar in spirit to the loan growth regression of Hancock and Wilcox, who studied the link between bank capital and bank loan growth while controlling for macroeconomic conditions. Our particular specification is closer, however, to that used by Kashyap and Stein (1995, 2000), who modeled the growth rate of bank loans in terms of supply and demand factors.

We model the growth rate of BHC loans ($\Delta\% LOAN_{i,t}$) as a function of its own lags, lags of aggregate economic growth ($\Delta\% GDP_t$), lags of the GDP price inflation rate ($\Delta\% INF_t$), lags of the change in the federal funds rate (ΔRFF_t), and lags of lending standards (STD_t). In addition, we include lagged BHC-specific characteristics such as a control for bank liquidity ($LIQU_{i,t}$), which is measured by the ratio of securities to total assets, and the ratio of net charge-offs to total assets ($CHG_{i,t}$). The last term we include is our estimates of bank capital surpluses/shortfalls ($Z_{i,t}$), which is the main variable of interest. Thus, we fit the following panel regression equation:

$$\begin{aligned}\Delta\% LOAN_{i,t} = &\sum_{s=1}^{4} \alpha_s \cdot \Delta\% LOAN_{i,t-s} + \sum_{s=1}^{4} \gamma_s \cdot \Delta\% GDP_{t-s} \\ &+ \sum_{s=1}^{4} \delta_s \cdot INF_{t-s} + \sum_{s=1}^{4} \beta_s \cdot \Delta RFF_{t-s} \\ &+ \sum_{s=1}^{4} \zeta_s \cdot STD_{t-s} + \psi \cdot LIQU_{i,t-1} + \chi \cdot CHG_{i,t-1} \\ &+ \psi \cdot Z_{i,t-1} + \epsilon_{i,t}. \end{aligned} \tag{3}$$

The model is specified in growth rates to deal with non-stationary variables; to avoid potential endogeneity issues, only lags of the explanatory variables are used. We also include quarterly dummies to control for possible seasonality in the data. In this specification, GDP growth is used to control for changes in loan demand, lending standards are used to control for credit supply changes that arise from changes in banks' lending behavior, and the federal funds rate controls for monetary policy changes. At the firm level, the liquidity variable is intended to capture the extent to which BHCs use their stock of securities to adjust their loan growth, all else equal (a point raised by Kashyap and Stein) while the fraction of charge-offs to total assets is a proxy for risk, which should act as a drag on loan growth. Equation (3) is estimated with bank fixed effects in order to allow for potential omitted factors that vary across BHCs and are constant over time.

Table 3 presents the estimation results for equation (3).[20] The first pair of columns gives results when total bank loans are used as the dependent variable, while the second pair of columns gives results when C&I loans are used as the dependent variable. For both total bank loans and C&I loans there is a positive and significant coefficient on capital surpluses, indicating that the growth rate of both total and C&I bank loans is larger for banks with greater amounts of excess capital. The effects, however, are small: The long-run impact of a capital surplus (shortfall) on both total and C&I BHC loan growth is roughly to increase (reduce) annualized loan growth by 0.25 percentage point when capital exceeds (falls short of) its target level by 1 percent.

Our estimate of the effects of a 1 percent capital surplus on loan growth can—with some manipulation—be compared to the effects found by Hancock and Wilcox, who looked at dollar amounts. Using a 10-percentage-point average capital-to-assets ratio and a 60 percent average share of loans (as a fraction of bank assets), our model estimates imply that a \$1 capital surplus results in a \$1.86 boost in loan volumes.[21] This is a touch higher than Hancock and Wilcox's estimates, which implied that a \$1 surplus leads to a \$1.50 boost

[20] Because an auxiliary equation is used to estimate the capital surplus variable, our estimation of equation (3) is subject to a generated regressor bias. We use a bootstrap procedure to address this problem.

[21] The calculation is performed as follows. The relationship between a percent capital surplus/shortfall and loan growth is $\frac{L_t - L_{t-1}}{L_{t-1}} = 0.25 \cdot \frac{K_t - K_t^*}{K_t^*}$. Assuming that the economy was initially in steady state implies that $\frac{L_t - L_{t-1}^*}{L_{t-1}^*} = 0.25 \cdot \frac{K_t - K_t^*}{K_t^*}$, which yields $(L_t - L_{t-1}^*) = 0.25 \cdot \frac{L^*}{A^*} \cdot \frac{A^*}{K^*}(K_t - K_t^*)$ (ignoring time sub-scripts on the ratios of steady-state variables). Since $\frac{L^*}{A^*} = 0.62$ and $\frac{A^*}{K^*} = 12$ in the data, this implies that $(L_t - L_{t-1}^*) = 1.86(K_t - K_t^*)$.

Table 3. Effect of Capital Surplus on BHC Loan Growth

	Total Loans		**C&I Loans**	
	Estimate	**χ^2-stat**	**Estimate**	**χ^2-stat**
Loan Growth Lags (Sum)	0.1598***	(15.49)	0.0731	(1.34)
Surplus Capital	0.0523***	(75.98)	0.0602***	(24.78)
Securities/Asset	0.0668***	(20.25)	0.0559**	(4.03)
Net Charge-offs/Asset	−1.6177***	(25.02)	−1.6410***	(8.59)
Lending Standards (Sum)	−0.0181***	(8.24)	−0.0274***	(6.47)
GDP Growth Lags (Sum)	0.8562**	(5.52)	2.1685***	(7.65)
Federal Funds Lags (Sum)	−0.4534	(1.83)	−0.0059	(0.00)
Inflation Lags (Sum)	−0.1918	(0.07)	−1.7849	(2.19)
Intercept	0.5080	(0.42)	1.5000	(0.99)
Within R^2	0.215		0.098	
N	7,793		7,760	

Notes: This table reports the fixed-effect regression estimates of equation (3) for the determinants of bank loan growth on a measure of BHCs' capital surplus/shortfall and control variables. The dependent variable is the quarterly growth rate of total loans and C&I loans. Explanatory variables are as follows: four lags of loan growth; the measure of capital surplus, calculated as percent deviations of actual capital level relative to estimated capital targets; the ratio of investment securities to total assets; the ratio of net charge-offs to total assets; three lags of lending standards (fraction of banks reporting tightening lending in the Senior Loan Officer Opinion Survey); four lags of quarterly GDP growth; four lags of the quarterly change in the federal funds rate; and four lags of the inflation rate (cumulative percent change of the Consumer Price Index during the previous three months). Regressions include quarterly dummies to control for seasonal factors. Because an auxiliary equation is used to estimate the capital surplus variable, standard errors are adjusted to correct for the generated regressor bias through bootstrap methods. χ^2-statistics are given in parentheses. *, **, and *** denote significance at the 10 percent, 5 percent, and 1 percent level, respectively.

in loan volumes, but it is not vastly different. Our estimate is, however, very different—i.e., about one-sixth to one-tenth the size—from the \$8 to \$12.50 increase in loans that the constant leverage ratio assumption would imply.

The results in table 3 suggest a notable role for demand shocks and exogenous changes in net charge-offs in determining both total and C&I loan growth. The results imply that in the long run a 1-percentage-point reduction in GDP growth lowers the annualized growth rate of total loans by 4.1 percentage points and lowers the annualized growth rate of C&I loans by 9.4 percentage points. Furthermore, changes in net charge-offs have a large and significant impact in both specifications. For example, a 1-percentage-point increase in the net charge-off ratio depresses (in the long run) the annualized growth rate of both total and C&I BHC loans by about 7 percentage points. Turning to the other regressors, although the coefficient on the securities-to-assets ratio is significant, the effect of increased holdings of securities on BHC loan growth is small: A 1-percentage-point increase in the securities-to-assets ratio leads to about a 0.3-percentage-point long-run increase in both total and C&I annualized loan growth. Finally, the coefficients on lending standards in both the total and C&I loan growth equations are negative (as expected) and statistically significant, but imply an impact on loan growth that is economically small: A 1-percentage-point increase in lending standards leads to about a 0.1-percentage-point decrease in annualized loan growth. The small effect of standards may be explained by the fact that the capital surplus measure and charge-off ratio likely already capture changes in banks' willingness to lend.

3.3 Estimation of the Effect of Bank Capital Ratios on Loan Growth

As we will discuss in the next section (when we consider possible identification issues present in our panel regression analysis), one potential problem with the Hancock-Wilcox approach is that the target capital-to-assets ratio equation could be misspecified. If we model the target poorly, our estimates of the effect of a capital surplus or shortfall on loans will also be poor, thereby biasing the estimated surplus/shortfall impact coefficient ψ. (The direction of this bias will likely be downward for the usual errors-in-variables reasons.) An alternative approach to investigating the link between bank capital and bank lending that does not suffer from this issue involves focusing on the relationship between capital ratios and bank loan growth. For this purpose, we replace our measure of the capital surplus in equation (3) with the lagged value of different measures of the capital ratio. That is, we estimate

$$\Delta\%LOAN_{i,t} = \sum_{s=1}^{4} \alpha_s \cdot \Delta\%LOAN_{i,t-s} + \sum_{s=1}^{4} \gamma_s \cdot \Delta\%GDP_{t-s}$$

$$+ \sum_{s=1}^{4} \delta_s \cdot INF_{t-s} + \sum_{s=1}^{4} \beta_s \cdot \Delta RFF_{t-s}$$
$$+ \sum_{s=1}^{4} \zeta_s \cdot STD_{t-s} + \psi \cdot LIQU_{i,t-1} + \chi \cdot CHG_{i,t-1}$$
$$+ \psi \cdot (K_{i,t-1}/A_{i,t-1}) + \epsilon_{i,t} \quad (4)$$

using several measures of the capital-to-assets ratio for $K_{i,t-1}/A_{i,t-1}$: the equity capital-to-assets ratio (the book leverage ratio), the total risk-based capital ratio, the tier 1 risk-based capital ratio, and the tangible common equity (TCE) ratio. We consider each capital ratio separately. The different capital ratios measure different dimensions of banks' balance sheet positions and thereby reflect different forces that act to influence lending. For example, by using the equity capital ratio, we capture the impact of capitalization decisions on loan growth, while using the two risk-based capital ratios in our regression captures the effect on loan growth of both capital adequacy and portfolio composition because the risk-based measures also include off-balance-sheet assets. The TCE ratio is a more conservative measure of bank capital, intended to represent the first-loss position in the event of a bank's failure. We include this alternative measure of bank capital primarily because it was the focus of market participants in the quarters after October 2008 in evaluating the financial strength of banking institutions. (We would note that the specification implied by this equation (4) is broadly similar to the loan growth equation estimated in our VAR model in section 4.)

The estimation results for the effect of capital ratio shocks to capital ratios on BHC loan growth are reported in table 4. As can be seen from the table, all variables enter with the expected sign, and most are statistically significant. For example, there are positive and significant coefficients on the BHC capital ratios, which is consistent with the growth rate of bank loans being larger for banks with higher capital ratios. However, regardless of the capital ratio we use as our main explanatory variable, the effects of shocks to capital on BHC loan growth are small. Depending on the capital ratio employed, our results in table 4 suggest that a 1-percentage-point increase in the capital ratio leads to a long-run increase in annualized BHC loan growth that is only between 0.7 and 1.2 percentage points.

Table 4. Effect of Capital Shocks on BHC Loan Growth

	Equity-to-Asset Ratio (1)		Risk-Based Capital Ratio (2)		Risk-Based Tier 1 Capital Ratio (3)		TCE Ratio (4)	
	Estimate	***t/F*-stat**	**Estimate**	***t/F*-stat**	**Estimate**	***t/F*-stat**	**Estimate**	***t/F*-stat**
Loan Growth Lags (Sum)	0.1648***	(33.04)	0.1583***	(24.22)	0.1571***	(24.07)	0.1743***	(37.66)
Capital Ratio	0.1450***	(3.23)	0.1572***	(4.69)	0.1674***	(4.98)	0.2521***	(5.75)
Securities/Asset	0.0556***	(5.64)	0.0431***	(2.76)	0.0421***	(2.66)	0.0498***	(4.84)
Net Charge-offs/Asset	−1.6835***	(−6.75)	−1.8524***	(−6.92)	−1.8346***	(−6.86)	−1.6832***	(−6.65)
Lending Standards (Sum)	−0.0135**	(10.46)	−0.0130***	(7.09)	−0.0130***	(7.09)	−0.0131***	(9.89)
GDP Growth Lags (Sum)	0.8628***	(10.41)	1.0884***	(9.61)	1.0684***	(9.25)	0.8508***	(10.67)
Federal Funds Lags (Sum)	−0.1164	(0.28)	−0.5391*	(2.88)	−0.5696*	(3.17)	−0.1792	(0.64)
Inflation Lags (Sum)	−1.1988***	(6.65)	−0.4573	(0.44)	−0.3482	(0.26)	−0.7072	(2.31)
Intercept	−0.1001	(−1.90)	−1.2893	(−1.64)	−1.1345	(−1.49)	−0.7761	(−1.28)
Within R^2	0.186		0.196		0.197		0.190	
N	8,549		6,658		6,658		8,549	

Notes: This table reports the fixed-effect regression estimates of equation (4) for the determinants of bank loan growth on BHCs' capital ratios and control variables. Columns 1–4 report estimation results for each of the different capital ratios. The dependent variable is the quarterly growth rate of total loans. Explanatory variables are as follows: four lags of total loan growth, the capital ratio as indicated in each column, the ratio of investment securities to total assets, the ratio of net charge-offs to total assets, three lags of lending standards (fraction of banks reporting tightening lending in the Senior Loan Officer Opinion Survey), four lags of quarterly GDP growth, four lags of the quarterly change in the federal funds rate, and four lags of the inflation rate (cumulative percent change of the Consumer Price Index during the previous three months). Regressions include quarterly dummies to control for seasonal factors. Robust t and F-statistics are given in parentheses. *, **, and *** denote significance at the 10 percent, 5 percent, and 1 percent level, respectively.

The largest estimated effect of capital shocks on BHC loan growth comes from the regression specification that uses the TCE ratio; as mentioned above, this measure received significant attention by market participants, especially during the weeks surrounding the bank stress tests in 2009. However, even this specification suggests that higher bank capitalization would have a very modest impact on BHC loan growth. In other words, in the event that BHCs managed to raise additional capital either by issuing more common stock or by converting preferred shares to common shares—as some proposals for regulatory reform have suggested—the resulting increase in common equity would have only a minor impact on the growth rate of bank loans.

Our estimate of the long-run effect of a 1-percentage-point increase in the equity capital ratio, which lies toward the lower end of the above range of estimates, can be compared with the effect obtained by Bernanke and Lown for their sample of New Jersey banks. Bernanke and Lown's reported estimates are notably larger than ours: For their full sample of 111 New Jersey banks, they found that a 1-percentage-point increase in the equity capital ratio resulted in a 2-percentage-point increase in loan growth, while the corresponding increase in loan growth for their sample of 90 small New Jersey banks was around 2.5 percentage points. Clearly, the size of the institutions considered is one reason why our estimates differ from Bernanke and Lown's; indeed, Bernanke and Lown's analysis does indicate a notably smaller degree of sensitivity of bank loans to equity capital ratios for larger banks. Note also that because our analysis considers the largest 165 BHCs in the United States while Bernanke and Lown's analysis only considers banks that operate within the state of New Jersey, even the smallest institutions in our sample are larger than the "large" banks in Bernanke and Lown's sample.[22] Of course, another reason why our estimate may be smaller

[22]In principle, relative bank size should also induce similar differences between the estimates reported in section 3.2 and those found by Hancock and Wilcox. However, we do not observe such differences. One possible reason is that Hancock and Wilcox perform their analysis on the entire United States and exclude commercial banks with assets less than $300 million (roughly $500 million in current dollars). In addition, they "roll up" the commercial banks in their sample to the BHC top holder. Hence, it is likely that the banks in the Hancock-Wilcox study are on average larger than those used by Bernanke and Lown.

than that of Bernanke and Lown's is that whereas our model of loan growth includes a number of controls (including bank-specific ones), Bernanke and Lown's model only includes the capital ratio, which could be capturing the effects of other variables.

While the effects of the equity capital ratio on bank loan growth that we obtain in our analysis are somewhat smaller than those estimated by Bernanke and Lown, they are notably smaller—on the order of one-tenth the size—than the loan growth impact that would be implied by the constant leverage ratio assumption. That is, the constant leverage ratio assumption would imply that the percentage-point long-run increase in loan growth implied by a 1-percentage-point increase in the equity capital ratio is equal to the capital ratio—i.e., 8 percentage points for an 8 percent equity capital ratio and 12 percentage points for a 12 percent equity capital ratio.

As was the case for our analysis of the effects of capital shortfalls on loan growth, the model's estimation results suggest that demand shocks and exogenous changes in net charge-offs play the most important role in determining total loan growth. The response of BHC loan growth to demand shocks (measured by changes in GDP growth), implies that a 1-percentage-point reduction in GDP growth leads to about a 4-percentage-point decline in annualized loan growth. Furthermore, the potential effects of continued write-downs through further increases in net charge-offs imply a large and significant impact in all capital ratio specifications that we considered, as a 1-percentage-point increase in the net charge-off ratio reduces the annualized growth rate of total BHC loans by about 8 percentage points. As before, the coefficient on the securities-to-assets ratio is significant, but the effect of increased holdings of securities is small and leads to only a 0.3-percentage-point increase in BHC loan growth for every percentage-point increase in the securities-to-assets ratio. Finally, the coefficient on lending standards is negative and significant but has also small impact on BHC loans. As before, this latter finding might be explained by the fact that the capital shock and changes in the charge-off ratio already capture changes in banks' willingness to lend.

3.4 Endogeneity of the Capital Measures

In the discussion of our estimation results in sub-sections 3.2 and 3.3, we interpreted our parameter ψ as reflecting the magnitude of the increase in loan growth implied by a higher capital surplus or a higher capital ratio. Doing this, however, requires us to rule out other possible interpretations that could result in a relationship between bank loan growth and bank capital.

One possible (and well-recognized) reason why we might observe a positive correlation between a bank's loan growth and its capital—whether lagged or not—is that this correlation is driven by a third variable that affects both independently. It is likely that capital ratios will be positively correlated with the state of the economy (because of higher profits growth), which is in turn positively correlated with loan growth. Including macroeconomic variables—especially, economic growth—in our loan growth equation does limit this possibility, and the fact that the coefficients on lags of real GDP growth are so large in our loan growth equations suggests that we are in fact successfully capturing movements in loan growth that arise from fluctuations in economic activity.

A positive relationship between loan growth and *lagged* capital measures could also arise if banks preemptively increase their capital levels in anticipation of an increase in loan volumes (and hence balance sheet size). This response of bank capital levels to anticipated changes in loan volumes raises the possibility of an endogeneity issue in the Bernanke and Lown loan growth equations. However, this endogeneity issue is not necessarily an overwhelming concern if we simply wish to determine whether the constant leverage assumption exaggerates the estimated response of loan growth to an increase in capital ratios, because this source of endogeneity would induce an upward bias in our estimates of ψ. Nevertheless, to examine this further we considered an alternative specification that also includes the ratio of the BHC's cash dividends to its net income. To the extent that BHCs change their dividend policy so as to boost their capital in anticipation of future lending opportunities (for example, banks may cut dividends when they expect further increases in loan growth), the dividend-to-net-income ratio would capture the endogenous variation in capital due to expected changes in loan volumes. Our estimates using this alternative specification of the Bernanke-Lown equation (not reported) imply that, depending on the capital ratio employed, a 1-percentage-point increase in the capital ratio leads to an increase in the annualized BHC loan growth rate of 0.5 to 1 percentage points. This is close to the 0.7- to

1.2-percentage-point increase in BHC loan growth that we find when the benchmark Bernanke-Lown specification is used.

To the extent that our implementation of the Hancock-Wilcox approach correctly models the target capital level $K^*_{i,t}$, this identification issue should be of limited concern. If the reason banks boost their capital is that they anticipate a future balance sheet expansion, this should also be reflected in an increase in their target capital, which implies no net change in the capital surplus or shortfall. Consequently, the possibility that banks increase their capital in response to an anticipated expansion in lending does not generate a positive relationship between loan growth and capital surpluses or shortfalls, and the interpretation of ψ is preserved. That said, the possibility remains that our model of banks' target capital is incorrect and that—perhaps because a forward-looking component is absent from equation (2)—we have not completely purged our parameter ψ of this potential source of bias. Naturally, it would be possible to address this problem by introducing forward-looking behavior to the target capital relation, although this would also require specifying simple forecasting equations for the explanatory variables, which could themselves be misspecified. Thus, we do not pursue this approach. We tolerate some possibility of an incorrectly modeled target capital inducing some modest amount of upward bias in the parameter ψ, because the misspecification of the target capital ratio equation is always present and it is not clear that omitting a forward-looking component from the equation is necessarily the most important omission.

The preceding discussion highlights why we use both the Hancock-Wilcox and Bernanke-Lown approaches to analyze the effect of bank capital on bank lending. While we generally consider the Hancock-Wilcox approach to be preferable inasmuch as it allows us to interpret the parameter ψ as capturing the effect of an exogenous increase in capital on loan growth, it is vulnerable to the misspecification of its target capital equation, which can potentially reduce estimates of ψ. In contrast, while the Bernanke-Lown approach is not vulnerable to misspecifying a target capital variable, it is susceptible to some endogeneity issues that could then limit the interpretation of the parameter ψ. Again, however, these do not necessarily represent an overwhelming concern when our goal is to determine whether the constant leverage assumption exaggerates the estimated response of loan growth to an increase in capital ratios.

3.5 Robustness Analysis

We conducted a number of robustness checks that ultimately leave our key conclusions unchanged.

First, we considered an alternative, atheoretic approach to calculating the capital surplus term used in the Hancock-Wilcox regressions. Specifically, instead of estimating a model to calculate banks' capital targets, we used a one-sided Hodrick-Prescott filter applied to each BHC's capital ratio. The capital surpluses and shortfalls implied by the filter's estimates are smaller and smoother than their model-based counterparts but are also less consistent with the conventional view of when the banking system was operating with capital surpluses or shortfalls. (Indeed, this is the main reason that we chose the model-based target measure as our primary specification.) We found that the estimated effect of capital shortfalls on loan growth under the alternative measure was statistically significant and larger, implying roughly a 0.7-percentage-point reduction in annualized loan growth when capital is 1 percent below its target level. Nevertheless, this is still small relative to the values that would be implied by the constant leverage assumption.

We also included four additional lags of our various capital measures in regression equations (3) and (4) to allow for a less parsimonious specification aimed at capturing richer dynamics and potentially more persistent effects of capital on loan growth. In both cases the coefficients on the capital terms were almost unchanged, so that we still obtained only a modest impact of capital on BHC loan growth.

We next performed rolling window panel regressions over the sample period to address the possibility that the effect of shocks to bank capital on loan growth has been declining.[23] We used fixed-size rolling windows of six, seven, and eight years and quarterly data starting in 1992:Q1. In all cases, we found that the coefficients on our various capital terms were relatively stable and therefore continued to imply a small impact of capital on lending.

[23] Research on the bank lending channel of monetary policy seems to suggest decreased sensitivity of lending to bank capital as a result of greater securitization. See Loutskina and Strahan (2006) and Altunbas, Gambacorta, and Marques-Ibanez (2009).

We also considered cross-sectional regressions that relate four-quarter changes in BHC loan growth between 2007:Q3 and 2008:Q3 to the bank-specific variables in equations (3) and (4), lagged one year. In both cases, we found that the coefficients on our capital measures are slightly different but still imply a modest impact of capital on lending: The coefficient for the capital surplus measure implies a 0.2-percentage-point reduction in annual loan growth for a 1 percent capital shortfall. The coefficient for the equity-to-asset ratio almost doubles but still only implies that a 1 percent reduction in the capital ratio leads to a 1.4-percentage-point contraction in the annual loan growth. Interestingly, in both cross-sectional specifications the coefficient on the net charge-offs ratio doubles too, implying a much stronger impact of bank risk perception on BHC loan growth.

We also considered the possibility that the effect of capital on lending could vary with the business cycle and that such effects could be diluted by estimating, as we do, the relationship between loan growth and capital over a long sample period. To test for this possibility in the full-sample model, we interact the coefficients on capital in equations (3) and (4) with the output gap (as measured by the U.S. Congressional Budget Office) to determine whether the size of the effect of capital on loan growth increases in slumps. Although we do obtain the correctly signed coefficient in both cases (bank capital seems to have a larger effect in "bad times"), the interaction term is statistically insignificant.

Finally, we examined the possibility that capital could have non-linear effects on BHC loan growth by adding quadratic and/or cubic terms of the different capital measures to our loan growth regressions. In every case we found that none of these terms was statistically significant. For the capital surplus/shortfall regressions we also considered differences in the impact of capital deviations on loan growth between BHCs operating with surpluses and those operating with shortfalls.[24] Here we found no statistically

[24] We did this in two ways (both yielded the same conclusion). The first involved splitting the sample according to the sign of the capital deviation and then reestimating equation (3) for each sub-group. The second involved constructing a dummy variable that depends on whether a shortfall or a surplus is present; this dummy variable was interacted with the $Z_{i,t}$ term in equation (3), thus allowing for different values of ϕ depending on whether $Z_{i,t}$ represents a surplus or shortfall.

significant difference between the separate surplus and shortfall coefficients.

3.6 Potential Effect of Omitting Firm-Level Variables

While our combined methodology is able to address most forms of endogeneity and survivorship bias, it cannot deal with biases that might arise if bank and firm matches are not randomly formed. Addressing this bias requires bank lending and firm borrowing to be considered jointly, which in turn implies that individual loans must be modeled rather than aggregate lending by banks (or aggregate borrowing by firms). A number of recent papers pursue this approach using loan-level data obtained from national credit registers; these include Jiménez et al. (2010), who use data collected by the Bank of Spain, and Albertazzi and Marchetti (2010), who use data collected by the Bank of Italy. These studies, which control for firm-level characteristics, find larger effects of low bank capitalization and scarce liquidity on credit supply.

Matching bank loan information to borrower characteristics has a number of advantages in helping to identify causal effects since it allows the cross-sectional heterogeneity of both banks and firms to be exploited. However, we believe that for considering the linkage between financial conditions and real activity, this approach is only informative in practice if the matched sample is large and representative of the universe of lenders and borrowers. In the United States, matched-sample analysis has only been applied to data on syndicated loans—specifically, bank loan-level data from the Loan Pricing Corporation's (LPC) Dealscan database, which can be matched with BHC data from the FR Y-9C form and data on firms from Compustat. In addition to representing a fairly narrow class of lending, these matched data have a number of problems such as important missing information and outliers, which may then further restrict (in some cases significantly) the number of available firm-bank-loan observations. That said, studies that use these firm and bank loan-level data also seem to suggest relatively small effects of bank capital on lending. For example, Santos and Winton (2010) find that a 1 percent change in a bank's capital-to-assets ratio results in only a 3-basis-point increase in loan rates, thus implying a relatively modest effect.

This is consistent with our finding of a small effect of changes in bank capital on lending behavior.

4. Bank Capital and Loan Growth at the Aggregate Level

As noted earlier, our panel estimation approach only uses BHCs that have existed over our entire sample period (or at least most of it). It could therefore potentially suffer from survivor bias, which would tend to lower the estimated effect of bank capital on loans.[25] To address this concern, in this section we use aggregate data and a vector autoregression (VAR) model to study the effects of commercial bank capital ratios on bank loan growth. In our panel regressions, we specified BHC loan growth equations that controlled for the effects of other supply and demand variables and so allowed us to interpret the coefficient ψ on the equations' capital terms as the effect of BHC capital positions on loan growth. Here we take a different approach; specifically, we attempt to identify structural innovations in bank capital ratios and then examine the dynamic impact of these structural innovations on bank loan growth.

An important determinant of the influence that structural innovations to bank capital ratios will have on bank loan growth is the effect that these innovations have on other variables in the model—such as real economic growth and bank lending standards—that in turn feed back to lending. These indirect channels need to be borne in mind in considering how changes in bank capital affect loans; in particular, they must be accounted for in comparing the magnitudes of the effects found in the time-series analysis with those obtained from the panel regressions. We therefore conduct additional exercises where we shut down these indirect effects.

4.1 Estimation and Identification of the VAR Model

The vector autoregression model that we use is a slightly modified version of the extended VAR model with bank capital considered by

[25]This bias could result if negative capital shocks cause some BHCs to leave the sample. Note that such survivor bias was of limited concern for both Bernanke and Lown (1991) and Hancock and Wilcox (1993, 1994), since these studies employed cross-sectional data over a single year.

Lown and Morgan (2006). Our VAR consists of six variables: real GDP growth and GDP price inflation (excluding food and energy), the federal funds rate, commercial bank and thrift loan growth, the aggregate capital-to-assets ratio of the commercial bank sector, and the net fraction of loan officers at commercial banks reporting a tightening of credit standards for C&I loans. The first three variables in the VAR are standard components of a monetary policy VAR, while the last three variables—commercial bank and thrift loan growth, the capital-to-assets ratio, and lending standards—are added to the model to allow us to study the interaction between banks and the macroeconomy. As in our panel regressions, we use the estimation period 1990:Q3 to 2008:Q3 for our VARs; that is, we end the VAR analysis just when the crisis intensified and just before TARP capital injections began to occur.

The residuals obtained from each of the estimated OLS regressions in the VAR system are combinations of underlying structural innovations. Importantly, the structural innovations have economic interpretations, such as exogenous shocks to monetary policy, technology, and the markup of prices over costs in the monetary policy portion of the model and exogenous shocks to bank lending standards (reflecting changes in cautiousness), bank capital ratios, and bank loan volumes (from changes in conditions in other credit markets) in the banking portion of the model. However, although structural innovations underlie the residuals from the estimated equation, they cannot, in the absence of any additional assumptions, be observed or inferred. A wide variety of assumptions can be used to identify structural innovations based on the estimated equation residuals; for our VAR we use a recursive identification scheme with the causal order given by the ordering of variables described above. We are primarily interested in separately identifying the three banking shocks in our model; for our purposes, it is sufficient to amalgamate the macroeconomic shocks into a single composite innovation.

Our identification procedure assumes that structural innovations to any of the model's banking variables do not contemporaneously affect the model's macro variables, but that innovations to the model's macroeconomic variables have a contemporaneous impact on the model's banking variables. One interpretation for this assumption as it relates to economic activity and inflation is that households, firms, and all levels of government can arrange their spending, production, and price-determination plans for the *current* period in such a way that surprise developments in the credit sector will not prevent these plans from taking place (at least in the aggregate). That said, such plans can be modified in response to surprise developments that are directly relevant for spending, production, and price determination (the macroeconomic shocks); moreover, these macro developments *will* contemporaneously affect variables in the banking sector. Likewise, monetary policymakers do not alter their interest-rate-setting decisions in response to within-quarter developments in the banking sector, although monetary policy surprises can have immediate effects on banking-sector variables.

Within the banking portion of the VAR, our identification procedure assumes that structural innovations to loan volumes can affect bank capital ratios and lending standards immediately but that innovations to these latter variables do not have contemporaneous effects on loans. The first part of this assumption suggests that contemporaneous structural innovations to bank loan growth originate from the non-banking part of the credit sector (as well as the real economy). Thus, a reasonable interpretation of these innovations (particularly in the context of the recent crisis) is that they reflect disruptions to segments of the financial system that can in turn impact bank loan volumes contemporaneously.[26] The second part of the assumption—that capital ratios and lending standards do not affect lending contemporaneously—reflects the fact that it takes time to arrange a loan. This may at first appear to be at odds with our assumption that changes in the real economy can affect lending contemporaneously; however, for that case we have in mind households and firms drawing down or repaying pre-arranged lines of credit. Clearly, credit limits on loans that take the form of "on demand" credit lines (such as HELOCs, credit cards, or C&I commitment loans) can eventually be modified by banks in response to

[26]Examples include developments in commercial paper (CP) and bond markets. In the recent crisis and its aftermath, developments in these securities markets have affected bank loan growth in important ways. For example, early in the crisis when CP markets were coming under strain, CP issuers drew down backup lines of credit at banks and this had a sizable positive impact on bank loan growth. More recently, improving conditions in bond markets appear to be having some negative effect on bank loan growth as firms that can issue bonds substitute away from intermediated credit.

capital ratio or standards innovations, but these changes are difficult to implement quickly.

We would note that it is reasonable to assume that changes in lending standards have no contemporaneous impact on loan volumes inasmuch as there are many ways through which banks can alter lending standards other than contemporaneous changes in loan volumes; for example, banks can change the maturity of loans, the cost of credit lines, the spreads on loan rates, any covenants on the loan, and the loans' collateralization requirements. Our identifying assumption therefore implies that banks only make these changes in the period in which standards are altered.

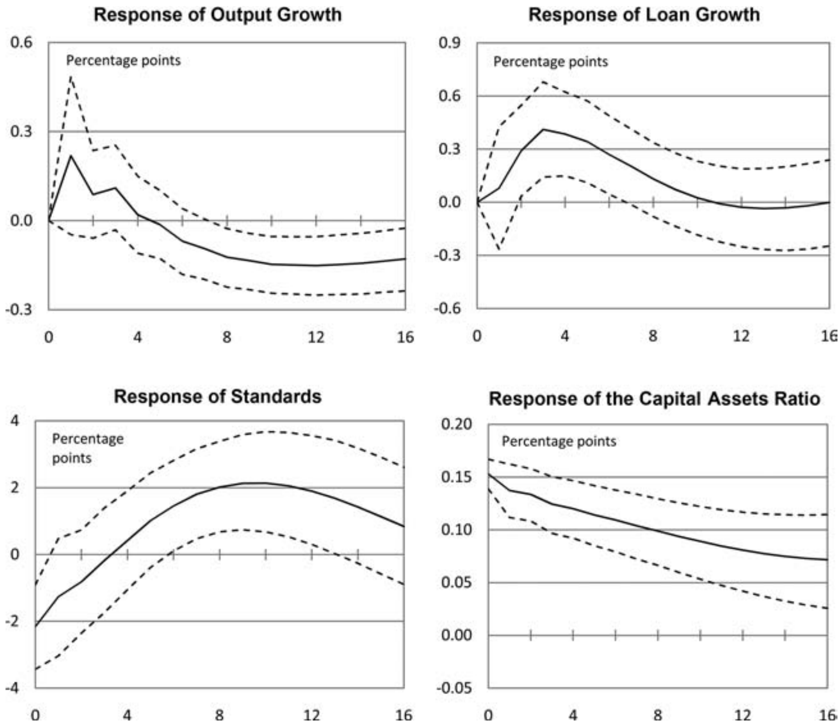
Our identification method also assumes that innovations to lending standards do not affect bank capital ratios contemporaneously, although bank capital ratios do affect standards. The first part of this assumption makes sense given that we assume that innovations to standards have no contemporaneous effect on loan volumes, while the second part of the assumption reflects the fact that current lending standards are the most easily altered variable in the VAR.

Finally, note that we only separately identify structural innovations to loan volumes, bank capital, and lending standards, but do not identify the model's macroeconomic shocks separately. Thus, in our results we only report the *combined* effects of the model's macroeconomic shocks.

4.2 The Effects of an Innovation to the Equity Capital Ratio

Figure 6 shows the responses of key model variables—specifically, output growth, loan growth, the capital-to-assets ratio, and standards—to a one-standard-deviation structural innovation to the capital-to-assets ratio. This innovation implies an immediate 16-basis-point increase in the capital-to-assets ratio, which then declines gradually back to its long-run level. Lending standards immediately loosen by about 2 percentage points in response to the increase in capital ratios, and loan growth is boosted for about two years with a maximum response of about 0.5 percentage point two quarters after the impact of the shock. The (statistically insignificant) response of GDP growth is rapid and short lived (about one year), peaking at 0.2 percentage point in the period immediately following the shock. One possible reason for the smaller and shorter-lived response of

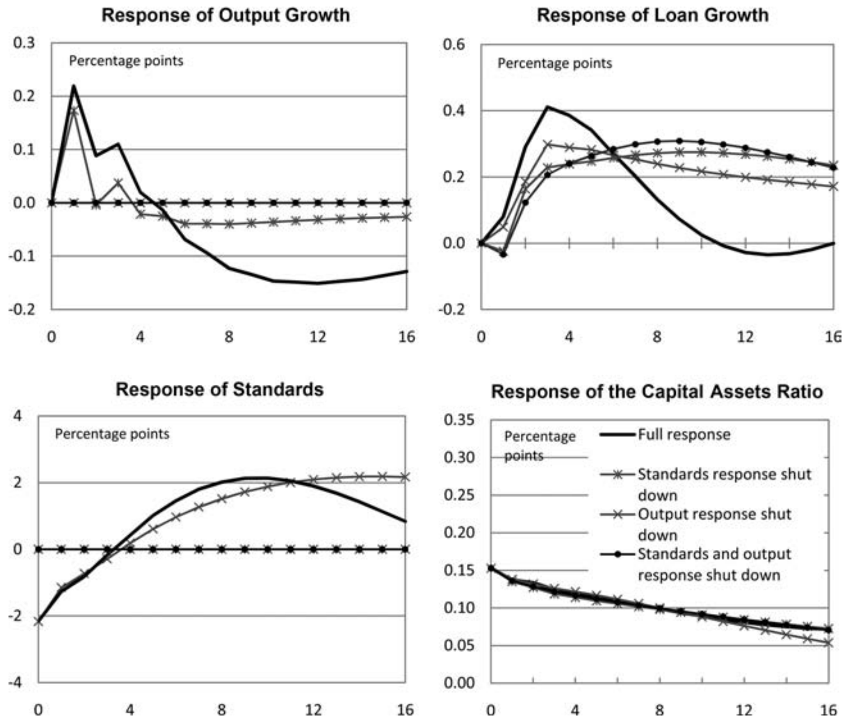
Figure 6. Response to a Capital-to-Assets Ratio Shock



output growth relative to loan growth is that some of the growth in loans may just represent substitution across institutions. Indeed, additional analysis (not shown) that uses the sum of bank and thrift loans *and* asset-backed securities issuers' holdings of mortgage and consumer credit in the VAR in place of bank and thrift loans seems to support this view.

Our results imply that a 1-percentage-point increase in the capital-to-assets ratio results in a 2.75-percentage-point increase in loan growth that peaks about a year after the initial shock. This is almost four times as large as the corresponding response from the panel regressions. The relatively larger response implied by the VAR model is perhaps not surprising given that the VAR approach allows for the endogenous response of several variables—most notably, standards and real GDP growth—that in turn have additional effects on

Figure 7. Response to a Capital-to-Assets Ratio Shock (Endogenous Responses Shut Down)



loan growth. Given this, it is instructive to shut down these additional effects in the VAR so as to permit a more direct comparison with the results obtained using institution-level BHC data.[27]

Figure 7 reports the responses to a one-standard-deviation capital-to-assets ratio shock that obtain when various endogenous responses of other model variables are eliminated. In particular, the light gray starred line shuts down the response of standards to the capital-to-assets ratio shock, the light gray crossed line shuts

[27]We shut down the different endogenous responses of standards and real GDP growth—both individually and in combination—using the approach followed by Bernanke, Gertler, and Watson (1997). Specifically, in addition to subjecting the model to a one-standard-deviation capital ratio shock, we also apply a sequence of other shocks to the model so as to hold the variable or variables in question constant over the period of the impulse response simulations.

down the response of real GDP growth to the capital-to-assets ratio shocks, and the dark gray dotted line shuts down both the standards and real GDP growth responses. These results suggest that in the absence of endogenous responses from other variables, bank loan growth should increase approximately 1.2 percentage points in response to a 1-percentage-point capital-to-assets ratio shock. This increase is about twice as large as the increase in loan growth implied by the panel regressions, but is still considerably smaller than what would be predicted under the constant leverage assumption that is implied by the scatter plot approach.

5. The Effects of Capital Injections on Loan Growth

We now use the models developed in sections 3 and 4 to estimate the likely effects of TARP-related capital injections on bank loan growth.

5.1 The Effects of Recent Capital Injections on BHC Loan Growth

The panel regression results from section 3 can be used to estimate the effect of the original TARP disbursement ($182 billion for our sample of large banks) on BHC loan volumes. The CPP capital infusions during 2008:Q4 raised equity capital, yielding an increase in the capital surplus term of about 14 percentage points on average. Taken together with the estimation results of section 3.2, which indicate that a 1-percentage-point increase in the capital surplus implies a 0.25-percentage-point increase in annualized loan growth, this implies that the CPP injections increased loan growth by 3.5 percentage points at an annual rate in our sample of banks. In ratio terms, the CPP injections raised the equity-to-assets ratio by about 1.4 percentage points and the risk-based capital ratio by about 1.7 percentage points on average. Given the estimation results of section 3.3, this implies that the CPP injections boosted BHC loan growth by 1.0 to 1.3 percentage points.

In dollar terms, the results from the surplus/shortfall regressions imply that the $182 billion increase in capital that large BHCs in our sample received would have resulted in a $273 billion increase in these BHCs' loans over the following year. The results from the capital ratio regressions imply an increase in loan volumes for large BHCs of between \$76 billion and \$100 billion over the following year (depending on the measure of capital considered). These volumes are significantly less than the boost in lending capacity that would be implied by the U.S. Treasury's assumptions. Specifically, since the Treasury assumed that a \$1 increase in capital would generate between \$8 and \$12 of lending, a \$182 billion capital injection should have resulted in a \$1.5 trillion to \$2.25 trillion increase in lending.

5.2 The Effects of Recent Capital Injections on Commercial Bank Loan Growth

We can also use the VAR model's impulse response functions to estimate the magnitude of the impact of CPP injections on commercial bank loan volumes. As noted above, capital injections under the CPP raised the equity-to-assets ratio by 1.4 percentage points. The impulse response functions reported in figure 6 imply that the CPP capital injections would boost loan growth by 3.7 percentage points over the first year. Cumulating the effect on loan growth implies a 5 percent increase in the level of loans, or a \$300 billion increase in commercial bank lending.

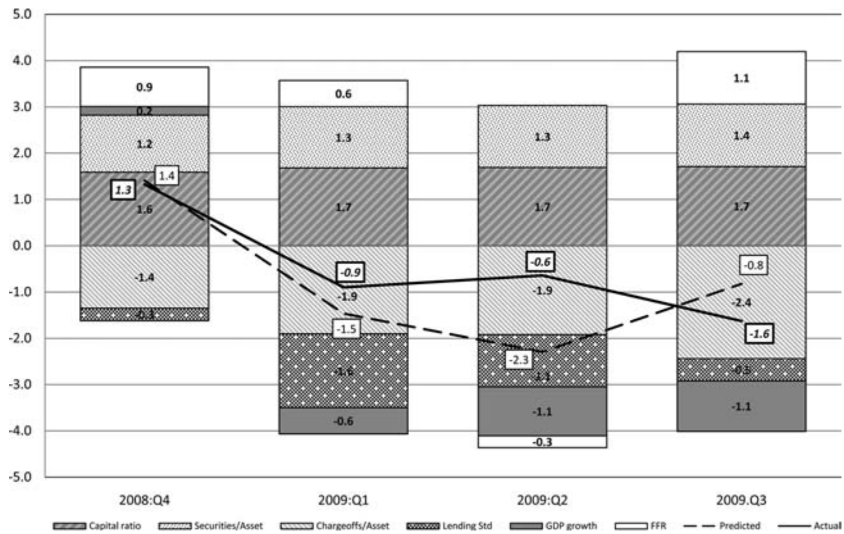
6. Understanding the Recent Decline in Loan Growth

We can also use these models to ascertain the effect of recent changes in capital ratios on loan growth. First, we use results from the panel regressions to understand the causes of the recent decline in bank loan growth and to gauge whether the large decline in lending is mainly explained by supply or demand factors. We then use our VAR estimates to decompose changes in commercial bank lending across the various banking and non-financial shocks that we estimate.

6.1 Decomposition of BHC Loan Growth

We conduct out-of-sample forecasts of BHC loan growth using both our capital surplus/shortfall and capital ratio panel regressions and decompose the quarterly growth rate of BHC loans into supply and demand factors such as capital, lending standards, risk, and GDP growth. Figure 8 reports the results from the capital ratio panel regressions.

Figure 8. Decomposition of BHC Quarterly Loan Growth



As can be seen in figure 8, our capital ratio panel regression model performs relatively well in explaining the recent drop in the quarterly growth rate of BHC loans. The decomposition of the predicted growth rate confirms the importance of loan demand shocks and increased risk in accounting for most of the decline in BHC loan growth. Out of the 2.9-percentage-point decline in the quarterly loan growth rate, from 1.3 percent in 2008:Q4 to –1.6 percent in 2009:Q3—of which the model explains 2.2 percentage points—1.3 percentage points is accounted for by changes in GDP growth (from 0.2 percent to –1.1 percent) and 1 percentage point is explained by changes in net charge-off rates (from –1.4 percent to –2.4 percent), which captures the deterioration in loan quality. Tighter lending standards significantly restrained BHC loan growth during the first quarter of 2009 (by about 1.3 percentage points), though this effect lessened in 2009:Q3. On net, then, lending standards account for only 0.2 percentage point of the 2008:Q4 to 2009:Q3 reduction in BHC loan growth (from –0.3 percent to –0.5 percent).

Figure 8 also confirms that changes in bank capital had only a small impact on BHC loan growth. In particular, the changes in BHC capital ratios associated with both TARP-related capital injections in 2008:Q4 and subsequent efforts to raise capital privately during the first half of 2009 made only a small positive contribution to bank loan growth (of only 0.1 percentage point, from 1.6 percent to 1.7 percent) and were more than offset by adverse factors such as reduced loan demand, increased risk, and somewhat tighter lending standards.

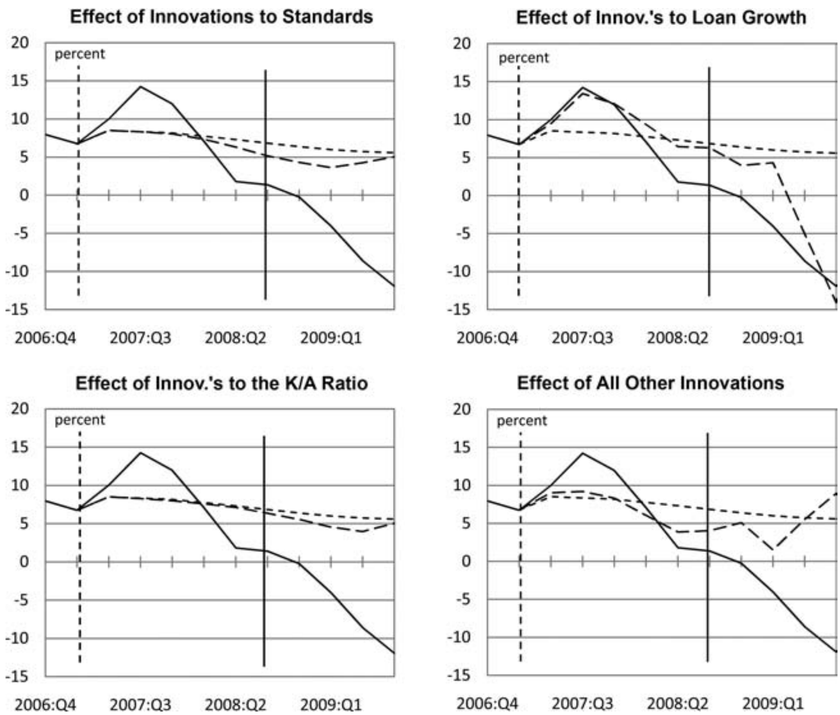
The results from the capital surplus/shortfall model (not shown) indicate that this model does not perform as well as the capital ratio model in terms of predicting the 2008:Q4 to 2009:Q3 decline in BHC loan growth. In particular, the surplus/shortfall model explains only 1.3 percentage points of the 2.9-percentage-point decline in loan growth over this period. That said, the conclusions obtained from the capital ratio model with respect to the relative importance of real GDP growth, net charge-off rates, and capital in explaining BHC loan growth continue to hold.[28]

6.2 Decomposition of Commercial Bank Loan Growth

The VAR model's identification procedure allows us to decompose movements in each variable into the portion attributable to the various shocks. This is done in figure 9. In each panel of the chart we show the actual path of loan growth over the past three years—the solid black line—as well as the model's unconditional forecast—the dotted line. We also show the path of loan growth that would be implied by the unconditional forecast and a specific set of structural innovations. We find that structural innovations to capital ratios explain very little of the below-par pace of loan growth over the

[28]It is interesting to note that even though our parameter estimates for the effect of a capital surplus or shortfall on BHC loan growth are similar to those in Hancock and Wilcox (1993, 1994), the size of deviations of capital from target are much smaller now than they were in the early 1990s. As a result, for a fairly similar contraction in bank loan growth in the 1990–91 and the most recent recessions, we find a more important role in the recent episode for loan demand conditions—proxied by the GDP growth rate—than for shocks to bank capital. This is not surprising given that the contraction in economic activity has been at least three times larger during the most recent recession than in the 1990–91 recession. Hence, despite similar parameter estimates, this difference leads us to conclude that the contribution of bank capital shortfalls to changes in loan growth is smaller than the Hancock-Wilcox findings.

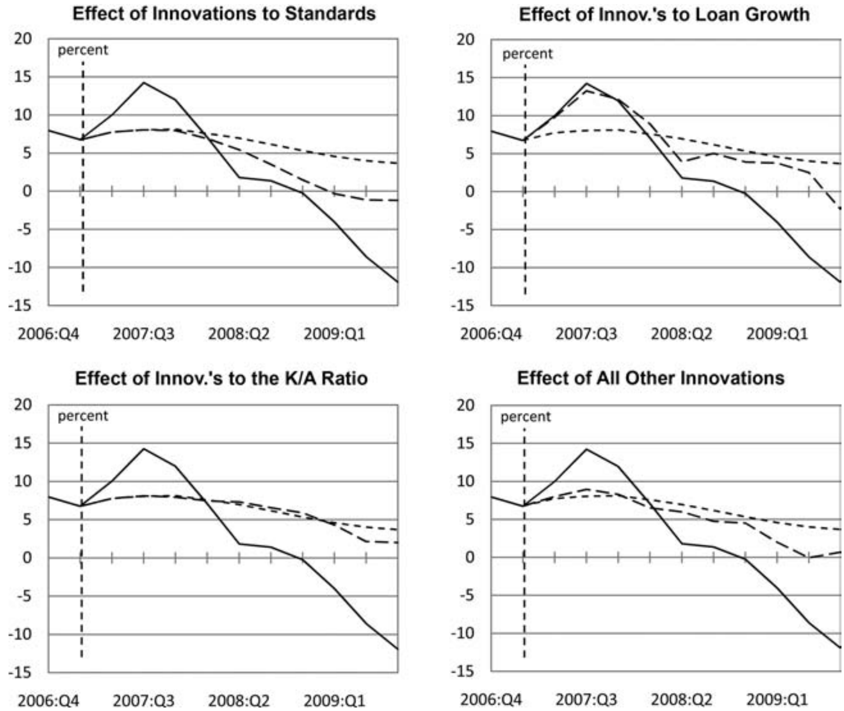
Figure 9. Decomposition of Loan Growth (Model Estimated through 2008:Q3)



past year (see the lower-left panel). Prior to the rebound in economic growth following the sharp contractions in activity in 2008:Q4 and 2009:Q1, macro innovations appear to have accounted for much of the decline in loan growth (the lower-right panel), with standards innovations accounting for a portion as well (the upper-left panel). Finally, over the last two quarters of the simulation period, own-variable innovations to loan growth account for the bulk of the series' decline (see the upper-right panel).

To some extent the large recent contribution of the own-variable loan growth innovations reflects the out-of-sample nature of our simulation exercise. The same exercise performed with a model estimated through to 2009:Q3—see figure 10—attributes much less of the recent decline in loan growth to own-variable shocks and much more to standards and macro shocks. Note, however, that capital

Figure 10. Decomposition of Loan Growth (Model Estimated through 2009:Q3)



ratio shocks continue to explain only a small portion of the decline in loan growth.

7. Summing Up

This paper has applied a number of different methods to examine how bank capital influences the extension of bank credit. An analysis based on panel data finds modest effects of BHC capital-to-asset ratios on BHC lending; using macroeconomic time series and aggregate commercial bank balance sheet data, we find larger—but still modest—effects of capital ratio shocks on loan growth.

These results stand in marked contrast to the constant leverage view, which has been quite influential of late in shaping forecasters' and policymakers' views regarding the size of the effect that bank capital changes have on loan growth. In particular, this view predicts effects of bank capital ratio changes on loan growth (that are six to ten times) larger than what we found in either our panel or time-series estimation results. We attribute the difference to the fact that the simple correlation between asset growth and leverage growth for commercial banks that underpins the constant leverage view is not very robust across time and especially fails to hold for the post-Basel sample period over which our analysis is conducted.

While our empirical results suggest relatively modest effects of both capital shortfalls and capital ratios on loan growth, we find more important roles for other factors such as economic activity and increased perception of riskiness by banks. One interpretation of this result is that banks and BHCs give relatively little consideration to their capital position when deciding whether to lend and instead allow other factors such as loan demand and risk to guide their decisionmaking. If this is the case, it would provide one explanation for the slowdown in loan growth that took place over the first three quarters of 2009 despite the sizable capital injections that occurred in 2008:Q4 under the CPP. However, another interpretation of our results is that existing measures of bank capital do not capture the "true" capital position of banks that is relevant for determining bank behavior. If true, this interpretation calls into question whether existing risk-based regulatory capital measures are valid, and would suggest that such measures need to be improved.

References

Adrian, T., and H. S. Shin. 2007. "Liquidity and Leverage." Mimeo.

Albertazzi, U., and D. Marchetti. 2010. "Credit Supply, Flight to Quality and Evergreening: An Analysis of Bank-Firm Relationships after Lehman." Working Paper No. 756, Bank of Italy.

Altunbas, Y., L. Gambacorta, and D. Marques-Ibanez. 2009. "Securitisation and the Bank Lending Channel." *European Economic Review* 53 (8): 996–1009.

Ashcraft, A. 2004. "Are Bank Holding Companies a Source of Strength to Their Banking Subsidiaries?" Federal Reserve Bank of New York Staff Report No. 189.

Berger, A., and C. H. S. Bouwman. 2009. "Bank Capital, Survival, and Performance around Financial Crises." Working Paper.

Berger, A., R. DeYoung, M. Flannery, D. Lee, and O. Öztekin. 2008. "How Do Large Banking Organizations Manage Their Capital Ratios?" *Journal of Financial Services Research* 34 (2): 123–49.

Berger, A., and G. Udell. 1994. "Did Risk-Based Capital Allocate Bank Credit and Cause a 'Credit Crunch' in the United States?" *Journal of Money, Credit, and Banking* 26 (3): 585–628.

Bernanke, B., M. Gertler, and M. Watson. 1997. "Systematic Monetary Policy and the Effects of Oil Price Shocks." *Brookings Papers on Economic Activity* 1: 91–116.

Bernanke, B., and C. Lown. 1991. "The Credit Crunch." *Brookings Papers on Economic Activity* 2: 205–47.

Ciccarelli, M., A. Maddaloni, and J.-L. Peydró-Alcalde. 2010. "Trusting the Bankers: Another Look at the Credit Channel of Monetary Policy." Mimeo.

De Nicolò, G., and M. Lucchetta. 2010. "Systemic Risks and the Macroeconomy." IMF Working Paper No. 29.

Elliott, D. 2010. "Quantifying the Effects on Lending of Increased Capital Requirements." Briefing Paper No. 7, Brookings Institution.

Flannery, M., and K. Rangan. 2008. "What Caused the Bank Capital Build-up of the 1990s?" *Review of Finance* 12 (2): 391–429.

Francis, W., and M. Osborne. 2009. "Bank Regulation, Capital and Credit Supply: Measuring the Impact of Prudential Standards." Occasional Paper No. 36, UK Financial Services Authority.

Giannetti, M., and A. Simonov. 2010. "On the Real Effects of Bank Bailouts: Micro-Evidence from Japan." CEPR Discussion Paper No. 7441.

Hancock, D., and J. Wilcox. 1993. "Has There Been a 'Capital Crunch' in Banking? The Effects on Bank Lending of Real Estate Market Conditions and Bank Capital Shortfalls." *Journal of Housing Economics* 3 (1): 31–50.

———. 1994. "Bank Capital and Credit Crunch: The Roles of Risk-Weighted and Unweighted Capital Regulations." *Journal of the American Real Estate and Urban Economics Association* 22 (1): 59–94.

Hatzius, J. 2007. "Leveraged Losses: Why Mortgage Defaults Matter." *Goldman Sachs U.S. Daily*, November 15, 2007.

———. 2008. "Beyond Leveraged Losses: The Balance Sheet Effects of the Home Price Downturn." *Brookings Papers on Economic Activity* 2: 195–227.

Houston, J., C. James, and D. Marcus. 1997. "Capital Market Frictions and the Role of Internal Capital Markets in Banking." *Journal of Financial Economics* 46: 135–64.

Jiménez, G., S. Ongena, J. Peydró-Alcalde, and J. Saurina. 2010. "Credit Supply: Identifying Balance-Sheet Channels with Loan Applications and Granted Loans." CEPR Discussion Paper No. 7655.

Judson, R., and A. Owen. 1999. "Estimating Dynamic Panel Data Models: A Practical Guide for Macroeconomists." *Economics Letters* 65 (1): 9–15.

Kashyap, A., and J. Stein. 1995. "The Impact of Monetary Policy on Bank Balance Sheets." *Carnegie-Rochester Conference Series on Public Policy* 42 (1): 151–95.

———. 2000. "What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?" *American Economic Review* 90 (3): 407–28.

Loutskina, E., and P. Strahan. 2006. "Securitization and the Declining Impact of Bank Finance on Loan Supply: Evidence from Mortgage Acceptance Rates." NBER Working Paper No. 11983.

Lown, C., and D. Morgan. 2006. "The Credit Cycle and the Business Cycle: New Findings Using the Loan Officer Opinion Survey." *Journal of Money, Credit, and Banking* 38 (6): 1575–97.

Peek, J., and E. Rosengren. 1995. "The Capital Crunch: Neither a Borrower Nor a Lender Be." *Journal of Money, Credit, and Banking* 27 (3): 625–38.

———. 1997. "The International Transmission of Financial Shocks: The Case of Japan." *American Economic Review* 87: 495–505.

Puri, M., J. Rocholl, and S. Steffen. Forthcoming. "Global Retail Lending in the Aftermath of the U.S. Financial Crisis: Distinguishing between Supply and Demand Effects." *Journal of Financial Economics.*

Santos, J., and A. Winton. 2010. "Bank Capital, Borrower Power, and Loan Rates." Working Paper.

Discussion of "The Effects of Bank Capital on Lending: What Do We Know, and What Does It Mean?"*

José-Luis Peydró
European Central Bank

1. Introduction

The events of the last few years suggest that the financial sector—the banking sector in particular—is a crucial determinant of business cycle fluctuations. The worst financial crisis in Europe and in the United States since World War II was followed by a severe economic recession. The key channel through which banks affect the economy is the provision of credit to fund private investment and consumption. In a crisis, a credit reduction may be the result itself of the grim economic outlook, via weaker credit demand and net worth of firms and households. However, credit may also decline from a reduction of credit supply due to banks' capital and liquidity problems.

The banking sector is not only at the center stage for controversy by being too stingy with credit on the downside of the business cycle, but banks are also accused of excessive credit creation and asset bubbles—excessive risk taking—on the upside of the cycle, both with important consequences for economic growth and financial stability.

The primary suspect in both scenarios is the credit supply channel; i.e., credit growth (or lack thereof) is being dictated by malfunctions in the credit supply process rather than economic fundamentals. In the downside of the cycle, in 2008, it was such a fear that forced central banks and governments around the world to intervene with hundreds of billions of euros in the banking industry. And it

*The views expressed are my own and do not necessarily reflect those of the European Central Bank or the Eurosystem. E-mail: jose.peydro@gmail.com or jose-luis.peydro-alcalde@ecb.europa.eu.

is also such a fear that has forced central banks to pay significantly more attention to macroprudential supervision and regulation when also we return to the upside of the cycle.

A common culprit of malfunctions in the credit supply process is the low level of capital by banks. On the upside, low levels of capital may cause excessive credit creation since banks' shareholders may have little capital at stake to lose in the event of bank losses and, moreover, they do not fully internalize all (potential) social losses.[1] On the downside, low level of capital may result in a credit crunch since the market may force banks to build up the capital ratio and also banks may hoard liquidity to use it in good investment opportunities arising from problems in other banks.[2]

What (and how important) are the effects of bank capital on bank lending? Despite the importance of the question, the empirical literature is thin. One reason is that in the macro literature there has not been much emphasis on the implications of bank capital and of credit supply in general.[3] Conversely, there has been considerable emphasis on the regulation of bank capital both as a key mechanism to limit significant risk taking by banks on the upside of the cycle and as a buffer for banks to continue lending in the downside, starting in Basel I in the 1980s to the very recent Basel III. In this sense, it should be noted that there has been recently approved in Basel a substantial increase of global minimum bank capital standards.[4]

Banks were instead opposed to the change in capital regulation since they argue that bank profits would decrease, forcing them in turn to reduce lending, especially to small and medium firms. The Institute of International Finance (2010), representing the world's big banks, issued a report concluding that phasing in an increase in bank capital requirements of 2 percent will lead to a drop of GDP

[1]For evidence of negative externalities in banking (contagion) and of higher social costs (than private) caused by bank failures, see Iyer and Peydró (2010). For systemic risk in banking, see Bandt, Hartmann, and Peydró (2009).

[2]See, among others, Holmstrom and Tirole (1997), Diamond and Rajan (2009), and Hanson, Kashyap and Stein (2010).

[3]See Adrian and Shin (2010) and Gertler and Kiyotaki (2010) for two recent exceptions. See also Freixas and Rochet (2008) and the references therein, where more emphasis is given to micro aspects.

[4]For the new bank rules, see www.bis.org/press/p100912.pdf?noframes=1.

of 3 percent in the United States, the euro area, and Japan. Conversely, banking regulators argue that the short-term costs of the new regulation are likely to be small and transitory, while the benefits of a stronger and healthier financial system will be around for years to come (the Basel Committee on Banking Supervision 2010 and Macroeconomic Assessment Group 2010). These studies assume different scenarios leading to very different real costs of an increase in bank capital. However, it is also very challenging to measure the relationship between bank capital and credit even when analyzing historical data.

The identification and measurement of the impact of bank capital on credit presents several key empirical challenges and, as I will argue later, both macro data and micro data at the bank (or firm) level (see, e.g., Berrospide and Edge 2010 or, more generally, papers that use bank-level regressions such as Kashyap and Stein 2000) are insufficiently precise to allow the identification of the causal impact of bank balance sheet shocks on economic activity through credit supply. The main reasons are that with these data sets one cannot distinguish between demand and supply of credit nor take into account the general equilibrium effects related to credit substitution. If this is true, one should be very cautious both in deriving policy implications (e.g., on bank bailouts) and in testing theories related to the credit channel from papers which only use aggregate- or bank-level data.

Berrospide and Edge (2010)—using bank-level data—find relatively small effects of bank capital on lending (they study the lending of large bank holding companies and use panel regression techniques following Bernanke and Lown 1991 and Hancock and Wilcox 1993, 1994). They also consider the effects of capital using a variant of Lown and Morgan's (2006) vector autoregression (VAR) model and, again, find a relatively modest impact of bank capital ratio changes on lending. Then, they use the estimated models to understand recent developments in bank lending and, in particular, to consider the role of TARP-related capital injections in affecting these developments.

My main comment is therefore that one should be very cautious in drawing policy implications regarding TARP derived from research using a data set based on bank-level and macro data since one cannot, in principle, identify the causal impact of bank capital on credit and the estimated coefficients may be biased upward or downward. Nonetheless, the paper is very interesting, rich in results, and useful for understanding the relationship between bank capital and lending. Moreover, it analyzes almost twenty years of data of the most important country in the world, the United States.

To tackle the different identification challenges, I will argue that one needs either exhaustive (firm-bank) loan-level data (see Jiménez, Mian, et al. 2010), even if possible with loan applications (Jiménez, Ongena, et al. 2010a, 2010b), or bank lending surveys from central banks, where decomposition between developments in credit demand and supply are given (see Ciccarelli, Maddaloni, and Peydró 2010). In the case of the United States, unfortunately, there is not an exhaustive credit register; however, there is a bank lending survey carried out by the Federal Reserve (see Lown and Morgan 2006).[5] Using the U.S. Senior Loan Officer Opinion Survey, Ciccarelli, Maddaloni, and Peydró (2010) also find a weak impact of bank capital on aggregate output during the recent crisis.

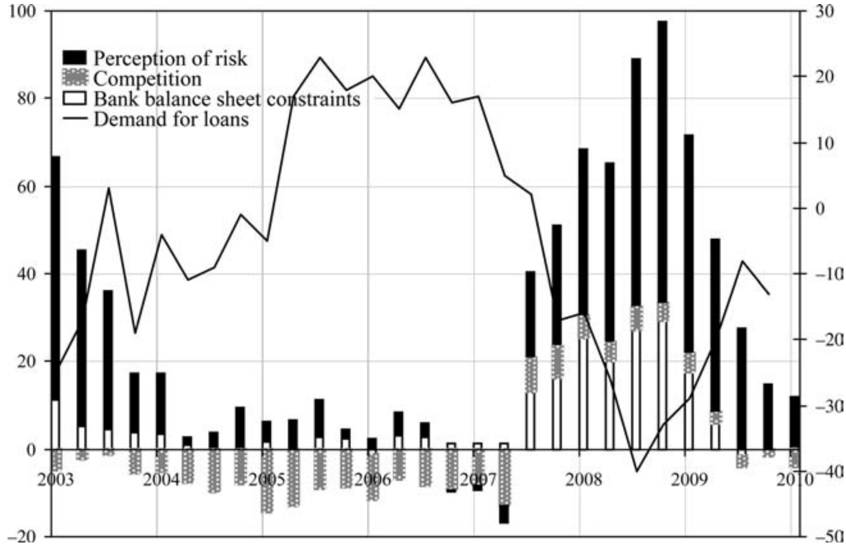
The rest of the discussion proceeds as follows. Section 2 addresses the main identification challenges of measuring the impact of bank capital on credit supply. Section 3 suggests an identification strategy that tackles the empirical challenges. Section 4 discusses other comments and suggestions. Section 5 concludes.

2. Identification Challenges

In the current crisis, banks started early on to tighten their lending standards both in the United States and in Europe. As figure 1 shows for the euro area, the tightening of lending standards started in 2007:Q3. It is important to note that at the same time that lending standards were tightened, credit demand was also reduced. Moreover, banks tightened their standards for loans due to bank balance sheet constraints (in capital and liquidity) but also because the borrowers were perceived riskier (with lower net worth, riskier collateral, and worse outlook). Hence, not only credit supply (due to

[5] In the United States there are loan-level data for large firms (syndicated loans). However, these data are not exhaustive even for the set of (large) firms, in turn making it difficult to separate credit demand from supply and to take into account the general equilibrium effects related to credit substitution.

Figure 1. Credit Demand and Supply Changes in the Euro Area



Note: This figure shows the lending standards (left axis) and credit demand (right axis) for corporate loans in the euro area from the Bank Lending Survey (BLS) based on a weighted average of the twelve (initial) euro-area countries as of 2004:Q4 (Austria, Belgium, France, Finland, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, and Spain). Lending standards for corporate loans are the net percentages of banks reporting a tightening of credit standards for loans to enterprises in the BLS. The darkest shaded part is the change in lending standards related to changes in economic conditions and borrower outlook and risk, the gray shaded part is the change in standards due to changes in bank competition, and the non-shaded (white) part is the change in standards due to changes in bank balance sheet strength (capital and liquidity). The black line is the change in loan demand from the BLS. See Maddaloni and Peydró (2010).

bank capital and liquidity problems) may have decreased but also credit demand and net worth of firms and households, which complicates any identification of bank capital implications on credit supply during the recent crisis.[6]

[6]Similar dynamics are observed in the United States (see www.federalreserve.gov/boarddocs/snloansurvey). The data for the euro area is from Maddaloni and Peydró (2010). For the sake of simplicity, I have plotted only the data related to corporate loans, but results are very similar for loans to households.

The main empirical challenges to identifying the causal impact of a bank capital shock on loan supply are the following. First, bank capital is endogenous to economic conditions; in particular, capital may be low in a crisis because investment opportunities in lending are worse.

Second, credit demand changes are correlated with credit supply changes (as figure 1 shows for the time series). Bank capital shocks are often correlated with changes in the economic environment: if, for example, economic outlook worsens, bank capital decreases, but firms' demand for credit also drops since there is a reduced need for investment. Therefore, it is highly likely that an economy experiences at the same time credit decline and lower demand and supply of credit. These initial two points suggest that by looking only at the time series it may be almost impossible to obtain the causal impact of capital on credit supply.

Third, changes in the quality (net worth and risk) of the borrower pool may also be correlated with changes in credit supply (see also figure 1 for the time series). In the time series—e.g., in worse times—firms are in principle riskier and have lower net worth, so banks face a lower-quality pool of borrowers overall. Hence credit may decline because of bank capital problems but also because of (borrower) firm and household capital and risk problems.

Furthermore, there can also be a correlation on the *cross-section*. Banks with lower capital may lend more on average to firms with worse risk or capital, which further complicates the analysis at the bank level.[7] For example, in crisis times, weaker firms may be more financially constrained and, hence, may need more bank finance (thus increasing credit demand). If these firms are therefore matched with banks with weaker capital, at the bank level one would find that banks with lower capital do not reduce lending, but this would be

[7]In theory, firm and bank balance sheet strengths could be correlated: the higher the agency problems between firms and banks due to the firms' moral hazard, the more fragile the banks will need to be (Diamond and Rajan 2000, 2001). Moreover, Peek and Rosengren (2005) and Caballero, Hoshi, and Kashyap (2008) document that, during the Japanese financial crisis, banks with capital ratios closer to the minimum binding levels lent more to zombie firms. Hence, the strength of the lending banks' balance sheets was positively correlated with those of the borrowing firms.

because these weaker banks are facing higher (countercyclical) credit demand, not because bank capital is not crucial.[8]

Fourth, there can be a substitution of the sources of credit, and banks with higher capital could seek finance firms who were borrowing from banks with lower capital (i.e., one needs to take into account the general equilibrium effects related to credit substitution). In this sense, a credit supply reduction in some banks would only translate into binding credit constraints if firms cannot compensate credit with other, less affected, banks. For economic activity, what matters most are the general equilibrium effects of lending, and these may be unaffected even if some banks are partially reducing lending. In consequence, an analysis at the bank level misses these general equilibrium effects since one cannot know whether the reduction in supply from one bank is compensated by another bank.

It should also be noted that it is increasingly recognized that documenting credit supply side failures at the bank level based on partial equilibrium analysis may not be relevant from a systemic risk perspective. What is strongly needed is a macroprudential approach that takes into account general equilibrium effects (Hanson, Kashyap, and Stein 2010). For example, a reduction in credit supply due to an adverse bank balance sheet shock may not have any negative impact if affected firms can go elsewhere to compensate for the loss in credit. Indeed, proponents of non-interventionist central banking argue that such general equilibrium effects are strong enough to let credit markets heal on their own. Unnecessary interventions, they argue, create more mischief by punctuating the virtuous cycle of creative destruction (see Jiménez, Mian, et al. 2010).

All in all, the above identification challenges suggest that obtaining the causal impact of bank capital on credit is tricky and that aggregate-level data and bank-level data are not enough to precisely estimate the elasticity of bank capital on credit. In this sense, Berrospide and Edge (2010) provide correlations rather than the causal coefficients and do not control for general equilibrium effects, at least in the first part of their analysis based on bank-level data. Moreover, the estimated coefficient could be in principle biased

[8]For similar reasons, an analysis at the firm level in the presence of firm-bank matching problems may yield biased estimates.

upward or downward depending on the covariance of the unobserved credit demand and supply shocks.

3. An Alternative, Potentially Better Way of Identification

Bernanke and Lown (1991) define a credit crunch as "a significant leftward shift in the supply curve for loans, holding constant both the safe real interest rate and the quality of potential borrowers." Given that borrowers may differ in quality both in observable and unobservable characteristics, a way to fully control for borrower quality and demand of credit is to analyze credit supply at the firm-bank level using firm fixed effects to fully capture the borrower demand and risk (Khwaja and Mian 2008). In this case, identification comes from comparing lending to the same firm from a bank with a positive shock vis-à-vis a bank with a negative shock.[9]

Jiménez, Ongena, et al. (2010a), to identify the credit supply effects of monetary policy, use firm*time fixed effects to control for unobserved time-varying heterogeneity in firm loan demand and risk. They use exhaustive loan-level data from Spain over twenty years. They find that bank-level regressions underestimate the bank lending channel (e.g., Kashyap and Stein 2000); hence there is a non-random matching problem between borrowers and lenders and, hence, regressions at the bank level are not well specified. Moreover, they analyze whether firms borrowing more from affected banks can substitute credit from other, less affected, banks, and find binding effects of bank capital. Moreover, they find that banks with higher capital lend more to existing clients than banks with lower capital; however, banks with lower capital tend to lend more on average to new clients (probably riskier customers for the bank). That is, the effect of capital on credit supply is totally different in the intensive vis-à-vis the extensive margin of lending.

There are still two problems using firm*time fixed effects with outstanding credit: First, one cannot perfectly analyze the extensive margin of lending to new clients (one cannot analyze changes in credit from different banks which were lending to the same firm

[9]In Khwaja and Mian (2008) the shock comes from a nuclear threat, which impacts bank liquidity differently across different banks.

since the firm is a new client); hence, with the above strategy of fixed effects and change in outstanding credit, it is only possible to fully identify the intensive margin of lending and extensive margin of terminated loans. Second, one cannot horse-race the non-financial borrower (e.g., firm) balance sheet channel vis-à-vis the bank balance sheet (lending) channel since the firm fixed effect fully captures the firm channel, but not adding the firm fixed effect implies not controlling for credit demand. Jiménez, Ongena, et al. (2010a), however, use loan applications for analyzing both the extensive margin and the firm vis-à-vis the bank balance sheet channels. In this case identification is possible, since the different loan applications from new clients are observed and firm fixed effects are less needed to control for credit demand (volume).

Credit composition is also crucial. Supply changes of composition of credit (bank risk taking) are analyzed by Jiménez, Ongena, et al. (2010b). Using the credit register from Spain, they analyze bank risk-taking effects of monetary policy. They use firm*time and bank*time fixed effects to account for both observed and unobserved time-varying firm and bank heterogeneity that proxy the firm demand, balance sheet, and bank lending channels. They find that a lower overnight interest rate induces lower capitalized banks to expand and prolong credit to riskier firms, and to lend to riskier new applicants, granting them loans that are larger and longer term. A lower long-term rate, however, has smaller or no such effects.

General equilibrium effects related to credit substitution are taken into account and formalized in Jiménez, Mian, et al. (2010). They formalize a new methodology for estimating the aggregate firm-level impact of the bank lending channel, and apply it to estimate the effect of securitization (capital and liquidity shock) on credit supply in Spain. At the bank level, they find that an increase in banks' ability to securitize real-estate assets leads to an increase in the quantity of credit extended to non-real-estate firms. However, their methodology shows that this effect is close to zero for most firms in general equilibrium due to a crowding out of existing bank credit. Nonetheless, securitization does lead to an expansion in credit for first-time clients that are significantly more likely to end up in default. Securitization also leads to a broader relaxation in credit terms, probably due to stiffer competition. Finally, while the collapse of the private securitization market in 2008 contracted credit from more securitization-dependent banks, aggregate firm-level impact of securitization-driven credit crunch is close to zero.

Firms can also substitute credit with other sources of finance—for example, with other debt claims such as trade credit and market debt. Iyer et al. (2010) analyze the impact of the interbank liquidity crunch on the credit crunch using credit register data from Portugal. They find that the bank liquidity shock of August 2007 implied a credit supply reduction from banks, and that firms could not substitute this bank credit reduction with other sources of credit.

The previous studies allow an identification of credit supply and can take into account general equilibrium effects related to credit substitution. However, difference-in-difference exercises cannot take fully all the general equilibrium effects. The micro identification cannot analyze the total effect of a bank capital shock on real activity, but only a difference-in-difference effect by comparing banks with different degrees of capital. Moreover, bank capital may affect economic activity through credit supply, and these changes in real activity may in turn affect bank capital. These cannot be fully captured by the difference-in-difference coefficients.

A way to control for these general equilibrium effects is provided by Ciccarelli, Maddaloni, and Peydró (2010). A key issue is how to obtain at the macro level time-varying information for both credit supply and demand since these are mostly unobserved. Bank lending surveys by central banks do, however, contain reliable quarterly information on credit supply and demand's quantity and quality, which serve to identify the credit demand and the firm, household, and bank balance sheet channels.

Ciccarelli, Maddaloni, and Peydró (2010) use the detailed answers of the confidential and unique Bank Lending Survey (BLS) for the euro area and of the Senior Loan Officer Opinion Survey (SLOOS) for the United States. Euro-area national central banks and regional Federal Reserve banks request from banks quarterly information on the lending standards that banks apply and on the loan demand that banks receive from firms and households. The detailed information reported in the surveys is very reliable, not least because the surveys are carried out by central banks, which are in most cases the bank supervisors and can cross-check the information received with exhaustive hard bank information.

They find that the credit channel amplifies a monetary policy shock on GDP and inflation, through the balance sheets of households, firms, and banks. For corporate loans, amplification is highest through credit supply; for households, demand is the strongest channel. Finally, in the euro area, a credit crunch for firms due to bank capital and liquidity problems reduced GDP significantly during the financial crisis. In the case of the United States, the authors do not find a strong effect of bank capital in the current crisis, consistent with the results of Berrospide and Edge (2010).

4. Other Comments and Suggestions

I would introduce, in Berrospide and Edge (2010), interactions between the business cycle (economic and monetary conditions) and bank capital to study whether the effect of bank capital on lending is stronger in bad vis-à-vis good times. For example, Jiménez, Ongena, et al. (2010a) find that conditioning on several loan applications of a firm in a month to new banks, banks with lower capital on average grant more loans in the extensive margin of lending to new clients, but in bad times the opposite occurs! Bernanke and Lown (1991) could also find stronger effects than Berrospide and Edge (2010) because, probably, Bernanke and Lown (1991) analyze the impact of bank capital on credit supply in bad times. Moreover, the effects in Adrian and Shin (2010) are different depending on the business cycle.

One may still find a small effect in loan volume in bad times due to loan evergreening (see, for example, evidence in Italy and Japan).[10] For example, because of fear of loan defaults, banks may renew their loans with their weakest borrowers. Hence, it is crucial to analyze the extensive vis-à-vis the intensive margin of lending, and also credit composition and not only credit volume.

I would also introduce time fixed effects at least once to truly analyze the cross-sectional implications of bank capital on lending (note that it is difficult to control for all the aggregate "demand factors" with observable variables such as current GDP, prices, etc.).

[10] See Albertazzi and Marchetti (2010) for evidence in Italy, and Peek and Rosengren (2005) and Caballero, Hoshi, and Kashyap (2008) for evidence in Japan.

For example, expectations on the business cycle also matter. I would also introduce bank fixed effects once to analyze the “within” results, i.e., how changes in capital in a bank affect its lending (important given the bank omitted variables).

Other suggestions are to control for bank total assets and to analyze the effect of capital if one does not put net charge-offs in the regressions, since the latter should be correlated with capital. In addition, it would be nice to control for the drawing down of committed loans and credit lines. Finally, a key issue is whether bank capital is well measured or not, especially given all the off-balance-sheet items and the bank capital regulatory arbitrage in the United States before the recent crisis (Acharya and Richardson 2010).

5. Summary

Berrospide and Edge (2010) is a very interesting paper, very rich in results and very useful for understanding the relationship between bank capital and lending. Moreover, they analyze almost twenty years of data of the most important country in the world—the United States. They find a weak relationship between total lending and bank capital.

My main discussion point is to argue that causality and identification of credit supply may not be perfectly achieved because of the data available. As I argued in the discussion, due to the data, Berrospide and Edge (2010) may underestimate or overestimate the effects of bank capital on credit;[11] therefore, one should be careful both in drawing policy implications and in rejecting theories based on the importance of credit supply and bank capital.

I argued instead that exhaustive credit registers are better to identify credit supply changes, to analyze general equilibrium effects related to credit substitution, and, hence, to test the theories of the credit channel, providing in turn policy implications (see e.g. Jiménez, Mian, et al. 2010). Finally, to take into account all the different general equilibrium effects, bank lending surveys carried out by central banks (in most cases bank supervisors) may be very

[11]The covariance between the credit demand and supply shocks may be very different even for the same country depending on the financial frictions one is analyzing (see Jiménez, Mian, et al. 2010 versus Jiménez, Ongena, et al. 2010a).

useful for analyzing the aggregate real effects of credit supply since surveys provide quarterly information on credit demand and supply developments (Ciccarelli, Maddaloni, and Peydró 2010).

References

Acharya, V. V., and M. Richardson. 2010. *Restoring Financial Stability: How to Repair a Failed System.* New York, NY: John Wiley & Sons.

Adrian, T., and H. S. Shin. 2010. "Financial Intermediaries and Monetary Economics." Forthcoming in *Handbook of Monetary Economics*, Vol. 3B, ed. B. M. Friedman and M. Woodford. New York, NY: Elsevier.

Albertazzi, U., and D. Marchetti. 2010. "Credit Supply, Flight to Quality and Evergreening: An Analysis of Bank-Firm Relationships after Lehman." Working Paper No. 756, Bank of Italy.

Bandt, O., P. Hartmann, and J.-L. Peydró. 2009. "Systemic Risk in Banking: An Update." In *The Oxford Handbook of Banking*, ed. A. N. Berger, P. Molyneux, and J. Wilson. Oxford University Press.

Basel Committee on Banking Supervision. 2010. "An Assessment of the Long-Term Economic Impact of the New Regulatory Framework." (August).

Bernanke, B., and C. Lown. 1991. "The Credit Crunch." *Brookings Papers on Economic Activity* 2: 205–47.

Berrospide, J. M., and R. M. Edge. 2010. "The Effects of Bank Capital on Lending: What Do We Know, and What Does It Mean?" *International Journal of Central Banking* 6 (4).

Caballero, R. J., T. Hoshi, and A. K. Kashyap. 2008. "Zombie Lending and Depressed Restructuring in Japan." *American Economic Review* 98 (5): 1943–77.

Ciccarelli, M., A. Maddaloni, and J.-L. Peydró. 2010. "Trusting the Bankers: A New Look at the Credit Channel of Monetary Policy." ECB Working Paper No. 1228.

Diamond, D. W., and R. G. Rajan. 2000. "A Theory of Bank Capital." *Journal of Finance* 55 (6): 2431–65.

———. 2001. "Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking." *Journal of Political Economy* 109 (2): 287–327.

———. 2009. "Fear of Fire Sales and the Credit Freeze." Booth School of Business.

Freixas, X., and J.-C. Rochet. 2008. *Microeconomics of Banking*. 2nd ed. Cambridge, MA: MIT Press.

Gertler, M., and N. Kiyotaki. 2010. "Financial Intermediation and Credit Policy in Business Cycle Analysis." Forthcoming in *Handbook of Monetary Economics*, Vol. 3B, ed. B. M. Friedman and M. Woodford. New York, NY: Elsevier.

Hancock, D., and J. Wilcox. 1993. "Has There Been a 'Capital Crunch' in Banking? The Effects on Bank Lending of Real Estate Market Conditions and Bank Capital Shortfalls." *Journal of Housing Economics* 3 (1): 31–50.

———. 1994. "Bank Capital and Credit Crunch: The Roles of Risk-Weighted and Unweighted Capital Regulations." *Journal of the American Real Estate and Urban Economics Association* 22 (1): 59–94.

Hanson, S., A Kashyap, and J. Stein. 2010. "A Macroprudential Approach to Financial Regulation." Forthcoming in *Journal of Economic Perspectives*.

Holmstrom, B., and J. Tirole. 1997. "Financial Intermediation, Loanable Funds, and the Real Sector." *Quarterly Journal of Economics* 112 (3): 663–91.

Institute of International Finance. 2010. "Interim Report on the Cumulative Impact on the Global Economy of Proposed Changes in the Banking Regulatory Framework." (June).

Iyer, R., S. Lopes, J.-L. Peydró, and A. Schoar. 2010. "The Interbank Liquidity Crunch and the Firm Credit Crunch: Evidence from the 2007–09 Crisis." Mimeo, MIT.

Iyer, R., and J.-L. Peydró. 2010. "Interbank Contagion at Work: Evidence from a Natural Experiment." Forthcoming in *Review of Financial Studies*.

Jiménez, G., A. Mian, J.-L. Peydró, and J. Saurina. 2010. "Estimating the Aggregate Impact of Credit Supply Channel: Evidence from Securitization in Spain." Available at http://ssrn.com/abstract=1674828.

Jiménez, G., S. Ongena, J.-L. Peydró, and J. Saurina. 2010a. "Credit Supply: Identifying Balance-Sheet Channels with Loan Applications and Granted Loans." CEPR Discussion Paper No. 7655.

———. 2010b. "Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say about the Effects of Monetary Policy on Credit Risk?" ECB Working Paper.

Kashyap, A., and J. Stein. 2000. "What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?" *American Economic Review* 90 (3): 407–28.

Khwaja, A. I., and A. Mian. 2008. "Tracing the Impact of Bank Liquidity Shocks." *American Economic Review* 98 (4): 1413–42.

Lown, C., and D. Morgan. 2006. "The Credit Cycle and the Business Cycle: New Findings Using the Loan Officer Opinion Survey." *Journal of Money, Credit, and Banking* 38 (6): 1575–97.

Macroeconomic Assessment Group of the Financial Stability Board and Basel Committee on Banking Supervision. 2010. "Assessing the Macroeconomic Impact of the Transition to Stronger Capital and Liquidity Requirements." Bank for International Settlements (August).

Maddaloni, A., and J.-L. Peydró. 2010. "Bank Risk Taking, Securitization, Supervision, and Low Interest Rates: Evidence from Euro Area and U.S. Lending Standards." ECB Working Paper.

Peek, J., and E. S. Rosengren. 2005. "Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan." *American Economic Review* 95 (4): 1144–66.

Banks' Financial Conditions and the Transmission of Monetary Policy: A FAVAR Approach*

Ramona Jimborean and Jean-Stéphane Mésonnier
Banque de France

We propose a novel approach to assess whether banks' financial conditions, as reflected by bank-level information, matter for the transmission of monetary policy, while reconciling the micro and macro levels of analysis. We include factors summarizing large sets of individual bank balance sheet ratios in a standard factor-augmented vector autoregression model (FAVAR) of the French economy. We first find that factors extracted from banks' liquidity and leverage ratios predict macroeconomic fluctuations. This suggests a potential scope for macroprudential policies aimed at dampening the procyclical effects of adjustments in banks' balance sheet structures. However, we also find that fluctuations in bank ratio factors are largely irrelevant for the transmission of monetary shocks. Thus, there is little point in monitoring the information

*We would like to thank Karim Barhoumi, Florin Bilbiie, Stefano Neri, Tao Zha, and an anonymous referee, as well as participants at the Second IJCB Financial Stability Conference on the Theory and Practice of Macroprudential Supervision hosted by the Banco de Espana (Madrid), the 2010 CEA Conference (Quebec), the 2010 Transmission Channel workshop of the Basel Committee Research Task Force (Paris), and seminars at the Banque de France and the European Central Bank, for valuable comments and suggestions. We are also grateful to Benoît Mojon and Dalibor Stevanovic for sharing their codes with us. Jocelyne Tanguy provided excellent research assistance. Last but not least, we thank our colleagues from the French Supervisory Authority for granting us access to the BAFI database and, in particular, Stéphane Chouard who coded the requested extraction numbers. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Banque de France. Corresponding author: Mésonnier. Address: Banque de France, Financial Economics Research Division, 41-1391 Recfin, 39 rue Croix des Petits-Champs, 75001 Paris. Author E-mails: Mésonnier: jean-stephane.mesonnier@banque-france.fr; Jimborean: ramona.jimborean@banque-france.fr.

contained in bank balance sheets, above the information already contained in credit aggregates, as far as monetary policy transmission is concerned.

JEL Codes: E44, E52, G21.

1. Introduction

The sub-prime crisis and the fears of a widespread credit crunch it has fueled over 2008–09 in most developed economies have highlighted the importance of sound financial conditions of banks for the ability of monetary policy rate cuts to effectively curb the contraction in credit supply to the economy. Over the last decade, the dominant view of the monetary policy transmission mechanism mainly pointed to the importance of the expectations channel of monetary policy, where monetary policy influences output and prices merely through the expected path of future short-term rates.[1] By contrast, the recent crisis helped dramatically to revive the complementary "credit" view, according to which banks' individual reactions to monetary policy decisions matter much for the overall level of activity. As in previous episodes of wide-ranging bank capital depletion, like in the United States in the early 1990s or in Japan later in that decade (cf. Bernanke and Lown 1991; Woo 2003), empirical assessments of the bank lending and bank capital channels have recently gained a heightened attention in both academic and policy circles (see, e.g., Adrian and Shin 2009c).

However, the practical relevance of the credit channel for monetary policy has been one of the most fiercely debated empirical issues in monetary policy for at least two decades.[2] Broadly speaking, empirical research has followed two main routes so far—one based on detailed individual bank information, the other based on measures of credit at the aggregate level—but both remain relatively inconclusive regarding the macroeconomic significance of financial frictions at the bank's level. We propose here a new approach that reconciles the use of both types of data: microeconomic bank data

[1] See, for instance, Blinder (1998), Bernanke (2004), and Woodford (2005).

[2] For a general perspective on the credit channel issue and the usual distinction between the so-called *bank lending channel* and the balance sheet channel of monetary policy transmission, see Bernanke and Gertler (1995). For a view of this debate at the euro-area level, see Angeloni, Kashyap, and Mojon (2003).

for a large population of French banks and a rich macroeconomic database for France. While the choice of this country is partly dictated by issues of data availability, testing the importance of banks' financial conditions for monetary policy transmission is of particular interest in the case of France because its financial system is (still) largely bank based rather than market based.

As hinted above, a first strand of the applied literature has endeavored to identify the role of bank heterogeneity and loan supply effects at the *micro* level, running panel data regressions on bank balance sheet data in order to investigate the determinants of individual credit fluctuations (see, for instance, Kashyap and Stein 1995; Ehrmann et al. 2001). These studies have highlighted the impact of several banks' characteristics, such as total assets' size, capitalization, and liquidity ratios, on a differentiated response of bank loans to monetary policy shocks. Typically, the traditional bank lending channel of monetary policy transmission appears then to be stronger for small, poorly capitalized, and/or less liquid credit institutions.[3] However, a limit to the policy relevance of this literature is that little can be inferred from the results of micro data studies about the macro consequences of bank balance sheet constraints. As argued by Ashcraft (2006), on the basis of such panel data regressions, one cannot tell whether the financial frictions at play in the bank lending channel and affecting, for instance, small banks really do account for a significant part of the dampening of real activity that follows a monetary policy tightening.

A second strand of the literature then relies on the estimation of small monetary vector autoregressions (VARs) at the *macro* level, following notably Bernanke and Blinder (1992).[4] Indeed, impulse response functions (IRFs) derived from simple structural VAR models that factor in a few macro variables (e.g., GDP, inflation, and a measure of the policy stance) provide a useful device for evaluating monetary policy transmission. By adding a credit aggregate variable to this basic framework, it should be easy to gauge the impact of

[3]See, for instance, Kishan and Opiela (2000), Gambacorta and Mistrulli (2004), and Engler et al. (2005) for the United States, Italy, and Austria, respectively, and Loupias, Savignac, and Sevestre (2002) for France.

[4]See also Ramey (1993), as well as Ehrmann and Worms (2004), Hülsewig, Winker, and Worms (2004), and Den Haan, Sumner, and Yamashiro (2007), for more recent contributions.

monetary policy shocks on total credit and the role of credit supply restrictions in economic downturns. In practice, however, things turn out to be trickier.

First, the estimated response of total bank loans to a monetary policy shock appears often to be muted and non-significant (Bernanke and Gertler 1995). A closer inspection of the dynamics of various aggregate bank credit series—contrasting, e.g., loans to households versus loans to non-financial firms, or short-term versus long-term loans—shows that this may result from a compensation effect of diverging responses of the main components in banks' loan portfolios (Den Haan, Sumner, and Yamashiro 2007; Mésonnier 2008). In turn, this hints that a small VAR including only one credit variable is probably misspecified. A solution to this misspecification problem could then be to add several aggregate loan series to the VAR; but, as is well known, the inclusion of additional variables in standard VARs is restricted by the degrees-of-freedom problems.[5] Second, the information basis contained in a standard VAR with a handful of macroeconomic and aggregate credit variables appears to be too narrow, so that a proper identification of credit supply effects remains out of reach. As a matter of fact, by using a simple VAR framework, it is generally not possible to tell whether credit contracts after an interest rate hike because banks face a deteriorated balance sheet and then ration some borrowers within a process of deleveraging (loan supply effect) or because the deteriorated outlook has shifted down the demand for bank credit (loan demand effect). Overall, these limitations suggest that an empirical strategy that would rely on a data-rich environment à la Stock and Watson (2002) and would exploit information on heterogeneity in individual banks' financial conditions and the way they change through time could be more appropriate to detect the potential active role of banks in the transmission of the monetary policy shocks.

In this paper, we propose to examine the strength of the credit channel while reconciling the micro and macro levels of analysis into an integrated estimation framework. Following Bernanke, Boivin, and Eliasz (henceforth BBE) (2005) and Boivin and Gianonni (2009), we employ a factor-augmented vector autoregression model

[5]Nevertheless, Giannone, Lenza, and Reichlin (2009) propose to overcome this dimensionality problem and estimate such a large-scale monetary VAR using Bayesian techniques.

(FAVAR) that we extend to explicitly include factors reflecting relevant fluctuations in a set of individual bank balance sheet ratios. A key feature of the BBE framework is to extract estimates of macroeconomic factors that affect the data of interest by exploiting the information contained in a large set of economic indicators. According to BBE (2005), the FAVAR framework leads to a better identification of the monetary policy shock compared with standard VARs, since it accounts explicitly for the large information set that central banks monitor in practice and also because it does not require taking an ex ante stand on the appropriate measure of economic concepts such as inflation or real activity, which are treated as latent common components. Finally, another appealing feature of the FAVAR approach is that the impulse response functions to a monetary policy shock can be computed for any variable included in the data set, while the dimensionality of the estimated VAR is kept reasonably low.

We implement this methodology in the case of France over the period 1993–2009, with quarterly data. The novelty of our approach consists in the parallel extraction of dynamic factors from large data sets of bank balance sheet indicators. Using French supervisory sources, we construct an original database of disaggregated bank financial ratios for a large set of French credit institutions, making up 70 percent of total domestic bank credit. We consider this way, in our application, the information contained both in a macroeconomic database comprising a large number of macroeconomic indicators and in a microeconomic database. We then use the FAVAR setup to quantify the specific impact of banks' financial conditions, if any, on the response of key macroeconomic variables to monetary policy shocks.

While our methodology closely follows that of Boivin and Gianonni (2009), who focus on the role of international factors in the transmission of U.S. monetary policy, our interest in the role of bank-level information is, to our knowledge, quite a novelty in the FAVAR literature. We are aware of only a few recent studies that go along a similar route. Gilchrist, Yankov, and Zakrajsek (2009) extract unobserved factors from a broad array of corporate bond spreads and study the macroeconomic impact of shocks to these measures of credit risk in a FAVAR model of the U.S. economy. Boivin, Gianonni, and Stevanovic (2009) perform a similar exercise but implement a different identification scheme of credit shocks which allows an economic interpretation of the (transformed) principal component analysis (PCA) factors. While close in spirit to ours, these studies do not, however, deal directly with monetary policy transmission and do not consider disaggregated bank data. By contrast, Dave, Dressler, and Zhang (2009) investigate the dynamic response of both credit aggregates and bank-level loan growth measures to a monetary policy shock using disaggregated U.S. bank data. However, they mainly focus on the differentiated responses of different types of loans, in the spirit of Den Haan, Sumner, and Yamashiro (2007), and do not use their FAVAR model to assess, as we do, whether fluctuations in banks' financial conditions (and their dispersion) significantly alter the transmission of monetary shocks to the broader macro economy.

Thematically, our work, of course, fits in with the abundant credit channel literature. However, more specifically, it can be related to a series of recent attempts to bridge the gap between microeconomic information about the health of financial institutions and the macro economy. Among these are recent studies by Adrian and Shin (2009a) that highlight the procyclical role of U.S. investment banks' leverage, as well as somewhat earlier research by Peek, Rosengren, and Tootell (1999, 2003), who use a summary of U.S. bank-level supervisory information to identify the effect of loan supply shocks on GDP and its main sub-components.[6] The latter notably find that bank supervisory information predicts macroeconomic fluctuations, and they provide evidence that this information is in fact used by the Federal Reserve to conduct monetary policy. However, they do not formally examine the consequences of fluctuations in their bank health indicator for the transmission of monetary shocks.

Our main results are twofold. First, it appears that the first two principal components extracted from banks' liquidity or leverage ratios are quite correlated with industrial production and housing market conditions and that they predict macroeconomic fluctuations. This suggests a potential scope for macroprudential policies aiming at dampening the procyclical effects of changes in banks' balance sheet structures. Second, we nevertheless find that the fluctuations in banks' financial conditions do not matter much per se for

[6]Their summary variable is the percentage share of bank assets that fall within the CAMEL 5 rating computed by the U.S. Federal Reserve (i.e., the share of the riskiest part of the regulated U.S. banking system).

the transmission of monetary policy shocks to the French economy. In other words, there is little point in monitoring the information contained in bank balance sheets above the information already contained in credit aggregates, as far as monetary policy transmission is concerned.

The rest of the paper is organized as follows. Section 2 reviews the econometric framework and the estimation approach, with a detailed presentation of the data used in our estimation. In section 3 we present the latent factors and the co-movements between the macro and micro factors. Section 4 investigates the role of the bank-level factors in the monetary policy transmission mechanism. Section 5 concludes.

2. Econometric Framework: FAVAR

We aim to evaluate the importance of individual banks' financial conditions for monetary policy transmission in the case of France, a country for which we have access to a rich supervisory database. In other words, we seek to estimate to what extent the specific response of banks' financial conditions enhances or mitigates the effect of monetary policy on the economy. In this section, we first describe the empirical model and the estimation approach. In so doing, we closely follow the lines of Boivin and Giannoni (2009).

2.1 Description of FAVAR

We consider an econometric framework based on a standard macro FAVAR model that we extend to include additional factors summarizing the financial health of individual banks. We assume that the macroeconomic conditions can be adequately summarized by a $K \times 1$ vector of unobserved components or factors, C_t, while another $K^* \times 1$ vector of factors C_t^* is enough to describe the financial conditions of the banking sector. Note that, in what follows, the variables related to microeconomic information on banks will always be denoted with a star ($*$). In practice, we can assess the state of the economy and the health status of resident banks using (i) a large vector of macroeconomic indicators (denoted by X_t) and (ii) a vector of individual bank balance sheet indicators for a large number of banks (denoted by X_t^*). These vectors are of dimension $N \times 1$ and $N^* \times 1$, respectively.

We assume that the macroeconomic indicators are related to the state of the economy and that disaggregated bank balance sheet ratios are related to the overall financial conditions of the banking sector according to the following observation equations:

$$X_t = \Lambda C_t + e_t \tag{1}$$

$$X_t^* = \Lambda^* C_t^* + e_t^*, \tag{2}$$

where Λ and Λ^* are matrices of factor loadings and the $N \times 1$ (and $N^* \times 1$) vectors e_t and e_t^* stand for (mean-zero) series-specific components. By construction, these specific terms are uncorrelated with the common components C_t or C_t^* within each equation but are allowed to be serially correlated and (weakly) correlated across indicators. Note that the number of common factors is assumed to be small relative to the number of indicators ($N > K$ and $N^* > K^*$). Within this framework, C_t and C_t^* represent two sets of components, common to all data series in each block and, in general, correlated across the two sides of the economy (macro conditions versus bank financial conditions).

The common factors should be understood as pervasive forces that drive the common dynamics of the data in each block, summarizing at each date either the state of the "real" economy or the financial strength of banks, as reflected by equations (1) and (2). The variables in X_t are then taken as noisy measures of the underlying unobserved factors C_t. This means, for instance, that GDP growth, which belongs to the vector of macro series, is a noisy measure of "real activity," while the liquidity or leverage ratio of a bank j, which belongs to the vector of banking sector series, is a noisy measure of the financial health of the overall banking sector. We note that, in principle, it is not restrictive to assume that X_t depends only on the current values of the factors, since C_t might capture arbitrary lags of some fundamental factors.[7]

The dynamics of the common factors—i.e., the transition equation—are modeled as a structural VAR:

$$\Phi_0 \begin{bmatrix} C_t^* \\ C_t \end{bmatrix} = \Phi(L) \begin{bmatrix} C_{t-1}^* \\ C_{t-1} \end{bmatrix} + \begin{bmatrix} v_t^* \\ v_t \end{bmatrix}, \tag{3}$$

[7] Stock and Watson (1999) refer to (1) as a dynamic factor model.

where Φ_0 is a matrix of appropriate size on which we later impose some restrictions, $\Phi(L)$ is a lag polynomial of finite order, and the "structural" shocks v_t and v_t^* are assumed to be i.i.d. with zero mean and diagonal covariance matrix Q and Q^*, respectively. These shocks are uncorrelated, but any one of them may affect common factors of the other block (French economy versus banks' financial conditions) immediately or over time, through the off-diagonal elements of Φ_0 and $\Phi(L)$. By pre-multiplying both sides of (3) by Φ_0^{-1}, the structural VAR has then the following reduced-form representation:

$$\begin{bmatrix} C_t^* \\ C_t \end{bmatrix} = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} C_{t-1}^* \\ C_{t-1} \end{bmatrix} + \begin{bmatrix} u_t^* \\ u_t \end{bmatrix}, \qquad (4)$$

where the reduced-form innovations u_t and u_t^* may be cross-correlated.

Since we are interested in characterizing the effects of monetary policy on the economy, we want to include in the vector of macroeconomic common components an observable measure of the monetary policy stance. As is commonly the case in VAR studies of monetary policy in European countries, we consider here that the three-month money market rate, R_t, (i.e., the three-month PIBOR before 1999 and the three-month EURIBOR afterwards) is an appropriate measure of the monetary policy stance in France over the period from 1993 to 2009. The short-term interest rate is thus allowed to have a pervasive effect throughout the economy and is considered a common component of all macro data series. This way, we can write

$$C_t = \begin{bmatrix} F_t \\ R_t \end{bmatrix}, \qquad (5)$$

where F_t is a vector of latent macroeconomic factors summarizing the behavior of the rest of the French economy.

2.2 Estimation

We estimate the empirical model using Boivin and Gianonni's (2009) variant of the two-step principal component approach developed notably by Stock and Watson (2002) and BBE (2005).

The first step consists of extracting separately principal components from X_t and X_t^* in order to obtain consistent estimates of the common factors under the structure laid out. As stated above, we impose the constraint that the short-term interest rate is one of the factors for the set of macroeconomic series. This guarantees that the other estimated latent factors recover dimensions of the common dynamics that are not captured by the short-term interest rate. Starting from an initial estimate of F_t, denoted by $F_t^{(0)}$ and obtained as the first $K-1$ principal components, we thus iterate through the following steps:

(i) Regress X_t on $F_t^{(0)}$ and R_t, to obtain $\hat{\lambda}_R^{(0)}$.
(ii) Compute $\tilde{X}_t^{(0)} = X_t - \hat{\lambda}_R^{(0)} R_t$.
(iii) Estimate $F_t^{(1)}$ as the first $K-1$ principal components of $\tilde{X}_t^{(0)}$.
(iv) Return to step (i).

As far as the common factors spanning individual bank balance sheet data are concerned, we do not impose any constraint in the first step. We simply estimate F_t^* as the first K^* principal components of X_t^*, where X_t^* collects series of individual bank ratios as explained below in the data section.

In the second step, the short-term rate is added to the estimated macroeconomic factors F_t and the VAR in C_t and C_t^* (equation (4)) is estimated. The matrix polynomial $\Psi_{21}(L)$ is then of particular interest, since it captures the lagged effect of banking conditions on macroeconomic factors. Note that the VAR coefficients $\Psi_{ij}(L)$ are identified, provided that the variance-covariance matrix of the innovations $[u_t^{*'}, u_t']'$ is non-singular. In particular, a sufficient condition for the coefficients in $\Psi_{21}(L)$ to be identified is that the factors standing for banking conditions do Granger-cause macroeconomic factors, an issue that we explore below.

2.3 Data Description

We use two distinct sets of data for the estimation of the FAVAR: a bank-level one and a macroeconomic one, both over the period 1993:Q2 to 2009:Q1, with quarterly frequency. Regarding the first data set, a distinctive feature of our study is that we make use of a large database of disaggregated bank balance sheet information as collected by the French supervisory agency (Commission bancaire). This database is particularly attractive because of its exhaustive coverage of all credit institutions chartered in France.[8] Nevertheless, its time depth is limited—bank balance sheet details are only available (with this broad coverage and under consistent reporting guidelines) since the first quarter of 1993. Although this tends to limit econometric investigations, it may also be noted that the year 1993 coincides with the adoption of the Bank of France's independence by law, as well as with the launch of the last stage of the convergence process toward the European Monetary Union (EMU). This way, the period from 1993 to 2009 may be seen more convincingly as a single monetary policy regime, without any significant structural break (a point that we will discuss more in depth below, while presenting the impulse response functions computed from the FAVAR).

The macroeconomic data set comprises sixty-eight macroeconomic series—sixty series for the French economy and eight series for the German economy. All macro series have been transformed to induce stationarity when necessary, as indicated in the appendix. Other details about data sources and definitions are also provided in the appendix. The inclusion of some key German series is motivated by the fact that the French monetary policy was largely tied to the monetary policy of the German Bundesbank during the run-up to the EMU and, notably, from 1993 on, within the frame of the European exchange rate mechanism. Including measures of both French and German activity, inflation, and money prevents the implicitly estimated monetary policy function within the FAVAR from suffering significant structural breaks due to the introduction of the euro and the consecutive delegation of the French monetary policy to the European Central Bank (ECB). In particular, we rely on the fact that both countries, taken together, account for roughly half of the GDP of the entire euro area and behave (partly by construction) in a quite similar way to the euro-area average, which is the relevant aggregation level for the ECB since 1999.

Our initial bank-level data set, with some 620 credit institutions at sample end, is not directly suitable for the purpose of factor analysis because of (i) the high degree of heterogeneity between

[8]Note that this balance sheet information is collected on a territorial basis, which means that credit granted by subsidiaries of French banking groups abroad is not reported.

institutions of different types and (ii) the wide-ranging process of concentration within the French banking sector during the 1990s: indeed, the number of institutions shrank by a factor of 2.5 over the period from 1993 to 2009, implying that many banks were not observable over the whole period. Selecting an adequate sub-sample of banks involves thus a difficult trade-off. The right balance has to be struck between, on the first hand, keeping enough individual banks to catch something of the dispersion of banks' financial conditions in the estimated factors, and, on the second hand, discarding enough atypical small institutions so that the few PC factors we extract from the final sample are economically meaningful.

To alleviate the heterogeneity issue and focus on "banks" as commonly understood (i.e., credit institutions whose business is to some large extent to collect deposits from the public and grant credits to non-financial agents), we first removed several categories of specialized credit institutions. These special categories include specialized financial institutions (such as leasing banks, customer credit institutions, factoring institutions, etc.), municipal credit institutions, and regional development institutions. We also dropped the regional branches of the three large mutualist and cooperative banks (in order to avoid double-counting when the group head also reports for the entire network), the regional savings banks affiliated with a large savings bank network (whose individual ratios were suspected to be affected by important within-network transfers), and the branches of foreign-held commercial banks (whose credit policy and financial health may be relatively immune to local economic conditions).[9] Finally, we also dropped commercial banks that operate only in the French overseas territories. Overall, this preliminary cleaning leaves us with a population of 105 banks at sample end, accounting for 77 percent of total customer credit and 79 percent of total assets in the initial database.[10]

At this stage of the database cleaning process, the population of banks remains still very heterogenous in terms of assets'

[9]For instance, Peek and Rosengren (2000) show that the lending capacity of branches of Japanese banks in the United States, in the late 1990s, was heavily constrained by the capital crunch faced by the mother institutions in Japan, leading to severe restrictions in the U.S. commercial mortgage market.

[10]Customer credit is defined throughout as loans to non-banks in the supervisory database.

size and business profile. A quarter of these institutions are very small banks—mainly private banks, local banks, or specialized institutions—with total assets below €500 million each at sample end. Since the unobserved factors C_t^* are extracted by unweighted principal components analysis, keeping so many small banks implies that the information from the small banks' ratios (that have little macroeconomic significance but are often more volatile) can have a disproportionate bearing on the estimated factors, which would be detrimental to their economic significativity. We thus chose to drop from the above selected categories the banks belonging to the first quartile in terms of mean total assets.[11] Besides, as the FAVAR approach requires a balanced data set, we are forced to keep only the banks for which we have data over the entire period 1993–2009.[12]

Combined, these two steps reduced the population to a sample of sixty banks, making up nearly 71 percent of total customer credit (and 71 percent of total bank assets). For each of these banks, we constructed three different financial ratios as defined below in more detail. Finally, a visual inspection of these ratios, cross-checked with the results of standard Bai and Perron tests of mean stability, revealed large statistical breaks for about half of these institutions, the bulk of them being again smaller banks accounting together for some 2 percent of total credit. We could check that these breaks are generally not explained by either acquisitions or changes of regulatory category but may reflect other sources of statistical noise like changes in capital detention or business lines, about which we have no information in the database. We thus decided to correct for the biggest breaks using a simple statistical procedure along the lines of Den Haan, Sumner, and Yamashiro (2007). More precisely, for each

[11] The twenty-six banks thus discarded have average total assets below €405 million over 1993–2009.

[12] In so doing, we obviously created a potential for a selection bias problem. An alternative, often considered in papers running panel regressions on individual bank data, consists of reconstructing mergers and acquisitions (M&As) backward in time, in order to keep information both from target and buying banks before the M&A (cf., for instance, Loupias, Savignac, and Sevestre 2002). This approach means in practice adding the corresponding balance sheet items of both banks prior to the merger. We think this method would be inappropriate for the purpose of our analysis, since it would end up creating individual measures of banks' financial health that did not exist in reality, adding uncontrolled sources of noise into the data.

ratio data set, we computed growth rates of the ratio series and identified outliers as values distant from the cross-sectional mean by more than 2.63 standard deviations. We replaced these outlier growth rates with the corresponding cross-sectional means. Index ratio series were then reconstructed using the corrected growth rates and starting from the initial level values of the ratios. Finally, we dropped eight banks, mainly market banks, that presented more than 20 percent outliers for at least one of the ratios.[13]

The resulting final sample consists of fifty-two commercial and cooperative banks accounting for 70 percent of total loans granted by all credit institutions at sample end (and 69 percent of total assets). Figure 1 shows the share of our sample in the total of bank loans over the whole period (1993:Q1–2009:Q1). This share increased somehow through time, staying between 66 percent and 70 percent since early 1997.

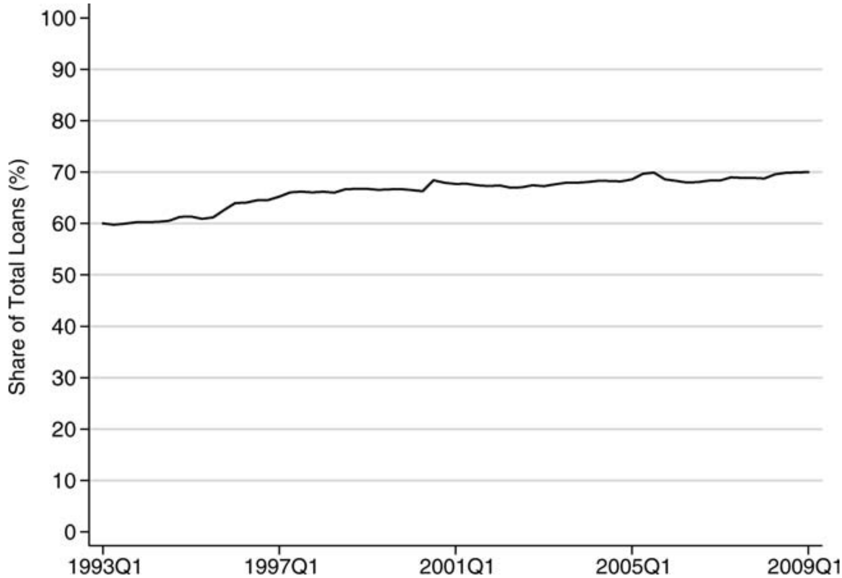
For each bank in the sample, we define three ratios that capture key dimensions of a bank's financial situation: one liquidity ratio and two leverage ratios—a total (or broad) leverage ratio and a credit (or narrow) leverage ratio. Both types of ratios have been identified in the empirical panel literature as important determinants of banks' reaction to monetary policy shocks, in line with the standard descriptions of the credit channel and the renewed versions of the same theory (like the bank capital channel of Van den Heuvel 2002). The introduction of total leverage ratios is also motivated by Adrian and Shin's (2009b) recent findings that the total leverage of at least some U.S. credit institutions is highly procyclical and their proposal that banks' leverage should be more closely and systematically monitored in a macroprudential perspective.[14]

We compute firstly an indicator of bank *liquidity* (labeled LIQ in the following), which we define as the ratio of liquid assets to total assets. Liquid assets are computed as the sum of cash, interbanking transactions, securities bought under repurchase agreements, and securities held in the trading portfolio.

[13]In the final ratio data sets, the outlier detection procedure implied a correction of 3 percent of the observations in the case of the broad leverage ratio and less than 2 percent for the other ratios.

[14]Regulatory capital ratios would have been equally interesting candidates, but this information was not available with enough time depth to be used in a time-series analysis.

Figure 1. Share of the Banks in Sample in Total Loans Granted by All Resident Credit Institutions, 1993–2009



Secondly, we define our broad measure of *leverage* (LEV1) as the ratio of total assets to tier 1 capital. This broad leverage ratio is thus the inverse of the capitalization ratio often considered in earlier panel data studies.[15] Finally, we also compute a narrow measure of leverage (LEV2), defined as the ratio of customer credit to tier 1 capital. Although less frequently used in the academic literature, this second leverage ratio is monitored by the French regulators on a regular basis, which motivates its inclusion in our study.[16]

Table 1 presents some descriptive statistics for our sample in 1999:Q1. Note that the sample, although relatively small and centered on banks of relatively close types, is still quite diverse along the standard dimensions explored by the credit channel literature, be it in terms of size (with total assets ranging from €268 million

[15] See, for instance, Kishan and Opiela (2000) and Loupias, Savignac, and Sevestre (2002) for the French case.

[16] Cf., for instance, Commission bancaire (2009, p. 69 et seq.)

Table 1. Descriptive Statistics (2009:Q1)

	Mean	Median	SD	Min.	Max.
Assets (Billions of Euros)	99.5	5.4	255	0.3	1,390
% of Total Bank Assets	1.3	0.1	3.4	0.0	18.6
Loans (Billions of Euros)	27.1	1.9	62.9	0.0	334
% of Total Bank Loans	1.4	0.1	3.1	0.0	16.6
Liquidity (LIQ)	0.22	0.29	0.23	0.01	0.99
Broad Leverage (LEV1)	20.5	29.1	26.1	1.18	124.7
Narrow Leverage (LEV2)	8.0	11.6	11.23	0.18	50.52

to €1,390 billion), liquidity (with a ratio between 1 percent and 99 percent), or leverage (with a broad ratio between 1.2 and 125).

Factor estimation using principal components requires stationary times series.[17] Although it seems reasonable to assume from an economic point of view that bank ratios should be stationary, standard unit-root tests reveal that it is not always the case from a statistical point of view in our sample.[18] Neglecting the results of such tests and keeping stochastic trends in the ratio data sets may lead to inconsistent estimates of the bank ratio factors and spurious correlations of these factors with the most persistent macroeconomic series in our database—in particular, interest rates. We thus took the first difference of the individual ratio series that showed a unit root at the 95 percent probability level, and left other ratios series unchanged.[19] Last, note that, following common practice, all macro and micro series were demeaned and standardized before extraction of the factors by PCA.

2.4 Specification of the FAVAR

The empirical model presented above is a dynamic factor model that links a large set of observable indicators to a small set of common

[17]Although simple methods have been proposed to estimate consistent factors from non-stationary panel data (see the PANIC methodology by Bai and Ng 2004), such methods are not suited for data sets mixing stationary and non-stationary series. Besides, the FAVAR approach requires stationary factors.

[18]We ran augmented Dickey-Fuller tests with a constant and a number of lags selected according to the Akaike information criterion.

[19]Between 64 percent and 83 percent of the ratios were considered as having a unit root, depending on the type of ratio.

components through the observation equations (1) and (2). The evolution of the common components is then specified by the transition equation (3) or its reduced-form representation (4). Theoretically, to the extent that we keep a sufficiently large number of common components from the PCA of each data block, the estimated factors collected in C_t and C_t^* span the same space as the unknown "true" factors or latent variables that drive the set of noisy indicators X_t and X_t^*. The issue of the number of factors selected is thus an important one in theory. In practice, however, there is still no clear consensus about the right analytical criteria for the choice of this number, and numerous applications rely on judgmental or empirical evidence like, for instance, the change induced to the estimated impulse response functions when new factors (accounting for a smaller part of the database variance) are added to the FAVAR model.[20]

In our case, it should be borne in mind that the class of specifications we can consider is severely constrained by the sample size (sixty-four quarters of observations), which especially limits the number of lags in (4) as the number of factors gets larger. We want to include more than one common component from a given bank ratio data set, in order to assess the potential impact of bank heterogeneity on the economy. Nevertheless, small sample size prevents us from including simultaneously common components from all three ratio data sets. We do then consider three distinct FAVAR models, replacing X_t^* with each of the three ratio data sets in turn.

We thus based the choice of the number of factors on two empirical criteria. Firstly, we computed Bai and Ng (2002) PCP2 and IC2 criteria, which indicated a maximum number of two factors for each of the three ratio data sets for the baseline sample of banks. Secondly, for each type of bank ratio, we estimated a FAVAR with up to six macroeconomic factors (including the short rate) and up to three common components from individual bank series. It appears that the form of the IRF to monetary policy shocks is quite robust to the inclusion of additional factors when at least four macro factors are included in the model. Whatever the bank ratio considered, our preferred specification of the corresponding FAVAR thus includes four

[20] This is the route followed, e.g., by BBE (2005), Boivin, Giannoni, and Mojon (2008), and Boivin and Giannoni (2009).

macroeconomic factors and two bank ratio factors, and the transition equation (4) has one lag.[21]

3. Bank Balance Sheet Factors and Macroeconomic Dynamics

We first aim at clarifying how the factors summarizing French macroeconomic dynamics relate to disaggregated bank-level factors, as extracted separately from three microeconomic data sets of individual liquidity and leverage ratios. In this section, we thus use the common factors extracted from our various data sets and first determine the fraction of fluctuations in indicators of real activity, inflation, credit aggregates, and interest rates that can be explained by macro and bank-level factors, respectively. This first simple look at correlations and Granger causalities suggests that there is potentially a scope for a macroprudential regulation of banks' leverage and liquidity with a view of limiting the extent of macroeconomic fluctuations induced by banks' behavior. In the next section, we will then compare the impulse responses of various key macroeconomic variables when bank ratio factors are allowed to interact with macro factors or alternatively when this additional feedback mechanism is artificially shut down.

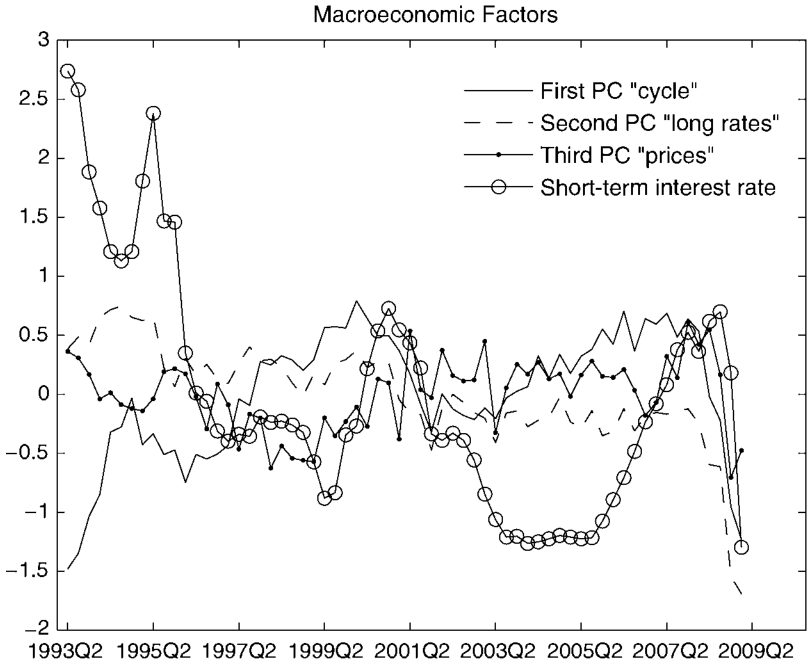
3.1 Interpreting the Latent Factors: A First Look

We start by examining how the macro and bank ratio or micro factors are correlated with each other and with key macro variables, in an attempt to roughly characterize these latent factors. Figure 2 shows the estimated macro factors, while the bank ratio factors are plotted in figures 3–5.

Table 2 reports the correlations of the first three macro factors (excluding of course the short-term interest rate, which we force to be the fourth macro factor as explained above) with a selection of macroeconomic variables. The first latent macro factor obviously

[21] Results from standard tests of lag selection were mixed. The Schwarz information criterion suggested one lag in the various FAVAR models, generally in line with the Hannan-Quinn criterion, while the Akaike information criterion suggested from two to five lags, depending on the model.

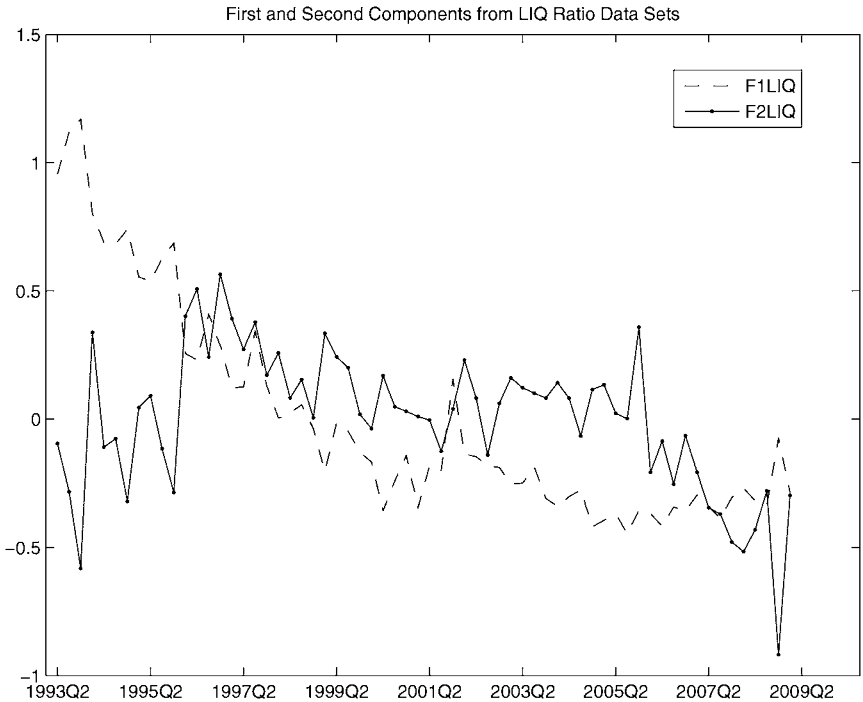
Figure 2. Macro Factors



stands for a measure of the business cycle, with a high positive correlation to GDP growth and its components and a negative correlation to unemployment. The second macro factor is also positively correlated with GDP growth but can be more easily characterized as driven by longer-term interest rates, while the third macro factor tends to capture the dynamics of inflation.

Table 3 reports the correlation of the micro factors, as extracted separately from each of the three bank ratio data sets, with the macro factors. The first LIQ component is strongly correlated with the business cycle macro factor (and, with opposite sign, with the short-term interest rate), while the second bank liquidity component is mostly correlated with the second macro "long-term rate" factor, and the correlation coefficient is smaller. The second LEV1 factor and the first LEV2 factor have similar correlation profiles and are strongly correlated with the short-term interest rate. The other two leverage factors are also correlated with the "interest rate" and

Figure 3. LIQ Factors

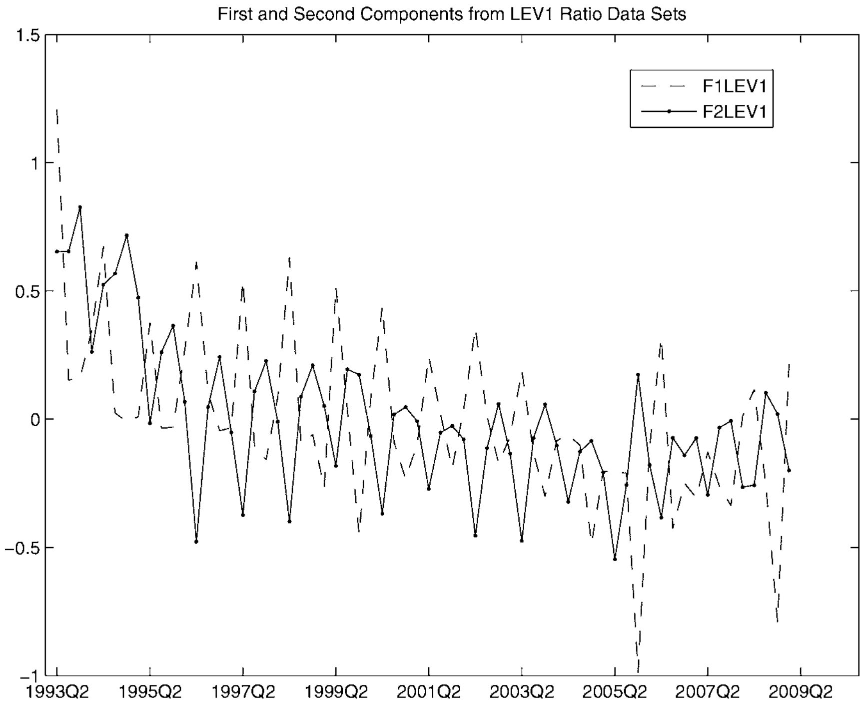


"business cycle" macro factors, but the correlation is weaker. Overall, these preliminary calculations confirm the intuition that banks adjust key dimensions of their balance sheets to fluctuations in real activity and market interest rate conditions. We do not know, however, to what extent such adjustments are active, as part of their asset-liability management policy, or passive, as an effect of changes in demand for credit. Neither can we determine on the basis of this evidence alone whether changes in bank conditions have an impact on macro conditions.

3.2 Co-Movements between Macro and Micro Factors

In a second step, we investigate to what extent French macroeconomic variables are explained by macro versus bank-based factors. To do this, we regress each macro variable on the three macro factors

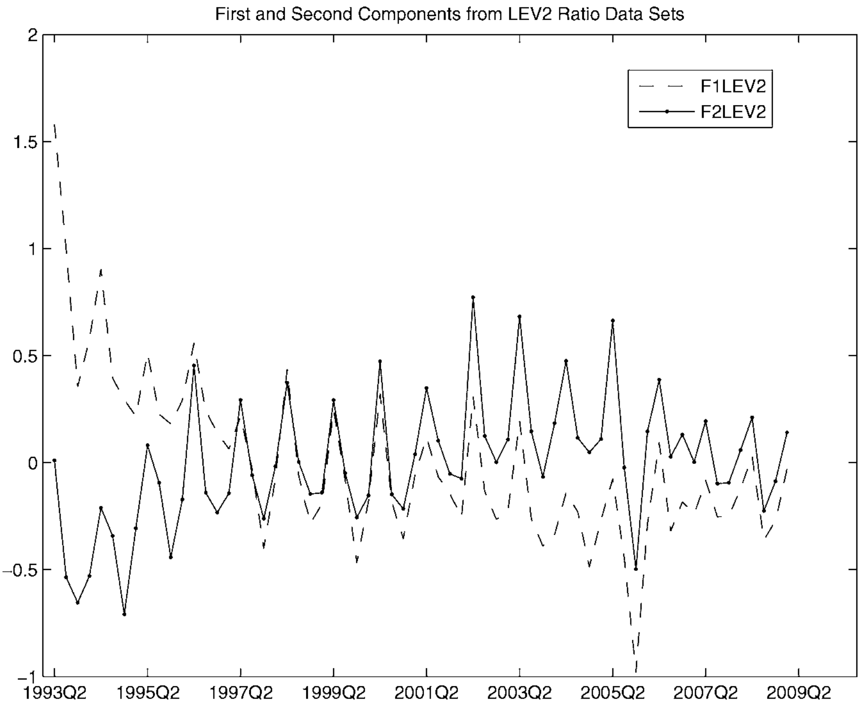
Figure 4. LEV1 Factors



(including the short-term interest rate) or the first two bank-based factors obtained for a given type of bank ratio, taking each type of ratio in turn. Table 4 reports the fraction of variance of the series listed (i.e., the R^2 of the least-squares regressions) that is explained by the macro factors and the bank ratio factors of each type (i.e., either related to liquidity or leverage, broadly or narrowly defined), respectively.

As is apparent in the first row of the table, the entire macroeconomic data set, X_t, is on average strongly correlated with the common factors. The R^2 of the macro factors is of 0.56, showing that the macro factors capture a good part of fluctuations in the French economy overall. As could be expected, the first three PCA factors of each set of individual bank ratios are less correlated with the macro series on average, with an R^2 of between 0.11 and 0.28, depending on the type of bank ratio considered.

Figure 5. LEV2 Factors



When looking more closely at selected macroeconomic indicators, we first find that quarterly growth rates of real GDP, industrial production, HICP inflation, employment, and non-residential investment present high correlations with the macro factors (R^2 statistics of 0.84, 0.87, 0.81, 0.80, and 0.67, respectively), while it is far less so for consumption (be it consumption of durable or non-durable goods). Macro factors also do a good job in tracking market and credit interest rates. The important point here is that most of the fluctuations in very cyclical variables are captured by only four macro factors. Note that the macro factors also explain some 62 percent of the variance in housing prices and 70 percent of the variance in housing loans.

Regarding the results of regressions on bank-level common components, we first find that bank liquidity and credit leverage factors explain a substantial part of the variance of housing prices (between

Table 2. Correlations between the Macro Factors and Selected Macro Variables

Variable	F1	F2	F3
Interest Rate	−0.49	0.50	0.15
GDP	0.57	0.69	−0.07
Industrial Production Index	0.79	−0.44	0.22
Employment	0.75	0.39	−0.23
Unemployment	−0.65	−0.42	−0.02
Consumption	0.33	0.18	−0.19
Consumption Durable	0.11	0.13	−0.14
Consumption Non-Durable	0.06	0.12	−0.10
Non-Residential Investment	0.67	0.43	−0.18
Inventories	0.78	−0.00	0.04
Residential Investment by Household	0.46	0.43	−0.02
Housing Prices	0.58	−0.02	0.25
HICP	0.16	0.17	0.87
GDP Deflator	0.20	−0.28	0.54
Total Loans	0.61	−0.20	0.33
Housing Loans	0.60	−0.47	0.30
Inv. Corporate Loans	0.54	−0.51	0.07
C&I Loans	0.59	0.01	0.02
France Ten-Year Yield	−0.48	0.71	−0.01
Interest Rate C&I Loans	−0.57	0.60	0.08
Interest Rate Invt. Loans	−0.72	0.55	0.13
Interest Rate Housing Loans	−0.63	0.67	−0.11

Table 3. Correlation between Bank Ratio Factors and Macro Factors

Variables	F1 LIQ	F2 LIQ	F1 LEV1	F2 LEV1	F1 LEV2	F2 LEV2
Fmacro1	−0.70***	0.07	−0.33***	−0.47***	−0.61***	0.29**
Fmacro2	0.61***	0.39***	0.40***	0.47***	0.47***	−0.35***
Fmacro3	−0.05	−0.21*	0.00	−0.01	0.01	0.01
Interest Rate	0.76***	−0.31**	0.38***	0.64***	0.67***	−0.49***

Notes: *, **, and *** denote significance at the 10 percent, 5 percent, and 1 percent level, respectively.

Table 4. R^2 for Regressions of Selected French Macro Indicators on Various Sets of Macro and Bank Ratio Factors (Sample 1993:Q2–2009:Q1)

	All Macro Factors (1)	LIQ Factors (2)	LEV1 Factors (3)	LEV2 Factors (4)
All France Data X_t (Average Over All French Data)				
Selected FR Indicators				
Interest Rate	1.00	0.68	0.56	0.69
GDP	0.84	0.16	0.02	0.00
Industrial Production Index	0.87	0.83	0.62	0.71
Employment	0.80	0.11	0.02	0.05
Unemployment	0.68	0.03	0.03	0.04
Consumption	0.19	0.05	0.01	0.03
Consumption Durable	0.05	0.01	0.03	0.01
Consumption Non-Durable	0.04	0.02	0.01	0.00
Non-Residential Investment	0.67	0.06	0.02	0.05
Inventories	0.69	0.27	0.23	0.26
Residential Inv. by Household	0.46	0.06	0.01	0.01
Housing Prices	0.62	0.32	0.22	0.33
HICP	0.81	0.01	0.00	0.00
GDP Deflator	0.41	0.14	0.11	0.08
Total Loans	0.56	0.28	0.14	0.16
Housing Loans	0.70	0.57	0.43	0.48
Inv. Corporate Loans	0.60	0.49	0.25	0.33
C&I Loans	0.45	0.13	0.08	0.08
France Ten-Year Yield	0.80	0.70	0.60	0.64
Interest Rate C&I Loans	0.89	0.71	0.62	0.78
Interest Rate Invt. Loans	0.95	0.82	0.69	0.83
Interest Rate Housing Loans	0.91	0.85	0.75	0.79

22 percent and 32 percent). The correlation with contemporaneous housing price growth, while saying nothing about the direction of causality, highlights the fact that the recent boom and bust in housing prices over the last decade was largely associated with changes

Table 5. Granger Causality Tests for Bank Factors Affecting Macro Factors

	LIQ	**LEV1**	**LEV2**
All Sample (1993–2009)			
F1	0.92	0.00	0.00
F2	0.00	0.15	0.08
F3	0.50	0.89	0.86
Interest Rate	0.73	0.00	0.00
Before 2007–09 Crisis			
F1	0.02	0.06	0.06
F2	0.00	0.01	0.00
F3	0.21	0.02	0.05
Interest Rate	0.21	0.00	0.01

Note: Table reports p-values.

in individual banks' balance sheets. Rolling regressions over an eight-year period (not reproduced here to save space) show that this correlation was in fact higher, above 60 percent, over the decade from 1995 to 2005, which corresponds to the boom episode in France. Second, bank liquidity ratios are highly correlated with average bank interest rates for new loans, notably housing loans and investment loans to non-financial firms.

3.3 Do Banks' Financial Conditions Predict Macro Fluctuations?

The correlations discussed so far shed some light on the interrelation between macroeconomic conditions and the balance sheets of individual credit institutions. However, we do not know so far whether changes in banks' balance sheets are passively driven by the macro economy or whether they actively contribute to shaping the business cycle, as the bank lending channel would suggest, or at least can help to predict it.

In a first attempt at identifying the information content of banks' financial conditions for future macroeconomic conditions, we computed standard Granger causality tests within each of the three FAVAR models with bank ratio factors. Table 5 reports in rows the results of tests of the joint significance of bank-level factors of a given type (as stated in the column heads) in a regression of each macro factor over all lags of all macro factors and lags of the bank ratio components. Under the null hypothesis, bank-level factors have no predictive power. The upper panel reports results of estimations over the whole sample, while the lower panel restricts to the period before the onset of the sub-prime crisis in 2007:Q3.

The results show that the three types of bank ratios do not have the same informational content for macroeconomic conditions. The information extracted from narrow (respectively, large) bank leverage predicts three (two) of the macro factors over the whole period, and up to four macro factors over the pre-crisis period (at the 10 percent level). In particular, bank leverage factors consistently predict the short-term interest rate and the "business cycle" factor one quarter ahead. In contrast, factors summarizing bank liquidity mainly predict the second macro factor and, over the pre-crisis period, the "business cycle" factor. By the way, these results confirm that the coefficients $\Psi_{21}(L)$ in the reduced-form model (equation (4)) are determinate, at least some of them.

While preliminary, the outcome of these causality tests suggests that microprudential regulations of liquidity or leverage of credit institutions should also matter in a macroprudential perspective.[22]

4. Implications for the Monetary Transmission Mechanism

We have documented so far that common components from key individual banks' balance sheet ratios co-move with selected macroeconomic variables and, to some extent, drive changes in broad macroeconomic conditions. A natural question that arises then is whether the endogenous reaction of individual banks to an unexpected monetary policy impulse significantly alters the response of aggregate variables of interest (like GDP or consumer price inflation). The standard theory about the credit channel of monetary policy transmission suggests that the endogenous response of banks may amplify

[22]Note, however, that our measure of leverage is the total asset ratio to book equity capital, which significantly differs from the ratio of capital to risk-weighted assets usually monitored by bank supervisors.

the effects of a monetary tightening, e.g., due to an increase in the external finance premium required by banks in the face of an induced deterioration of borrowers' creditworthiness (financial accelerator effects) or, similarly, a rise in the external finance premium faced by banks following an induced deterioration in their own assets' value (bank capital channel). Alternatively, theories of credit rationing suggest that capital shortages or liquidity constraints on the side of banks may, on the contrary, dampen the response of bank credit to monetary policy attempts at loosening overall financial conditions.

We investigate here this issue within the FAVAR framework presented above in section 2, following a general approach initiated by Boivin and Gianonni (2009). More precisely, we compare the impulse response functions (IRFs) of selected macroeconomic variables to a 100-basis-points monetary policy shock under alternative hypotheses regarding the coefficients $\Psi_{21}(L)$ in equation (4), which links macro factors to lagged bank-level factors. The difference between the IRF when this block of coefficients is set to zero and when it is left unrestricted provides a measure of how important the endogenous response of individual banks' balance sheets is for the monetary transmission mechanism in France. In other words, the larger the difference, the more a model of monetary policy transmission which includes only money and credit aggregates is misspecified.

A preliminary important issue is, however, whether the launch of the euro and the changeover from the Bank of France to the ECB from January 1999 on implies a regime shift for monetary policy in France or not. If it were the case, then non-linearities should arise (at least) in the short-term interest rate equation of our FAVAR, and we would not be allowed to investigate credit channel issues on the basis of linear VARs estimated over a time period that includes the date of EMU inception.

We think, however, that the assumption of no regime shift is amply vindicated in the case of France since 1993, the year when the Bank of France gained formal independence by law for the conduct of monetary policy. We base our position on both institutional and statistical arguments. First, as hinted above in section 2.3, French monetary policy was closely anchored to the policy conducted by the German Bundesbank and, indirectly, to German economic conditions between 1993 and 1999, only because of the commitment of the Bank of France to peg the French franc to the deutsche mark in order

Table 6. Multivariate Chow Tests of a Structural Break in 1999:Q1

	p-value (%)
Macro Factors Only	0.26
Macro + LIQ Factors	0.56
Macro + LEV1 Factors	0.58
Macro + LEV2 Factors	0.57

Note: Bootstrapped p-values (5,000 replications).

to meet nominal convergence requirements during the run-up phase to the EMU. To reflect this, we included some key macroeconomic German series in our macro database, as detailed above in the data section. Second, French macro aggregates tend to co-move strongly with the (reconstructed) euro-area average since the mid-1990s. Since France and Germany both account for about 50 percent of overall euro-area GDP, one should be comfortable with the idea that our macro factors are both quite relevant as summary ingredients of the reaction function of the Bank of France before 1999 and highly correlated with the euro-area measures of activity and inflation the ECB is likely to respond to since 1999. Third, Boivin, Gianonni, and Mojon (2008) have shown that the launch of the euro did not significantly affect the transmission of monetary policy shock in France and Germany. Fourth and last, we carried out standard breakpoint tests for our FAVAR models, positing the first quarter of 1999 as a possible break date. Table 6 presents the results of multivariate Chow sample-split tests. As it is well known that such tests tend to over-reject the null of no break in samples of common sizes, we followed Candelon and Lütkepohl (2001) and computed bootstrapped p-values (with 5,000 replications). The results show that the null of no break in the FAVAR coefficients due to the inception of the euro is confirmed in all cases, at the 26 percent level for the "purely macro" model without bank factors and at levels above 56 percent for the models with LIQ, LEV1, and LEV2 bank-level factors.

Finally, figures 6, 7, and 8 show the estimated impulse response functions of selected macroeconomic indicators to an unexpected tightening of monetary policy. In each figure, the solid lines represent the responses computed for the FAVAR model based on the

Figure 6. Impulse Responses to an Identified Monetary Policy Shock: Model with Macro Factors Only (Solid Line) vs. Model Augmented with Bank LIQ Factors (Dashed Line)

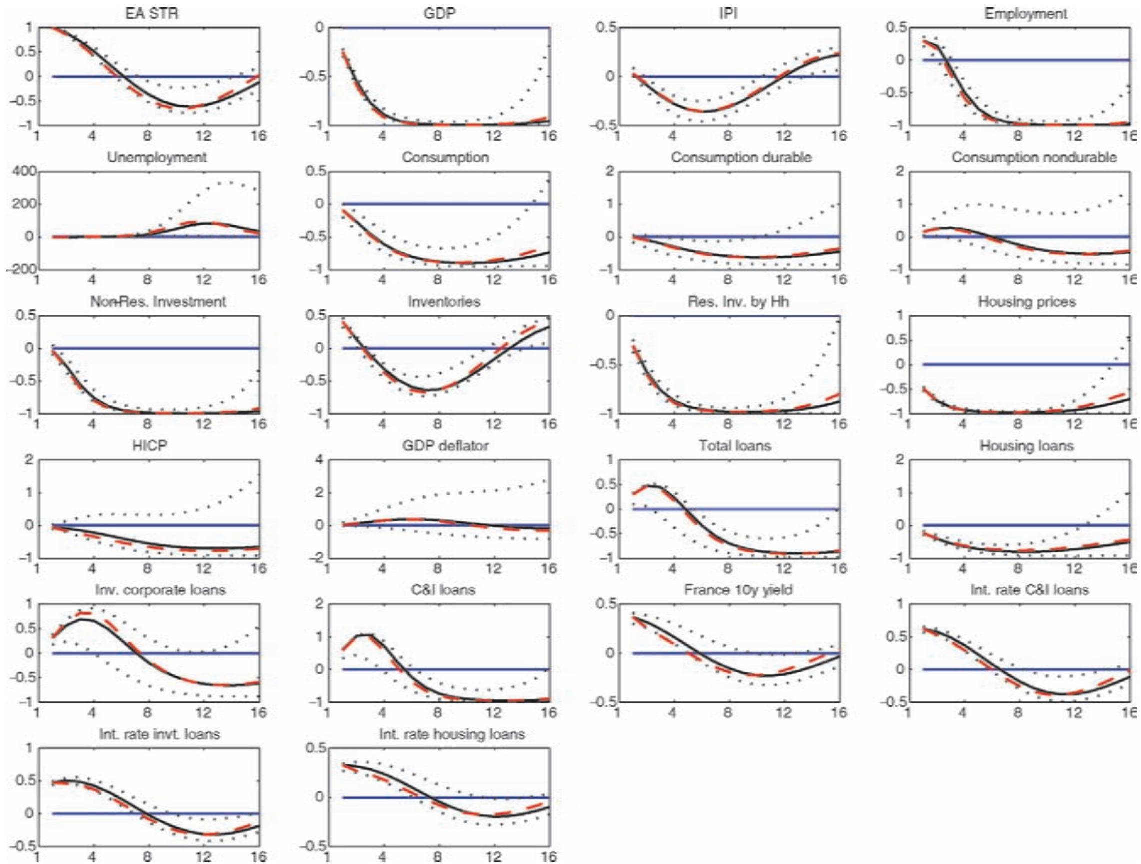


Figure 7. Impulse Responses to an Identified Monetary Policy Shock: Model with Macro Factors Only (Solid Line) vs. Model Augmented with Bank LEV1 Factors (Dashed Line)

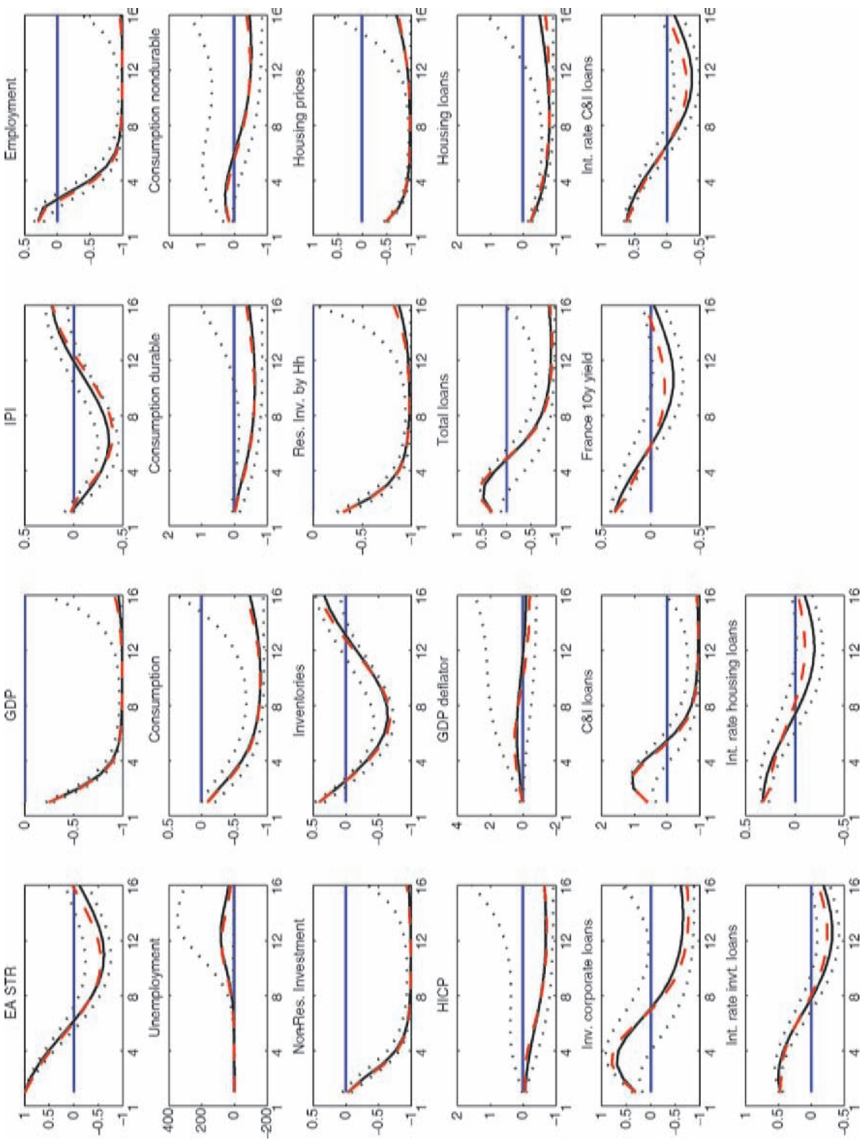
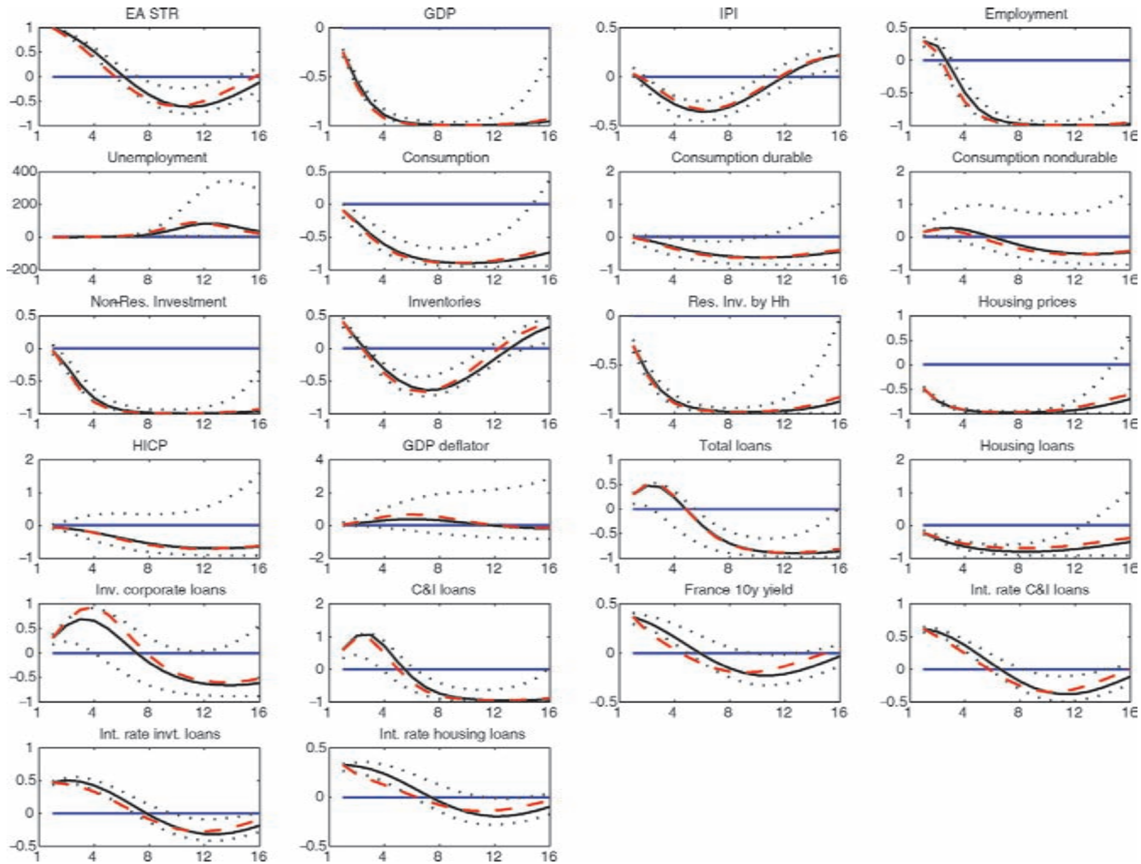


Figure 8. Impulse Responses to an Identified Monetary Policy Shock: Model with Macro Factors Only (Solid Line) vs. Model Augmented with Bank LEV2 Factors (Dashed Line)



sole macro factors, while the dashed lines stand for the responses when the macro model is augmented with two common components extracted from one of the individual bank ratio data sets (LIQ, LEV1, and LEV2, respectively). The impulse responses are plotted along with the 70 percent confidence intervals.[23]

Regarding first the responses computed for the FAVAR model limited to the macro factors, the results look in line with usual findings and economic intuition. Following an unexpected monetary policy tightening, activity declines over the first six quarters and resumes slowly thereafter. Industrial production gets back to its original level within three years, while GDP reverts more slowly. Investment, either residential or not, and inventories react more than consumption, while within consumption, consumption of durables is more negatively affected by an interest rate hike than is the consumption of non-durable goods. The rate of unemployment reacts sluggishly, and employment, which reaches a low after three years, reverts very slowly to the original level, which may be consistent with conventional wisdom for France over this period. Interestingly, consumption prices as measured by the HICP decrease slowly over the first three years without the initial upswing or "price puzzle" that is often obtained within small macro VAR models. The response of the GDP deflator exhibits some price puzzle, but it is also more muted and globally non-significant. Housing prices react vigorously and on impact to an interest rate hike and reach their low within two years. Long-term government bond yield as well as the various bank loan interest rates also react positively on impact to the monetary policy tightening. Interest rates on commercial and industrial (C&I) loans, which are mostly short-term loans indexed on short-term market rates, adjust almost completely, while the pass-through of the short rate to interest rates on housing loans, which are in France mostly long-term fixed-rate mortgage loans, is significantly positive but muted, in line with previous findings for this country.[24] Regarding the response of the various types of bank loans at the aggregate

[23]We use Kilian's (1998) bootstrap procedure to compute the confidence intervals. Note that we bootstrap both the estimation of the factors and of the coefficients, so that the confidence intervals also account for estimation uncertainty about the unobserved factors.

[24]Cf., for instance, Coffinet (2005).

level, housing loans decrease over the first two years, while corporate loans, and notably shorter-term C&I loans, react positively in the short run and recede thereafter.

The "puzzling" positive response of short-term C&I loans to a monetary tightening has already been documented on U.S. data (cf. Kashyap and Stein 1995; Morgan 1998; Den Haan, Sumner, and Yamashiro 2007). Several types of explanations for this temporary increase can be found in the literature. A first line of reasoning points to a demand effect by firms, which may have to finance an inventory buildup following a monetary tightening or have to bear temporarily a higher cost for their working capital (cf. Bernanke and Gertler 1995). Other authors look for supply effects by banks themselves, which may want to optimize the return on their credit portfolio and/or adjust their (risk-weighted) assets structure to keep complying with capital regulation in spite of the adverse effects of the monetary tightening on their interest revenues and hence on their equity base (cf. Van den Heuvel 2002; Den Haan, Sumner, and Yamashiro 2007). These banks would therefore shift their portfolio toward short-term loans and out of longer-term credit, which either typically yields fixed interest rates (like mortgage loans) or requires a higher capital coverage (like long-term loans to non-financial firms, at least under Basel I bank capital regulations). However, even if banks aim at reducing loan supply, this may be delayed by prevailing loan commitments to the benefits of larger firms, which account for the bulk of commercial bank credit. Indeed, large firms frequently borrow from commercial banks under loan commitment contracts so as to secure the volume and conditions of the loans they have over a pre-agreed period. As Morgan (1998) shows in the U.S. case, loans without commitments do contract after a tightening monetary policy shock, while small firms also complain about tighter credit conditions offered by banks. Meanwhile, loans under commitments do not falter, or they even increase.[25] In our case, the significant positive short-term response of inventories points toward a dominant role of credit demand by firms in the positive response of C&I loans to a monetary shock. Besides, the similar responses of these

[25]Note that the argument about loan commitments may also be relevant to explain the shape of the response of bank loans for investment purpose.

loans in the macro FAVAR and in the model augmented for common components extracted from total leverage ratios (LEV1) hints that the potential bearing of loan supply effects due to bank capital constraints is limited here.

Overall, figures 6, 7, and 8 show that the responses in the models with bank-level ratio factors are very close, and at least not significantly different at the 70 percent level of confidence, to the responses obtained in the simpler models that include only macroeconomic information. Our exercise suggests thus that the specific reaction of individual banks to a monetary policy shock and the feedback of the induced changes in bank balance sheets on macroeconomic variables do not significantly alter the transmission process of monetary policy to the macro economy. This does not mean that the banking system and the way it interacts with non-financial private agents is a pure veil, but merely that the information already included in monetary macro variables like aggregate flows of bank credit for housing or corporate investment purpose is sufficient to capture the macroeconomic consequences of the credit channel. To that extent, the differentiated reactions to monetary policy shocks that are associated with heterogeneities in individual banks' balance sheet structures appear to be largely irrelevant from a macroeconomic point of view.

How do our findings relate with earlier literature? Considering the vast amount of studies on the credit channel, we find it more useful to focus this discussion on the differences between this study and two contributions by Ramey (1993) and Peek, Rosengren, and Tootell (1999, 2003), which have particularly close connections with two dimensions of our approach.

In a somewhat older paper, Ramey (1993) did a counterfactual exercise that is formally close to the one we conducted in this section. Using a small-scale VAR model of the U.S. economy with four variables (output, money, credit, and the federal funds rate), she sets alternatively to zero the coefficients of the policy variable in either the money or credit equation, which is equivalent, she claims, to shutting down either the money or credit channel of monetary transmission. She then compares the impulse responses of output to a policy shock she obtains with either of the restricted models to the impulse response from the unrestricted VAR and concludes that the credit channel is unimportant in explaining monetary transmission. In his discussion of her paper, Bernanke (1993) sharply criticizes

Ramey's reading of her results, pointing out notably that (i) they are consistent with both the money and the credit views, as both views imply a quick reduction in bank liabilities, and that (ii) this device alone cannot solve the age-old identification issue of bank loan supply versus demand effects. Although we agree with Bernanke's points, we think that they do not apply to the results presented in this paper. First, we are not interested in assessing the relative strength of the money channel (or, to put it in more modern terms, of the interest rate channel) vis-à-vis the credit channel in the particular case of the French economy. Since our macroeconomic database includes a list of credit aggregates and bank loan interest rates as well as monetary aggregates, our (restricted) baseline FAVAR with only macro factors does not exclude the possibility of operative bank lending and balance sheet channels. Our point is merely to assess the macroeconomic relevance of findings of the empirical literature in the vein of Kashyap and Stein (2000), showing that individual banks with different characteristics in terms of, notably, liquidity and capitalization react differently to monetary shocks, which creates a potential for credit restrictions by at least some banks. Neither do we claim that we can identify (the absence of) loan supply effects using our methodology. Indeed, we cannot tell a priori whether changes in banks' financial conditions as captured by the factors are exogenous or driven by some other shocks to loan demand: even innovations to the second factors, which take more account of changes in heterogeneity of the ratios across banks, can reflect the adjustment to idiosyncratic (e.g., industry specific) shocks that affect the customers of some banks more than others and thus the demand for credit addressed to those banks. To conclude on this, we thus do not read our results as proving that loan supply effects are unimportant. Instead, we just conclude that, at least in normal times, this heterogeneity in banks' financial conditions does not matter much for explaining monetary transmission at the macro level.

More recently, Peek, Rosengren, and Tootell (1999, 2003) also tried to reconcile micro bank information with macro outcomes as they used detailed confidential supervisory information to construct an aggregate indicator of banks' financial health (the share of assets held by banks viewed by bank regulators as likely to fail, i.e., those with a "CAMEL" rating of 5). Running univariate regressions, they first find that their bank health variable (but no other summaries of bank leverage and liquidity ratios) contains useful marginal information to forecast unemployment and inflation up to four quarters ahead, which suggests that the Federal Reserve should look carefully at such bank-level information from supervisory sources to conduct monetary policy. On the basis of probit models of the Federal Reserve's target rate decisions, they also find that the Federal Reserve does actually take into account this information. In their second paper, they then provide evidence that shocks to their CAMEL indicator do reflect shocks to bank credit supply, which implies that part of the forecasting power of the bank health indicator for output has a causal interpretation. Our results contrast with theirs on two points. Firstly, contrary to part of their first findings, our Granger causality tests reported above suggest that even simple bank leverage ratios do help to improve forecasts of macroeconomic activity. Secondly and more importantly, Peek, Rosengren, and Tootell (2003) do not formally examine, as we do, the consequences of monetary policy shocks, although they point in their conclusion to the relevance of their study for quantitative assessments of the bank lending channel. Instead, they focus on the (difficult) task of identifying loan supply shocks per se and show that such effects matter for U.S. macroeconomic fluctuations. However, the mere fact that loan supply shocks exist and are important is a necessary but not a sufficient condition to prove that banks' reactions to monetary policy shocks, as a consequence of financial frictions that constrain adjustments to their balance sheets, do amplify the effects of the policy moves. Indeed, as Peek, Rosengren, and Tootell (2003) show that their findings are robust when the bank health indicator has been priorily orthogonalized with respect to the federal funds rate (and other state variables), one may think that it is the exogenous part in innovations to banks' health that matters (i.e., true loan supply shocks), not necessarily the endogenous reaction to other shocks. To conclude, their results do not contradict ours as far as the credit channel is concerned, but instead suggest interesting avenues for further research on the effects of shocks to banks' financial conditions using our FAVAR framework.

5. Conclusion

In this paper we aim to quantify whether changes in banks' financial conditions at the microeconomic level matter at the macroeconomic level, notably by altering the monetary policy transmission mechanism. Using a unique and comprehensive database of individual bank balance sheets, we set up a FAVAR framework that allows us to summarize both overall macroeconomic conditions in the French economy and the financial conditions of banks resident in France with a small number of factors.

Within this framework, we first provide evidence that the information contained in three types of individual bank financial ratios—capitalization, liquidity, and leverage ratios—explains a substantial part of macroeconomic fluctuations in some aggregate variables, most notably those related to the housing market (housing prices, residential investment, and housing loans). Moreover, we find that the first two principal components extracted from individual bank leverage and liquidity ratios have a significant predictive power for macroeconomic conditions, which suggests that there is potentially a scope for active macroprudential policies aimed at constraining changes in these ratios.

Finally, we compare the impulse response functions of alternative FAVAR models that either allow for or restrict the feedback effects of bank-level factors on macroeconomic ones. We find that the information contained in individual bank ratios, including changes in balance sheet heterogeneity among banks, does not matter much for the transmission of monetary policy shocks.

This work could be extended in at least three ways. First, instead of following the methodology of Boivin and Gianonni (2009), we could implement the recently developed dynamic hierarchical factor model of Ng, Moench, and Potter (2009), whose advantage consists in distinguishing series-specific variations from two types of common variations: those from factors that are common to units within a block, and those from factors that are common across blocks. This could allow us to simultaneously consider the inclusion of all three kinds of bank ratios within the same FAVAR model. Second, it could be interesting to look at the effects of real demand shocks and see whether bank factors are relevant for the transmission of such shocks to the economy (as per the financial accelerator hypothesis). Finally, taking stock of the results of causality tests presented here and as suggested above by our discussion of Peek, Rosengren, and Tootell (2003), we could investigate within a FAVAR framework the macroeconomic consequences of shocks to trends or dispersion in banks' leverage or liquidity conditions, in the spirit of the tests developed by Gilchrist, Yankov, and Zakrajsek (2009). This is left for further research.

Appendix. Data Sets

Macroeconomic Series

The format contains the series number; data span (in quarters); transformation code; and series description as appears in the database. The transformation codes are: 1 - no transformation; 5 - first difference of logarithm. The series were taken from the Monetary Statistics database of the Banque de France, the Bank for International Settlements (BIS), the Eurostat database, the IN/IP INSEE database, and the OI/OP OECD database.

	France		
1	1993:1-2009:1	1	EMU 3-month EURIBOR, total, end of period
2	1993:1-2009:1	5	Total loans of French credit institutions, SA
3	1993:1-2009:1	5	Total loans to NFIs, SA
4	1993:1-2009:1	5	Loans to NFIs for cash needs, SA
5	1993:1-2009:1	5	Loans to NFIs for investing needs, SA
6	1993:1-2009:1	5	Total loans to households, SA
7	1993:1-2009:1	5	Mortgage loans to households, SA
8	1993:1-2009:1	5	Monetary aggregate M3, outstanding amounts at the end of the period (stocks)
9	1993:1-2009:1	5	Gross domestic product at market price, chain linked volumes, reference year 2000, SA
10	1993:1-2009:1	1	Production of total industry, index, SA
11	1993:1-2009:1	5	Final consumption of households and NPISH's (private consumption), chain linked volumes, reference year 2000, SA
12	1993:1-2009:1	5	Private consumption of households, durable goods, chain linked volumes, reference year 2000, SA

13	1993:1-2009:1	5	Private consumption of households, non-durable goods, chain linked volumes, reference year 2000, SA
14	1993:1-2009:1	5	Exports of goods and services, chained volume estimates, SA
15	1993:1-2009:1	5	Imports of goods and services, chained volume estimates, SA
16	1993:1-2009:1	5	Gross fixed capital formation of financial institutions, goods and services, SA
17	1993:1-2009:1	5	Gross fixed capital formation of public services, goods and services, SA
18	1993:1-2009:1	5	Gross fixed capital formation of households, goods and services, SA
19	1993:1-2009:1	5	Gross fixed capital formation of households, building and civil engineering, SA
20	1993:1-2009:1	5	Gross fixed capital formation of households, real estate services, SA
21	1993:1-2009:1	5	Gross fixed capital formation of NFIs, goods and services, SA
22	1993:1-2009:1	5	Gross fixed capital formation of NFIs, building and civil engineering, SA
23	1993:1-2009:1	5	Gross fixed capital formation of all sectors, goods and services, SA
24	1993:1-2009:1	5	Employees, full time equivalent, SA
25	1993:1-2009:1	5	Unemployment rate, BIT definition, SA
26	1993:1-2009:1	1	Increase in stocks, end of period (%)
27	1993:1-2009:1	5	Construction costs, total, cost of materials, NSA, Index, 1953 Oct = 100
28	1993:1-2009:1	5	Cost of construction: multiple dwellings, end of period
29	1993:1-2009:1	5	CAC40, end of day
30	1993:1-2009:1	5	Oil price, brent crude - 1 month forward, level
31	1993:1-2009:1	5	GDP Deflator, index publication base SA
32	1993:1-2009:1	5	Consumer price index, harmonised, SA
33	1993:1-2009:1	5	Consumer price index, end of period
34	1993:1-2009:1	5	CPI (households, base 1998) - Food and non-alcoholic drinks
35	1993:1-2009:1	5	CPI (households, base 1998) - Alcoholic drinks

36	1993:1-2009:1	5	CPI (households, base 1998) - Clothing and footwear
37	1993:1-2009:1	5	CPI (households, base 1998) - Housing, water, gas, electricity and other combustibles
38	1993:1-2009:1	5	CPI (households, base 1998) - Furniture, domestic equipment and house keeping
39	1993:1-2009:1	5	CPI (households, base 1998) - Health
40	1993:1-2009:1	5	CPI (households, base 1998) - Transportation
41	1993:1-2009:1	5	CPI (households, base 1998) - Communications
42	1993:1-2009:1	5	CPI (households, base 1998) - Leisure and culture
43	1993:1-2009:1	5	CPI (households, base 1998) - Hotels, cafes and restaurants
44	1993:1-2009:1	5	CPI (households, base 1998) - Other goods and services
45	1993:1-2009:1	5	PPI - Buldings
46	1993:1-2009:1	5	PPI - Extractive industry, energy, water, wastes management and depolluting
47	1993:1-2009:1	5	PPI - Manufacturing industry
48	1993:1-2009:1	1	Long-term interest rate on government bonds
49	1993:1-2009:1	1	Average rate on loans to NFIs $<$ 1 year, new contracts
50	1993:1-2009:1	1	Average rate on loans to NFIs $>$ 1 year, new contracts
51	1993:1-2009:1	1	Average rate on consumer loans to households, new contracts
52	1993:1-2009:1	1	Average rate on mortgage loans to households, new contracts
53	1993:1-2009:1	1	Consumer confidence indicator, SA
54	1993:1-2009:1	1	Food-processing industry - production capacity utilization, %, SA
55	1993:1-2009:1	1	Consumption goods industry - production capacity utilization, %, SA
56	1993:1-2009:1	1	Automobile industry - production capacity utilization, %, SA
57	1993:1-2009:1	1	Equipment industry - production capacity utilization, %, SA

58	1993:1-2009:1	1	Intermediary goods industry - production capacity utilization, %, SA
59	1993:1-2009:1	1	Business climate in industry
60	1993:1-2009:1	1	Business climate in trade services
	Germany		
61	1993:1-2009:1	5	Monetary aggregate M2, outstanding amounts at the end of the period (stocks)
62	1993:1-2009:1	5	Gross domestic product at market price, chain linked volumes, reference year 2000, SA
63	1993:1-2009:1	5	Employees, persons (Thousands, SA)
64	1993:1-2009:1	5	Unemployed persons (Thousands, SA)
65	1993:1-2009:1	5	Gross domestic product, implicit price deflator, SA
66	1993:1-2009:1	5	HICP, SA
67	1993:1-2009:1	1	Consumer confidence indicator, SA
68	1993:1-2009:1	1	Long-term interest rate on government bonds

*NFI - non-financial institutions; NPISH - non-profit institutions serving households; PPI - Producer Price Index; SA - seasonally adjusted.

Disaggregated Bank Balance Sheet Series

The format contains the series number; data span (in quarters); bank identification code (CIB); and the name of the credit institutions as appears in the database. The ratios were computed using balance sheet information from the BAFI database of the French Banking Commission, Banque de France.

Liquidity and Leverage Ratios

No.	CIB	Period	Credit Institution
1	7	1993:2-2009:1	CUMUL BQ POP HORS AGREMENT COLLECTIF SCM
2	10057	1993:2-2009:1	STE BORDELAISE DE CIT IND ET COMMERCIAL
3	10096	1993:2-2009:1	LYONNAISE DE BANQUE L,B,
4	10178	1993:2-2009:1	BANQUE CHAIX
5	10188	1993:2-2009:1	BANQUE CHALUS

6	10228	1993:2-2009:1	BANQUE LAYDERNIER
7	10268	1993:2-2009:1	BANQUE COURTOIS
8	10468	1993:2-2009:1	BANQUE RHONE-ALPES
9	10558	1993:2-2009:1	BANQUE TARNEAUD
10	10638	1993:2-2009:1	CREDIT COMMERCIAL DU SUD-OUEST
11	11188	1993:2-2009:1	RCI BANQUE
12	11449	1993:2-2009:1	BANQUE THEMIS
13	11808	1993:2-2009:1	BANQUE FEDERATIVE DU CREDIT MUTUEL
14	12280	1993:2-2009:1	SOCRAM BANQUE
15	12869	1993:2-2009:1	BANQUE ACCORD
16	12939	1993:2-2009:1	BANQUE DUPUY DE PARSEVAL
17	13259	1993:2-2009:1	BANQUE KOLB
18	13539	1993:2-2009:1	BANQUE SOLFEA
19	17290	1993:2-2009:1	DEXIA CREDIT LOCAL
20	17679	1993:2-2009:1	STE DE BANQUE ET D'EXPANSION-SBE (2EME)
21	18029	1993:2-2009:1	BNP PARIBAS PERSONAL FINANCE
22	18189	1993:2-2009:1	CIE GLE DE CIT AUX PARTICULIERS CREDIPAR
23	18359	1993:2-2009:1	OSEO FINANCEMENT
24	18370	1993:2-2009:1	BANQUE FINAMA
25	18609	1993:2-2009:1	CAISSE CENTRALE CIT IMMOB DE FRANCE-3CIF
26	18839	1993:2-2009:1	B F T BANQUE DE FINT ET DE TRESORERIE
27	18889	1993:2-2009:1	CORTAL CONSORS
28	19239	1993:2-2009:1	NATIXIS TRANSPORT FINANCE
29	19269	1993:2-2009:1	GENEBANQUE
30	19870	1993:2-2009:1	STE DES PAIEMENTS PASS - S2P
31	22040	1993:2-2009:1	CONFEDERATION NATIONALE DU CREDIT MUTUEL
32	30002	1993:2-2009:1	CREDIT LYONNAIS
33	30003	1993:2-2009:1	STE GENERALE
34	30004	1993:2-2009:1	BNP PARIBAS
35	30027	1993:2-2009:1	BANQUE SCALBERT DUPONT - CIN

36	30047	1993:2-2009:1	CREDIT INDUSTRIEL DE L OUEST
37	30056	1993:2-2009:1	HSBC FRANCE
38	30066	1993:2-2009:1	CREDIT INDUSTRIEL ET COMMERCIAL - CIC
39	30076	1993:2-2009:1	CREDIT DU NORD
40	30087	1993:2-2009:1	BANQUE CIC EST
41	30488	1993:2-2009:1	FORTIS BANQUE FRANCE
42	30568	1993:2-2009:1	BANQUE TRANSATLANTIQUE
43	30958	1993:2-2009:1	BNP PARIBAS LEASE GROUP
44	31489	1993:2-2009:1	CALYON
45	39996	1993:2-2009:1	GROUPE CREDIT AGRICOLE
46	40168	1993:2-2009:1	BANQUE DE BRETAGNE
47	41199	1993:2-2009:1	BANCO POPULAR FRANCE
48	42959	1993:2-2009:1	ELECTRO BANQUE
49	43799	1993:2-2009:1	BANQUE DE GESTION PRIVEE INDOSUEZ - BGPI
50	43899	1993:2-2009:1	UNION DE BANQUES ARABES ET FRSES U B A F
51	44449	1993:2-2009:1	LIXXCREDIT
52	50140	1993:2-2009:1	CMP-BANQUE

References

Adrian, T., and H. S. Shin. 2009a. "Financial Intermediaries and Monetary Economics." Federal Reserve Bank of New York Staff Report No. 398 (October).

———. 2009b. "Financial Intermediaries, Financial Stability, and Monetary Policy." In *Maintaining Stability in a Changing Financial System*, 287–334. Proceedings of the Federal Reserve Bank of Kansas City Economic Symposium at Jackson Hole, Wyoming, August 21–23, 2008.

———. 2009c. "Prices and Quantities in the Monetary Policy Transmission Mechanism." Federal Reserve Bank of New York Staff Report No. 396 (October).

Angeloni, I., A. Kashyap, and B. Mojon, eds. 2003. *Monetary Policy Transmission in the Euro Area.* Cambridge: Cambridge University Press.

Ashcraft, A. 2006. "New Evidence on the Lending Channel." *Journal of Money, Credit, and Banking* 38 (3): 751–76.

Bai, J., and S. Ng. 2002. "Determining the Number of Factors in Approximate Factor Models." *Econometrica* 70 (1): 191–221.

———. 2004. "A PANIC Attack on Unit-Roots and Cointegration." *Econometrica* 72 (4): 1127–77.

Bernanke, B. S. 1993. "How Important Is the Credit Channel in the Transmission of Monetary Policy? A Comment." *Carnegie-Rochester Conference Series on Public Policy* 39 (1): 47–52.

———. 2004. "The Logic of Monetary Policy." Remarks before the National Economists Club, Washington, DC, December 2.

Bernanke, B. S., and A. S. Blinder. 1992. "The Federal Funds Rate and the Channels of Monetary Transmission." *American Economic Review* 82 (4): 901–21.

Bernanke, B. S., J. Boivin, and P. Eliasz. 2005. "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach." *Quarterly Journal of Economics* 120 (1): 387–422.

Bernanke, B. S., and M. L. Gertler. 1995. "Inside the Black Box: The Credit Channel of Monetary Policy Transmission." *Journal of Economic Perspectives* 9 (4): 27–48.

Bernanke, B., and C. Lown. 1991. "The Credit Crunch." *Brookings Papers on Economic Activity* 2: 205–39.

Blinder, A. 1998. *Central Banking in Theory and Practice.* Cambridge, MA: MIT Press.

Boivin, J., and M. P. Giannoni. 2009. "Global Forces and Monetary Policy Effectiveness." In *International Dimensions of Monetary Policy*, ed. J. Galí and M. Gertler, 429–78. Chicago: University of Chicago Press.

Boivin, J., M. P. Giannoni, and B. Mojon. 2008. "How Has the Euro Changed the Monetary Transmission?" In *NBER Macroeconomics Annual 2008*, ed. D. Acemoglu, K. Rogoff, and M. Woodford. Chicago: NBER and University of Chicago Press.

Boivin, J., M. Gianonni, and D. Stevanovic. 2009. "Dynamic Effects of Credit Shocks in a Data-Rich Environment." Unpublished Manuscript, University of Montreal.

Candelon, B., and H. Lütkepohl. 2001. "On the Reliability of Chow-Type Tests for Parameter Constancy in Multivariate Dynamic Models." *Economics Letters* 73 (2): 155–60.

Coffinet, J. 2005. "The Single Monetary Policy and the Interest Rate Channel in France and in the Euro Area." Banque de France Bulletin Digest 139 (July).

Commission bancaire. 2009. *Rapport annuel de la Commission bancaire 2008*. Paris: Banque de France.

Dave, C., S. J. Dressler, and L. Zhang. 2009. "The Bank Lending Channel: A FAVAR Analysis." Working Paper No. 4, Villanova School of Business Economics.

Den Haan, W. J., S. W. Sumner, and G. Yamashiro. 2007. "Bank Loan Portfolios and the Monetary Transmission Mechanism." *Journal of Monetary Economics* 54 (3): 904–24.

Ehrmann, M., L. Gambacorta, J. Martinez-Pagés, P. Sevestre, and A. Worms. 2001. "Financial Systems and the Role of Banks in Monetary Policy Transmission in the Euro Area." ECB Working Paper No. 105.

Ehrmann, M., and A. Worms. 2004. "Bank Networks and Monetary Policy Transmission." *Journal of the European Economic Association* 2 (6): 1148–71.

Engler, P., T. Jokipii, C. Merkl, P. Rovira Kaltwasser, and L. Vinhas de Souza. 2005. "The Effect of Capital Requirement Regulation on the Transmission of Monetary Policy: Evidence from Austria." Working Paper No. 99, Central Bank of Austria.

Gambacorta, L., and P. E. Mistrulli. 2004. "Does Bank Capital Affect Lending Behavior?" *Journal of Financial Intermediation* 13 (4): 436–57.

Giannone, D., M. Lenza, and L. Reichlin. 2009. "Monetary Analysis with Disaggregated Variables." Unpublished Manuscript, European Central Bank.

Gilchrist, S., V. Yankov, and E. Zakrajsek. 2009. "Credit Market Shocks and Economic Fluctuations: Evidence from Corporate Bond and Stock Markets." *Journal of Monetary Economics* 56 (4): 471–93.

Hülsewig, O., P. Winker, and A. Worms. 2004. "Bank Lending in the Transmission of Monetary Policy: A VECM Analysis for Germany." *Jahrbücher für Nationalökonomie und Statistik* (*Journal of Economics and Statistics*, Justus Liebig University Giessen, Department of Statistics and Economics) 224 (5): 511–29.

Kashyap, A. K., and J. C. Stein. 1995. “The Impact of Monetary Policy on Bank Balance Sheets.” *Carnegie-Rochester Conference Series on Public Policy* 42 (1): 151–95.

———. 2000. “What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?” *American Economic Review* 90 (3): 407–28.

Kilian, L. 1998. “Small-Sample Confidence Intervals for Impulse Response Functions.” *Review of Economics and Statistics* 80 (2): 218–30.

Kishan, R. P., and T. P. Opiela. 2000. “Bank Size, Bank Capital and the Bank Lending Channel.” *Journal of Money, Credit, and Banking* 32 (1): 121–41.

Loupias, C., F. Savignac, and P. Sevestre. 2002. “Is There a Bank Lending Channel in France? Evidence from Bank Panel Data.” Notes d’Etudes et de Recherche No. 92, Banque de France.

Mésonnier, J.-S. 2008. “Bank Loan Portfolios, Bank Heterogeneity and the Bank Lending Channel: New Macro Evidence for France.” Unpublished Manuscript.

Morgan, D. P. 1998. “The Credit Effects of Monetary Policy: Evidence Using Loan Commitments.” *Journal of Money, Credit, and Banking* 30 (1): 102–118.

Ng, S., E. Moench, and S. Potter. 2009. “Dynamic Hierarchical Factor Models.” Working paper presented at the NBER annual conference.

Peek, J., and E. Rosengren. 2000. “Collateral Damage: Effects of the Japanese Bank Crisis on Real Activity in the United States.” *American Economic Review* 90 (1): 30–45.

Peek, J., E. Rosengren, and G. Tootell. 1999. “Is Bank Supervision Central to Central Banking?” *Quarterly Journal of Economics* 114 (2): 629–53.

———. 2003. “Identifying the Macroeconomic Effect of Loan Supply Shocks.” *Journal of Money, Credit, and Banking* 35 (6): 932–46.

Ramey, V. 1993. “How Important Is the Credit Channel in the Transmission of Monetary Policy?” *Carnegie-Rochester Conference Series on Public Policy* 39 (1): 1–45.

Stock, J. H., and M. W. Watson. 1999. “Forecasting Inflation.” *Journal of Monetary Economics* 44 (2): 293–335.

———. 2002. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business and Economic Statistics* 20 (2): 147–62.

Van den Heuvel, S. K. 2002. "Does Bank Capital Matter for Monetary Policy Transmission?" *Economic Policy Review* (Federal Reserve Bank of New York) 8 (1): 259–65.

Woo, D. 2003. "In Search of 'Capital Crunch': Supply Factors behind the Credit Slowdown in Japan." *Journal of Money, Credit, and Banking* 35 (6): 1019–38.

Woodford, M. 2005. *Central Bank Communication and Policy Effectiveness.* Proceedings of the Federal Reserve Bank of Kansas City Economic Symposium at Jackson Hole, Wyoming, August 25–27.

Discussion of "Banks' Financial Conditions and the Transmission of Monetary Policy: A FAVAR Approach"

Donald P. Morgan
Federal Reserve Bank of New York

This paper has much to recommend it. The authors, Ramona Jimborean and Jean-Stéphane Mésonnier, make a solid contribution to the emerging literature on factor-augmented vector autoregression (FAVAR) models. They deftly handle their data and circumvent potential problems in the form of too many banks (to fit their framework) and the launch of the European Union (EU). They motivate and exposit the paper very well. Most importantly, I agree with their findings if not all inferences they draw from their findings.

My comments come in three parts: their contribution to the FAVAR literature, their contribution to the broader literature that tries to identify causal/structural bank lending and balance sheets and the transmission of monetary policy, and last but not least, their contribution to macroprudential supervision, the topic of this conference.

1. Contribution to FAVAR Literature

If this were Hollywood, FAVAR would be described as principal components (PC) analysis meets vector autoregressions (VAR). Given the lengths of most time series, VAR models are limited to half a dozen or so variables. Such parsimonious specifications clearly leave room for omitted variables and thus misidentification of shocks. Lately, macro econometricians have begun to use PC analysis to extract common components, or factors, from micro data sets (on banks or bonds, for example) and then add those factors to a VAR comprising macro factors or ordinary macro time series.

Given data-rich (and data-hungry) central bankers, FAVAR may become, if it is not already, the standard for macro and monetary econometrics. The FAVAR literature is small, however, because FAVAR programs are not available off the shelf. The Jimborean and Mésonnier paper is one of only about half a dozen such papers to date, including most notably Bernanke, Boivin, and Elias (2005). While Jimborean and Mésonnier are not the first to look for evidence of a bank lending channel using FAVAR models, they are the first to look at two bank factors that figure prominently in the literature on macroprudential supervision and in popular narratives of the crisis: liquidity and leverage.[1]

Their main findings are that (i) bank liquidity and leverage factors predict macro variables, including housing market proxies, and (ii) the impact of a monetary impulse on macro variables is invariant to bank factors. I will discuss their second finding more later in my comments. A curious aspect of their first finding is that the liquidity factor appears to predict macro factors *less* during the crisis. In their table 5, the liquidity factors are only significant in the pre-crisis period. By contrast, the leverage factors are significant over the full sample and pre-crisis period. That goes against the narrative that the U.S. bank crisis was a liquidity crisis, not a solvency (leverage) crisis.

Two cautions are worth noting in interpreting their results. First, to fit the FAVAR framework and work with a balanced panel, the authors wound up studying only 52 of their initial 620 institutions. Because they include only banks that are in the sample over the full sample period, they may have, as they admit in a footnote, sample selection bias. As a robustness check, they might be advised to follow Dave, Dressler, and Zhang (2009) and take a random sample of banks.

Second, while they are data rich, they are still degrees-of-freedom poor. As a result, they cannot include all the bank liquidity and leverage factors simultaneously in the FAVAR models. Thus, they cannot answer the natural question of whether liquidity matters given leverage, or vice versa.

[1]The FAVAR analysis in Dave, Dressler, and Zhang (2009) adds disaggregated bank lending data to investigate a lending channel of monetary policy in the United States.

2. Contribution to Literature Identifying Bank Lending and Balance Sheet Channels

The authors' second finding is that the impact of a monetary impulse on macro variables is invariant to bank factors. This finding seems to fit more with the literature investigating the causal/structural link between monetary and financial impulses and the macro economy, and I think it is here that their paper contributes the least. Their paper is essentially a FAVAR version of the VAR analysis by Ramey (1993).[2] Like the authors, Ramey tries to identify the credit channel of monetary policy by shutting down that channel (by zeroing out certain off-diagonal coefficient elements), then seeing whether the response of output to a monetary impulse changes appreciably. Like Jimborean and Mésonnier, Ramey finds that the impact of a monetary policy shock barely changes when most proxies for the credit channel are shut down. Her bottom line is as follows: "the marginal effect of some of the leading credit channels is negligible" (Ramey 1993, p. 43).

In his comment on Ramey, then-professor Bernanke (1993) admonishes that trying to uncover structural relationships using reduced-form timing relationships is futile. He admits that the same point applies to much of the 1990s' literature on credit channels, including Bernanke and Blinder (1992) and Romer and Romer (1990).

In light of the Bernanke (1993) critique of Ramey (1993), Jimborean and Mésonnier are careful not to infer too much from their reduced-form timing relationships; they merely conclude that central bankers need not monitor bank liquidity and leverage positions when forecasting how their actions will affect the macro economy. To be safe, I wish the authors had included a warning in their conclusion *not* to make any structural inferences from their findings on the transmission of monetary policy.

The authors were a little light with the literature review. Except for a paper by Ashcraft, they neglect recent studies by EU researchers, such as Jiménez et al. (2010), that use bank- and borrower-level data to identify bank loan and balance sheet effects. The identification in those borrower-level studies is very fine, and

[2] Actually, Ramey (1993) used a vector error-correction augmented VAR.

the authors consistently turn up evidence of credit effects. Including those studies in the literature review would give the audience the proper perspective on their strictly time-series evidence Jimborean and Mésonnier provide.

3. Contribution to Macroprudential Supervision

The authors conclude that the predictive power of bank factors suggests a potential scope for "macroprudential policies aimed at dampening the procyclical effects of wide-ranging changes in banks' balance sheet structure." That conclusion seems premature. Finding a predictive relationship between liquidity and leverage does not, following the logic above, imply a causal relationship, so a central banker would not necessarily want to embark on liquidity and leverage policies based on those predictive relationships. On a different point, the authors tell central bankers to monitor bank liquidity and leverage factors to better predict macro outcomes. If they are advising central bankers to pay *more* attention to those factors, it would be useful if the authors could tell central bankers to what factors they can pay less attention. Central bankers, like everyone else, have limited attention. In that regard, it would be helpful if the authors could find some variables that are insignificant, given liquidity and leverage factors.

Actually, their second finding does suggest something central bankers can ignore in their deliberations: the feedback from monetary policy to the bank factors and thence the macro economy. There seems to be a little tension between their first result and the second. The first result says the bank factors do predict macro outcomes, so macroprudential supervision (of those factors) is advised. The second result says those factors do not affect how monetary transmission is associated with macroeconomic outcomes. It seems curious to say to the central banker in charge that tight liquidity or high leverage may be a drag on economic activity, but that drag can be ignored in setting the course for monetary policy.

At first blush, there also seems to be tension between their results and those of Peek, Rosengren, and Tootell (1999). Peek, Rosengren, and Tootell (1999) find that confidential, supervisory measures of U.S. bank health predict U.S. macro variables and that that those bank health variables predict U.S. monetary policy. The first result in Jimborean and Mésonnier seems consistent with the first result in Peek, Rosengren, and Tootell (1999). Their second result appears contradictory but in fact is not because Jimborean and Mésonnier are essentially looking at a second, or cross-derivative, comparative dynamic result while Peek, Rosengren, and Tootell (1999) are looking at a first derivative.[3]

In sum, I would say that Jimborean and Mésonnier make valuable contributions to the FAVAR and macroprudential literature. They contribute less to the literature on identifying structural/causal relationships between monetary policy and the macro economy, but they were careful not to promise too much in that dimension. I recommend their paper.

References

Bernanke, B. 1993. "How Important Is the Credit Channel in the Transmission of Monetary Policy? A Comment." *Carnegie-Rochester Conference Series on Public Policy* 39 (1): 47–52.

Bernanke, B., and A. Blinder. 1992. "The Federal Funds Rate and the Channels of Monetary Transmission." *American Economic Review* 82 (4): 901–21.

Bernanke, B., J. Boivin, and P. Elias. 2005. "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach." *Quarterly Journal of Economics* 120 (1): 387–422.

Dave, C., S. Dressler, and L. Zhang. 2009. "The Bank Lending Channel: A FAVAR Analysis." Working Paper No. 4, Villanova School of Business, Department of Economics and Statistics.

Jiménez, G., S. Ongena, J.-L Peydró, and J. Saurina. 2010. "Credit Supply: Identifying Balance-Sheet Channels with Loan Applications and Granted Loans." ECB Working Paper No. 1179.

Peek, J., E. Rosengren, and G. Tootell. 1999. "Is Bank Supervision Central to Central Banking?" *Quarterly Journal of Economics* 114 (2): 629–53.

[3] Letting M denote the monetary policy variable, Y economic activity, and H the bank health or factors, Peek, Rosengren, and Tootell (1999) look at $\delta Y/\delta M$ and $\delta M/\delta H$, while Jimborean and Mésonnier look at $\delta^2 Y/\delta^2 M$.

Ramey, V. 1993. "How Important Is the Credit Channel in the Transmission of Monetary Policy?" *Carnegie-Rochester Conference Series on Public Policy* 39 (1): 1–45.

Romer, C., and D. Romer. 1990. "New Evidence on the Monetary Transmission Mechanism." *Brookings Papers on Economic Activity* 21 (1): 149–213.

Banking Crises and Real Activity: Identifying the Linkages

Mark Gertler
New York University

I interpret some key aspects of the recent crisis through the lens of macroeconomic modeling of financial factors.

JEL Codes: E0, E50.

1. Introduction

The recent recession—now known as the "Great Recession"—featured a disruption of financial intermediation of the like not seen since the Great Depression. One of the great challenges macroeconomists face is to adjust existing models to account for these events, as well as the implications for economic policy. Efforts are rapidly under way. A good deal of this new research is surveyed in Gertler and Kiyotaki (2010).

In this commentary I would like to first give an overview of the research that links banking crises to macroeconomic activity. I will then describe how recent data suggests these linkages were at work during the crisis. Finally, I will briefly interpret some of the findings in this session's papers in light of the way these linkages work.

One theme of my comments is that the recent modeling work suggests that credit spreads are likely to be a more reliable indicator than credit quantities of financial distress that feeds into the real economy.

2. Modeling Real/Financial Linkages

In this section I sketch how recent macro modeling incorporates financial factors. I begin by describing the basic way a financial crisis can have detrimental effects on real activity. Key to this notion is the existence of limits to arbitrage which, roughly speaking, permit a gap to emerge between the expected return to capital and the riskless rate that is too large to be explained by risk preferences. The excess required return to capital implies a higher cost of borrowing. In this way, financial factors affect real activity.

I then describe the circumstances in which the origins of a financial crisis might be a banking crisis. In this situation, limits to arbitrage allow a wedge to emerge between the bank lending rate and the riskless rate. The excess bank lending rate in turn pushes up the required cost of capital. Next, I describe how endogenously determined balance sheet constraints (stemming from information/enforcement problems) can be the source of limits to arbitrage. Finally, I discuss the mechanism by which a crisis induces a tightening of these balance sheet constraints, which in turn forces up the spread between the expected return to capital and the riskless rate. The sharp rise in the excess return to capital, in turn, depresses real activity.

2.1 Financial Crises: Basic Concepts

Let R_{kt+1} be the gross rate of return to risky capital, R_{t+1} the gross riskless interest rate, and $\Lambda_{t,t+1}$ the representative household's stochastic discount factor. Then under frictionless financial markets, arbitrage ensures that the difference between the expected discounted return to capital and the discounted safe rate is zero:

$$E_t \Lambda_{t,t+1}(R_{kt+1} - R_{t+1}) = 0. \tag{1}$$

Equation (1) is a basic feature of conventional quantitative macroeconomic models that abstract from financial market frictions. Standard procedure is to log-linearize this equation. This yields up to a first-order equality between the expected return to capital and the riskless rate, where both variables are expressed in terms of deviations from their respective steady-state values. How monetary policy affects investment demand in these models then works as follows: Due to nominal rigidities, the central bank is able to manipulate the riskless rate. Then by arbitrage, the required expected return to capital changes one for one with the riskless rate. Investment demand then moves inversely with changes in the required return to capital.

With capital market frictions, the picture can change substantially. For simplicity, assume that households are able to perfectly insure idiosyncratic consumption risk so that we can still work with the representative household's stochastic discount factor. However, suppose that there exist frictions in the process of channeling funds from households to non-financial firms that impede perfect arbitrage. Then, in general, the expected discounted return to capital can exceed the discounted riskless rate:

$$E_t\Lambda_{t,t+1}(R_{kt+1} - R_{t+1}) \geq 0. \quad (2)$$

The basic idea underlying macro models with financial frictions is to incorporate mechanisms that move this rate gap countercyclically. Then the way financial propagation mechanisms work to enhance business fluctuations is to push up the cost of capital relative to the riskless rate in downturns. This magnifies the overall investment drop, which in turn magnifies the recession. In booms, the mechanism works in reverse. This notion of how financial factors propagate real activity dates back to Bernanke and Gertler (1989).

Within this framework a financial crisis is manifested by a sharp increase in $E_t\Lambda_{t,t+1}R_{kt+1}$ relative to $E_t\Lambda_{t,t+1}R_{t+1}$. The increase in the spread is a product of an explicitly modeled disruption of financial markets. The sharp increase in the cost of capital produces a collapse in durable goods spending.

2.2 Banking Crises

Up to this point we have said nothing about financial institutions. Now suppose that we introduce financial intermediaries—"banks" for short—that transfer funds between households and non-financial firms. Further, let R_{bt+1} be the (possibly state-contingent) bank lending rate.

Then with frictionless financial markets,

$$E_t\Lambda_{t,t+1}R_{kt+1} = E_t\Lambda_{t,t+1}R_{bt+1} = E_t\Lambda_{t,t+1}R_{t+1}. \quad (3)$$

In this case, to arbitrage ensures that the expected discounted return to capital equals expected discounted bank loan rate, and in turn that the latter equals the discounted riskless rate.

With capital market frictions, the following set of inequalities holds:

$$E_t\Lambda_{t,t+1}R_{kt+1} \geq E_t\Lambda_{t,t+1}R_{bt+1} \geq E_t\Lambda_{t,t+1}R_{t+1}. \quad (4)$$

In this instance, there may be impediments in the flow of funds between households and banks, as well as between banks and non-financial borrowers. That is, limits to arbitrage can introduce a wedge between $E_t\Lambda_{t,t+1}R_{bt+1}$ and $E_t\Lambda_{t,t+1}R_{t+1}$, and also between $E_t\Lambda_{t,t+1}R_{kt+1}$ and $E_t\Lambda_{t,t+1}R_{bt+1}$.

As before, a financial crisis is manifested by a sharp increase in the gap between $E_t\Lambda_{t,t+1}R_{kt+1}$ and $E_t\Lambda_{t,t+1}R_{t+1}$. The source of the increase in this gap, however, could either be a disruption of the flow of funds between non-financial borrowers and banks (i.e., an increase in $E_t\Lambda_{t,t+1}R_{kt+1} - E_t\Lambda_{t,t+1}R_{bt+1}$) or between banks and depositors (i.e., an increase in $E_t\Lambda_{t,t+1}R_{bt+1} - E_t\Lambda_{t,t+1}R_{t+1}$), or both.

In a banking crisis, there is a sharp increase in $E_t\Lambda_{t,t+1}(R_{bkt+1} - R_{t+1})$.

2.3 Banking Crises and the Capital Constraint

At the core of a banking crisis are limits to arbitrage in the flow of funds between banks and depositors. We next explore how these limits may come about.

Let L_t be loans, N_t bank equity, ϕ_t the bank's maximum feasible leverage ratio (assets to equity), ϕ^R the regulatory maximum, μ_t the discounted excess return to capital $E_t\Lambda_{t,t+1}(R_{kt+1} - R_{t+1})$, and σ_t the standard deviation of returns to the bank's portfolio.

To the extent banks have private information about their activities and/or it is costly for depositors to enforce repayment, the quantity of (uninsured) deposits a bank can attract will depend on its equity capital. Roughly speaking, with agency/enforcement problems present, how much a bank can borrow will vary positively with its ability to cushion creditor losses. Equity capital provides such a cushion. Beyond these natural market forces, there may be regulatory capital requirements. In the United States, for example, regulatory capital requirements are imposed on commercial banks to offset the incentives for risk taking afforded by deposit insurance. Investment banks, by contrast, did not face formal capital requirements.

The capital constraint may be represented as follows:

$$L_t \leq \phi_t N_t, \tag{5}$$

with

$$\phi_t = \min[\phi(\mu_t, \sigma_t), \phi^R]$$
$$\phi_1 > 0, \phi_2 < 0.$$

Overall, the ratio of the bank's assets to equity cannot exceed ϕ_t, which is the minimum of the natural limit due to agency/enforcement problems $\phi(\mu_t, \sigma_t)$ and the regulatory maximum ϕ^R. In general, the latter depends positively on the excess return μ_t and inversely on the degree of risk, measured by σ_t. Intuitively, if depositors perceive the bank can earn high excess returns, they are willing to tolerate more leverage. However, as the perceived risk increases, they tolerate less.

When the leverage constraint is binding, arbitrage between the bank rate lending and the riskless rate is precluded. The economy is in a situation with

$$E_t\Lambda_{t,t+1}R_{kt+1} \geq \; E_t\Lambda_{t,t+1}R_{bt+1} > E_t\Lambda_{t,t+1}R_{t+1},$$

where the inequality between $E_t\Lambda_{t,t+1}R_{bt+1}$ and $E_t\Lambda_{t,t+1}R_{t+1}$ is strict. The excess in the bank lending rate over the riskless rate translates into an excess return on capital over the riskless rate.

In a banking crisis, the limits to arbitrage tighten. This can occur because bank equity N_t drops sharply (as a result of losses on bank assets) and/or because there is a significant increase in risk. The former produces a sharp drop in bank lending, given the maximum leverage ratio ϕ_t. The latter produces a decline in the maximum leverage ratio, and thus causes bank lending to drop, for any given level of bank equity.

In the general equilibrium, the contraction in lending produces a fall in capital prices, thus leading to an increase in excess returns $\mu_t = E_t\Lambda_{t,t+1}(R_{kt+1} - R_{t+1})$. The rise in the required return to capital, of course, depresses durable goods spending. This is the way that banking crises affect real activity.

Note also that the increase in the excess return can raise the maximum leverage ratio. Thus, though declines in bank equity and increases in perceived risk work to reduce bank lending, the rise in the maximum leverage ratio works to moderate this effect to some degree.

Keep in mind, however, that what is critical for the real economy is not the overall level of lending per se, but rather the overall credit market distortion as measured by the excess return $E_t\Lambda_{t,t+1}(R_{kt+1} - R_{t+1})$.

2.4 The Volatility of Bank Equity

The last piece of the puzzle involves the evolution of bank equity. Let D_t be deposits. Then, by the accounting identity that assets must equal liabilities, bank loans equal the sum of bank equity and deposits:

$$L_t = N_t + D_t.$$

Suppose, as is largely consistent with evidence, banks find it prohibitively expensive to issue new equity and instead accumulate net worth via retained earnings. Bank equity then evolves as the difference between the gross return on assets and the gross cost of liabilities, net any dividend payments, Div_t:

$$N_t = R_{bt}L_{t-1} - R_tD_{t-1} - Div_t.$$

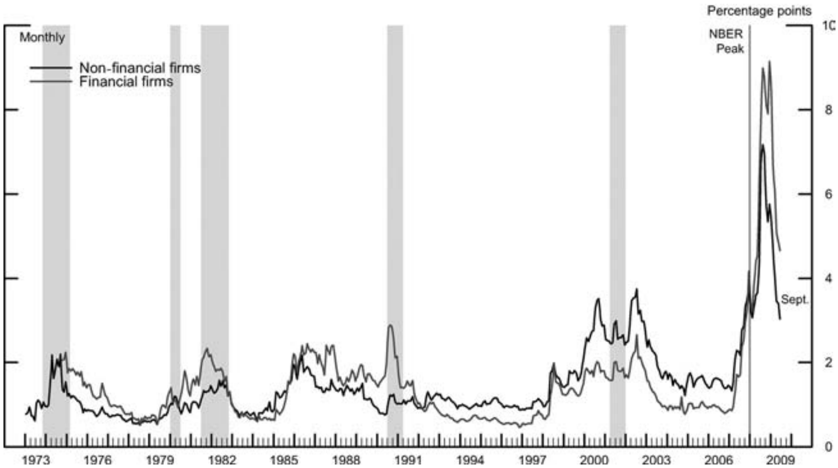
Given $L_{t-1} = \phi_{t-1}N_{t-1}$,

$$N_t = [(R_{bt} - R_t)\phi_{t-1} + R_t]N_{t-1} - Div_t. \tag{6}$$

Overall, volatility in the ex post net return on assets, $R_{bt} - R_t$, induces volatility in N_t. The overall effect is magnified by the size of the leverage ratio ϕ_{t-1}. Seen from this vantage, equations (5) and (6) capture in a simple way how the recent financial crisis played out. Losses on sub-prime mortgages at highly leveraged investment banks and (to a lesser degree) commercial banks induced a sharp contraction in the equity capital of these institutions. This in turn forced a sharp contraction in the assets of these institutions, leading to a sharp increase in return spreads.

It is true that the contraction in N_t can be moderated to some degree by a reduction in dividends Div_t. However, since dividends cannot turn negative, it is not possible to offset a sharp drop in N_t. In addition, for reasons we don't completely understand (e.g., signaling, etc.), banks seem reluctant to cut dividends, even when they are in distress.

Figure 1. Credit Spreads on Senior Unsecured Bonds



3. Credit Spreads vs. Credit Quantities during the Crisis

The underlying theory suggests that credit spreads should rise sharply during crises. It is less clear about the implications for credit quantities. We address each of these phenomena in turn.

3.1 Credit Spreads

Credit spreads normally move countercyclically. These movements are not necessarily signs of unusual distress. However, during the Great Recession the increase in credit spreads was off the charts.

Figure 1 plots indices of spreads between corporate debt and similar maturity government debt. The data is from Gilchrist, Yankov, and Zakrasjek (2010). For the 1981–82 recession, the most severe post-war recession until the Great Recession, credit spreads climbed only modestly, roughly 100 basis points for financial firms and about 50 basis points for non-financial firms. Note that financial distress is not thought to have been a significant factor in this downturn, which was driven mainly by tight monetary policy aimed at disinflating the economy.

The behavior of spreads is radically different in the current recession. For non-financial firms, spreads rose from under 200 basis points on average to nearly 700 in the wake of the collapse of Lehman Brothers in September 2008. Spreads for financial firms rose even further, climbing from roughly 100 basis points before the crisis to roughly 900 basis points around the Lehman collapse. There is also a slight lead in the increase in the financial firm spread.

The behavior of spreads is consistent with the notion of a banking crisis developed in the previous section. The sharp deterioration of financial institutions' balance sheets brought about by losses on sub-prime loans along with the associated increase in uncertainty curtailed the ability of these intermediaries to obtain funds. The net effect was a sharp increase in the cost of credit these institutions faced, which was presumably passed along to non-financial borrowers. Note that non-financial corporations that issue bonds directly on the open market still typically rely on commercial banks for working capital finance either directly or indirectly by using bank back-up lines of credit as collateral for open-market credit. Undoubtedly, the stress on financial firms contributed significantly to pushing up the non-financial spread.

To the extent that the movement in these spreads reflected increasing credit costs, they help account for how the financial crisis contributed to the collapse in durable goods spending. Again, the overall behavior of these spreads is consistent with the mechanism linking banking crises to real activity described in the previous section. Conversely, it would seem difficult to explain the movement in these spreads in a setting with frictionless financial markets. In Gertler (2009) I discuss some other evidence that points to a banking crisis being at work.

3.2 Credit Quantities

As I have noted, the theoretical mechanism does not have sharp implications for the behavior of credit quantities. With this in mind, I interpret recent events.

It is first important to consider investment banks along with commercial banks. Overall, the events were consistent with a "capital crunch" in investment banking. Losses on mortgage-backed securities induced significant depletion of equity in highly leveraged investment banks. Losses in equity values combined with increased uncertainty limited the ability of these banks to obtain funds, along the lines that the previous section suggested. As a result, securitized lending collapsed. Credit spreads on these instruments shot up, again consistent with the theory discussed earlier.

For commercial banks, the dynamics played out differently: These banks entered the crisis well capitalized. As events progressed, however, they experienced a combination of equity losses and increases in the demand for bank credit. The former stemmed from exposure to mortgage-backed securities, though not to the same extent as investment banks. The latter resulted from (i) a take-down of credit lines (Ivashina and Scharfstein 2009) by firms short on cash as the crisis unfolded and (ii) the absorption of assets being shed by investment banks, stemming from either explicit or implicit commitments to these institutions. For both these reasons there was an initial increase in commercial bank lending. As Ivashina and Scharfstein note, it would clearly be incorrect to interpret this increase in bank lending as reflecting the absence of a financial crisis.

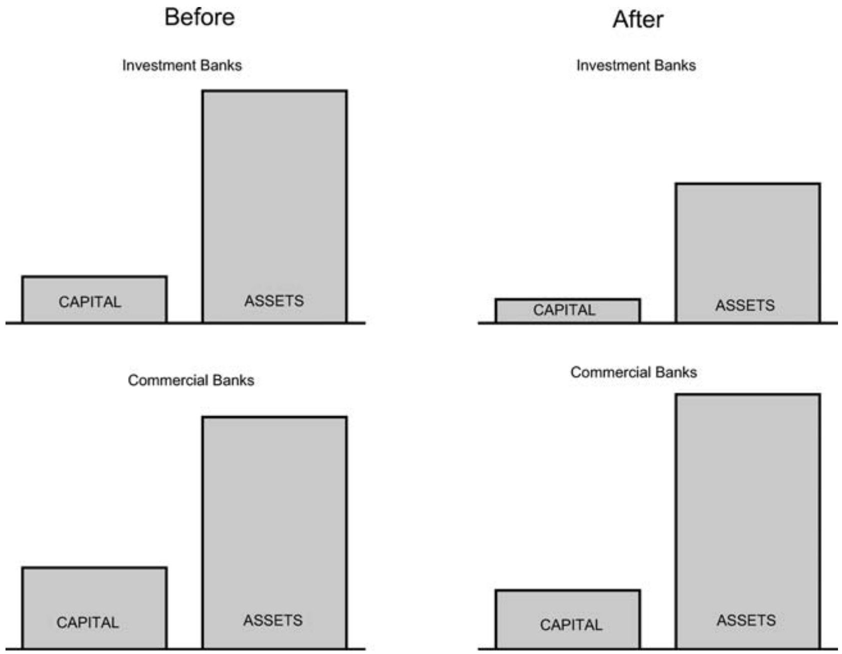
Figure 2 portrays the balance sheets of investment and commercial banks as the crisis unfolded. As the figure makes clear, even though commercial banks entered the crisis well capitalized, their capital was not sufficient to easily absorb assets sold off by investment banks. The net effect was a fire sale of securitized assets, leading to falling prices and rising expected returns to assets. Limits to arbitrage, of the type we described earlier, kept these prices and returns from quickly returning to normal.

4. Some Remarks on the Session Papers

4.1 The Impact of Equity Injections on Bank Lending

It is clear that the equity injections under the Troubled Asset Relief Program (TARP) did not lead to a proportionate increase in bank lending. Here I agree with Berrospide and Edge (this issue) that banks do not maintain constant leverage ratios. Indeed the theory laid out in section 2 suggests that, in general, banks will not have constant leverage ratios, even if they are capital constrained. Sorting out the impact of the TARP on lending will ultimately require a formal model where one can ask the questions of what would have happened to lending had the TARP not been enacted. I think a

Figure 2. Changes in the Level and Composition of Bank Lending and Capital over the Crisis



credible view is that there would have been an even greater contraction in lending.

Beyond these considerations, I think the authors' estimates of the effect of capital on lending may be too conservative. The sample period they study does not contain much variation in bank capital. Missing from the sample are the two periods where bank capital shortage was thought to be a problem: the post-Lehman-collapse period and the capital crunch of the late 1980s. In addition, the way they control for loan demand may be problematic. It fails to capture the initial countercyclical increase in loan demand at the onset of a recession as firms desire to borrow to meet fixed payments as cash flows begin to decline. However, even if we accept Berrospide and Edge's estimates, there is reason to think that bank capital played an important role in the current recession. The disruption of interbank and other liquidity markets likely enhanced the value of capital. In addition, many observers credit the TARP for stabilizing credit markets, as evidenced by the across-the-board reduction in credit spreads that followed this and related interventions.

4.2 The Forecasting Power of Bank Capital Asset Ratios

If one can identify exogenous shifts in bank capital asset ratios, then the theory suggests these shifts should help forecast real output. Jimborean and Méssonier (this issue) present strong evidence of this forecasting power. Berrospride and Edge (this issue) present somewhat weaker evidence. Interestingly, these authors show that a one-standard-deviation innovation decrease in the capital asset ratio leads to a roughly 0.2 percent decline in output growth. The drop in the capital asset ratio during the Great Recession was nearly ten standard deviations, implying a 2.0 percent decline in output growth, a rather substantial effect.

Of course, the identification approach in each case (Choleski decomposition) does not rule out possible unobservable factors being at work. In either case, these facts are interesting to interpret.

References

Bernanke, B., and M. Gertler. 1989. "Agency Costs, Net Worth and Business Fluctuations." *American Economic Review* 79 (1): 14–31.

Gertler, M. 2009. "Discussion of 'Financial Crises and Economic Activity'." In *Financial Stability and Economic Policy.* Proceedings of the Federal Reserve Bank of Kansas City Annual Economic Symposium, Jackson Hole, Wyoming, August 20–22.

Gertler, M., and N. Kiyotaki. 2010. "Financial Intermediation and Credit Policy in Business Cycle Analysis." Forthcoming in *Handbook of Monetary Economics,* Vol. 3B, ed. B. Friedman and M. Woodford. New York, NY: Elsevier.

Gilchrist, S., V. Yankov, and E. Zakrasjek. 2010. "Credit Market Shocks: Evidence from Stock and Bond Markets." Mimeo.

Ivashina, V., and D. Scharfstein. 2009. "Bank Lending during the Financial Crisis of 2008." Mimeo.

Procyclicality of Capital Requirements in a General Equilibrium Model of Liquidity Dependence*

Francisco Covas[a] and Shigeru Fujita[b]
[a]Division of Monetary Affairs, Board of Governors of the Federal Reserve System
[b]Research Department, Federal Reserve Bank of Philadelphia

This paper quantifies the procyclical effects of bank capital requirements in a general equilibrium model where financing of capital goods production is subject to an agency problem. At the center of this problem is the interaction between entrepreneurs' moral hazard and liquidity provision by banks as analyzed by Holmstrom and Tirole (1998). We impose capital requirements under the assumption that raising funds through bank equity is more costly than raising it through deposits. We consider the time-varying capital requirement (as in Basel II) as well as the constant requirement (as in Basel I). Importantly, under both regimes, the cost of issuing equity is higher during downturns. Comparing output fluctuations under the Basel I and Basel II economies with those in the no-requirement economy, we find that the regulations have relatively minor *average* effects on output fluctuations (measured by the differences in the standard deviations). However, the effects are more pronounced around business cycle peaks and troughs.

JEL Codes: E32, G21, G28.

*This is a substantially revised version of the paper formerly titled "Time-Varying Capital Requirements in a General Equilibrium Model of Liquidity Dependence." We are particularly grateful to an anonymous referee and the editor Rafael Repullo for their constructive comments. We also thank Xavier Freixas, David Jones, Ryo Kato, Garey Ramey, Javier Suarez, Giorgio Valente, Skander Van den Heuvel, and the participants of the ECB/CEPR conference on "Procyclicality and Financial Regulation" and the HKIMR/BIS conference on "Financial Stability: Towards a Macroprudential Approach" for helpful discussions and comments. Remaining errors are ours. The views expressed in this paper are ours and do not necessarily reflect the views of the Federal Reserve Board, the Federal Reserve Bank of Philadelphia, or the Federal Reserve System.
Author e-mails: francisco.b.covas@frb.gov and shfujita@gmail.com.

1. Introduction

There has been a strong interest in understanding interactions between bank capital regulation and macroeconomic fluctuations among policymakers and academic researchers. The interest has become even stronger in the aftermath of the recent financial crisis. One of the key concerns, especially from a macroeconomic perspective, is that bank capital regulation can induce significant "procyclicality," meaning that bank capital regulation can amplify macroeconomic fluctuations. The procyclical effect was recognized under the first bank capital regulation—i.e., Basel I—in which banks are required to hold a constant fraction of equity. The procyclicality issue has received significantly more attention under the so-called risk-sensitive regulation (Basel II). Under Basel II, the risk weight associated with each loan is negatively related to the borrower's credit quality; therefore, during an economic downturn when overall credit quality deteriorates, capital requirements become more stringent. This further limits banks' lending capacity. Our interest is to quantify the procyclical effects using a general equilibrium macroeconomic model.

There are many papers with similar interests. Blum and Hellwig (1995) examine the procyclical effects of fixed capital requirements under Basel I. Using a simple reduced-form macroeconomic framework, they argue that it is likely to amplify macroeconomic fluctuations. Heid (2007) goes one step further by studying the implications of risk-sensitive capital requirements in a similar reduced-form environment. More recently, Zhu (2008) studies the effects of bank capital regulation on banks' behavior by applying the industry model of Cooley and Quadrini (2001) to a banking sector that is subject to risk-sensitive capital requirements. Finally, Repullo and Suarez (2009) develop a microfounded partial equilibrium model of relationship banking and analyze the banks' behavior under risk-sensitive capital requirements. They show that the procyclicality under Basel II can be sizable.

Relative to these previous studies, we examine the business cycle implications of bank capital requirements in a general equilibrium macroeconomic model. Using a general equilibrium framework allows us to quantify the impact of bank capital regulation on macroeconomic variables. In our model, the financing of capital goods production is subject to an agency problem, as in Carlstrom and Fuerst (1997). The financing problem, however, is characterized by entrepreneurs' moral hazard and liquidity provision by financial intermediaries. This framework is proposed by Holmstrom and Tirole (1998) and adapted by Kato (2006) to a dynamic stochastic general equilibrium (DSGE) environment. We extend Kato's work along several dimensions so that we can examine the quantitative impact of bank capital regulation. We are aware that there are several alternative approaches. The first alternative would be the costly state verification framework developed by Townsend (1979) and popularized in macroeconomics by Bernanke and Gertler (1989), Carlstrom and Fuerst (1997), and Bernanke, Gertler, and Gilchrist (1999). The second alternative is the double moral hazard framework of Holmstrom and Tirole (1997) that is adapted to the macroeconomic environment by Chen (2001) and Meh and Moran (2010).[1] We have adopted the framework of Holmstrom and Tirole (1998) because the model generates countercyclical liquidity dependence; firms tend to rely more heavily on lines of credit to finance their liquidity needs during downturns, and this countercyclical liquidity dependence underscores the important role banks play in the economy. The importance of credit lines in bank financing is very well known. For instance, loans made under a credit line amount to almost 80 percent of total commercial and industrial (C&I) loans.[2] As Schuermann (2009) mentions, during economic downturns, market finance becomes scarce, and firms increase their liquidity dependence on banks by drawing down the loan commitments pre-arranged with banks.[3] These empirical observations have led us to use the liquidity dependence framework instead of other popular frameworks mentioned above. This paper attempts to quantify the interaction between liquidity dependence and bank capital regulation.

[1] In addition to entrepreneurs' moral hazard, this framework features moral hazard of the banking sector with respect to project monitoring, thus endogenously generating fluctuations in the bank capital ratio. However, Chen (2001) and Meh and Moran (2010) do not model regulatory requirements.

[2] See the Survey of Terms of Business Lending (FRB E.2 Release).

[3] There are several papers on the importance of loan commitments for bank risk management and corporate liquidity management. See, for example, Kashyap, Rajan, and Stein (2002) and Sufi (2009), respectively.

We impose bank capital requirements assuming that raising funds through equity is more costly than raising them through deposits, as in Repullo and Suarez (2009). We assume that the capital requirement ratio increases inversely with the business cycle under Basel II, while it is fixed at a pre-specified level of 8 percent under Basel I. Another key ingredient of our paper is the assumption that equity issuance is more costly in recessions. It is well known that new equity issuance during a downturn can be very costly.[4] More generally speaking, as noted by Kashyap and Stein (2004), bank equity becomes scarcer in downturns, raising its (shadow) cost. By introducing the time-varying equity issuance cost, we distinguish between the two regulatory regimes depending on whether only the equity issuance cost is time varying (Basel I) or both the equity issuance cost and the capital requirement ratio are time varying (Basel II). Adopting these parsimonious specifications allows us to assess the procyclicality effects in a stylized DSGE environment.

The model is calibrated by using relevant observable information such as the utilization rate of credit lines. We specify the time-varying capital requirement in our model by using the actual Basel II risk-weight formula. To calibrate the equity issuance cost variable, we utilize the evidence by Kashyap and Stein (2004) that the additional cyclical pressure on bank capital positions under Basel II is of almost the same order of magnitude as the effect under Basel I. Further, we also consider cases in which the equity issuance cost responds more sharply to the business cycles. This is motivated by Repullo and Suarez's (2009) claim that the shadow cost of bank equity can be very high at the time of financial distress. It is shown that, across various plausible calibrations, Basel I and Basel II contribute to increasing the standard deviation of output fluctuations by around 5 basis points and 10 basis points, respectively. These numbers suggest that the capital requirements have *on average* relatively minor procyclical effects on the aggregate economy. However, we also show that the effects can be more pronounced around business cycle peaks and troughs. For example, around the bottom of business cycles, output in the Basel I economy can be lower than that

[4]See, for example, the discussions and references in Repullo and Suarez (2009).

in the no-requirement economy by 10–15 basis points. This magnification effect can be as large as 20–40 basis points when the Basel II economy is compared with the no-requirement economy.

Qualitatively speaking, the procyclicality in our model is driven by the feature that capital requirements directly affect the bank's ability to provide lines of credit to entrepreneurs. Under Basel II, when the availability of lines of credit is more valuable (during downturns), the cost of making those loans is higher as a result of the higher capital requirement: A more strict capital requirement implies that it is more difficult for the bank to meet entrepreneurs' financing needs and, consequently, more positive net present value (NPV) projects (that would otherwise be implemented) are destroyed. The same mechanism works when the equity issuance cost rises during downturns. Banks face both of the two constraints under Basel II, whereas only the time-varying equity issuance cost induces procyclicality under Basel I.

While our main focus is on the volatility effects of capital requirements, we also look at the steady-state effects of a permanently higher capital requirement ratio. Since our model omits any welfare-improving aspects of capital requirements, the higher capital requirement necessarily entails only negative effects on the macroeconomy. Having recognized this limitation, the transition dynamics to the new steady state exhibit some interesting patterns among macro variables. First, our model naturally implies that output, consumption, and investment all converge to permanently lower levels. However, before reaching this new steady state, investment goes down significantly more in the first few years, while household consumption increases: The higher capital requirement reduces the supply of capital goods, causing the price of capital goods to rise over time, and this induces the substitution into consumption away from savings (investment) in the short run.

This paper proceeds as follows. In section 2, we describe our model. In section 3, we calibrate the model. Section 4 focuses on the steady-state effects of permanently higher capital requirement without aggregate fluctuations of the economy. Section 5 presents the main result of the paper, namely the business cycle implications of Basel I and Basel II. The section considers both the "average" effects and the effects around business cycle peaks and troughs. Section 6 concludes the paper.

2. Model

The model structure is similar to that in Kato (2006), who provides an important alternative to Carlstrom and Fuerst (1997) and Bernanke, Gertler, and Gilchrist (1999) in modeling the financial frictions in macroeconomic models. The latter two papers embed costly state verification (CSV) into an otherwise-standard real business cycle (RBC) model. Instead of CSV, Kato (2006) adopts the financial contract developed by Holmstrom and Tirole (1998), who emphasize the importance of liquidity provision by financial intermediaries and its interaction with entrepreneurs' moral hazard. We deviate from Kato (2006) in the following two ways. First, we allow for non-zero liquidation value when projects are abandoned. The second way, which is the focus of our paper, is that we impose capital requirements on financial intermediaries.

2.1 Environment

The economy is populated by four types of agents: a fixed mass $1 - \eta$ of households, a fixed mass η of entrepreneurs, banks, and firms. Both households and entrepreneurs supply labor and rent out capital to the firms that produce the consumption good. Entrepreneurs differ from households with respect to their ability to produce the capital good. Entrepreneurs borrow funds from the banks, which funnel households' savings. The intermediation is subject to the agency problem of Holmstrom and Tirole (1998). Further, banks are also constrained by the capital requirement. The sequence of events, which is similar to that in Carlstrom and Fuerst (1997), is summarized in table 1.

2.2 Financial Contract

The financial contract starts and ends within a period. The general equilibrium of the economy influences the contract only through the level of net worth n and the price of capital q. These variables are thus treated parametrically in this sub-section.

The contract involves two parties, a bank and an entrepreneur. Both parties are risk neutral. Entrepreneurs are endowed with technology that converts the consumption good into the capital good. Let

Table 1. Sequence of Events within a Period

1.	The aggregate technology shock (ϵ) is realized.
2.	Firms hire labor and rent capital from households and entrepreneurs and produce the consumption good.
3.	Households earn their labor and capital income and make the consumption-saving decision.
4.	Banks use the resources obtained from households to provide loans to entrepreneurs. The optimal contract is described in sub-section 2.2.
5.	Entrepreneurs borrow $i - n$ consumption goods from the bank and place all of them together with their entire net worth n into capital-creation projects.
6.	The idiosyncratic liquidity shocks (ω) are realized. Projects with $\omega \leq \overline{\omega}$ are financed through credit lines. Otherwise, projects are abandoned and the bank obtains the liquidation value of τi.
7.	Outcomes of the continued projects are realized. The entrepreneurs with successful projects distribute a part of the return to the bank.
8.	Entrepreneurs make the consumption-saving decision.

i be the investment size (measured in the consumption good), which yields Ri units of the capital good when the project is successful.[5] The success probability is p_j, where $j \in \{H, L\}$. The entrepreneur, whose net worth is n, borrows $i - n$ units of the consumption good from the bank.

The project proceeds in three stages (0, 1, and 2). At stage 0, the investment (i) is put in place. At stage 1, the exogenous "liquidity shock," $\omega \in [0, \infty)$, is realized. The shock ω measures a per-unit-of-investment liquidity infusion (in units of the capital good) that is necessary to continue the project. ω is assumed to be i.i.d. cross-sectionally and over time, and distributed according to $\Phi(\omega)$ with density $\phi(\omega)$. Without the cash infusion, the project is abandoned and liquidated. When the project is abandoned, the salvage value, τi, is transferred to the lender.[6] The last stage, in which the project is actually undertaken, is subject to moral hazard of the entrepreneur. He can exert effort or shirk. Exerting effort yields the success

[5] When it fails, the return is zero.

[6] Kato (2006) assumes that the liquidation value is zero. In contrast, we allow for a positive liquidation value, and the parameter τ is calibrated referring to the empirical evidence.

probability of p_H and no private benefit, and shirking results in the lower success probability of $p_L(< p_H)$ and yields a private benefit of Bi.

As Holmstrom and Tirole (1998) show, socially optimal financing in this environment is characterized by the cutoff rule that the project is abandoned if and only if $\omega \geq p_H R \equiv \omega_1$. This level of the liquidity shock is called the first best cutoff.

2.2.1 Capital Requirements and Equity Issuance Cost

The bank can raise funds through either deposits or equity (e), but holding equity involves the equity issuance cost c. We assume for analytical convenience that this cost is linear with respect to the size of equity issuance:

$$c = \gamma(A)e, \tag{1}$$

where $\gamma(A)$ represents the per-unit cost of holding equity, which is assumed to depend on the aggregate economic condition. More specifically, we take the total factor productivity (TFP) process (A) as representing this aggregate economic condition. The TFP process and the functional form for $\gamma(A)$ will be specified in the calibration section.

Under the presence of the equity issuance cost, the bank's zero profit condition can be written as follows:

$$\begin{aligned} i - n + qiE(\omega|\omega \leq \overline{\omega})\Phi(\overline{\omega}) = qi \int_0^{\overline{\omega}} p_H(R - R^e(\omega))\phi(\omega)d\omega \\ + qi(1 - \Phi(\bar{\omega}))\tau - c, \end{aligned} \tag{2}$$

where the left-hand side represents the total number of loans (i.e., the sum of initial project loans and credit lines in the middle stage).[7] On the right-hand side, R^e is the return to the entrepreneur, and thus the first term represents the expected return to the bank. The

[7] The left-hand side also equals the total resource (i.e., saving) that the household owns at the banking sector over the course of the financial contract.

second term gives the liquidation value when the project is terminated, and the last term gives the equity issuance cost. The bank is, by regulation, required to maintain a certain level of equity; the regulator imposes the "capital requirement" in terms of the size of equity relative to the total number of loans, which is written as

$$e = \theta(A)[i - n + qiE(\omega|\omega \leq \overline{\omega})\Phi(\overline{\omega})], \quad (3)$$

where $\theta(A)$ is the capital requirement ratio, which can depend on the aggregate TFP process. Note that because equity is more costly than deposits, the capital requirement holds with equality, which is imposed in equation (3). Combining equations (1), (2), and (3) results in

$$[1 + \Theta(A)][i - n + qiE(\omega|\omega \leq \overline{\omega})\Phi(\overline{\omega})] \\ = qi \int_0^{\overline{\omega}} p_H(R - R^e(\omega))\phi(\omega)d\omega + qi(1 - \Phi(\bar{\omega}))\tau, \quad (4)$$

where $\Theta(A) \equiv \theta(A)\gamma(A)$. Our main interest is to examine the procyclical effects of the two forms of capital requirements: (i) a flat requirement in which θ is fixed regardless of the aggregate condition and (ii) a "procyclical" requirement in which θ is higher (lower) during downturns (booms). Each of the two capital requirement schemes is intended to mimic the capital requirement under Basel I and Basel II, respectively. Note also that regardless of which regime the economy is in, the time-varying feature of the equity issuance cost is maintained.

We would like to stress here a few issues regarding our modeling strategy for the capital requirements and the equity issuance cost. In our model, banks are perfectly competitive intermediaries that simply channel funds from households to capital-producing entrepreneurs. Operating this service per se does not require bank equity, and the requirement of holding a certain level of bank equity is modeled simply as a "tax" on financial intermediation. Furthermore, we impose this cost exogenously, meaning that regulation plays no useful role in the economy. That is, since the intra-period financial contract is, by construction, not subject to the aggregate risk, the bank failure never occurs in our model. The bank capital requirement thus plays no role in this regard.[8] Another way to motivate capital regulation is to consider an environment like the one considered by Van den Heuvel (2008), in which the capital requirement plays the role of mitigating the bank's moral hazard created by deposit insurance. We leave these important issues aside, given that our main interest is on quantifying the volatility effects.

We also make an assumption that the equity issuance cost fluctuates exogenously over the business cycle. More specifically, we assume that $\gamma'(A) < 0$. It seems quite plausible that the cost of maintaining a certain level of bank equity is countercyclical. It is well known that issuing equity is much more expensive during recessions (see Repullo and Suarez 2009). More generally, Kashyap and Stein (2004) argue that bank capital becomes scarcer during bad times, raising the "shadow cost of capital."[9] This scarcity reflects the fact that higher loan losses and lower operating income deplete the stock of bank capital more rapidly during recessionary times. Ideally, we would like to endogenously incorporate this feature into the model. However, our model is arguably too stylized and thus we take a shortcut.[10] In section 3 below, we will discuss more specific features of this exogenous process.

2.2.2 *Optimal Contract*

The optimal contract maximizes the entrepreneur's expected payoffs by choosing (i) the size of the project i, (ii) the return to the entrepreneur, R^e when the project is successful,[11] and (iii) the cutoff liquidity shock ($\bar{\omega}$). The problem is subject to the bank's

[8] One could imagine a model with an *inter*-period contract that cannot be contingent on aggregate risk. In such an environment, bank capital requirements can play the role of avoiding bank failure.

[9] See also the references therein for further discussion of this point.

[10] Covas and Den Haan (2010) also specify a similar exogenous process for the equity issuance cost to analyze debt and equity finance over the business cycle, although their focus is on the behavior of the non-financial corporate sector.

[11] This is equivalent to choosing the division of the total return R between the two parties.

break-even constraint and the entrepreneur's incentive compatibility constraint:[12]

$$\max qi \int_0^{\bar{\omega}} p_H R^e(\omega)\phi(\omega)d\omega - n, \tag{5}$$

subject to equation (4) and the entrepreneur's incentive compatibility constraint,

$$p_H R^e(\omega) \geq p_L R^e(\omega) + B. \tag{6}$$

Observe that in the problem above, the return to the entrepreneur is allowed to be contingent on ω. However, as Holmstrom and Tirole (1998) show, equation (6) binds in the optimal contract and thus the dependence of the return to the entrepreneur on ω vanishes; i.e., $R^e(\omega) = \frac{B}{p_H - p_L}$. This result arises because it is optimal that the entrepreneur borrow as much as possible up to the so-called borrowing limit, which is in turn achieved by the borrower taking the minimum payment necessary to exert high effort. The return to the lender is then written as $p_H\left(R - \frac{B}{p_H - p_L}\right) \equiv \omega_0$, which is called pledgeable income. Using the binding incentive compatibility constraint, equation (6), in equation (4) results in

$$i = \frac{1}{1 - qh(\bar{\omega}, \Theta(A))} n, \tag{7}$$

where

$$h(\bar{\omega}, \Theta(A)) = \frac{\Phi(\bar{\omega})\omega_0 + (1 - \Phi(\bar{\omega}))\tau}{1 + \Theta(A)} - E(\omega|\omega \leq \overline{\omega})\Phi(\overline{\omega}). \tag{8}$$

These expressions make clear that investment is directly influenced not only by net worth and the price of capital as in Kato (2006) but also by the capital requirement ratio and the cost of bank equity. Using equation (7) in equation (5) and maximizing the resulting expression with respect to $\bar{\omega}$ results in the following first-order condition:

$$q \int_0^{\bar{\omega}} \Phi(\omega)d\omega = 1 - \frac{q\tau}{1 + \Theta(A)}. \tag{9}$$

[12] Note that a binding capital requirement is a priori imposed below.

We can solve this optimality condition for the cutoff liquidity shock $\bar{\omega}$, given the levels of q and $\Theta(A)$.

As in Kato (2006), we define the degree of liquidity dependence as follows:

$$\text{Liquidity dependence} \equiv \frac{i\int_0^{\bar{\omega}} \omega d\Phi(\omega)}{ip_H R\Phi(\bar{\omega})} = \frac{E(\omega|\omega \leq \overline{\omega})}{\omega_1}. \tag{10}$$

Equation (10) captures the dependence on the bank's liquidity provision relative to the size of investment. In Kato (2006), where $\tau = 0$ and $\Theta(A) = 0$, the second term in the right-hand side of equation (9) drops out, and it is particularly simple to see in that case that q and $\overline{\omega}$ are negatively related. The negative relationship then implies countercyclical liquidity dependence in equation (10) (given that q is procyclical). As we will see later, the countercyclical liquidity dependence will also be true in our model. Note, however, that equation (9) alone cannot tell how changes in Θ influence $\overline{\omega}$ and thus liquidity dependence.[13] To assess this effect, we will later consider an experiment of changing Θ without changing A and see how other endogenous variables respond to it. There we will see that higher Θ causes the supply of funds to drop and thus raises the price of capital. When the price of capital is higher, $\overline{\omega}$ and thus liquidity dependence go down.

2.3 Households

The representative household maximizes the discounted sum of their utility derived from consumption (c_t) and leisure (l_t):

$$E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, l_t), \tag{11}$$

where β is the discount factor. As in Carlstrom and Fuerst (1997) and Kato (2006), we assume that the utility function is

[13] In particular, equation (9) does not imply that non-zero τ is necessary for the capital requirement to have any effects in the model. For example, equations (7) and (8) indicate that Θ directly affects the relationship between net worth and investment, which in turn comes from the effect of Θ on the bank's balance sheet (equation (4)).

additively separable in consumption and leisure, and labor supply is indivisible:

$$u(c_t, l_t) = \frac{c_t^{1-\psi}}{1-\psi} + \nu(1 - l_t), \tag{12}$$

where ψ is the coefficient of relative risk aversion and ν is a normalizing constant.[14] The decisions are subject to the following budget constraints:

$$c_t + s_t = r_t k_t + w_t(1 - l_t), \tag{13}$$

$$k_{t+1} = (1 - \delta)k_t + \frac{1}{q_t} s_t, \tag{14}$$

where s_t is the household saving at the bank; k_t is the capital stock held by the household; w_t and r_t are, respectively, wage and interest rates paid by the firm; and δ is the depreciation rate of the capital stock. Note that the household saving s_t equals the total number of loans which correspond to the sum of the two terms appearing in the left-hand side of equation (2).[15]

The first-order conditions to this problem are standard and written as follows:

$$q_t = \beta E_t \left(\frac{c_t}{c_{t+1}} \right)^{\psi} [r_{t+1} + (1 - \delta) q_{t+1}], \tag{15}$$

$$-\nu c_t^{\psi} = w_t. \tag{16}$$

2.4 Entrepreneurs

In modeling the behavior of entrepreneurs, we make several assumptions similar to those in Carlstrom and Fuerst (1997), which are also adopted by Kato (2006). First, they are risk neutral. Second, they

[14]Our results below are not sensitive to the choice of the functional form. For instance, we have experimented with cases where the utility function is nonseparable and where labor supply is divisible, and obtained results similar to those reported below.

[15]The saving yields gross interest of unity as in Carlstrom and Fuerst (1997). This outcome is supported by the assumption that the financial contract starts and ends within the same period.

discount the future more heavily ($\beta^e < \beta$) than the households. This latter assumption is to avoid the self-financing equilibrium. In other words, absent this assumption, entrepreneurs quickly accumulate enough wealth by postponing consumption, so that the project can be financed through internal funds only. By lowering the discount factor for them, we prevent this self-financing equilibrium from arising.[16] The third assumption is that entrepreneurs supply labor inelastically to the firm.

The entrepreneur maximizes the discounted sum of their utility derived from consumption (c_t^e):

$$E_0 \sum_{t=0}^{\infty} (\beta^e)^t c_t^e. \tag{17}$$

As described in table 1, the entrepreneur first receives labor and capital incomes from the consumption goods sector at the start of each period. He then enters into the financial contract with net worth n_t:

$$n_t = (1-\delta) q_t z_t + r_t z_t + w_t^e, \tag{18}$$

where z_t is the capital stock held by the individual entrepreneur and w_t^e is the wage payment to the entrepreneur. The entrepreneur's net worth consists of the undepreciated part of the capital stock and the two flow incomes from the consumption goods sector. That the entrepreneur earns labor income is important because it allows the entrepreneur to enter into the financial contract with non-zero net worth even when he starts with zero capital stock.[17] When the entrepreneur withstands the liquidity shock (i.e., $\omega \leq \overline{\omega}$) and the project turns out to be successful, the consumption-saving decision is subject to the following constraint:

$$c_t^e + q_t z_{t+1} = \frac{q_t R^e}{1 - q_t h(\bar{\omega}_t, \Theta(A_t))} n_t. \tag{19}$$

[16]The alternative way to avoid this situation is to assume that a certain fraction of entrepreneurs die each period and sell their accumulated wealth to the household. See, for example, Carlstrom and Fuerst (1997) and Bernanke, Gertler, and Gilchrist (1999) for more details.

[17]He starts the current period with no capital stock when the project was abandoned or failed in the previous period.

The right-hand side of equation (19) represents the total return to the entrepreneur. The numerator gives the per-project return and the remaining term gives the size of investment as a multiple of his net worth, which was derived in equation (7). The entrepreneur's consumption-saving decision obeys the following Euler equation:

$$q_t = \beta^e E_t[q_{t+1}(1-\delta) + r_{t+1}] \frac{q_{t+1}(\omega_1 - \omega_0)\Phi(\bar{\omega}_{t+1})}{1 - q_{t+1} h(\bar{\omega}_{t+1}, \Theta(A_{t+1}))}. \qquad (20)$$

When the project is abandoned or unsuccessful, the entrepreneur gets nothing; thus, $c_t^e = z_{t+1} = 0$. Note that the entrepreneur's project succeeds with probability $p_H \Phi(\overline{\omega}_t)$. Aggregating the capital stocks held by η entrepreneurs, we can write

$$Z_{t+1} = \frac{(\omega_1 - \omega_0)\Phi(\bar{\omega}_t)}{1 - q_t h(\bar{\omega}_t, \Theta(A_t))} \eta n_t - \frac{\eta c_t^e}{q_t},$$

where Z_{t+1} is the aggregate capital stock held by the entrepreneurial sector.

2.5 Firms

The representative firm produces the consumption good by using the following constant-returns-to-scale technology:

$$Y_t = A_t K_t^{\alpha} H_t^{\iota} J_t^{1-\alpha-\iota}, \qquad (21)$$

where A is TFP, K is aggregate capital stock, H is labor supply by the household sector, and J is labor supply by the entrepreneurial sector. TFP evolves according to the following AR(1) process:

$$\ln A_{t+1} = \rho \ln A_t + \epsilon_{t+1}. \qquad (22)$$

The technology shock ϵ_t is distributed as standard normal $N(0, \sigma_\epsilon)$. Labor and capital rental markets are assumed to be competitive. The firm thus hires two types of labor and rents capital according to $r_t = \alpha A_t K_t^{\alpha-1} H_t^{\iota} J_t^{1-\alpha-\iota}$, $w_t = \iota A_t K_t^{\alpha} H_t^{\iota-1} J_t^{1-\alpha-\iota}$, and $w_t^e = (1-\alpha-\iota) A_t K_t^{\alpha} H_t^{\iota} J_t^{-\alpha-\iota}$.

2.6 General Equilibrium

The following market clearing conditions close the model:

$$H_t = (1-\eta)(1-l_t) \tag{23}$$

$$J_t = \eta \tag{24}$$

$$Y_t = (1-\eta)c_t + \eta c_t^e + \eta i \left(1 + q_t E(\omega|\omega \leq \overline{\omega})\Phi(\overline{\omega}) + q_t \frac{\Theta(A_t)\Phi(\bar{\omega}_t)\omega_0 - (1-\Phi(\bar{\omega}_t))\tau}{1+\Theta(A_t)}\right) \tag{25}$$

$$K_{t+1} = (1-\delta)K_t + \eta i \omega_1 \Phi(\overline{\omega}_t). \tag{26}$$

The first two equations above clear the two labor markets, the third equation clears the market for the consumption good, and the last equation clears the market for the capital good. The last term in equation (25) properly accounts for the resource cost of the capital requirement net of the liquidation value of the failed projects.

3. Calibration

We now discuss the baseline calibration of the model. One period in the model is assumed to be one quarter. The parameter values used in our baseline calibration are summarized in table 2.

3.1 Parameters Set Externally

The discount factor for the household (β) is set equal to 0.99. The discount factor for the entrepreneur (β^e) needs to be set to a lower value to avoid the self-financing equilibrium and is selected to be 0.94. This value is commonly used in this literature (e.g., Carlstrom and Fuerst 1997). The constant relative risk aversion (CRRA) parameter of the household (ψ) is set to 1.5. The firm's production technology is Cobb-Douglas, as shown in (21), with the capital share α equal to 0.33, the household's labor share ι equal to 0.66, and the entrepreneur's labor share equal to 0.01.[18] These numbers are all in

[18]Note that the small share is chosen so that entrepreneurs enter into the financial contract with at least some wealth. See Carlstrom and Fuerst (1997) and the discussion in sub-section 2.4.

Table 2. Baseline Calibration

Discount Factor of Households	β	0.99
Discount Factor of Entrepreneurs	β^e	0.94
Relative Risk Aversion of Households	ψ	1.50
Labor Supply Parameter	ν	2.68
Capital Share	α	0.33
Household Labor Share	ι	0.66
Depreciation Rate	δ	0.025
Standard Deviation of Liquidity Shock	σ_ω	0.44
First Best Cutoff	ω_1	2.75
Pledgeable Income	ω_0	1.93
Recovery Rate Parameter	τ	0.60
Mean Level of the Capital Requirement	θ_0	0.08
Elasticity of the Capital Requirement	θ_1	−8
Mean Level of the Equity Issuance Cost	γ_0	0.05
Elasticity of the Equity Issuance Cost	γ_1	−8
Persistence of Aggregate TFP Shock	ρ	0.95
Standard Deviation of Aggregate TFP Shock	σ_ϵ	0.007

Note: Under the flat capital requirement (Basel I) regime, the dependence of the capital requirement ratio on aggregate TFP is suppressed, i.e., $\theta_1 = 0$.

line with the previous literature. The depreciation rate of the capital stock δ is 0.025. The aggregate TFP process (A_t) is assumed to have the persistence parameter ρ equal to 0.95 and the conditional standard deviation σ_ϵ equal to 0.007. Finally, we set the fraction of entrepreneurs in the population η to 0.3 as in Kato (2006).[19]

3.2 Parameters Set Internally

First, the normalizing parameter ν of the labor supply function is chosen to be 2.68, such that the household spends one-third of its time on working, given all other parameter values. We assume that

[19]This parameter has little impact on the aggregate behavior of our model; we experimented with the case of $\eta = 0.2$ and found that the model's aggregate properties changed little. See Dorefeenko, Lee, and Salyer (2008), who show that this parameter works simply as "normalization" in the Carlstrom and Fuerst (1997) model. In footnote 7 and the appendix of that paper, they spell out this issue in detail.

Table 3. Selected Moments: Data vs. Model

Moments	Data (%)	Model (%)
LGD	39.8	35.4
PD	0.5	0.6
Utilization Rate of Credit Lines	32.5	36.0
Ratio of Unused Commitments over Total Loans	86.0	91.5

Notes: LGD (loss given default) equals the ratio of the amount of losses to loans outstanding at the time of default. The reported number is an average over 1982 through 1999 for a large U.S. bank, reported by Araten, Jacobs, and Varshney (2004). PD (the probability of default) is from Moody's default rate series. The reported number is an average over 1982 through 2004. The utilization rate is equal to the ratio of used revolving credits over the committed amount. It is taken from Sufi (2009), who uses a sample of 300 firms with debt outstanding over 1996 through 2003. The ratio of unused commitments to total loans is calculated from the series in Call Reports, RCFD3423 and RCFD1400, which cover all U.S. commercial banks over 1990 through 2004.

the distribution of liquidity shocks $\Phi(\omega)$ is log-normal with mean equal to one and a standard deviation of σ_ω.[20] We assign σ_ω together with the three parameters—the first-best cutoff ω_1, the pledgeable return from the investment ω_0, and the liquidation value parameter τ—to match the following four moments from the data: (i) loss given default (LGD) on bank loans, (ii) probability of default (PD), (iii) utilization rate on lines of credit, and (iv) the ratio of unused commitments to total loans. These four moments are summarized in table 3.

3.2.1 LGD

First, note that Moody's and Standard & Poor's (S&P) have databases with recovery rates for various debt instruments. However, most of the defaults in these two databases are for corporate bonds and not for bank loans. There are various differences between these two instruments, and in particular, corporate bonds are unsecured

[20]The mean of the distribution is set to one as normalization. In principle, we could use either the steady-state level of the price of capital or the mean as normalization.

and bank loans are often secured loans at the time of default. This results in significant differences in average recovery rates. Specifically, corporate bonds have significantly lower recovery rates compared with bank loans. We thus make use of the information provided by Araten, Jacobs, and Varshney (2004) to calibrate average LGD. This study was based on the default experience of a single large U.S. bank in the period between 1982 and 1999. Given the large size of this bank's portfolio, we believe that this series is more representative than the information based on a limited number of defaulted bank loans available in the Moody's or S&P databases. The average LGD of this bank over this eighteen-year period is 39.8 percent. We target this value in the steady state by associating it with the following concept in the model:

$$\text{LGD} = 1 - \frac{\tau q i}{i - n},$$

where the second term in the right-hand side is the recovery rate of the initial bank loan.

3.2.2 PD

For the empirical default rate, we rely on the default rate series published by Moody's that covers the period between 1982 and 2004. In Moody's data, the average default rate over the twenty-two-year period is 0.50 percent per quarter, which is taken to be our target. In the model, the corresponding default rate is simply the probability that the liquidity shock ω is higher than the cutoff value $\overline{\omega}$.[21]

3.2.3 Utilization Rate

We use the evidence gathered by Sufi (2009) to calibrate the average ratio of the used amount of the credit line to the committed amount. For a sample of 300 firms with debt outstanding, the average utilization rate is 32.5 percent over the period 1996 through 2003. This

[21]Recall that, in our model, entrepreneurs' projects can fail with probability $1-p_H$ even after the projects survive the liquidity shock. This variable p_H, however, does not need to be separately calibrated; this variable is embedded in ω_0 and ω_1. In terms of interpretation, we assume implicitly that p_H is very close to 1.

evidence is taken to be our target. In the model, we can define the utilization rate as follows:

$$\text{Utilization Rate} = \frac{\int_0^{\overline{\omega}} \omega d\Phi(\omega)}{\overline{\omega}\Phi(\overline{\omega})} = \frac{E(\omega|\omega \leq \overline{\omega})}{\overline{\omega}}.$$

3.2.4 Unused Commitments

The ratio of unused commitments to total loans is available on the regulatory filings of all commercial banks. The information is collected as part of the Call Reports. The sample period for this series is 1990 through 2004. The average ratio of unused commitments over this period amounts to 86 percent. In the model, this ratio is defined as follows:

$$\begin{aligned}\text{Ratio of Unused Commitments} &= \frac{qi\int_0^{\overline{\omega}}(\overline{\omega}-\omega)d\Phi(\omega)}{i-n+qi\int_0^{\overline{\omega}}\omega d\Phi(\omega)} \\ &= \frac{qi[\overline{\omega}-E(\omega|\omega \leq \overline{\omega})]\Phi(\overline{\omega})}{i-n+qiE(\omega|\omega \leq \overline{\omega})\Phi(\overline{\omega})}.\end{aligned}$$

Table 3 presents the steady-state performance in matching these four moments. While the model is unable to match these moments exactly, three moments appear reasonably close to their empirical counterparts.

3.3 Capital Requirement and Equity Insurance Cost

It remains to specify the processes for the regulatory capital requirement (θ) and the equity issuance cost (γ). We assume that these variables follow simple log-linear functions of the TFP process A_t:

$$\theta_t = \theta_0 A_t^{\theta_1}, \tag{27}$$

$$\gamma_t = \gamma_0 A_t^{\gamma_1}, \tag{28}$$

where θ_0, θ_1, γ_0, and γ_1 are the parameters to be calibrated. Putting these two costs together, we obtain

$$\Theta_t = \theta_0\gamma_0 A_t^{\theta_1+\gamma_1}. \tag{29}$$

As we mentioned above, we consider two regulatory regimes: (i) the fixed capital requirement regime and (ii) the procyclical capital requirement regime. In the former regime, θ_t is fixed at θ_0 and θ_1 is set equal to zero. Thus, only the time-varying equity issuance cost is maintained. This regime mimics the Basel I regulation. In the latter regime, both variables are time varying. The time-varying capital requirement in this regime is intended to mimic the Basel II regulation.

3.3.1 Capital Requirement Ratio

Let us first discuss the calibration of the capital requirement ratio. Under the procyclical regulation regime, the capital requirement ratio increases (decreases) when the economy is in a downturn (boom), implying that $\theta_1 < 0$. We calibrate θ_0 and θ_1 by using the Basel II risk-weight formula. Note first that Basel II requires banks to calculate the risk weight for each loan based on the formula, which is detailed in the appendix. Roughly speaking, the formula takes PD, LGD, and maturity of the loan and yields the risk weight for that loan. We calculate the average risk weight (i.e., the average capital requirement ratio) by using average values of these three variables. LGD is taken to be 40 percent, a level consistent with the evidence mentioned before. Average maturity is taken to be 2.5 years. Last, we use the Moody's default rate series for PD. We then obtain a time series of the economy-wide risk weight over the period 1982 and 2004. This series fluctuates over time because of the time variations in the default rate series. We calibrate θ_0 and θ_1 to replicate the cyclical features of this series. Specifically, the aggregate risk-weight series has a mean level of 8 percent and volatility of 0.10 after applying the HP filter with smoothing parameter of 1,600. We thus set θ_0 to 0.08. We also obtain $\theta_1 = -8$ as a value that matches the volatility of the HP-filtered series. Under the fixed capital requirement regime, $\theta_1 = -8$ is set equal to zero.

3.3.2 Equity Issuance Cost

First, consider the mean level of this cost γ_0. In the macro literature, Cooley and Quadrini (2001) assume that raising new equity involves a 30 percent premium over debt financing. In an application of the same model to the banking sector, Zhu (2008) also uses the same value; he argues that, referring to Van den Heuvel (2009), the 30 percent premium (relative to deposit financing) is in line with the size of the tax burden on bank profits, which can be a part of bank equity. It is, however, somewhat misleading to apply this value to our model directly, because this estimate is supposed to be costs for new equity, whether in the form of bank profits or new equity issuance. While the parameter γ_0 in our model corresponds to the cost of issuing new equity, it seems naive to adopt the literal interpretation to calibrate this parameter because in reality banks do not issue new equity every period. Instead, it seems more appropriate to interpret γ as overall costs for maintaining a certain level of bank equity. We thus refer to the evidence on the return on equity (ROE) in the banking sector. According to the Call Report, the average annual ROE in the period between 1985 and 2008 is 11.7 percent. In particular, between 1994 and 2006 the average ROE is more than 14 percent. Our reduced-form specification should include both the cost of internal equity and the cost associated with issuing new equity. Given that the latter cost may be much higher than the former as mentioned above, we set γ_0 at 0.05, which roughly corresponds to the annual cost of 20 percent.

To calibrate the elasticity parameter γ_1, we refer to the evidence provided by Kashyap and Stein (2004) in which they analyze the cyclicality of "capital charges," which represent the total amount of required bank capital. Based on their empirical results, they argue that the cyclical pressure on bank capital positions can be accounted for roughly equally by the higher requirement ratio induced by the Basel II risk-sensitive regulation and the higher shadow cost of capital.[22] This finding leads us to set $\gamma_1 = \theta_1 = -8$ in the baseline calibration. However, their result is based on the observations between 1998 and 2002, the period which is arguably a calm period in terms of financial distress. As argued by Repullo and Suarez (2009), the shadow cost of bank equity can become very high during the time of

[22]More precisely, Kashyap and Stein (2004) write, "The added cyclical pressure on bank capital positions associated with Basel II is of almost the same order of magnitude as the pre-existing baseline effect under Basel I." Recall that the shortfall of the capital positions under Basel I can be thought of as solely due to the increase in the shadow cost of capital.

financial distress. Therefore, we also consider the cases (as plausible alternatives) in which γ_1 is set to larger negative values (-12 and -15).

3.4 Solving the Model

We solve the model non-linearly by applying the projections' PEA (parameterized expectation algorithm). Specifically, we parameterize the conditional expectations in the two Euler equations (15) and (20) by a tensor product of the Chebyshev polynomials of the three state variables Z_t, K_t, and A_t. Given the parameterized values of the expectations, we solve for all endogenous variables. We can then evaluate the conditional expectations and see if they are close to the initially postulated conditional expectations. We iterate on this process until the coefficients of the parameterized functions converge.[23]

4. Permanently Higher Capital Requirement

Before discussing our main results, this section focuses on the steady-state effects of a permanent increase in the capital requirement ratio while suppressing fluctuations of aggregate productivity.[24] This helps us gain insights into the underlying mechanism of the main results in the subsequent section.

4.1 Level Effects

We first consider the effects when the capital requirement ratio is raised from 8 percent to a *permanently* higher level, 12 percent. The effects on the key macro variables are summarized in table 4. Since the higher capital requirement imposes a higher cost for financial intermediation, it reduces the total number of loans and liquidity

[23]The integrals associated with the aggregate shock are numerically calculated by using the Gauss-Hermite quadrature with five nodes. The integrals associated with the liquidity shock are calculated by using Simpson's rule with fifty-one nodes, since the distribution is truncated by $\overline{\omega}$.

[24]Because business cycle fluctuations of the economy are suppressed (i.e., $A_t = 1$) in this section, the difference between Basel I and Basel II is not relevant.

Table 4. Steady-State Effects of Higher Capital Requirement

Output	−0.07	Net Worth	−0.07
Capital Stock	−0.29	Liquidity Dependence	−0.02
Investment	−0.29	PD	+0.01
Total Loans	−0.27	Price of Capital	+0.22

Notes: Table reports percentage changes of each variable (except PD) when the capital requirement ratio is raised from 8 percent to 12 percent. The change in PD is expressed as percentage points.

dependence. The increase in PD reflects the fact that the higher capital requirement makes it more difficult for the bank-entrepreneur relationship to withstand the liquidity shock, meaning that more positive NPV projects are abandoned. This directly affects aggregate investment and the capital stock. The effect on aggregate output amounts to 0.07 percent. Finally, because supply of the capital good declines, the price of capital increases.[25]

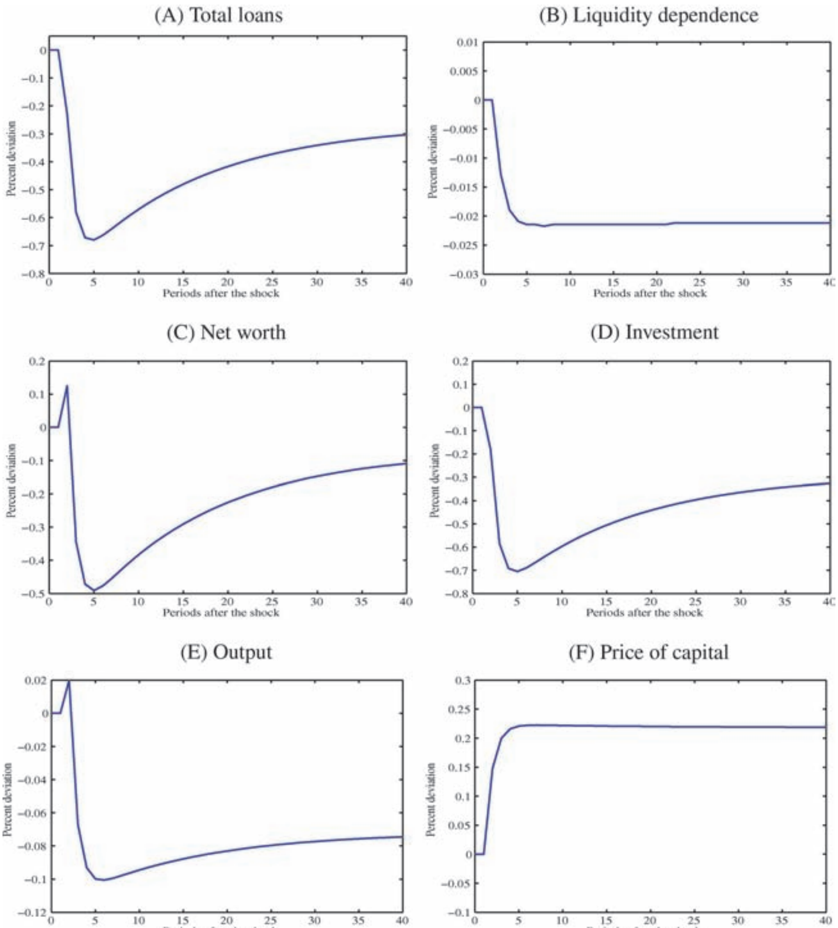
We would like to stress that the level effects highlighted here should be interpreted with caution because, as we mentioned above, our model takes into account only the costly aspects of capital requirements. For instance, higher capital requirements advocated in the aftermath of the recent financial crisis play a useful role in avoiding bank failures, which could result in significant negative effects on the macroeconomy. Our model has nothing to say about the possible benefits of higher capital requirements because there is no bank failure in our setup.

4.2 Transition Dynamics

Figure 1 presents the transition dynamics from the initial steady state to the new steady state. Interestingly, before reaching the new steady state, some variables go through larger declines, exhibiting (inverse) hump-shaped patterns. In particular, additional declines

[25]To save space, we did not report the effects on consumption and the interest rate. But not surprisingly, aggregate consumption declines and the interest rate increases.

Figure 1. Responses to a Permanently Higher Capital Requirement



Notes: Plotted are transition paths in response to a permanent increase in the capital requirement ratio from 8 percent to 12 percent.

in investment and total loans are most noticeable, whereas the path of output shows only a modest hump. The reason for this larger decline in investment is due to the household's substitution from savings into consumption. As mentioned in footnote 25, aggregate consumption eventually reaches a permanently lower level.

However, it increases initially because the substitution effect is large enough.[26] This substitution effect results mainly from the increasing price of capital over the first few quarters, as shown in panel F.

5. Procyclicality of Bank Capital Regulation

This section presents the main results of this paper, quantifying the procyclical effects of bank capital requirements. We first briefly summarize the quantitative properties of the model without the capital requirement. The procyclicality of capital requirements is then analyzed in the two regulatory regimes.

5.1 Dynamics in the No-Requirement Economy

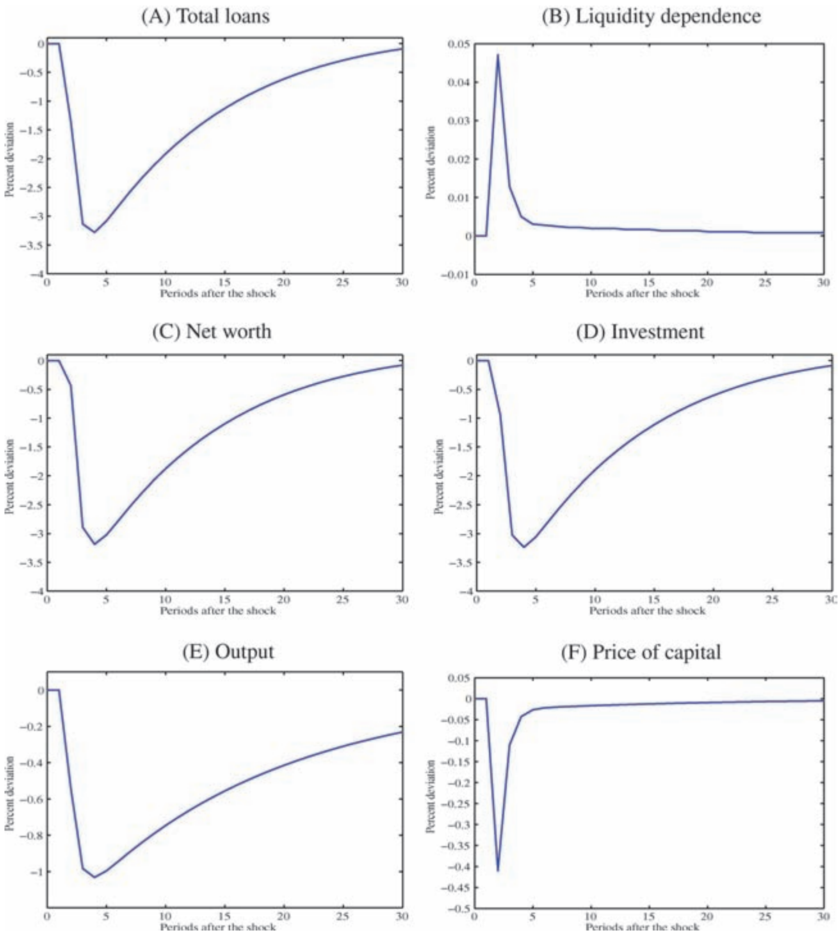
We first briefly discuss the dynamics of the model without imposing the capital requirement. Given our assumption that equity issuance is costly (with no benefits), banks hold no equity in this case.[27] Figure 2 presents the responses to a one-standard-deviation negative aggregate productivity shock in this economy.

Since this economy is already analyzed by Kato (2006), we will be brief and emphasize the results that are particularly relevant to our paper. In our view, a key feature of this model is the behavior of liquidity provision by banks in the wake of an adverse shock. In particular, observe that the firm's liquidity dependence *increases* when the negative shock occurs. That is, it is optimal for the bank to increase the liquidity dependence, allowing entrepreneurs to better withstand their liquidity needs in bad times. Naturally, aggregate loan demand falls in response to the negative shock since the marginal product of capital falls. However, the increase in liquidity dependence dampens the impact of the productivity shock on

[26] Aggregate consumption is largely driven by households' consumption because the share of entrepreneurs' consumption is very small in the model.

[27] Our calibration assumes the presence of the capital requirement and associated equity issuance cost. To solve the no-requirement economy, we simply set the capital requirement to zero. This had minimal impact on the steady-state values of the model presented in table 3.

Figure 2. Responses to a Productivity Shock

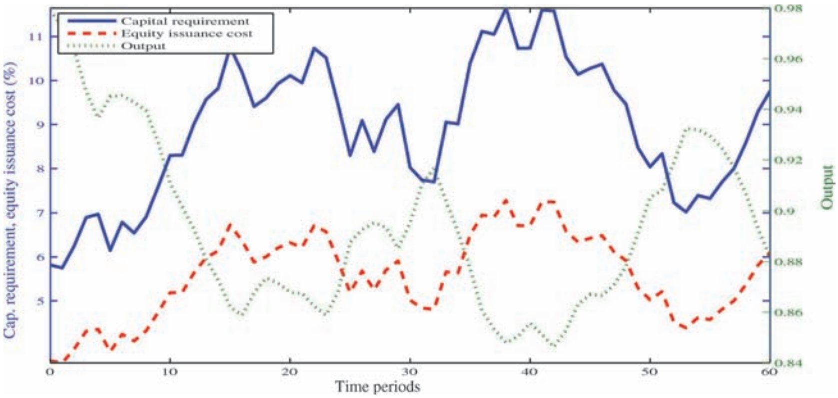


Notes: Plotted are responses to a one-standard-deviation negative productivity shock. The capital requirement ratio and the equity issuance cost are held fixed at 8 percent and 5 percent, respectively.

output. As emphasized by Kato (2006), this "smoothing" mechanism generates a hump-shaped response in aggregate output.[28]

[28] Kato (2006) compares output responses in his model and the standard RBC model and shows that the RBC model implies a larger initial response of output. While our model differs from Kato's in terms of calibration as well as the model itself, we also obtain the hump-shaped response.

Figure 3. Sample Paths of Capital Requirements and Aggregate Output



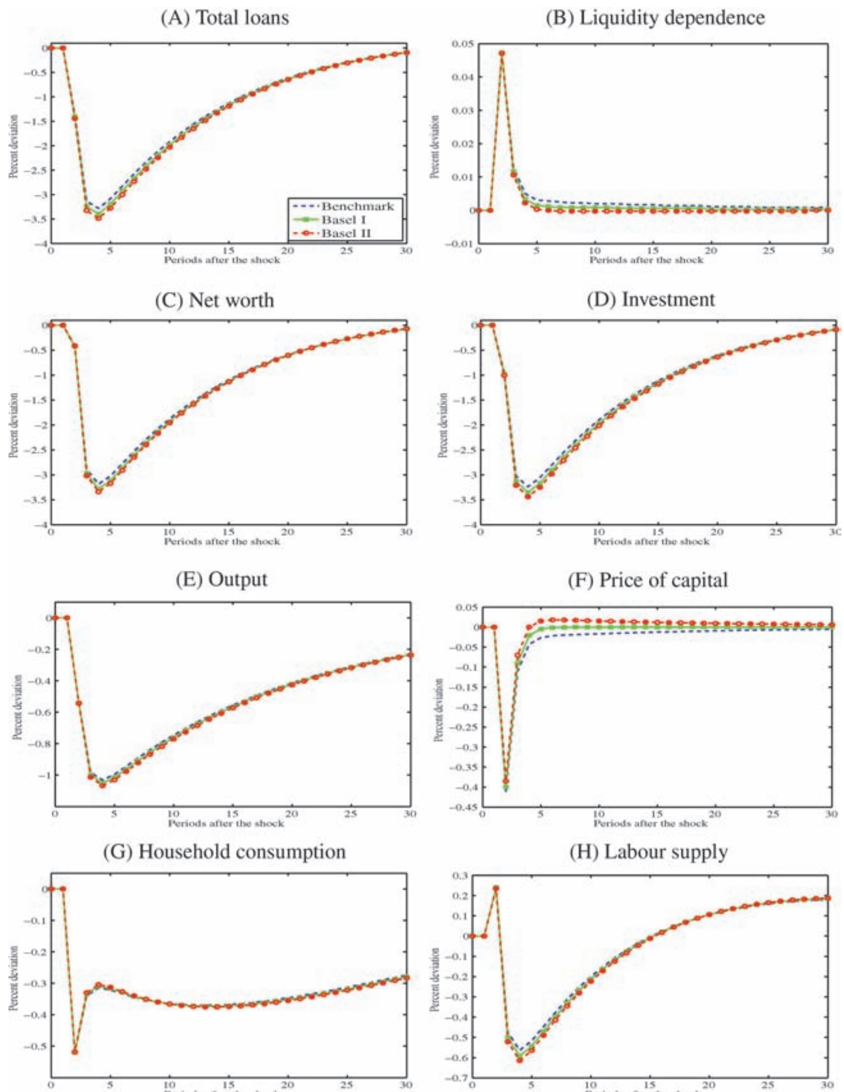
Notes: Plotted are simulated series of the capital requirement ratio (applied only in the Basel II regime), the equity issuance cost (applied in both regimes), and the aggregate output series. The output series is the one generated under the economy with no capital requirement and measured along the right axis.

5.2 Impulse Responses

We now consider the exercises to quantify the procyclicality induced by the two forms of capital requirements. We link the TFP series with the capital requirement ratio and the equity issuance cost through equations (27) and (28), respectively. Remember that these specifications assume that the two variables are perfectly correlated (negatively) with the aggregate productivity series. It is informative to see how they behave relative to aggregate output in our economy. Figure 3 compares the two series with aggregate output series in the no-requirement economy. The figure shows that the Basel II capital requirement ratio takes its highest value at around 11 percent and its lowest value at around 6 percent over the fifteen-year simulation period. The variability of this series appears to be plausible, since it is obtained using the available empirical evidence in the risk-weight formula of Basel II. The figure also plots our simulated series for the equity issuance cost. As implied by our calibration, it has the same proportional amplitude as the capital requirement ratio.

Figure 4 shows how the economy responds to a negative TFP shock under the different regulatory regimes. Recall that the

Figure 4. Procyclical Effects of Basel I and Basel II



Notes: Plotted are responses to a one-standard-deviation negative TFP shock under the three economies: in the benchmark economy, $\theta_t = 0.08$ and $\gamma_t = 0.05$; in the Basel I economy, $\theta_t = 0.08$ and $\gamma_t = 0.05A_t^{-8}$; and in the Basel II economy, $\theta_t = 0.08A_t^{-8}$ and $\gamma_t = 0.05A_t^{-8}$.

benchmark economy imposes no capital requirements. Figure 4 plots responses of household consumption and labor supply as well as the same six variables considered before. Panel A shows that the number of loans falls more under the presence of the capital requirements; at the lowest point in the responses, which occur in the third quarter, the differences from the benchmark economy amount to 12 basis points and 20 basis points under Basel I and Basel II, respectively. Panel B shows the differences in liquidity dependence. Remember that the model's key feature is that the entrepreneurs become more dependent on credit lines in bad times. Tighter capital requirements make the liquidity dependence more difficult. Interestingly, the peak responses under the three cases, which occur in the impact period, are very close to each other. However, capital requirements play a role in depressing the persistence of this variable.

The second row of figure 4 shows the responses of the entrepreneur's net worth and investment in the three economies. In line with the larger declines in lending under Basel II, declines in entrepreneurial net worth and investment are also more accentuated in that case. Because equity is more costly than deposits, the increases in the capital requirement and equity issuance cost in bad times raise the costs of external funds to entrepreneurs. Consequently, entrepreneurial profits and thus net worth are reduced by more under Basel II. The differences in declines in investment are significant; the declines are amplified by 12 basis points and 20 basis points at its trough, which occurs in the third quarter after the shock hits the economy.

The third row presents the responses of output and the price of capital. The output responses are hump-shaped across all cases. Again, the fall of output is most significant under Basel II. At the trough of the economy, the differences from the benchmark model amount to 2 basis points and 3 basis points in Basel I and Basel II, respectively. This can be considered relatively small. One important issue to remember, however, is that these are the differences with respect to a one-standard-deviation shock. In a more severe recession, the differences can be significantly larger. We will come back to this issue again below. As for the behavior of the price of capital, observe that the pattern of the responses is reversed. That is, declines in the price of capital are smallest under Basel II. It even goes above its steady-state level several periods after the shock. Remember that

Table 5. Output Volatility

	No Requirement	Basel I	Basel II
Baseline ($\gamma_1 = -8$)	1.84	1.87	1.92
	—	(1.016)	(1.043)
$\gamma_1 = -12$	—	1.89	1.94
	—	(1.027)	(1.054)
$\gamma_1 = -15$	—	1.91	1.97
	—	(1.038)	(1.071)

Notes: Results are based on 500 replications of 200 observations (after randomization of the initial condition). The standard deviations are based on logged HP-filtered series with a smoothing parameter of 1,600. Numbers in parentheses report relative volatility compared with that under the economy with no capital requirement.

the higher capital requirement itself raises the price of capital, since it reduces the supply of capital goods, as indicated by the exercise in section 4. Thus, declines in the price of capital in the face of the negative productivity shock are mitigated under this regime.

The responses of household consumption, presented in panel G, are more nuanced. First of all, the differences are difficult to observe; second, after sharp declines in the impact period, which are of the same magnitude across all three cases, consumption recovers more under Basel II. This stems from the behavior of the price of capital; upward trajectories of the price of capital from the second period on are most pronounced in that case, inducing the steepest recovery in consumption (see the discussion in sub-section 4.1). However, the differences in the behavior of consumption are relatively small. Last, panel H shows that the largest declines in labor supply occur under Basel II.

5.3 Output Volatility

The first row of table 5 translates the impulse responses in output into unconditional standard deviations. The numbers are based on logged HP-filtered series. When the economy moves from the benchmark no-requirement economy to the Basel I economy, output volatility increases from 1.84 percent to 1.87 percent. Moving to the Basel II economy raises the volatility to 1.92 percent.

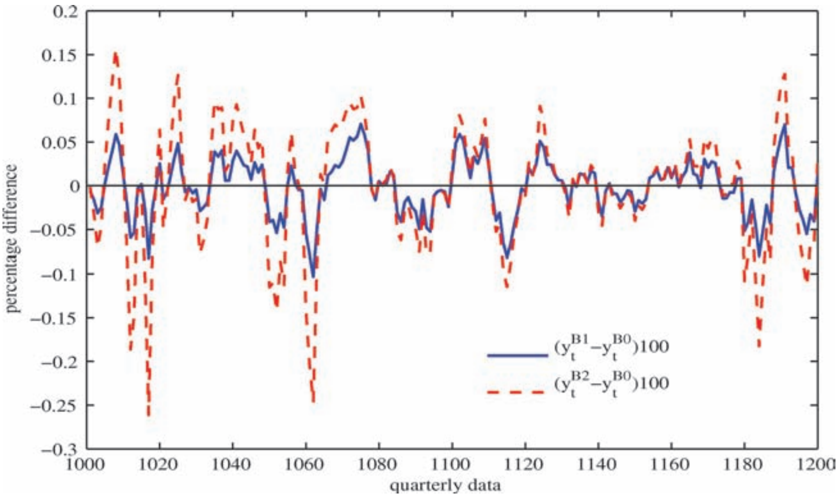
The remaining rows of table 5 present the comparisons with cases where the elasticity of the equity issuance cost is set to larger negative values. Note first that in the baseline calibration with $\gamma_1 = -8$, the equity issuance cost fluctuates within a relatively narrow range, as can be seen in figure 3; the standard deviation of the series is 0.9 percent and a three-standard-deviation upward move raises the equity issuance cost to roughly 8 percent from 5 percent. When the fluctuations of the equity issuance cost are raised by using alternative elasticities -12 and -15, the three-standard-deviation upward moves from the mean correspond, respectively, to 9.3 percent and 10.4 percent. The table shows that the volatility difference between the no-requirement economy and the Basel II economy (Basel I economy) reaches 10 (5) basis points when the elasticity is set larger than $\gamma_1 = -12$.

5.4 Effects during Booms and Downturns

Note that differences in the standard deviations convey only the information regarding the average effect. Perhaps more interesting are the differences in output paths under different regulatory environments during booms and downturns. More specifically, we simulate a long time series (10,000 observations) of aggregate output in the three regulatory environments—economies with (i) no capital requirement, (ii) the Basel I regulation, and (iii) the Basel II regulation—feeding the same sequence of the underlying driving process. We then calculate differences in logged HP-filtered output series at each period. Let y_t^{B0}, y_t^{B1}, and y_t^{B2} represent these output series in the no-requirement economy, the Basel I economy, and the Basel II economy, respectively. Figure 5 plots a randomly chosen sample of 200 quarterly observations of $(y_t^{B1} - y_t^{B0})100$ and $(y_t^{B2} - y_t^{B0})100$ under the baseline calibration ($\gamma_1 = -8$). These series capture the effects of Basel I and Basel II relative to no capital requirement. As can be seen, the average differences can be relatively small: the mean absolute deviations of the 10,000 observations are 4 basis points and 7 basis points, respectively. However, the graph also shows that there are several occasions in this twenty-year period in which the deviations are quite pronounced. For example, there are two instances where the effect of Basel II exceeds 25 basis points.

To systematically quantify these infrequent yet important events, table 6 presents the 1st, 5th, 95th, and 99th percentiles of the

Figure 5. Sample Paths of Differences in Output ($\gamma_1 = -8$)



Notes: The graph plots the sample paths (200 quarterly observations) of output differences, expressed as log-differences times 100. y_t^{B0}, y_t^{B1}, and y_t^{B2} are logged HP-filtered output series in the no-requirement economy, the Basel I economy, and the Basel II economy, respectively.

distributions of differences in output. Note that the lower part of the distribution (1st percentile and 5th percentile) capture the effects during downturns and the upper part corresponds to the effects during booms.[29] Let us focus on the lower tail of the distributions. Under the baseline calibration with $\gamma_1 = -8$, the 1st percentile of the distribution amounts to 12 basis points when Basel I and no-requirement economies are compared. The magnitude goes up to 40 basis points in the comparison of Basel II and no-requirement economies. It is true that these numbers correspond to the effects during the "rare" episodes.[30] However, even when we consider the

[29]Observe that the distributions are somewhat asymmetric. This asymmetry simply comes from the fact that the equity issuance cost is bounded by zero by construction of the process.

[30]Note that it is somewhat misleading to interpret 1st and 99th percentiles as once-every-100-quarter (twenty-five-year) events because these events tend to occur in a cluster given that the driving process is persistent. These events, however, do occur every forty to fifty years in our simulated series.

Table 6. Output Differences in Booms and Downturns

Percentiles		1	5	95	99
Baseline	$(y_t^{B1} - y_t^{B0})100$	−0.12	−0.08	0.06	0.09
	$(y_t^{B2} - y_t^{B0})100$	−0.40	−0.18	0.15	0.25
	$(y_t^{B2} - y_t^{B1})100$	−0.27	−0.11	0.09	0.17
$\gamma_1 = -12$	$(y_t^{B1} - y_t^{B0})100$	−0.22	−0.12	0.10	0.15
	$(y_t^{B2} - y_t^{B0})100$	−0.61	−0.24	0.20	0.38
	$(y_t^{B2} - y_t^{B1})100$	−0.39	−0.13	0.11	0.25
$\gamma_1 = -15$	$(y_t^{B1} - y_t^{B0})100$	−0.32	−0.16	0.13	0.21
	$(y_t^{B2} - y_t^{B0})100$	−0.83	−0.30	0.25	0.53
	$(y_t^{B2} - y_t^{B1})100$	−0.51	−0.14	0.14	0.34

Notes: The table reports output paths under the three regulatory environments that are hit by the same sequence of the underlying shock. Results are based on a long time series (10,000 observations) of logged and HP-filtered data. Each reported number is the log-difference (multiplied by 100) at each percentile of the distribution. y_t^{B0}, y_t^{B1}, and y_t^{B2} are logged HP-filtered output series under no capital requirement, the Basel I capital requirement, and the Basel II capital requirement, respectively.

5th percentile, output differences are not trivial, especially when the Basel II economy is compared with the no-requirement economy. The second and third panels of table 6 present the results from the same calculations under larger elasticities. The results indicate that the procyclicality induced by capital requirements is indeed sizable, especially during severe downturns. Overall, the procyclical effects of Basel I and Basel II amount to 10–15 basis points and 20–40 basis points, respectively.

6. Conclusion

In this paper, we have quantified the procyclical effects of bank capital requirements. Relative to previous studies, our analysis is based on a general equilibrium model where capital goods production is subject to the agency problem studied by Holmstrom and Tirole (1998). We find that, across various plausible calibrations, Basel I and Basel II contribute to increasing the standard deviation of output fluctuations by around 5 basis points and 10 basis points, respectively. The magnification effect can be significantly larger around business cycle peaks and troughs.

Many simplifying assumptions we made allowed us to quantify the macroeconomic effects of bank capital requirements in a general equilibrium model. Our model thus misses several aspects that are considered important in the earlier literature. First, as mentioned elsewhere, the way we impose capital requirements is very stylized. In particular, our model lacks a mechanism that generates a precautionary capital buffer, and thus capital requirements are always binding. As explored in Repullo and Suarez (2009), incorporating this "smoothing" motivation can alter the quantitative assessment of the procyclical effects. Second, we a priori assumed the existence of capital requirements without modeling why bank capital regulation exists to begin with. In this regard, capital requirements can be motivated as a device limiting the bank's moral hazard under the presence of deposit insurance as in Van den Heuvel (2008). Introducing the welfare-improving aspects of capital requirements could allow us to conduct a more serious welfare analysis. Extending the model along these dimensions is an important avenue for future research.

Appendix

The internal ratings based (IRB) approach uses the probability of default, the loss given default, the exposure at default, and maturity for each exposure to calculate the bank's capital requirement for each loan (see Basel Committee on Banking Supervision 2004). The capital requirement ratio is derived from the following formula:

$$\begin{aligned}&\text{Capital Requirement Ratio}_t\\&\quad = LGD \times \left[N\left(\frac{N^{-1}(PD_t) + \sqrt{C_t} \times N^{-1}(0.999)}{\sqrt{1-C_t}}\right) - PD_t\right]\\&\qquad \times \left(\frac{1+(M-2.5)\times b_t}{1-1.5\times b_t}\right),\end{aligned}$$

where $N(.)$ is the distribution function of standard normal, PD is the probability of default, LGD is loss given default, and M is average maturity of loans. C is called the correlation factor and is linked with PD through the following expression:

$$C_t = 0.12 + 0.12 \times \exp^{-50\times PD_t}.$$

Finally, there is a parameter, b, which adjusts the maturity of the loan according to its risk as found in various quantitative impact studies conducted by banks as requested by national supervisors and the Basel Committee. This parameter is defined as follows:

$$b_t = [0.11852 - 0.0547 \times \log(PD_t)]^2.$$

References

Araten, M., M. Jacobs Jr., and P. Varshney. 2004. "Measuring LGD on Commercial Loans: An 18-Year Internal Study." *The Risk Management Association Journal* 86: 96–103.

Basel Committee on Banking Supervision. 2004. "International Convergence of Capital Measurement and Capital Standards: A Revised Framework." (June).

Bernanke, B., and M. Gertler. 1989. "Agency Costs, Net Worth, and Business Fluctuations." *American Economic Review* 79 (1): 14–31.

Bernanke, B., M. Gertler, and S. Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics*, Vol. 1A, ed. J. Taylor and M. Woodford, chapter 21, 1341–93. Elsevier.

Blum, J., and M. Hellwig. 1995. "The Macroeconomic Implications of Capital Adequacy Requirements for Banks." *European Economic Review* 39 (3–4): 739–49.

Carlstrom, C., and T. Fuerst. 1997. "Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis." *American Economic Review* 87 (5): 893–910.

Chen, N. 2001. "Bank Net Worth, Asset Prices and Economic Activity." *Journal of Monetary Economics* 48 (2): 415–36.

Cooley, T., and V. Quadrini. 2001. "Financial Markets and Firm Dynamics." *American Economic Review* 91 (5): 1286–1310.

Covas, F., and W. Den Haan. 2010. "The Role of Debt and Equity Finance over the Business Cycle." Unpublished Manuscript (March).

Dorefeenko, V., G. Lee, and K. Salyer. 2008. "Time-Varying Uncertainty and the Credit Channel." *Bulletin of Economic Research* 60 (4): 375–403.

Heid, F. 2007. "The Cyclical Effects of the Basel II Capital Requirements." *Journal of Banking and Finance* 31 (12): 3885–3900.

Holmstrom, B., and J. Tirole. 1997. "Financial Intermediation, Loanable Funds, and the Real Sector." *Quarterly Journal of Economics* 112 (3): 663–91.

———. 1998. "Private and Public Supply of Liquidity." *Journal of Political Economy* 106 (1): 1–40.

Kashyap, A., R. Rajan, and J. Stein. 2002. "Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking." *Journal of Finance* 57 (1): 33–74.

Kashyap, A., and J. Stein. 2004. "Cyclical Implications of the Basel II Capital Standards." *Economic Perspectives* (Federal Reserve Bank of Chicago) 28 (Q1): 18–31.

Kato, R. 2006. "Liquidity, Infinite Horizons and Macroeconomic Fluctuations." *European Economic Review* 50 (5): 1105–30.

Meh, C., and K. Moran. 2010. "The Role of Bank Capital in the Propagation of Shocks." *Journal of Economic Dynamics and Control* 34 (3): 555–76.

Repullo, R., and J. Suarez. 2009. "The Procyclical Effects of Bank Capital Regulation." Unpublished Manuscript, CEMFI.

Schuermann, T. 2009. "Hearing on Commercial & Industrial and Commercial Real Estate Lending." Testimony before the Congressional Oversight Panel for the Troubled Asset Relief Program, New York, May 28.

Sufi, A. 2009. "Bank Lines of Credit in Corporate Finance: An Empirical Analysis." *Review of Financial Studies* 22 (3): 1057–88.

Townsend, R. 1979. "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory* 21 (2): 265–93.

Van Den Heuvel, S. 2008. "The Welfare Cost of Bank Capital Requirements." *Journal of Monetary Economics* 55 (2): 298–320.

———. 2009. "The Bank Capital Channel of Monetary Policy." Unpublished Manuscript (August).

Zhu, H. 2008. "Capital Regulation and Banks' Financial Decisions." *International Journal of Central Banking* 4 (1): 165–211.

Discussion of "Procyclicality of Capital Requirements in a General Equilibrium Model of Liquidity Dependence"

Javier Suarez
CEMFI and CEPR

1. Introduction

The paper that motivates this discussion belongs to a literature tradition (centered on capturing the effects of financial frictions in otherwise "standard" dynamic macroeconomic models) whose evolution and interest has been boosted by the current crisis. The crisis has made suddenly evident that conventional macroeconomic models offered an insufficient understanding, if at all, of banking, liquidity, and credit market phenomena whose development have been of great importance to the world economy during the last three years.

I will take the discussion of the paper as an excuse to give some general thoughts on the literature of reference. I will first frame the discussed paper in one particular branch of the tradition, then comment on alternative avenues and developments explored by the wider literature of reference, and finally come back to the discussed paper, with a brief summary of its modeling strategy and my critical reading of its main results. In that last part I will elaborate on why, in this type of model, the assessment of the likely effects of a significant rise in capital requirements (say, due to Basel III) may significantly differ from those obtained in models that give explicit consideration to the frictions that affect the dynamics of bank capital accumulation.

2. Placing the Current Paper in the Literature

The analysis provided in the article by Francisco Covas and Shigeru Fujita is based on a model in the tradition of Carlstrom and Fuerst

(1997). It explores the implications of incorporating ingredients taken from a specific well-known microeconomic model of financial imperfections into an otherwise standard real business cycle model. In Carlstrom and Fuerst (1997), the microeconomic setup was taken from Bernanke and Gertler (1989), while in the paper discussed here, the core microeconomic setup is taken from Holmstrom and Tirole (1998), henceforth HT. One might say that the authors of this paper fit the micro model of HT in a macro setup using essentially the same modeling "tricks" as Carlstrom and Fuerst (1997).

HT develops a three-date model that emphasizes the liquidity provision role of banks for financially constrained firms that, after obtaining some initial funding for an investment project, may suffer shocks that imply some additional liquidity need prior to the terminal maturity of the project. HT characterize the optimal contract that may govern the relationship between each firm and the intermediary that finances it. The contract can be naturally interpreted as a combination of an initial loan and a credit line that can be used or not along the remaining life of the lending relationship. HT is a good reference model if one wants to emphasize the liquidity provision role of credit lines—which the authors of the discussed paper do in the introduction and, paradoxically, not so much when discussing their results.

In fact, as recognized by Covas and Fujita, the idea of exploring the implications of the HT model into a real business cycle framework was already undertaken by Kato (2006), which is then the true closest reference for this paper. The distinctive and novel incremental contribution with respect to Kato (2006) yields at the explicit consideration of bank capital and its excess cost.

3. Incorporating Financial Frictions in Dynamic Macroeconomics

Many papers in the wider literature of reference can be described as attempts to explore dynamic macroeconomics mechanisms whose importance and microeconomic details had so far only been explored in rather static (and frequently also partial equilibrium) corporate finance and banking models.

Introducing the financial imperfections that affect firms' or households' investment decisions into the real business cycle or, more recently, into dynamic stochastic general equilibrium (DSGE) models is not immediate. It requires adopting what I have provocatively denoted as "tricks": whatever more or less convincing assumptions that one adopts with the main purpose of embedding the microfinance model into the macroeconomic model without altering "too much" the structure of any of them.

One first set of tricks is required in order to insert the typical risk-neutral, static (or at most two- or three-period) optimizing agents of the micro models into the dynamic structure of the macro models, where typically there is a representative household that is risk averse and optimizes over an infinite horizon.

The big alternative to this approach would be to try to derive the details of the micro part in the macroeconomic structure directly and in a fully consistent way. So far this avenue is not the most prevalent, especially among central bank researchers, but it has been explored already. Albuquerque and Hopenhayn (2004) and Clementi and Hopenhayn (2006), for instance, study infinite-horizon optimal financial contracting into particular general equilibrium setups with fully optimizing agents using a reasonably tractable "recursive utility" formulation. The problem with this approach (more satisfactory in terms of microfoundations and overall consistency) is that the techniques required to determine the equilibrium dynamic contracts are at odds with the current standards of application of DSGE models, especially at central banks.

Specifically, the models with complex dynamic financing problems typically do not yield closed-form solutions for the equilibrium contracts and tend to require considering additional state variables for the recursive representation of equilibrium. For example, decision rules may be the outcomes of Bellman equations whose solution one must determine, with numerical methods, at the same time as the whole dynamic general equilibrium is solved. Additionally, these models easily produce heterogeneity among the agents subject to financial constraints and to possibly idiosyncratic non-perfectly diversifiable or non-insurable shocks. Frequently, the full distribution of wealth among the wealth-constrained agents will have to be treated as an additional state variable.

Hence the models with fully fledged dynamic contracting problems stemming from financial frictions force the analyst to move away from solution techniques based on log-linearizing a set of manageable equilibrium conditions around some non-stochastic steady state. And this poses a problem from the perspective of the current standards of application of DSGE models.

Not surprisingly, most of the tricks found in the papers developed by central bank researchers and their academic periphery adopt the strategy of just "fitting" some essentially static problem with financial constraints into the dynamic macro model. And there are various available sub-strategies for doing this. One possibility is to simply add successive generations of short-lived agents that are around for at least (but not much more than) two dates, i.e., that operate for one period and then die or exit the economy. For example, these agents may be entrepreneurs who are born penniless or with limited wealth, get funding from the representative household (or from a bank that collects savings from such household), develop some specific one-period investment project, and die after its completion. At that point they are replaced by a new generation of entrepreneurs who enter the economy.

Many papers in this literature (starting with the seminal contribution of Bernanke and Gertler 1989) rely on simply incorporating some entrepreneurial sector whose essentially static funding is subject to imperfections described by a well-known micro model (costly state verification, moral hazard, adverse selection, etc.). Such entrepreneurial sector is typically assumed to be funded by a competitive financial market or banking sector that "lends" to the entrepreneurs the funds saved by more conventional infinitely lived households that truly optimize over time.

The models resulting from adopting this very trick are a good start but produce too little history dependence. In particular, they do not capture the dynamics of accumulation of wealth or net worth by the financially constrained entrepreneurial sector. To get more history dependence at a small modeling cost, one possibility is to introduce warm-glow bequests so as to make some entrepreneurial wealth pass from each generation of entrepreneurs to the next (see, for instance, Aghion and Bolton 1997). The trick here is to allow for intergenerational wealth transmission without making the entrepreneurs of each generation fully internalize the utility of the next generation. If they internalize this, say under standard altruistic preferences, then they will have to solve a fuller dynamic optimization problem, and the resulting model will have the same tractability problems as the models associated with the full dynamic-contracting approach.

Another possible approach is to think of infinitely lived risk-neutral entrepreneurs that somehow along the equilibrium path (and insofar as they remain active entrepreneurs) are always financially constrained and, consequently, are always saving as much as possible (an example along these lines is Kiyotaki and Moore 1997). In such an environment, entrepreneurs will tend to find it optimal to keep accumulating any earnings resulting from prior investments as net worth used for the self-financing (or as collateral in the external financing) of their subsequent investments. In this approach, entrepreneurs are (at least implicitly) assumed to solve a dynamic problem but one whose solution is (or is assumed to be) trivial in the part that refers to their wealth accumulation decisions.

In fact, in these models one typically needs to adopt some additional trick to prevent the internally accumulated net worth of the entrepreneurial sector from growing unboundedly. The typical trick here is to add some assumption (perhaps an exogenous shock) whose effect is to force (some) entrepreneurs to decumulate wealth. For instance, one can assume that some shocks arrive (say, following a Poisson process) that make entrepreneurs die or retire (or lose access to their investment projects), in which case they lose or decide to consume their accumulated wealth. One can alternatively assume that, for some reason (perhaps obtaining output from their projects which cannot be sold in the market), entrepreneurs are forced into something equivalent to paying out (or consuming) a fraction of their earnings.

In some other models, the process of net worth accumulation by entrepreneurs gets endogenously bounded by assuming that entrepreneurs are more impatient than the typical saving household (perhaps, but not necessarily, in a setup where households are assumed to differ in their discount factors so that, in equilibrium, some of them act as lenders while others act as borrowers).

A more sophisticated construction along these lines is the one developed in Gertler and Karadi (2009) and Gertler and Kiyotaki (2010), where the representative household has the special class of financially constrained entrepreneurs (and also some financially constrained "bankers") as some of its members. The representative household is assumed to provide consumption insurance to these agents which, effectively, then behave as risk neutral with respect to the supposedly diversifiable idiosyncratic risk of their investment activities. In this setup, agents are somewhat "schizophrenic" since for the purposes of consumption smoothing they are part of the representative household, while as entrepreneurs, they obtain some wealth from the representative household when starting up and, then, keep managing that wealth separately, on their own, until they fail or exit (in which case any residual wealth reverts back to the representative household). So the household insures entrepreneurs' consumption needs but does not insure or back their businesses (i.e., the financial needs related to their investment projects). Therefore, if entrepreneurs have too little (or run out of) wealth in their business activity, the household to which they belong does not "recapitalize" them. Hence, this sophisticated structure is still one full of tricks.

Papers in this literature need further tricks in connection with the final goal of exploring the quantitative implications of the resulting models in a way that facilitates comparison with mainstream quantitative macro models. This may require describing the production process of, say, the final consumption good using a Cobb-Douglas production function, so that labor shares and capital-to-output ratios can be matched to their empirical counterparts according to standard practice.

To make such a feature compatible with the producing role assigned to entrepreneurs, many papers introduce (following Bernanke and Gertler 1989) an explicit capital-producing sector in which the entrepreneurs act as the producers. Hence, instead of assuming, like in the canonical neoclassical growth model, that the consumption good can be transformed into the capital good using a frictionless, linear, reversible technology, models with financial frictions typically assign the role of transforming the consumption good into the capital good to the wealth-constrained entrepreneurs. The entrepreneurs produce capital out of projects subject to moral hazard problems or some other type of agency, contract, or informational imperfection that justifies the frictions that affect their financing.

4. Banks, Bank Capital, and the Discussed Paper

Prior to the current crisis, the macro literature had formally accepted these and similar tricks as part of their stock of knowledge, but I think that many macroeconomists felt uncomfortable about the practice of continually adding tricks to the basic frameworks. Perhaps this explains the little progress made after the synthesis provided by Bernanke, Gertler, and Gilchrist (1999), henceforth BGG, which incorporated a canonical financially constrained entrepreneurial sector model (based on costly state verification frictions) into a New Keynesian model with nominal frictions.

The arrival of the financial crisis led to the sudden discovery that the mainstream dynamic model and even the BGG model did not have banks. The latter, in particular, had financially constrained entrepreneurs and nice credit spreads that could move with the business cycle but no specific bank or banks, and hence no specific role for bank capital.

Recent research efforts have led to the emergence of a new generation of models that attempt to incorporate banks in the analysis. These new models include Meh and Moran (2010) and the previously mentioned Gertler and Karadi (2009) and Gertler and Kiyotaki (2010). Their strategy can be summarized as consistent on adding a second layer of financially constrained agents which are the banks or their owner-managers (the "bankers").

Among the references in the microeconomic literature relevant for this task, Holmstrom and Tirole (1997) occupies a prominent position. In that paper, moral hazard problems affect the incentives of both the entrepreneurs and the banks that monitor some of them. Specifically, the banks can ameliorate the incentive problems of the entrepreneurs in some intermediate net worth range. However, in order to undertake their own costly monitoring activity, bankers need to have the right incentives, and this essentially requires them to own a stake in the success returns of the funded projects. Such a stake justifies the incentive role of bank capital: bankers contribute their own wealth to the funding of the projects in exchange for a share in the success returns. So in this model bank capital plays for banks' funding essentially the same role as entrepreneurial net worth for entrepreneurial funding: it reduces the deadweight costs of external financing. This model is also nice in that it provides an intuitive rationale for (market-based) bank capital requirements.

Meh and Moran (2010) use Holmstrom and Tirole (1997) as the microeconomic model of reference and formalize the dynamics of bank capital in essentially the same way that the dynamics of entrepreneurial net worth had been modeled by the inherited literature tradition. In Gertler and Karadi (2009) and Gertler and Kiyotaki (2010), the rationale for bank capital comes from the existence of an enforceability problem (bankers might run away with a fraction of the resources under their management) that can be ameliorated by making bankers contribute their wealth to the funding of the bank. Bank capital also accumulates as the result of earnings retention by bankers.

The approach taken by Covas and Fujita in the discussed paper departs from this line of research in that it does not incorporate bankers who accumulate wealth so as to contribute it as equity. In fact, in the model, banks operate repeatedly for just a period and there are no meaningful bank capital dynamics. Regulation imposes that banks must finance a fraction θ of their lending with equity capital. Equity capital is provided to the bank by the representative household out of its savings essentially in the same way as it also provides deposit funding. However, equity funding is assumed to involve some "issuing cost": a resource cost equal to a proportion γ of the equity used in each period. Hence, the regulatory capital requirement is to all effects equivalent to a proportional tax on bank lending with a tax rate equal to $\gamma\theta$.

The paper explores the steady-state and cyclical implications of Basel I and Basel II capital requirements, taking into account the possibility that the excess cost of equity γ is sensitive to the cyclical position of the economy. Under Basel I, θ is assumed to be constant while under Basel II θ is described as a smooth increasing function of a total factor productivity parameter A, whose assumed random evolution is the final source of (cyclical) fluctuations in the model economy. The cyclical variability of γ is also captured in reduced form by making γ a smooth function of A.

The model is then carefully calibrated and its steady-state and cyclical properties are analyzed with state-of-the-art techniques. A few tables and graphs containing numerous impulse response functions summarize the main results. One section is devoted to analyzing the effects of permanent and transitory increases in a (Basel I type) cyclically invariant capital requirement. Another section examines the cyclicality added by a (Basel II type) cyclically variant requirement.

The results are nicely explained in the paper and yield a picture of qualitative effects that go well in line with intuition. Quantitatively, however, the effects of both increasing capital requirements and making them more cyclical seem to me quite tiny. For instance, in the baseline parameterization, a permanent unanticipated increase in capital requirements from 8 percent to 12 percent produces a deviation in total bank lending and investment of less than −0.7 percent in the short run and about −0.3 percent in the long run. The deviation in total output is of less than −0.1 percent.

One possible reading of the results is that capital requirements do not matter much after all. However, I think a fairer assessment of the results is that capital requirements cannot have important effects in this very type of model unless its calibration were stretched too much. Specifically, it turns out that if the excess cost of equity funding is calibrated to have a mean value of 5 percent (like in table 2 in the paper), moving the capital requirement from 8 percent to 12 percent essentially moves the implicit "tax" on bank lending from 0.4 percent to 0.6 percent and, eventually, it is the incidence of this "tax" the force that moves everything else in the exercise. Are these numbers reasonable?

On the one hand, an excess cost of equity funding of 5 percent is really a big number, since the model is calibrated on a quarterly basis and the model implies that the bank reissues its entire equity capital base at the beginning of every period, which in practice means that one unit of equity funding has an attributed yearly excess cost of 20 percent! On the other hand, when this number is combined with capital requirements of 8 percent and 12 percent, it implies a yearly tax on bank lending of 1.6 percent and 2.4 percent, respectively. And these numbers are important but arguably not very large.

In fact, I think that the numbers are possibly too large as an estimate of the long-run implications of having capital requirements in place, since a typical industry estimate of the required rate of return on bank equity is 10 percent per year (not 20 percent!). However, I also think that the results in the paper are likely to underestimate the (shadow) cost of equity capital and its contractive impact on bank lending during the transition from the steady state with $\theta = 8\%$ to the steady state with $\theta = 12\%$. Let me explain why.

Accommodating such a change in capital requirements with just a tiny decline in bank lending (as in the results presented in the paper) means that banks should be able to increase their equity funding by about 50 percent in a single quarter. Fifty percent is a huge increase by historical standards and one that, possibly, has never before been accommodated, at an industry-wide level, by issuing all the required extra equity in the market. I suspect that raising all that capital in a single quarter might not be feasible or only at a very high, perhaps prohibitive cost.

I would expect that, in practice, banks in the real world would start accumulating earnings at perhaps a higher speed than normal (say, by sacrificing payouts to their shareholders) so as to make their equity funding gradually converge to some new desired long-term level of capitalization. So in the transition, the internally accumulated equity funding of the banks (the wealth of the bankers in some of the models commented above) would be scarcer than usual and its shadow value would then be higher than usual. This will have two main effects: (i) it will make bank lending more expensive and smaller in quantity than in the new steady state, and (ii) it will make bank earnings per unit of equity capital temporarily higher than in the new steady state.

The first effect points to a temporary credit crunch effect that the current model cannot capture (see Repullo and Suarez 2009 for a model with its own tricks that captures temporary credit crunches and in which banks hold buffers of excess capital to partially prevent the crunches from happening). The second points to a force that will tend to endogenously accelerate the process of convergence to the new steady state (since it should allow speeding up the process of internal accumulation of equity funding) and is also missed in the current analysis.

So the main deficit of the model proposed by Covas and Fujita is the explicit treatment of bank equity capital as a stock variable that can be increased via earnings retention and, perhaps, also by issuing equity but definitely not at the same cost. The perfect model in this field is still to be produced. Ideally, we would like to have endogenously determined earnings retention, equity issuance, and payout policies. This is a challenge not only because of the difficulties involved in finding "tricks" with which to add them to a macro model but also in terms of the available microfoundations. Indeed the understanding of the distinction between inside and outside equity funding in corporate finance and banking literatures is still far from perfect, but that would be a story for another discussion.

References

Aghion, P., and P. Bolton. 1997. "A Theory of Trickle-Down Growth and Development." *Review of Economic Studies* 64 (2): 151–72.

Albuquerque, R., and H. Hopenhayn. 2004. "Optimal Lending Contracts and Firm Dynamics." *Review of Economic Studies* 71 (2): 285–315.

Bernanke, B., and M. Gertler. 1989. "Agency Costs, Net Worth, and Business Fluctuations." *American Economic Review* 79 (1): 14–31.

Bernanke, B., M. Gertler, and S. Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics*, Vol. 1A, ed. J. Taylor and M. Woodford, 1341–93. Elsevier.

Carlstrom, C., and T. Fuerst. 1997. "Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis." *American Economic Review* 87 (5): 893–910.

Clementi, G. L., and H. Hopenhayn. 2006. "A Theory of Financing Constraints and Firm Dynamics." *Quarterly Journal of Economics* 121 (1): 229–65.

Gertler, M., and P. Karadi. 2009. "A Model of Unconventional Monetary Policy." Mimeo, New York University.

Gertler, M., and N. Kiyotaki. 2010. "Financial Intermediation and Credit Policy in Business Cycle Analysis." Forthcoming in *Handbook of Monetary Economics*, Vol. 3B, ed. B. M. Friedman and M. Woodford. Elsevier.

Holmstrom, B., and J. Tirole. 1997. "Financial Intermediation, Loanable Funds, and the Real Sector." *Quarterly Journal of Economics* 112 (3): 663–91.

———. 1998. "Private and Public Supply of Liquidity." *Journal of Political Economy* 106 (1): 1–40.

Kato, R. 2006. “Liquidity, Infinite Horizons and Macroeconomic Fluctuations.” *European Economic Review* 50 (5): 1105–30.

Kiyotaki, N., and J. Moore. 1997. “Credit Cycles.” *Journal of Political Economy* 105 (2): 211–48.

Meh, C., and K. Moran. 2010. “The Role of Bank Capital in the Propagation of Shocks.” *Journal of Economic Dynamics and Control* 34 (3): 555–76.

Repullo, R., and J. Suarez. 2009. “The Procyclical Effects of Bank Capital Regulation.” Mimeo, CEMFI.

Capital Regulation and Risk Sharing*

Douglas Gale
New York University

Capital requirements are the principal tool of macroprudential regulation of banks. Bank capital serves both as a buffer and as a disincentive to excessive risk taking. When general equilibrium effects are taken into account, however, it is not clear that higher capital requirements will reduce the level of risk in the banking system. In addition, an increase in the required capital ratio can force banks to take on more risk in order to achieve target rates of return.

JEL Codes: G01, G21, G28.

1. Introduction

Minimum capital requirements are one of the three "pillars" of macroprudential regulation. As a result of the financial crisis of 2008–09, there have been proposals to increase the amount of capital banks are required to hold. A capital structure that contains a substantial amount of equity has a number of advantages. It reduces the bank's vulnerability to market freezes; it reduces the risk of contagion to other financial institutions; it reduces the subsidy provided by deposit insurance; and, as we have recently seen, shareholders are less likely to be bailed out by government than debt holders. But while it may be optimal to have a substantial amount of equity in the capital structure, the crucial question is, "How much?" Will

 The ideas in this commentary were included in some form in presentations at the Federal Reserve Bank of San Francisco Symposium on Banking and Financial Reform, the GWU-IMF Conference on Financial Regulation and Supervision: Lessons from the Crisis, and the IJCB's Second Conference on Financial Stability at the Banco de España in Madrid. I am grateful to the participants for their comments, especially to Mark Flannery, David Martinez, and Rafael Repullo. Author contact: Department of Economics, New York University, 19 W. 4th Street, New York, NY, USA 10012. E-mail: douglas.gale@nyu.edu. Web site: www.nyu.edu/econ/user/galed/.

banks choose the right capital structure, left to themselves, or does the government have to force them to raise more capital?

One of the main arguments for increasing capital requirements is that it will provide banks with incentives to take less risk. More precisely, it is argued that if shareholders have a larger equity stake in the bank (more "skin in the game"), the incentives to engage in risk shifting or asset substitution will be reduced. But while this may be true in some circumstances, the idea that it is universally true must be treated with some skepticism. There may be a case that higher capital requirements will reduce risk, but the case has not yet been made. The theoretical literature gives us grounds for doubting that increased capital requirements will necessarily reduce risk taking, whatever other benefits it might have.

In this note, I attempt to do three things. First, I review some general equilibrium studies of the impact of capital requirements. Analyses of the impact of capital requirements on risk taking are often presented in a partial equilibrium framework. In a general equilibrium context, there may be complex feedbacks that lead to unexpected comparative static results. Second, I revisit the classical risk-shifting argument that underlies the claim that capital reduces risk. This too is a partial equilibrium argument. In particular, it ignores the factors that determine the supply and cost of capital. When the cost of capital is high, forcing banks to raise more capital may have the result of increasing the probability of default. In the third part of this note, I present a sketch of a model in which managers have target rates of return which force them to "reach for yield." In this model, the effects of greater capital on risk taking are turned upside down. A short conclusion sums up my views on capital requirements and risk.

2. General Equilibrium Perspectives

2.1 Incomplete Markets

Broadly speaking, there are two functions of bank capital. First, capital provides a buffer that allows for the orderly disposal of assets and shields debt holders from losses. Without an adequate buffer, assets might be sold in a "fire sale" that would amplify the losses to depositors or the deposit insurance corporation. This is the *risk-sharing function* of capital. Second, the fact that shareholders suffer the first loss gives the shareholders, or the managers acting on their behalf, an incentive to avoid unnecessary risks. This is the *incentive function* of capital.

Financial institutions have an incentive to adopt an optimal capital structure, one that will, among other things, improve risk sharing and reduce agency costs. In the absence of some kind of market failure, there is no need to impose regulation to force banks to hold the right amount of capital. If markets are incomplete, however, there may be a divergence between privately and socially optimal capital structures. In fact, it is generically possible to find a welfare-improving intervention (Geanakoplos and Polemarchakis 1986). This is the approach followed in Gale (2004) and Gale and Özgür (2005). It turns out that it is possible to improve on the laissez-faire capital structure, but the optimal policy does not necessarily take the form of a minimum capital requirement.

Gale (2004) and Gale and Özgür (2005) use a variant of the model developed in Allen and Gale (1998). There are three dates: $t = 0, 1, 2$. Investments and contracting decisions are made at date 0; assets pay off and consumption occurs at dates 1 and 2.

There is a single good that can be used for consumption or investment. Financial intermediaries can invest in two assets, a *short* asset and a *long* asset. The short asset is represented by a storage technology: one unit invested at date t yields one unit at date $t+1$. The long asset is represented by a constant returns to scale technology that is only available at date 0: one unit invested in the long asset at date 0 produces $R > 1$ at date 2.

Depositors have Diamond-Dybvig preferences: with probability λ a depositor is an early consumer, who only values consumption at date 1, and with probability $1 - \lambda$ he is a late consumer, who only values consumption at date 2. So the expected utility of a depositor who is promised c_1 if he withdraws at date 1 and c_2 if he withdraws at date 2 is

$$\lambda U(c_1) + (1 - \lambda) U(c_2).$$

For each intermediary i, the proportion of early consumers is a random variable λ_i with mean λ. To keep things simple, one can assume that the risk is purely idiosyncratic, so that the economy-wide fraction of early consumers is constant and equal to λ:

$$\int_0^1 \lambda_i di = \lambda.$$

All uncertainty is resolved at the beginning of the second date, when depositors learn whether they are early or late consumers and each intermediary observes the fraction of early consumers.

The Allen-Gale model, in the tradition of Bryant (1980) and Diamond and Dybvig (1983), assumes that free entry and competition force intermediaries to offer contracts that maximize the welfare of their depositors. There is a perfectly elastic supply of capital at the exogenously determined opportunity cost of ρ. The advantage of this assumption is that it allows one to characterize the intermediary's optimal behavior as the solution to a contracting problem in which the intermediary maximizes the depositors' welfare subject to the intermediary's zero-profit constraint and the investors' participation constraint.

Capital is assumed to be "expensive" in the sense that the opportunity cost of capital ρ is greater than the return on the long and short assets:

$$\rho > R.$$

Intermediaries offer completely contingent, incentive-compatible risk-sharing contracts to consumers and equity contracts to investors, taking as given the opportunity cost of capital ρ. Since there is no aggregate uncertainty, the payments to depositors and investors are contingent only on intermediary i's idiosyncratic liquidity shock, that is, the fraction λ_i of early consumers at intermediary i.

An intermediary has one unit of deposits and k units of capital at date 0 and invests these in a portfolio consisting of y units of the short asset and $1 + k - y$ units of the long asset. At date $t = 1, 2$ in state λ_i, intermediary i promises $c_t(\lambda_i)$ units of consumption to depositors who withdraw at date t and promises investors dividends of $e_t(\lambda_i)$. If the liquidity shock is λ_i, the intermediary has to supply $\lambda_i c_1(\lambda_i)$ to early consumers who withdraw at date 1 and pay a dividend $e_1(\lambda_i)$. The return from the short asset is y, so in the event that $\lambda_i c_1(\lambda_i) + e_1(\lambda_i) > y$, it will be necessary to sell some of the long asset. Since there is no aggregate uncertainty, the price of the long asset at date 1 will be known with certainty at date 0.

Even though intermediaries maximize their depositors' welfare, equilibrium is *constrained inefficient* when markets are incomplete. Whereas individual intermediaries take the price of the long asset as given, a central planner, by contrast, would take into account the impact of their choices on the price of the long asset. When markets are complete, a small change in the asset price will help some intermediaries and hurt others but have no impact on general welfare. When markets are incomplete, by contrast, these pecuniary externalities will generically have an effect on (ex ante) welfare.

Allen and Gale (2004) show that if the depositors' coefficient of risk aversion is high enough, an increase in the price of the long asset will increase welfare. This could be accomplished by requiring all intermediaries to hold more liquidity (more of the short asset). An increase in the minimum capital requirement has the opposite effect. In order to minimize the burden of raising more capital, the intermediary invests most of the additional capital in the long asset, which raises the ratio of the long asset to the short asset and lowers the price of the long asset at date 1. Thus, imposing a minimum capital requirement that raises capital above the laissez-faire level may reduce welfare.

2.2 Charter Values

A bank's charter value is the difference between its value as a going concern and its liquidation value. Since the charter value will be lost if the bank fails, its existence gives managers an incentive to avoid excessive risk taking that might lead to bankruptcy. This may offset the effect of deposit insurance (Marcus 1984; Keeley 1990). Charter value is reflected in the value of equity, and the conventional wisdom is that charter value and bank capital have similar effects on risk taking.

Competition may reduce charter value and increase financial instability. Allen and Gale (2001), henceforth AG, provide a theoretical analysis of the relationship between competition and financial stability. AG assume that banks engage in Cournot competition in the deposit market. Individual banks choose the quantity of deposits they want, and the upward-sloping aggregate supply curve for deposits determines the market-clearing deposit rate. Banks take into account the effect of an increase in their own demand for deposits on the interest they must pay. The larger the number of banks, the smaller the impact of a single bank on the equilibrium deposit rate. Thus, as the number of banks increases, the quantity of deposits increases and the deposit rate rises to clear the market. An increase in the deposit rate reduces profits and, hence, the charter value. As the charter value falls, the incentive for risk shifting increases and so does the risk of bank failures.

Boyd and De Nicolò (2005), henceforth BDN, criticize AG for ignoring the impact of competition on risk taking by *firms*. They extend the AG model by introducing a market for loans. Instead of investing directly in safe and risky assets, banks lend money to firms. Each firm wants to undertake a risky project that requires an investment of one unit. Firms have heterogeneous opportunity costs but are otherwise identical. This assumption gives rise to a downward-sloping demand curve for loans. Banks compete in Cournot fashion. In the loan market, each bank chooses the amount it wishes to lend, and the downward-sloping demand curve determines the equilibrium loan rate. In the deposit market, as in AG, each bank chooses the amount of deposits it wants to obtain, and the upward-sloping supply curve determines the equilibrium deposit rate.

Firms are subject to moral hazard as in Stiglitz and Weiss (1981), henceforth SW. Each firm is assumed to choose from a continuum of projects that vary in the probability of success. A project yields a payoff R if successful and 0 otherwise; the probability of success, $p(R)$, depends on the payoff R. If the loan rate is r, the entrepreneur chooses R to maximize his expected return $p(R)(R - r)$. A higher loan rate encourages greater risk taking, as SW showed.[1]

[1]The first- and second-order conditions for a local optimum are

$$p'(R)(R - r) + p(R) = 0$$

and

$$p''(R) + 2p'(R) < 0.$$

The implicit function theorem tells us that

$$\frac{dR}{dr} = \frac{p'(R)}{p''(R) + 2p'(R)} > 0$$

because $p'(R) < 0$ and the denominator is negative by the second-order condition.

An increase in the number of banks leads to greater competition in the loan and deposit markets, raising the deposit rate and lowering the loan rate. Both changes will have the effect of lowering the charter value of the banks. But the level of risk in the economy is determined solely by the loan rate and, as we have noted, a fall in the loan rate will cause firm owners to choose less risky projects.

The essential difference between the AG and BDN models is that AG assigns the choice of risk (i.e., the choice between safe and risky assets) to banks, whereas BDN assigns the choice of risk (i.e., the risk of the project) to firms. In general, both effects are present, and a reduction in charter value may be associated with both greater risk shifting by banks and less risk taking by firms. The net effect of greater competition will depend on the balance of these two forces.

Hakenes and Schnabel (2007) combine the two effects to investigate the effect of capital requirements on risk taking. Like BDN, they assume that banks lend to firms and that firms choose the riskiness of the projects in which they invest. Like AG, they allow banks to have a direct hand in setting the level of risk by choosing the correlation of the risks in their loan portfolio. The treatment of capital is similar to Gale (2004) and Gale and Özgür (2005), but there is no risk-sharing function. So, in equilibrium, banks hold the minimum amount of capital required by regulation.

Increasing the capital requirement has four different effects. It has a direct effect, reducing charter value and increasing the incentive for banks to engage in risk shifting. Capital requirements also increase the bank's costs and reduce the level of activity. This will have indirect effects on the loan and deposit markets. It will lower the demand for deposits and hence lower the deposit rate, which increases the charter value and reduces the incentive to take on risk. It will lower the supply of loans and hence increase the loan rate. A higher loan rate increases the charter value, causing the bank to take less risk, but it also causes borrowers to take on more risk.

While these effects do not all go in the same direction, it can be shown that, under plausible conditions, the net effect of higher capital requirements will be to raise the riskiness of the bank's portfolio.

2.3 Imperfectly Correlated Risks

As we saw with the Hakenes-Schnabel model, the interaction of competition, charter values, and risk taking is affected by the correlation of risks. Many of the models in the banking literature assume that asset returns are perfectly correlated. Martinez and Repullo (2008) develop a single-factor model of risk and show that it can reverse the results obtained from models with perfectly correlated risks. In a single-factor model, an increase in the loan rate has two effects. It has the usual Stiglitz-Weiss effect of encouraging greater risk taking by borrowers, but it also has the effect of increasing the income from solvent borrowers. When risks are perfectly correlated, all borrowers survive or fail together. In a single-factor model, some fail and some survive, and the interest paid by the survivors can offset the losses of the failures. Thus, it is possible that, although greater competition reduces loan rates and, hence, reduces the incentives for borrowers to take risks, it also makes banks more vulnerable to failure because it reduces the buffer provided by the interest payments made by the surviving firms.

Martinez (2009) has used this model to study the impact of capital regulation on bank failures. As noted above, higher loan rates have two effects: they increase risk shifting by borrowers and raise buffers that offset potential losses. In Martinez's model, higher capital requirements reduce banks' leverage and, for a given asset risk, reduce the probability of bank failure. But higher capital requirements increase the cost of funding, which leads to higher loan rates and, possibly, riskier loans. Numerical simulations show that the relationship between capital requirements and the risk of bank failure is non-monotonic.

2.4 The Cost of Capital

The assumption that capital is in perfectly elastic supply at an exogenous opportunity cost ρ is obviously a simplification. A more satisfactory approach would recognize the endogeneity of the supply and cost of capital. This is one of several innovations found in a recent paper by De Nicolò and Lucchetta (2009), henceforth DNL. In their model, agents decide whether to become entrepreneurs, who set up banks to provide intermediation services, or to become investors, who purchase debt from the banks. This occupational choice determines the equilibrium capital asset ratio, and this ratio in turn has an important impact on the level of risk taking in the economy.

DNL assume there is a continuum of agents, $q \in [0, 1]$, where an agent of type q has an initial endowment qW. An agent can use his endowment to create $k \in (0, W)$ units of capital or he can keep his entire endowment qW and use it to buy bank debt. Since the amount of capital an agent can create is independent of the size of his endowment, an agent with a high (low) value of q has a comparative advantage in becoming an investor (entrepreneur). In equilibrium, there will be a cutoff q^* that determines who becomes an entrepreneur. Agents with $q < q^*$ choose to be entrepreneurs and those with $q > q^*$ choose to be investors.

An entrepreneur who has set up a bank can invest in projects that yield X per unit if successful and 0 otherwise. The probability of success p is a choice variable. Success requires costly effort. The cost of achieving success with probability p is assumed to be $\frac{1}{2}p^2$.

The division of returns is determined by the banks' ability to extract rents. This is measured by the return $\hat{R}$ per unit invested that is promised to investors. If L is the amount of debt issued by the representative bank, the entrepreneur's return is

$$p((1 - \hat{R})L + k)X - \frac{1}{2}p^2.$$

Maximizing this expression with respect to p gives us the first-order condition and decision rule:

$$p = ((1 - \hat{R})L + k)X.$$

Thus, the entrepreneur's equilibrium payoff is $\frac{1}{2}((1 - \hat{R})L + k)^2X^2$. Since the marginal type q^* must be indifferent between the two occupations, we have

$$\frac{1}{2}((1 - \hat{R})L + k)^2X^2 = pX\hat{R}q^*W.$$

Now, in order for the debt market to clear, the total debt issued by bank entrepreneurs must equal the investors' endowment, that is,

$$q^*L = \int_{q^*}^{1} qW = \frac{1}{2}(1 - q^{*2})W.$$

These two equations can be solved for the equilibrium value of q^*, which then determines L and the payoff of each agent.

DNL interpret the parameter $\hat{R}$ as a measure of the degree of competition in the banking sector. They argue that greater competition will raise the investors' share $\hat{R}$ toward the constrained optimum R^* that maximizes total surplus. By definition, as $\hat{R}$ increases, the entrepreneur's share of output falls. It can also be shown that q^* falls. But this does not mean that the entrepreneur's payoff falls. Fewer agents choose to be entrepreneurs and more choose to be investors, so L increases. In fact, L increases so much that the entrepreneur's payoff increases in spite of the increase in $\hat{R}$. DNL show that *both* the capital asset ratio and the level of risk *fall* as $\hat{R}$ rises to R^*. Risk can only fall if p increases and p is equal to the banker's return conditional on success. Thus, the banker's return conditional on success must be rising even though the investor's share $\hat{R}$ is also rising. The explanation of this counter-intuitive result is that, as the ratio of bankers to investors falls, the *scale* of each bank increases. The increase in scale for each bank leads to greater efficiency by both reducing the average fixed cost of setting up a bank and increasing the banker's incentive to take effort.

To sum up, whereas greater competition in AG and BDN is identified with an increase in the number of banks, in DNL it is identified with a fall in banks' ability to extract rents from investors. Since the marginal type q^* must be indifferent between becoming a bank entrepreneur and an investor, there is an increase in the returns to *both* investing and entrepreneurship. To achieve this, the number of banks decreases and the scale of each bank increases. The result of greater competition is an *increase* in charter value, and it is this increase in charter value that motivates the reduction in risk shifting, in spite of the fall in the capital asset ratio.

3. Risk Shifting Revisited

In reviewing these contributions to the literature on capital regulation, I have focused on the general equilibrium impact of capital on risk, without questioning the basic moral hazard model that underlies the view that more capital reduces the incentive for risk shifting. Now it is time to review the risk-shifting argument itself. There is a long literature on the role of capital structure on risk taking. In his classic paper on the option value of deposit insurance, Merton (1977) pointed out the rationale for capital requirements. He concluded that banks should be required to have more "skin in the game" as a counterweight to the incentive for excessive risk taking provided by (mispriced) deposit insurance. There followed a debate about the effects of capital on risk taking. A series of papers argued that in a mean-variance framework, higher capital ratios (lower leverage) do not necessarily lead a utility-maximizing manager to choose a portfolio with a smaller probability of default (Kahane 1977; Blair and Heggestad 1978; Koehn and Santomero 1980; and Kim and Santomero 1988). Furlong and Keeley (1989) and Keeley and Furlong (1990) argued that these authors had neglected the option value of deposit insurance and that in a model that included the option value of deposit insurance, it could be shown that increased levels of capital always lead to lower risk taking (as measured by the probability of default). Merton's analysis seems to be generally accepted these days and is certainly valid on its own terms, but that does not mean that there is no value in reconsidering the logic of the risk-shifting argument.

Consider the example in SW. A risk-neutral entrepreneur wants to undertake a risky venture that requires an investment of one unit. The entrepreneur has an initial endowment $0 < w < 1$ and raises the rest of the required investment by issuing debt with a face value of d. We assume that the debt is issued to risk-neutral investors whose opportunity cost of funds is one. Projects have two possible outcomes, success and failure, with returns R and 0, respectively. The probability of success, $p(R)$, is a decreasing function of the return R. The investors will be willing to purchase the firm's debt if it satisfies their participation constraint

$$p(R)d = 1 - w.$$

The entrepreneur chooses the project's target return $\hat{R}$ to maximize $p(R)(R - d)$. A revealed preference argument shows that $R^* \leq \hat{R}$ and $p(\hat{R}) \leq p(R^*)$, where R^* is the efficient project that maximizes

$p(R)R$.[2] Typically these inequalities are strict and the difference depends in a non-trivial way on the face value of the debt. Then, increasing w, the amount of "skin in the game," reduces the face value of the debt d and this in turn reduces the target return $\hat{R}$ and increases the probability of success $p(\hat{R})$.

The example is very simple, but it captures the essential logic of the argument that higher capital requirements reduce risk-shifting behavior. As the face value of the debt falls, the entrepreneur captures more of the total return and therefore internalizes more of the costs and benefits of his decision. How well does this correspond to reality? We have already seen that taking the supply of capital as exogenous can be misleading. From this point of view there are several aspects of the model that are special. First, the supply of capital is exogenous. The entrepreneur has a fixed amount of capital and invests it all in the risky project. Second, the entrepreneur has no alternative use for this capital and hence no opportunity cost. Third, it is assumed that, over the relevant range, risk and expected returns are negatively related. In other words, risk is mispriced. Finally, there is no room in this example for the common view that it is "expensive" to raise capital or that banks must "reach for yield" in order to satisfy shareholders.

Admati and Pfleiderer (2010) have argued that it is meaningless to say that capital is "expensive." Their argument is based on the Modigliani-Miller theorem (MMT), according to which the value of the firm is independent of the capital structure. The MMT is a useful counter-example to the idea that capital is universally "expensive," but the MMT itself is not universally valid. In particular, in a world of incomplete participation, where the purchasers of debt and equity are disjoint, the cost of equity finance may well be higher

[2]The efficient project R^* maximizes the expected return $p(R)R$, so

$$p(R^*)R^* \geq p(\hat{R})\hat{R}$$

and, since $\hat{R}$ maximizes the entrepreneur's expected return,

$$p(R^*)(R^* - d) \leq p(\hat{R})(\hat{R} - d).$$

These two inequalities imply that

$$-p(R^*)d \leq -p(\hat{R})d.$$

than the cost of debt finance, either because the purchasers of equity have better alternative investment opportunities or because equity holders are more risk averse at the margin.

3.1 A Model of "Return-Driven" Risk Taking

Suppose that equity and debt are held, at the margin, by different economic agents. Equity is held by a large number of risk-neutral investors, and debt is held by a large number of risk-averse consumers. Investors are assumed to have access to alternative investment opportunities that yield a return of ρ, so the opportunity cost of providing capital to the banking sector is ρ.

Banks invest in a constant returns to scale investment technology. The return to a unit investment is given by the usual SW technology: the return is R with probability $p(R)$ and 0 otherwise. We assume that capital is "expensive" in the sense that the returns on the banks' investments are lower than the investors' opportunity cost:

$$\rho > \sup_R p(R)R.$$

Since the scale of operation is immaterial, we assume that the representative bank raises one unit of deposits and k units of capital and invests the sum $1 + k$ in the investment technology with target return R.

To simplify the analysis, it is assumed that deposits are fully insured at no cost to the bank. Depositors are assumed to have an opportunity cost equal to one and supply funds elastically at this rate.

The interests of bank managers are not aligned with those of the shareholders. The bank's managers receive private benefits B if the bank survives and nothing otherwise. This ensures that managers have an incentive to avoid risk subject to the constraint that they must satisfy the shareholders.[3] The managers choose the amount of capital k, the face value of deposits d, and the portfolio return R

[3] Alternatively, instead of treating the cost of capital as a constraint on the manager's choice, we could assume that the shareholders provide high-powered incentives to the manager (e.g., stock options or shares) to achieve the desired return on equity.

to maximize $p(R)B$ subject to the participation constraints of the shareholders

$$p(R)(R-d)(1+k) \geq \rho k \tag{1}$$

and the debt holders $d \geq 1$. In equilibrium, these participation constraints will hold with equality, so we might as well set $d = 1$ from the outset.

Since capital is expensive and does not improve risk sharing (thanks to free deposit insurance), it is optimal to set $k = k_0$, where k_0 is the minimum amount of capital allowed under the Basel Accords. Then the manager will minimize R subject to the investors' participation constraint (1). Suppose that $f(R, k_0) = p(R)(R-1)(1+k_0) - \rho k_0$ is a concave function of k_0. For any feasible level of k_0, the participation constraint (1) will be satisfied for values of R that lie in an interval $[R_{\min}(k_0), R_{\max}(k_0)]$, where $R_{\min}(k_0)$ and $R_{\max}(k_0)$ are the values of R at which the graph of $f(R, k_0)$ intersects the horizontal axis. The manager chooses the lowest value of R that yields a non-negative value of $f(R, k_0)$; that is, he will set $R = R_{\min}(k_0)$. An increase in k_0 shifts the curve down, and the left intersection point $R_{\min}(k_0)$ with the horizontal axis shifts right. Thus, an increase in the minimum capital requirement forces the manager to take on more risk. See figure 1.

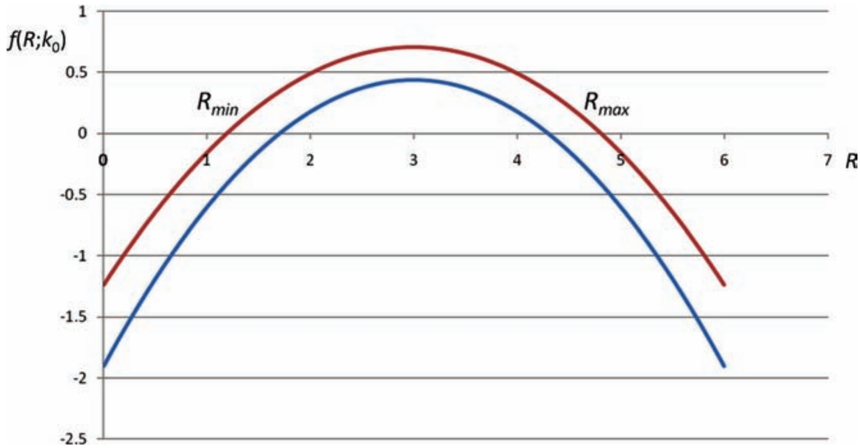
3.2 Risk-Weighted Capital

It might be argued that a more sophisticated system of capital requirements would take care of this problem. For example, if riskier investments require more capital, the manager will be forced to internalize the cost of increasing risk. But even if the risk of individual assets is held constant, the risk of the entire portfolio may be increased by choosing assets with more highly correlated returns.

Suppose there are two risky assets with returns R_i, for $i = 1, 2$. The returns are assumed to be independently and identically distributed, with distribution

$$R_i = \begin{cases} R & \text{w. prob. } p, \\ 0 & \text{w. prob. } 1-p. \end{cases}$$

Figure 1. Values of R Satisfying the Investors' Participation Constraint for Different Minimum Capital Requirements



The bank invests a fraction $0 \leq \alpha \leq \frac{1}{2}$ in asset 1 and the complementary fraction $1-\alpha$ in asset 2. Then the distribution of the bank's returns is given by

$$(\alpha R_1 + (1-\alpha)R_2)(1+k) = \begin{cases} R(1+k) & \text{w. prob. } p^2, \\ \alpha R(1+k) & \text{w. prob. } (1-p)p, \\ (1-\alpha)R(1+k) & \text{w. prob. } (1-p)p, \\ 0 & \text{w. prob. } (1-p)^2. \end{cases}$$

The banker's decision problem is to minimize the probability of default, subject to the investors' participation constraint. The minimum possible probability of default is achieved with $\alpha = \frac{1}{2}$. Suppose that $\frac{1}{2}R > 1$. Then it will be optimal under laissez-faire to set $k = 0$, and the probability of default will be $(1-p)^2$. If $k_0 > 0$, the banker will set $k = k_0$. For sufficiently small values of k_0, say $k_0 \leq k_0^*$, the probability of default will be $(1-p)^2$ as before but, for $k_0 > k_0^*$, we will have $\alpha R(1+k) < 1$ and the probability of default will be $(1-p)^2 + (1-p)p = 1-p$. In the first case, the participation constraint will require

$$p^2(R-1) + (1-p)p((1-\alpha)R-1) + (1-p)p(\alpha R-1) \geq \rho\frac{k_0}{1+k_0}$$

or

$$p^2(R-1) + (1-p)p(R-2) \geq \rho\frac{k_0}{1+k_0},$$

which is independent of α. Thus, the critical value of α is determined by the condition $\alpha^* R - 1 = 0$. For values of $k_0 > k_0^*$, we must have $\alpha < \alpha^*$, and the probability of default is $1-p$. Then the participation constraint is

$$p^2(R-1) + (1-p)p((1-\alpha)R-1) \geq \rho\frac{k_0}{1+k_0}.$$

This shows that increasing capital requirements can increase the risk of the bank's portfolio, even when all assets come from the same risk bucket.

3.3 A "Condominium" Theory of Banking

The preceding example is very stark, but it suggests a general perspective that may be worth pursuing further. Banking, as it has developed historically, requires people with very different risk tolerances to share the same institution. The problem is exacerbated by the transition from utility banking to casino banking, driven by high target returns on equity. Equity funding has benefits as well—e.g., improved risk sharing—but even then general equilibrium effects may imply a positive relation between capital and risk.

4. Conclusion

Tougher capital requirements may have positive benefits—they may reduce the consequences of market freezes, they may encourage banks to become smaller to avoid "systemic" capital requirements, and they may reduce contagion—but can they be relied on to reduce the risk of bank failure? Theory gives us a lot of contradictory results, and there is no overarching framework that holds out the promise of a clear, simple answer. Other tools may be needed and, possibly, structural changes in the banking industry, to avoid financial crises in the future.

References

Admati, A., and P. Pfleiderer. 2010. "Increased-Liability Equity: A Proposal to Improve Capital Regulation of Large Financial Institutions." Unpublished, Stanford University Graduate School of Business.

Allen, F., and D. Gale. 1998. "Optimal Financial Crises." *Journal of Finance* 53 (4): 1245–84.

———. 2001. *Comparing Financial Systems.* Cambridge, MA: MIT Press.

———. 2004. "Financial Intermediaries and Markets." *Econometrica* 72 (4): 1023–61.

Blair, R., and A. Heggestad. 1978. "Bank Portfolio Regulation and the Probability of Bank Failure: A Note." *Journal of Money, Credit, and Banking* 10 (1): 88–93.

Boyd, J., and G. De Nicolò. 2005. "The Theory of Bank Risk Taking and Competition Revisited." *Journal of Finance* 60 (3): 1329–43.

Bryant, J. 1980. "A Model of Reserves, Bank Runs, and Deposit Insurance." *Journal of Banking and Finance* 4 (4): 335–44.

De Nicolò, G., and M. Lucchetta. 2009. "Financial Intermediation, Competition, and Risk: A General Equilibrium Exposition." IMF Working Paper No. 09/105.

Diamond, D., and P. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91 (3): 401–19.

Furlong, F. T., and M. C. Keeley. 1989. "Capital Regulation and Bank Risk-Taking: A Note." *Journal of Banking and Finance* 13 (6): 883–91.

Gale, D. 2004. "Notes on Optimal Capital Regulation." In *The Evolving Financial System and Public Policy*, ed. Pierre St-Amant and Carolyn Wilkins. Ottawa: Bank of Canada.

Gale, D., and O. Özgür. 2005. "Are Bank Capital Ratios Too High or Too Low? Risk Aversion, Incomplete Markets, and Optimal Capital Structure." *Journal of the European Economic Association* 3 (2–3): 690–700.

Geanakoplos, J., and H. Polemarchakis. 1986. "Existence, Regularity, and Constrained Suboptimality of Competitive Allocations when the Asset Market Is Incomplete." In *Essays in Honor* *of Kenneth J. Arrow*, Vol. 3, ed. W. Heller, R. Starr, and D. Starrett. Cambridge: Cambridge University Press.

Hakenes, H., and I. Schnabel. 2007. "Bank Competition and Capital Regulation." Unpublished, Max Planck Institute, Bonn.

Kahane, Y. 1977. "Capital Adequacy and the Regulation of Financial Intermediaries." *Journal of Banking and Finance* 1 (2): 207–18.

Keeley, M. C. 1990. "Deposit Insurance, Risk, and Market Power in Banking." *American Economic Review* 80 (5): 1183–1200.

Keeley, M. C., and F. T. Furlong. 1990. "A Reexamination of Mean-Variance Analysis of Bank Capital Regulation." *Journal of Banking and Finance* 14 (1): 69–84.

Kim, D., and A. M. Santomero. 1988. "Risk in Banking and Capital Regulation." *Journal of Finance* 43 (5): 1219–33.

Koehn, M., and A. M. Santomero. 1980. "Regulation of Bank Capital and Portfolio Risk." *Journal of Finance* 35 (5): 1235–44.

Marcus, A. J. 1984. "Deregulation and Bank Financial Policy." *Journal of Banking and Finance* 8 (4): 557–65.

Martinez, D. 2009. "Capital Requirements and Bank Failure." Unpublished, CEMFI.

Martinez, D., and R. Repullo. 2008. "Does Competition Reduce the Risk of Bank Failure?" CEMFI Working Paper No. 0801.

Merton, R. 1977. "An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees: An Application of Modern Option Pricing Theory." *Journal of Banking and Finance* 1 (1): 3–11.

Stiglitz, J., and A. Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71 (3): 393–410.

Are Banks Too Big to Fail? Measuring Systemic Importance of Financial Institutions*

Chen Zhou
De Nederlandsche Bank and Erasmus University Rotterdam

This paper considers three measures of the systemic importance of a financial institution within an interconnected financial system. The measures are applied to study the relation between the size of a financial institution and its systemic importance. Both the theoretical model and empirical analysis reveal that, when analyzing the systemic risk posed by one financial institution to the system, size should not be considered as a proxy of systemic importance. In other words, the "too big to fail" argument is not always valid, and measures of systemic importance should be considered. We provide the estimation methodology of systemic importance measures under the multivariate extreme value theory (EVT) framework.

JEL Codes: G21, C14.

1. Introduction

During financial crises, authorities have an incentive to prevent the failure of a financial institution because such a failure would pose a significant risk to the financial system, and consequently to the broader economy. A bailout is usually supported by the argument that a financial firm is "too big to fail": that is, larger banks exhibit higher systemic importance. A natural question arising from the debate on bailing out large financial firms is why particularly large

*The author thanks Markus Brunnermeier, Gabriele Galati, Douglas Gale, Miguel Segoviano, Stefan Straetmans, Casper de Vries, Haibin Zhu, participants at the second Financial Stability conference of the International Journal of Central Banking, and an anonymous referee for useful discussion and suggestions. Views expressed do not necessarily reflect official positions of De Nederlandsche Bank. Author postal address: Economics and Research Division, De Nederlandsche Bank, 1000AB Amsterdam, The Netherlands. E-mail: C.Zhou@DNB.nl.

banks should be favored: are banks really too big to fail? An equivalent question might also be posed: does the size of a bank really matter for its systemic impact if it fails? The major difficulty in answering such a question is to design measures on the systemic importance of financial institutions. More specifically, we need to measure to what extent the failure of a particular bank will "contribute" to the systemic risk.

This paper deals with the problem in four steps. Firstly, we discuss the potential drawbacks of existing measures of systemic importance and propose alternatives that overcome these drawbacks. Secondly, we construct a theoretical model to assess whether larger banks correspond with higher systemic importance. Thirdly, we employ statistical methodology in estimating such measures within a constructed system consisting of twenty-eight U.S. banks. Finally, we use the estimated systemic importance measures and the size measures to empirically test the "too big to fail" statement.

Although the term "too big to fail" appears frequently in support of bailout activities, its downside is well acknowledged in the literature. Besides the distortion of the market discipline, the preference given to large financial firms encourages excessive risk-taking behavior, which potentially imposes more risk. Therefore, using the "too big to fail" argument to support intervention will result in a moral hazard problem: large firms that the government is compelled to support were among the greatest risk takers during the boom period. Furthermore, such a moral hazard problem will provide an incentive for firms to grow in order to be perceived as "too big to fail" (see Stern and Feldman 2004 for more discussions on the moral hazard problem).

Recently, both policymakers and academicians have begun to distinguish the size of a financial institution from the systemic importance it has by introducing new terms focusing on what the potential systemic impact might be if that particular institution fails. For example, Bernanke (2009) addresses the problem of financial institutions that are deemed "too interconnected to fail"; Rajan (2009) uses the term "too systemic to fail" to set the central focus of new regulation development. This urges the design of alternative measures on systemic importance. Measuring the systemic importance of financial institutions is particularly important for policymakers. It is the key issue in both financial stability assessment and macroprudential supervision. On the one hand, during crises, it is necessary to have such measures in order to justify bailout actions. On the other hand, it is crucial to supervise and monitor banks with higher systemic importance during regular periods. Policy proposals for stabilizing the financial system always rely on such measures. For instance, charging a deposit insurance premium is an alternative proposed by Acharya, Santos, and Yorulmazer (2010) for defending systemic risk. Systemic importance measures can serve as an indicator for pricing the corresponding insurance premium or taxation.

A few applicable measures of systemic importance have appeared in recent empirical studies. Adrian and Brunnermeier (2008) proposed the conditional value-at-risk (CoVaR) for measuring risk spillover. Similar to the value-at-risk measure, which quantifies the unconditional tail risk of a financial institution, the CoVaR quantifies how the financial stress of one institution can increase the tail risk of others. These measures demonstrate the bilateral relation between the tail risks of two financial institutions. The setup of the CoVaR measure indicates that it is designed for bilateral risk spillover. When applying CoVaR to assess the systemic importance of one financial institution to the system, it is necessary to construct a system indicator on the status of the system and then to analyze the bilateral relation between the system indicator and a specific bank. However, the complexity of the financial system is of a higher order than bilateral relations. Thus, a general indicator of the system is usually difficult to construct. Furthermore, the CoVaR measure is difficult to generalize into a systemic context in order to analyze multiple financial institutions all together. As an alternative, Segoviano and Goodhart (2009) introduced the "probability that at least one bank becomes distressed" (PAO). Comparing these two measures, we observe that the CoVaR is a measure of conditional quantile, while the PAO is a measure of conditional probability, which acts as the counterpart of conditional quantile. Within a probability setup, generalization from bivariate to multivariate is possible. However, the PAO measure focuses on the probability of having a systemic impact that there will be at least one extra crisis, without specifying how severe the systemic impact is. Therefore, the measure may not provide sufficient information on the systemic importance of a financial institution. Our empirical results partially reflect the less informative feature of the PAO: the PAO measures remain at a constant level across different financial institutions and across time.

Extending the PAO measure while staying within the multivariate context, we propose the systemic impact index (SII), which measures the expected number of bank failures in the banking system given a situation in which one particular bank fails. Clearly, the SII measure emphasizes more the systemic impact. We also consider a reversed measure: the probability that a particular bank fails, given that there exists at least one other failure in the system. This we refer to as the vulnerability index (VI).

To test "too big to fail," we first consider a theoretical model from which both size and systemic importance measures can be explicitly calculated. The model stems from the literature on systemic risk. In the literature, two categories of models consider crisis contagion and systemic risk in the banking system: banks are systemically linked via either direct channels such as interbank markets or indirect channels such as similar portfolio holdings in bank balance sheets. Studies in the first category focus on the contagion effect; i.e., a crisis in one financial institution may cause crises in the others. The second category of studies focuses on modeling systemic risk; i.e., crises of different financial institutions may occur simultaneously. For the first category of studies, particularly on modeling the interbank market, see Allen and Gale (2000), Freixas, Parigi, and Rochet (2000), and Dasgupta (2004). Cifuentes, Shin, and Ferrucci (2005) consider two channels: similar portfolio holdings and mutual credit exposure. They show that contagion is mainly driven by changes in asset prices. Hence the indirect channel dominates. For models focusing on the indirect channel, Lagunoff and Schreft (2001) assume that the return of one bank's portfolio depends on the portfolio allocation of other banks, and they show that crises can either spread from one bank to another or happen simultaneously due to forward-looking behavior. De Vries (2005) starts from the fat-tail property of the underlying assets shared by banks, and he argues that this creates the potential for systemic breakdown. For an overview of contagion and systemic risk modeling, we refer to de Bandt and Hartmann (2001) and Allen, Carletti, and Babus (2009).

The contagion literature focuses mainly on explaining the existence of a contagion effect: how a crisis in one financial institution leads to a crisis in another. Thus, the models usually consider the risk spillover between only two banks. To address the financial system as a complex entity, several studies have considered network models combined with bilateral spillover. Following those theoretical studies, empirical analyses, such as the CoVaR measure, were then designed for measuring bilateral relations. In order to address the systemic importance issue within a systemic context, we must consider a multibank approach. Thus, we establish our theoretical model based on the indirect channel models: similar portfolio holdings lead to the possibility of simultaneous crises. The fundamental intuition is that banks are interconnected due to the common exposures on their balance sheets. Thus, the systemic importance of a particular bank is closely associated with the number of different risky banking activities in which the bank participates. This, in turn, may not be directly associated with the bank's size. In other words, "too big to fail" is not always valid. Our model further shows that banks concentrating on few specific activities can grow large without increasing their systemic importance.

We acknowledge a potential downside of considering the indirect channel model: it does not provide a model on the causality effect. Nevertheless, even without addressing the contagion effect, investigating systemic importance based on the systemic risk of having simultaneous crises is an important issue. This comes from the "too many to fail" phenomenon discussed by Acharya and Yorulmazer (2007). They consider a game theory approach and show that because regulators would bail out a bank in distress only if a large part of the system suffers from distresses, individual banks would have an incentive to hold similar portfolios in order to increase the possibility of being rescued when a crisis occurs. Such a "too many to fail" phenomenon is in line with the intuition that when evaluating systemic importance of a financial institution, it is necessary to evaluate to what extent the crisis of the financial institution is accompanied by crises of others. Therefore, we choose to build the systemic importance measure on an indirect channel model.

The theoretical finding that "too big to fail" is not always valid contributes to policy discussions on micro-level risk management and macro-level banking supervision. Since diversification is the usual tool in micro-level risk control, financial institutions, particularly the large ones, tend to take part in more banking activities in order to diversify away their individual risk. This may, however, increase their systemic importance. It is important to acknowledge the trade-off between managing individual risk and maintaining independence within the entire banking system. Portfolio construction toward reduction in individual risk may imply a transfer of risk to systemic linkage and thus increase the systemic importance. Therefore, prudential regulations which limit individual risk taking, such as Basel I and II type regulations, are not sufficient for maintaining stability of the entire financial system. Macroprudential supervision which considers the system as a whole is necessary for achieving financial stability. A macroprudential approach requires careful consideration of both individual risk taking and the systemic importance of each individual financial institution.

Next, we demonstrate how to empirically estimate the proposed systemic importance measures. We adopt the multivariate extreme value theory (EVT) framework for empirical estimation. In any investigation of crises, or rare events, the major difficulty is the scarcity of observations on crisis events. Since our intention is to address the interconnectedness of the banking crises, which is in effect a joint crisis, the difficulty with regard to the shortfall in observations is further enhanced. A modern statistical instrument—EVT—fills the gap. The essential idea of EVT is to model the intermediate-level observations, which are close to extreme, and extrapolate the observed properties into an extreme level. Therefore, the interconnectedness of crises can be approximated by the interconnectedness of tail events, which are not necessarily at a crisis level. Univariate EVT has been applied in value-at-risk assessment for individual risks (see, e.g., Embrechts, Klüppelberg, and Mikosch 1997). Recent developments on multivariate EVT provide the opportunity to investigate extreme co-movements, which serves our purpose. For instance, EVT was applied to measure risk contagions across different financial markets in Hartmann, Straetmans, and de Vries (2004) and Poon, Rockinger, and Tawn (2004). An application of multivariate EVT in analyzing bilateral relations within the banking system can be found in Hartmann, Straetmans, and de Vries (2005). Beyond bivariate relations, the *Global Financial Stability Report* published by the International Monetary Fund in April 2009 (IMF 2009) demonstrates—using EVT analysis—the interconnection of financial distress within a system consisting of three banks.

Our empirical estimation of the systemic importance measures uses multivariate EVT without restricting the number of banks under consideration.

We provide an empirical methodology with which to estimate the systemic importance measures—PAO, SII, and VI—under the multivariate EVT framework. We conduct an exercise for a constructed system consisting of twenty-eight U.S. banks and test the correlation between the systemic importance measures and different measures on size. We find that, in general, systemic importance measures are not correlated with all bank size measures. Hence, the size of a financial institution should not be considered as a proxy of its systemic importance without careful justification. This agrees with our theoretical model. Overall, we conclude that it is necessary to have proper systemic importance measures for identifying the systemically important financial institutions.

2. Systemic Importance Measures

We consider a banking system containing d banks with their status indicated by $(X_1, \ldots, X_d)$, where an extremely high value of X_i indicates a distress situation or a crisis in bank i. Potential candidates for such an indicator might include the loss of equity returns, loss returns on total asset, or credit default swap (CDS) rates.

To define a crisis, it is necessary to consider what constitutes a properly high threshold. Our approach takes value-at-risk as such a threshold. For the distress status X of a financial institution, a VaR at a tail probability level p is defined by

$$P(X > VaR(p)) = p.$$

Prudential regulations consider the p-level as 1 percent or 0.1 percent in order to evaluate risk-taking behavior of an individual institution. We say that a bank is in crisis if $X > VaR(p)$ with an extremely low p. Here, we do not specify the level p explicitly. Instead, we impose a restriction that the p-level in the definition of banking crises is constant across all banks. Notice that banks may differ in their risk profiles, which results in different endurability on risk, i.e., different $VaR(p)$. Thus, a unified level of loss for crisis definition may not fit the diversified situation of different financial institutions. Instead, an extreme event $X > VaR(p)$ corresponds to a return frequency as $1/p$. Fixing such a frequency for crisis definition takes account of the diversity of bank risk profiles. Furthermore, such a definition is aligned with the usual crisis description: for example, with yearly data, a p equal to $1/50$ corresponds to "a crisis once per fifty years."

The systemic importance measures consider the impact on other financial institutions when one of them falls into crisis. We start from the measure proposed by Segoviano and Goodhart (2009): the conditional probability of having at least one extra bank failure, given that a particular bank fails (PAO). In our model, this measure is the following probability:

$$PAO_i(p) := P(\{\exists j \neq i, \;\; \text{s.t.} \;\; X_j > VaR_j(p)\}|X_i > VaR_i(p)). \quad (1)$$

We argue that the PAO measure may not provide sufficient information for identifying the systemically important banks. Consider the following example. Suppose we have a banking system with banks categorized into two separate groups. Banks within each group are strongly linked, while banks from the two groups are independent of each other. One group contains only two banks, X_1 and X_2, while the other group contains more banks, $X_3, \ldots, X_d$, $d > 4$. In other words, X_1 and X_2 are highly related; $X_3, \ldots, X_d$ are highly related; and X_i and X_j are independent for any $1 \leq i \leq 2$ and $3 \leq j \leq d$.

Then the PAO measure for X_1, PAO_1, will be close to 1 since a crisis of X_1 will be accompanied by a crisis of X_2. However, the PAO measure for X_3, PAO_3, will also be close to 1 because of similar reasoning. When d is high, it is clear that X_3 is more systemically important than X_1 because it is associated with a larger fraction of the entire banking system. However, this will not be reflected by the comparison between PAO_1 and PAO_3. In this example, PAO_1 and PAO_3 should be at a high, comparable level.

Generally speaking, the PAO measure only provides the probability of a systemic impact when one particular bank fails—that is, an extra crisis occurring in other financial institutions. It does not specify the size of such an impact—that is, the number of extra crises in the entire system. Hence, if every institution in the system is connected to a certain fraction of the system, the PAO measures of all should stay at a high, comparable level. With indistinguishable PAO measures, it is not sufficient to identify the systemically important financial institutions.

A natural extension of the PAO measure is to consider the expected number of failures in the system, given that a particular bank fails. This is defined as our systemic impact index (SII). Using the notation above, it can be written as

$$SII_i(p) := E\left(\sum_{j=1}^{d} 1_{X_j > VaR_j(p)} | X_i > VaR_i(p)\right), \tag{2}$$

where 1_A is the indicator function that is equal to 1 when A holds, and is 0 otherwise.

Since the PAO and SII measures characterize the outlook of the financial system when a particular bank fails, a reverse question is what the probability of a particular bank failure is when the system exhibits some distress. To characterize the system distress, we use the same term as in the PAO measure: there exists at least one other bank failure. Hence, we define a vulnerability index (VI) by swapping the two items in the PAO definition as follows:

$$VI_i(p) := P(X_i > VaR_i(p) | \{\exists j \neq i, \ \text{s.t.} \ X_j > VaR_j(p)\}). \tag{3}$$

From the definitions, all three measures summarize specific information on the risk spillover in the banking system. It is necessary to consider all of them when assessing the systemic importance of financial institutions.

3. Extreme Value Theory and Systemic Importance Measures

3.1 The Setup of Extreme Value Theory

Consider our d-bank setup. Modeling the crisis of a particular financial institution i corresponds to modeling the tail distribution of X_i. Moreover, modeling the systemic risk—i.e., the extreme comovements among $(X_1, \ldots, X_d)$—corresponds to modeling the tail dependence structure of $(X_1, \ldots, X_d)$. Extreme value theory provides models for such a purpose.

To assess VaR with a low probability level p, univariate EVT can be applied in modeling the tail behavior of the loss. Since our focus is on systemic risk, we omit the details on univariate risk modeling (see, e.g., Embrechts, Küppelberg, and Mikosch 1997). Multivariate EVT models consider not only the tail behavior of individual X_i but also the extreme co-movements among them.

The fundamental setup of multivariate EVT is as follows. For any $x_1, x_2, \ldots, x_d > 0$, as $p \to 0$, we assume that

$$\frac{P(X_1 > VaR_1(x_1 p) \text{ or } \cdots \text{ or } X_d > VaR_d(x_d p))}{p} \to L(x_1, x_2, \ldots, x_d), \quad (4)$$

where VaR_i denotes the value-at-risk of X_i, and L is a finite positive function.[1] The L function characterizes the co-movement of extreme events that X_i exceeds a high threshold $VaR_i(x_i p)$. $(x_1, \ldots, x_d)$ controls the level of high threshold, which in turn controls the direction of extreme co-movement. For the property on the L function, see de Haan and Ferreira (2006).

The value of L at a specific point, $L(1, 1, \ldots, 1)$, is a measure of the systemic linkage of banking crises among the d banks. From the definition in (4), we have

$$L(1, 1, \ldots, 1) = \lim_{p \to 0} \frac{P(X_1 > VaR_1(p) \text{ or } \ldots \text{ or } X_d > VaR_d(p))}{p}. \quad (5)$$

In the context of our banking system, when p is at a low level, it approximates the quotient ratio between the probability that there exists at least one bank in crisis and the tail probability p used in the definition of individual crisis. For the bivariate case, this was considered by Hartmann, Straetmans, and de Vries (2004) in measuring systemic risk across different financial markets.

[1] Notice that considering the union of the events—i.e., using "or" in (4)—is simply a result of the definition of the distribution function. Define $F(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d)$ as the distribution function of $(X_1, \ldots, X_d)$. In order to consider the tail property, the assumption is made on the tail part $1 - F$, which is the probability of the union of extreme events as in relation (4).

Note that the L function is connected with the modern instrument of dependence modeling—the copula. Denote the joint distribution function of $(X_1, \ldots, X_d)$ as $F(x_1, \ldots, x_d)$ while the marginal distributions are denoted as $F_i(x_i)$ for $i = 1, \ldots, d$. Then there exists a unique distribution function $C(x_1, \ldots, x_d)$ on $[0, 1]^d$, such that

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)),$$

where all marginal distributions of C are standard uniform distributions. C is called the *copula*. By decomposing F into marginal distributions and copula, we separate the marginal information from the dependence structure summarized in the copula C. Condition (4) is equivalent to the following relation. For any $x_1, x_2, \ldots, x_d > 0$, as $p \to 0$,

$$\frac{1 - C(1 - px_1, \ldots, 1 - px_d)}{p} \to L(x_1, x_2, \ldots, x_d).$$

Hence the L function characterizes the limit behavior of the copula C at the corner point $(1, \ldots, 1) \in [0, 1]^d$. In other words, the L function captures the tail behavior of the copula C.

Linking the L function to the tail behavior of copula yields the two following views. Firstly, since it is connected to the copula, the L function does not contain any marginal information. Thus, in modeling the linkage of banking crises, the L function is irrelevant to the risk profile of the individual bank. Secondly, in characterizing the tail behavior of a copula, the L function does not contain dependence information at a moderate level, as in the copula C. Instead, L only contains tail dependence information. To summarize, the L function contains the minimal amount of required information in modeling extreme co-movements. Therefore, models on L are flexible to accommodate all potential marginal risk profiles and potential moderate-level dependence structures. Compared to Segoviano and Goodhart (2009), who consider the Consistent Information Multivariate Density Optimizing (CIMDO) approach on estimating the copula C, since models on the copula C incorporate the interconnection of banking systems in regular time, estimating a copula model may misspecify the tail dependence structure. Because we intend to model the interconnection of banking crises, considering the L function in the multivariate EVT approach is sufficient and less restrictive.

3.2 Systemic Importance Measures under Multivariate EVT

Under the multivariate EVT setup, the limit of the three systemic importance measures can be directly calculated from the L function. Notice that in the definitions of these measures, the probability level p for defining crisis is considered. However, we prove that, as $p \to 0$, the systemic importance measures can be well approximated by their limits.

The following proposition shows the limit of the PAO measure. The proof is in appendix 1.

PROPOSITION 1. *Suppose* $(X_1, X_2, \ldots X_d)$ *follows the multivariate EVT setup. With the definition of PAO in (1), we have*

$$PAO_i := \lim_{p \to 0} PAO_i(p) = L_{\neq i}(1, 1, \ldots, 1) + 1 - L(1, 1, \ldots, 1), \quad (6)$$

where L *is the* L *function characterizing the tail dependence of* $(X_1, \ldots, X_d)$*, and* $L_{\neq i}(1, 1, \ldots, 1)$ *is the* L *function characterizing the tail dependence of* $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots X_d)$.

Notice that L is defined on $\mathbb{R}^d$, while $L_{\neq i}$ is defined on $\mathbb{R}^{d-1}$. Moreover,

$$L_{\neq i}(1, 1, \ldots, 1) = L(1, 1, \ldots, 1, 0, 1, \ldots, 1),$$

where 0 appears at the i-th dimension.

Proposition 1 shows that when considering a low-level p, the measure $PAO_i(p)$ is close to its limit denoted by PAO_i. For calculating PAO_i, it is sufficient if the L function is known. Therefore, we could have a proxy of the PAO measure with low-level p by estimating the L function. In a theoretical model, the L function can be explicitly calculated. For an empirical analysis, the L function can be estimated from historical data. We provide a practical guide for estimating the L function in appendix 2. For more discussions, see de Haan and Ferreira (2006).

Analogous to that of PAO, the limit of $VI(p)$ exists under the multivariate EVT setup. We present the result in the following proposition but omit the proof.

PROPOSITION 2. *Suppose* $(X_1, X_2, \ldots X_d)$ *follows the multivariate EVT setup. With the definition of VI in (3), we have*

$$VI_i := \lim_{p \to 0} VI_i(p) = \frac{L_{\neq i}(1,1,\ldots,1) + 1 - L(1,1,\ldots,1)}{L_{\neq i}(1,1,\ldots,1)}, \qquad (7)$$

with the same notation defined in proposition 1.

From propositions 1 and 2, we get the following corollary.

COROLLARY 1. $PAO_i > PAO_j$ *holds if and only if* $VI_i > VI_j$*.*

Corollary 1 implies that when considering the ranking instead of the absolute level, the VI measure is in fact as informative as the PAO measure.

The following proposition shows how to calculate the limit of SII under the multivariate EVT setup. The proof appears in appendix 1.

PROPOSITION 3. *Suppose* $(X_1, X_2, \ldots X_d)$ *follows the multivariate EVT setup. With the definition of SII in (2), we have*

$$SII_i := \lim_{p \to 0} SII_i(p) = \sum_{j=1}^{d} (2 - L_{i,j}(1,1)), \qquad (8)$$

where $L_{i,j}$ *is the L function characterizing the tail dependence of* (X_i, X_j)*.*

Notice that

$$L_{i,j}(1,1) = L(0,\ldots,0,1,0,\ldots,0,1,0,\ldots,0),$$

where 1 appears only at the i-th and j-th dimensions. We remark that $2 - L_{i,j}(1,1)$ is in fact a measure of bilateral relation between the crises of banks X_i and X_j. Thus, the SII measure is an aggregation of measures on bilateral relations. This is parallel to the spillover index studied in Diebold and Yilmaz (2009) when measuring volatility spillover in a multivariate system: after measuring the volatility spillover between each pair, the spillover index is an aggregation of the measures on bilateral relations.

Again, proposition 3 shows that SII_i is a good approximation of $SII_i(p)$ when p is at a low level. And the estimation of SII_i is only based on the estimation of the L function. From the calculation of PAO and SII, it is clear that the two measures provide different information on systemic importance. A ranking based on PAO does not necessarily imply the same ranking on SII. Thus, it is still necessary to look at both of the measures in order to obtain a complete picture on the systemic importance of a bank.

To summarize, the multivariate EVT setup provides the opportunity to evaluate all three systemic importance measures when the L function is known. Since the L function characterizes the tail dependence structure in $(X_1, \ldots, X_d)$, all the systemic importance measures can be viewed as characterizations of the tail dependence among banking crises.

4. Are Banks "Too Big to Fail"? A Theoretical Model

We construct a simple model showing that large banks might have a lower level of systemic importance compared with small banks: banks are not necessarily too big to fail.

We start by reviewing a simple model in de Vries (2005), which explains the systemic risk within a two-bank system.

Consider two banks (X_1, X_2) holding exposures on two independent projects (Y_1, Y_2), as in the following affine portfolio model:

$$\begin{cases} X_1 = (1-\gamma)Y_1 + \gamma Y_2, \\ X_2 = \gamma Y_1 + (1-\gamma)Y_2, \end{cases} \tag{9}$$

where $0 < \gamma < 1$, (Y_1, Y_2) indicates the loss returns of the two projects. To measure the systemic risk, de Vries (2005) considers the following measure:

$$\lim_{s \to \infty} E(\kappa | \kappa \geq 1) := \lim_{s \to \infty} \frac{P(X_1 > s) + P(X_2 > s)}{P(X_1 > s \text{ or } X_2 > s)}. \tag{10}$$

Intuitively, $E(\kappa|\kappa \geq 1)$ is the expected number of bank crises in the two-bank system, given that at least one bank is in crisis. Here, the crisis of X_i is defined as $X_i > s$. It is proved that when Y_i, $i = 1, 2$ are normally distributed, $\lim_{s \to \infty} E(\kappa|\kappa \geq 1) = 1$. Thus, given that there exists at least one bank in crisis, the expected total number of crises is 1. Hence, there is no extra crisis except the existing one. This is called a weak fragility case. In other words, the systemic impact does not exist. To the contrary, suppose Y_i, $i = 1, 2$ follow a heavy-tailed distribution on the right tail. The result differs. The heavy-tailed distribution is defined as

$$\begin{cases} P(Y_i > s) = s^{-\alpha} K(s), \ i = 1, 2, \\ P(Y_i < -s) = o(P(Y_i > s)), \end{cases} \tag{11}$$

where $\alpha > 0$ is called the *tail index* and $K(s)$ is a *slowly varying function* satisfying

$$\lim_{t \to \infty} \frac{K(ts)}{K(s)} = 1,$$

for all $s > 0$. De Vries (2005) proved that for $\gamma \in [1/2, 1]$,

$$\lim_{s \to \infty} E(\kappa | \kappa \geq 1) = 1 + (1/\gamma - 1)^{\alpha} > 1.$$

This is called the strong fragility case because one existing crisis will be accompanied by potential extra crises. The empirical literature has extensively documented that the losses of asset returns follow heavy-tailed distributions. Therefore, the latter model based on heavy-tailed distributions reflects the empirical observations and explains the systemic risk existing in the financial system.

We remark that when assuming the heavy-tailedness of (Y_1, Y_2), and the affine portfolio model in (9), it is a direct consequence that (X_1, X_2) follows a two-dimensional EVT setup.[2] Moreover, if Y_1 and Y_2 are identically distributed, for a fixed tail probability p, the VaRs of X_1 and X_2 are equal; i.e., $VaR_1(p) = VaR_2(p)$. Replacing s with $VaR_i(p)$ in the definition of the systemic risk measure (10), and asking $p \to 0$, we get

$$\begin{aligned} \lim_{p \to 0} E(\kappa | \kappa \geq 1) &:= \lim_{p \to 0} \frac{P(X_1 > VaR_1(p)) + P(X_2 > VaR_2(p))}{P(X_1 > VaR_1(p) \text{ or } X_2 > VaR_2(p))} \\ &= \frac{2}{L(1,1)}. \end{aligned}$$

[2]For a formal proof, see Zhou (2008, ch. 5).

Therefore, the setup in de Vries (2005) imposes a multivariate EVT setup, and the measure on the systemic risk is essentially based on $L(1,1)$.

We point out that within this two-bank, two-project model, it is not possible to differentiate the systemic importance of the two banks. From the model and from proposition 1, we get

$$SII_i = 3 - L(1,1),\ \ i = 1,2.$$

Hence the two banks have equal systemic importance measured by SII. Similar results hold for the other two measures, PAO and VI. Intuitively, within a two-bank setup, the linkage of crises is a mutual bilateral relation. Hence, one could not distinguish the systemic importance of the two banks. In order to construct a model in which it is possible to compare the systemic importance at different levels, it would be necessary to generalize the de Vries (2005) model to a system consisting of at least three banks.

Next, we consider the size issue. In the de Vries (2005) two-bank model, suppose that both of the two banks have total capital 1, and both of the two projects receive capital 1. According to the affine portfolio model (9), the capital market is clear. In this case, the two banks have the same size in terms of total assets. In order to differentiate the sizes of the banks, a more complex affine portfolio model is necessary.

Addressing the two above-mentioned points, we consider a model with three banks (X_1, X_2, X_3) and three independent projects (Y_1, Y_2, Y_3). Suppose X_1 holds capital 2 for investment, while X_2 and X_3 hold capital 1 each. Moreover, suppose the project Y_1 demands an investment 2, while Y_2 and Y_3 each have a capital demand 1. Then the market is clear, with the following affine portfolio model:

$$\begin{cases} X_1 = (2-2\gamma)Y_1 + \gamma Y_2 + \gamma Y_3 \\ X_2 = \gamma Y_1 + (1-\gamma-\mu)Y_2 + \mu Y_3, \\ X_3 = \gamma Y_1 + \mu Y_2 + (1-\gamma-\mu)Y_3, \end{cases} \tag{12}$$

where $0 < \gamma, \mu < 1$, and $\gamma + \mu < 1$. Clearly, this is not the only possible allocation for market clearance. Nevertheless, it is sufficient to demonstrate our argument regarding the "too big to fail" problem. Notice that X_1 is a larger bank compared with X_2 and X_3. Here, the size refers to the total investment in the risky projects. We intend to compare the systemic importance of X_1 with that of X_2 and X_3.

The two parameters γ and μ are interpreted as the control of similarity in portfolio holdings across the three banks. The parameter γ controls the similarity between the large bank and the small banks. When γ is close to 1, the strategy of the large bank is different from that of the two small banks, while the two small banks hold similar portfolios. When $\gamma = 1/2$, the large bank has exposures on all three projects proportional to their capital demands. Hence, the large bank is involved in all projects. When γ is close to 0, the large bank is again different from the two small banks. In the latest case, the similarity of the two small banks is further controlled by the parameter μ: a μ lying in the middle of $(0, 1-\gamma)$ shows that the two small banks are similar in portfolio holding, while a μ lying close to the two corners of $(0, 1-\gamma)$ corresponds to different strategies between the two small banks.

Suppose all Y_i follow a heavy-tailed distribution defined in (11) for $i = 1, 2, 3$. Then, similar to the two-bank case, (X_1, X_2, X_3) follows a three-dimensional EVT setup. Instead of discussing all possible values on the parameters (γ, μ), we focus on three cases: γ is close to 1, $\gamma = 1/2$, and γ is close to 0. The results from comparing the SII measures are in the following theorem. The proof is again in appendix 1.

THEOREM 1. *Consider a three-bank, three-project model with the affine portfolio given in (12). Suppose the losses of the three projects exhibit the same heavy-tailed feature as in (11), with $\alpha > 1$. We have the following relations.*

CASE 1: $\frac{2}{3} \leq \gamma < 1$

$$SSI_1 < 1 = SSI_2 = SSI_3.$$

CASE 2: $\gamma = \frac{1}{2}$

$$SSI_1 \geq SSI_2 = SSI_3.$$

The equality holds if and only if $\mu = 1/4$.

CASE 3: $0 < \gamma < \frac{1}{3}$

There exists a $\mu^* < \frac{1-\gamma}{2}$ *such that for any* μ *satisfying* $\mu^* < \mu < 1 - \gamma - \mu^*$,

$$SSI_1 < SSI_2 = SSI_3.$$

On the other hand, for any μ *satisfying* $0 < \mu < \mu^*$ *or* $1 - \gamma - \mu^* < \mu < 1 - \gamma$, *we have*

$$SSI_1 > SSI_2 = SSI_3.$$

When $\mu = \mu^*$ *or* $\mu = 1 - \gamma - \mu^*$,

$$SII_1 = SII_2 = SII_3.$$

The following lemma shows that the comparison among the PAO measures follows the comparison among the SII measures in the three-bank model.

LEMMA 1. *With the assumptions in theorem 1, the order of PAO follows the order of SII; i.e., for any* $i \neq j$, $PAO_i > PAO_j$ *holds if and only if* $SII_i > SII_j$.

Combining lemma 1 and corollary 1, we see that the order of VI also follows the order of SII in the simple three-bank model. Notice that the three-bank model is a very specific and simple case. The result in lemma 1 does not hold in a general context when the number of banks is more than three. Therefore, for empirical study within a multibank system, it is still necessary to estimate all three measures, which may provide different views.

We interpret the results in theorem 1 as follows.

If γ is close to 1, the large bank X_1 focuses on the two smaller projects Y_2 and Y_3, while small banks X_2 and X_3 focus on the large project Y_1. In this case, the balance sheet of the large bank is quite different from that of the small banks, while the two small banks are holding similar portfolios. Therefore, the large bank has less systemic linkage to the other two small banks. We observe that the large bank is less systemically important compared with the others; i.e., the large bank is not "too big to fail."

If $\gamma = 1/2$, the large bank X_1 invests $(1, 1/2, 1/2)$ in three projects. Hence it is involved in all three projects, which creates the linkage to the other two small banks. In this case, it is "too big to fail." The inequality becomes an equation if and only if $\mu = 1/4$. For $\mu = 1/4$, the three banks all invest in three projects proportional to their capital demands. They have exactly the same strategy in managing their portfolios. A crisis in any of the three banks will be accompanied by crises in the other two. Therefore, they are equally systemically important. Excluding $\mu = 1/4$, the large bank will be the most systemically important bank.

If γ is close to 0, the large bank X_1 focuses on the large project Y_1, while it still has exposures on Y_2 and Y_3. The small banks X_2 and X_3 focus on the two small projects Y_2 and Y_3. Now it matters how similar their portfolios are. If μ is in the middle ($\mu^* < \mu < 1-\gamma-\mu^*$), then the balance sheet composition of two small banks is relatively similar. Hence, they are more systemically important compared with the large bank. If μ is close to the corner ($0 < \mu < \mu^*$ or $1 - \gamma - \mu^* < \mu < 1 - \gamma$), then the two small banks differ in their balance sheets. Since the large bank still has exposures on Y_2 and Y_3 equally, it is the most systemically important bank. It is worth mentioning that the systemic importance of bank X_1 is determined not only by its own risk positions but also by the risk-taking behavior of the others. Even though the portfolio of bank X_1 is fixed by fixing γ, the change of the portfolios hold by the other two banks can still result in a change of the systemic importance of X_1.

To summarize, we observe that "too big" is not necessarily the reason for being "too systemically important." Instead, a bank having a balance sheet that is exposed to more risky projects would cause it to become more systemically important. Here, we regard Y_i, $i = 1, 2, 3$ as different risky projects. One may also regard them as different risky banking activities. Therefore, a bank that is more diversified in banking activities may turn out to be "too big to fail."

Notice that having a diversified balance sheet is the usual way of managing individual risk. In order to obtain the diversification, banks, particularly large banks, will be spurred on to take part in more banking activities. The above discussion shows that this will simultaneously result in a "too big to fail" problem. Conversely, a large bank specialized in a limited number of banking activities might be risky as an individual but at the same time less systemically important. There is a trade-off between managing individual risk and keeping a sense of independence within the entire banking system. For maintaining the stability of the financial system, it is necessary for the regulators to recognize such a trade-off and impose proper regulations in order to give banks incentives to balance their individual risk position and systemic importance.

5. Empirical Results

5.1 Empirical Setup and Data

We apply the three proposed measures of systemic importance to an artificially constructed financial system consisting of twenty-eight U.S. banks. After estimating the three measures, we calculate the correlation coefficients between these measures and the measures of the size of the banks. From the test on correlation coefficients, we can empirically test whether larger banks exhibit larger systemic importance, thereby testing the "too big to fail" argument. We also consider a moving window approach, which demonstrates the variation of the systemic importance measures across time.

The data set for constructing the systemic importance measures consists of daily equity returns of twenty-eight U.S. banks listed on the New York Stock Exchange (NYSE) from 1987 to 2009 (twenty-three years).[3] The chosen banks are listed in table 1 with the descriptive statistics on their stock returns.

Regarding the size of the banks, the data set consists of various measures. We consider total assets, total equity, and total debt for the twenty-eight banks.[4] The data that appear are reported in a yearly frequency from 1987 to 2009. For each bank, we present the end-of-2009 values as well as the average values across the twenty-three years in table 2.

From the descriptive statistics of the equity returns, we observe that all daily returns exhibit high kurtosis compared with the

[3]The data are obtained from Datastream. Three selection criteria are applied: the financial institutions should be classified in the sector "Banks"; they should be traded primarily on the NYSE (DS code starting with "U:"); and the time series should be active from the beginning of 1987 till the end of 2009. The selection procedure results in twenty-eight banks.

[4]The data on total assets, total equity, and total debt are obtained from Datastream, with item code WC02999, WC03501, and WC03255, respectively. Notice that total debt consists of short-term debt, current portion of long-term debt, and long-term debt.

Table 1. Descriptive Statistics on Daily Stock Returns of Twenty-Eight U.S. Banks

Banks	Mean	Std. Dev.	Min.	Max.	Skew.	Kurtosis
BancorpSouth	0.040	2.28	−15.12	19.71	0.23	8.63
Bank of America	0.031	2.70	−34.12	30.21	−0.35	32.80
Bank of Hawaii	0.045	1.78	−25.51	12.95	−0.63	19.77
BB&T	0.041	2.09	−26.61	21.20	0.12	19.53
Central Pacific	0.009	3.44	−66.87	69.31	0.28	68.85
Citigroup	0.020	2.98	−49.47	45.63	−0.50	42.92
City National	0.029	2.29	−18.92	20.21	0.13	11.15
Community Bank[a]	0.041	2.30	−14.22	16.55	0.37	9.11
Comerica	0.038	2.13	−22.69	18.81	−0.16	18.68
Cullen/Frost Bankers	0.054	2.17	−21.46	19.78	0.17	13.52
First Horizon National	0.033	2.36	−44.11	25.54	−1.18	45.41
FNB	0.024	2.52	−25.53	19.57	−0.19	14.23
JP Morgan	0.033	2.55	−32.46	22.39	−0.12	17.58
KeyCorp	0.011	2.56	−40.55	43.34	−0.51	50.04
M&T Bank	0.054	1.75	−17.59	22.83	0.33	24.60
Marshall & Ilsley	0.017	2.53	−30.15	32.93	−0.23	40.12
Old National Bancorp	0.019	1.74	−19.47	15.84	0.31	16.55
PNC	0.031	2.28	−53.44	31.55	−1.30	68.26
Regions Financial	0.001	2.76	−52.88	39.48	−0.62	56.56
Sterling	0.017	2.27	−21.26	19.39	0.27	12.36
Suntrust Banks	0.023	2.40	−31.71	26.67	−0.39	29.95
Synovus Financial	0.017	2.78	−30.07	24.86	−0.02	17.68
TCF Financial	0.047	2.35	−17.65	23.53	0.49	13.73
US Bancorp	0.054	2.17	−20.05	25.76	0.22	18.69
Valley National Bancorp	0.034	2.22	−17.48	21.71	0.29	11.39
Webster Financial	0.024	2.56	−23.51	31.03	0.07	20.03
Wells Fargo	0.060	2.35	−27.21	28.34	0.66	27.47
Wilmington Trust	0.026	2.11	−23.12	27.62	0.05	20.85

Notes: The sample period runs from January 2, 1987, to December 31, 2009 (sample size 6,000). All values except the skewness and kurtosis are in percentages. The list consists of all banks that are preliminarily traded on the NYSE during the sample period.
[a]Community Bank N.A. (CBNA) is a bank holding company in Upstate New York. Its predecessor bank was founded in 1866, and it is the wholly owned banking subsidiary of Community Bank System, Inc. (CBSI).

kurtosis from the normal distribution, which is 3. This indicates that the daily stock returns may follow heavy-tailed distributions. Moreover, most of the skewness measures are negative or close to 0. This

Table 2. Descriptive Statistics on Yearly Size Measures of Twenty-Eight U.S. Banks

	Total Asset		Total Equity		Total Debt	
Banks	**2009**	**Average**	**2009**	**Average**	**2009**	**Average**
BancorpSouth	13.2	6.5	1.3	0.6	1.0	0.5
Bank of America	2223.3	648.4	193.6	51.4	764.7	199.3
Bank of Hawaii	12.4	11.3	0.9	0.9	1.7	2.2
BB&T	165.2	55.0	16.2	5.1	29.5	11.3
Central Pacific	4.9	2.4	0.2	0.2	0.9	0.3
Citigroup	1856.6	1145.3	152.4	73.1	561.3	346.8
City National	20.9	8.6	1.8	0.8	1.4	0.8
Community Bank	5.4	2.3	0.6	0.2	0.9	0.3
Comerica	59.2	38.4	4.8	3.1	11.5	6.9
Cullen/Frost Bankers	16.3	7.4	1.9	0.6	0.9	0.6
First Horizon National	26.1	18.4	2.1	1.3	6.8	4.6
FNB	8.7	3.9	1.0	0.3	1.2	0.5
JP Morgan	2032.0	655.8	157.2	45.9	625.3	173.5
KeyCorp	92.7	69.7	7.9	5.2	13.6	16.6
M&T Bank	68.9	28.5	7.0	2.7	12.7	5.5
Marshall & Ilsley	56.2	26.0	5.3	2.4	7.5	5.3
Old National Bancorp	7.9	6.0	0.8	0.5	1.0	1.0
PNC	269.9	87.3	21.6	6.9	39.3	18.4
Regions Financial	142.3	48.8	14.1	5.3	22.1	8.0
Sterling	2.1	1.2	0.1	0.1	0.3	0.2
Suntrust Banks	174.2	92.8	17.5	8.2	22.9	17.3
Synovus Financial	32.8	14.6	1.9	1.4	2.2	1.4
TCF Financial	17.9	9.7	1.2	0.7	4.8	2.4
US Bancorp	281.2	109.6	24.5	9.6	63.9	26.9
Valley National Bancorp	14.3	6.9	1.3	0.5	3.3	1.1
Webster Financial	17.6	9.0	1.5	0.7	2.0	2.1
Wells Fargo	1240.4	295.0	103.3	23.3	242.8	69.8
Wilmington Trust	10.9	6.9	1.0	0.6	1.0	1.2

Notes: The sample period is from 1987 to December 2009 (twenty-three years). The list consists of all banks that are preliminarily traded on the NYSE during the sample period. The numbers are in millions of U.S. dollars.

indicates that the heavy-tailedness may come from the downside of the distribution, the losses. Hence, our heavy-tailed assumption on the tail of losses is valid for the data set employed.

From the descriptive statistics of the size measures, we observe a large variation on the size measures of the selected banks. The top three largest banks in the list—Bank of America, JP Morgan, and

Citigroup—are approximately 1,000 times larger than the smallest bank in the list, Sterling Banc, in all aspects. Although the criterion that the selected banks have to be active in the stock market for twenty-three years may result in a sample selection bias, since such banks are more likely to be large banks, the variation of the size measures shows that the constructed banking system contains both large and small banks.

Using the stock returns is a natural choice for our approach in analyzing the systemic importance. The restriction imposed by the methodology of estimating the L function is that the sample size has to be sufficient; see appendix 2. Moreover, since we intend to perform a moving window analysis, the restriction on the length of the time series is further enhanced. Therefore, daily or higher frequency is necessary for a full non-parametric approach. This limits us to using financial market data. Equity returns are the most convenient choice. Other high-frequency indicators such as CDS spreads are also possible. Nevertheless, the CDS data do not go back for a sufficiently long period, which keeps us from performing a moving window analysis. It is also possible to apply the proposed methodology with low-frequency data, such as return on asset from bank balance sheet. In that case, a full non-parametric estimate on the L function is not applicable. Instead, further modeling on the L function should be considered. In this study we intend to illustrate the methodology without modeling the L function. Hence, we stick to the equity return data.

5.2 Estimation of the Systemic Importance Measures

By estimating the L function (for details, see appendix 2), we obtain the estimates of the three systemic importance measures (SII, PAO, and VI) across the full sample period, as shown in table 3. We start with the PAO measure proposed by Segoviano and Goodhart (2009). A general observation is that all of the estimates are quite high (above 60 percent). This is in line with our prediction that the PAO measures of all banks in a system are at a relatively high level and do not differ much from each other. Since the PAO measure is directly connected to the VI measure as shown in corollary 1, a similar feature is observed for the VI measures. In fact, the order of the VI measures follows that of the PAO measures as proved in corollary 1. To name

Table 3. Estimated Systemic Importance Measures: Full Sample Analysis

Banks	SII	PAO	VI
BancorpSouth	6.72	59.44%	6.73%
Bank of America	10.84	94.44%	10.28%
Bank of Hawaii	10.44	84.44%	9.30%
BB&T	10.88	86.11%	9.46%
Central Pacific	8.68	65.00%	7.31%
Citigroup	10.59	90.56%	9.90%
City National	9.30	76.11%	8.46%
Community Bank	6.53	59.44%	6.73%
Comerica	12.02	92.78%	10.12%
Cullen/Frost Bankers	8.05	73.89%	8.23%
First Horizon National	10.84	83.33%	9.19%
FNB	7.41	57.78%	6.55%
JP Morgan	9.76	86.67%	9.52%
KeyCorp	12.44	93.33%	10.18%
M&T Bank	11.10	86.67%	9.52%
Marshall & Ilsley	11.92	93.89%	10.23%
Old National Bancorp	9.36	77.78%	8.63%
PNC	10.73	86.11%	9.46%
Regions Financial	11.91	90.56%	9.90%
Sterling	8.69	65.56%	7.37%
Suntrust Banks	12.11	92.78%	10.12%
Synovus Financial	10.11	83.89%	9.24%
TCF Financial	10.57	81.11%	8.96%
US Bancorp	10.32	78.89%	8.74%
Valley National Bancorp	7.82	65.00%	7.31%
Webster Financial	9.87	82.22%	9.07%
Wells Fargo	11.25	90.00%	9.85%
Wilmington Trust	10.91	83.33%	9.19%

Notes: SII is the systemic importance index, defined as the number of expected bank failures given a particular bank fails; see (2). PAO is the probability of having at least one extra bank failure when a particular bank fails, defined in (1). VI is the vulnerability index, defined as the probability of failure given there exists at least another bank failure in the system; see (3).

a few banks with the highest PAO and VI measures, Bank of America reports the highest PAO at 94.44 percent, followed by Marshall & Ilsley with 93.89 percent and KeyCorp with 93.33 percent. The corresponding VI measures are 10.28 percent, 10.23 percent, and 10.18 percent, respectively. At the bottom of the list ranked by the PAO measure, we have FNB, BancorpSouth, and Community Bank.[5]

The SII measure introduced in this paper gives a somewhat different outlook compared with the PAO measure. The three lowest SII banks are the same as those with the lowest three PAO measures, although with a different order. The lowest SII measure comes from Community Bank, which is 6.53. This means that if Community Bank is experiencing a crisis, it will be accompanied by an average of 5.53 extra failures in this system. Compared with the size of the system, twenty-eight banks, this is not a high systemic impact. The highest estimated SII measure is 12.44 from KeyCorp, which is almost twice as high as the lowest value. A crisis of KeyCorp will be accompanied by an average of 11.44 extra crises in this system, twice the systemic impact of Community Bank. Hence, we observe a variation of the SII measure across different banks. To name a few with the highest SII measures, the top three are KeyCorp (12.44), Suntrust Banks (12.11), and Comerica (12.02). They are different from the banks with the top-three highest PAO. In general, ranking the PAO measures is different from ranking the SII measures. For example, the bank with the highest PAO, Bank of America, is only ranked at tenth place among all banks when considering the SII measure.

To summarize, the comparison between the three measures shows that although they have different economic backgrounds, the PAO measure and the VI measure are equally informative in terms of ranking the systemic importance of financial institutions. The SII measure, in contrast, provides information on the size of the systemic impact corresponding to the failure of a particular bank. It therefore provides a different view than the other two. Across different banks, the SII measures vary while the PAO measures remain at a high, comparable level. This agrees with our theoretical prediction. Therefore, the SII measure is more informative in distinguishing the systemic importance of financial institutions.

---

[5] Community Bank N.A. (CBNA) is a bank holding company in Upstate New York. Its predecessor bank was founded in 1866, and it is the wholly owned banking subsidiary of Community Bank System, Inc. (CBSI).

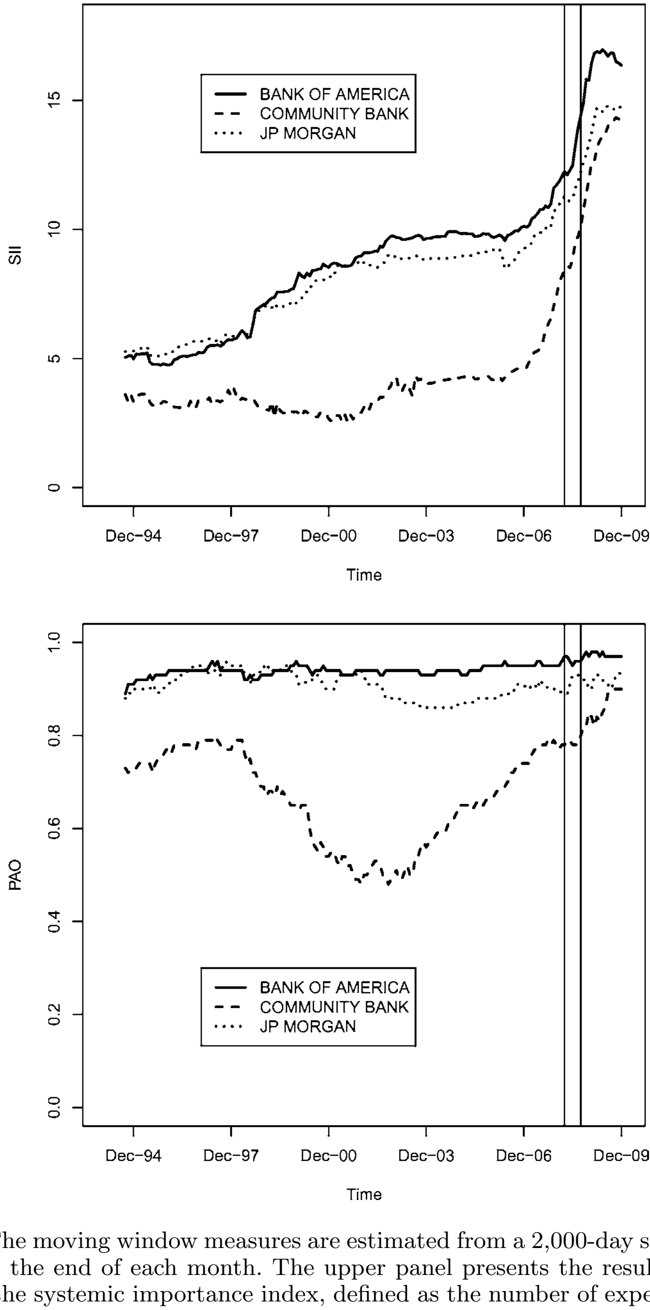
Besides estimating the three systemic importance measures from the full sample period, we consider sub-samples for the estimation and perform a moving window approach. By moving the sub-sample window, we could obtain time-varying estimation on the systemic importance measures. We consider the estimation window as 2,000 days (approximately eight years), and then move the estimation window forward month by month. The first possible window ends at September 1994. In other words, the first estimation considers data ending at September 30, 1994, and going back 2,000 days. From then on, we take the end of each month as the ending day of each estimation window and use the data going back 2,000 days. By moving the estimation window month by month, we observe the estimates at the end of each month from September 1994 to December 2009. For simplicity's sake, we only plot the results for selected banks,[6] as shown in figure 1. The upper panel shows the moving window SII measures, and the bottom panel shows the moving window PAO measures. The two vertical lines in the two figures correspond to the failures of two large investment banks: Bear Stearns (March 2008) and Lehman Brothers (September 2008).

From the moving window SII estimates, we observe that the SII measures gradually increased from 1998 to 2003, then remained relatively stable until the end of 2006. From 2007, there was a sharp rise. The sharp rise of SII started before the failure of Bear Stearns and continued with the failure of Lehman Brothers, until early 2009. From mid-2009, the SII measures became stable and with a slight downward slope. In contrast, the PAO measures are stable across time, particularly for the large banks. Only for the least systemically important bank can some variation be observed. This is due to the fact that the PAO measures of large banks were already at a high level in the early period of our sample. It would thus be difficult to obtain a further rise to a higher level.

The observations from the moving window approach again confirm our theoretical prediction that the PAO measures always stay at a high, comparable level, while the SII measure varies across time

[6]We select the least systemically important bank, Community Bank, and the two largest banks, Bank of America and JP Morgan, in the plots. According to the full sample analysis, Bank of America is the most systemically important bank in terms of PAO.

Figure 1. Moving Window Results on Systemic Importance Measures



Notes: The moving window measures are estimated from a 2,000-day sub-sample ending at the end of each month. The upper panel presents the results for SII, which is the systemic importance index, defined as the number of expected bank failures given a particular bank fails; see (2). The bottom panel presents the results for PAO, which is the probability of having at least one extra bank failure when a particular bank fails, defined in (1). The two vertical lines refer to the failures of Bear Stearns (March 2008) and Lehman Brothers (September 2008).

and across institutions. The sharp rise of the SII measures addresses the crisis starting from 2008 and hence is more informative in analyzing the systemic risk in a financial system. Although we observe an early rise before the crisis, we do not emphasize that the SII measure is a predictor of the crisis. The sharp rise of SII measures might either be a predictor of the crisis or an ex post consequence caused by the crisis. The intuition for the latter possibility is as follows. Banks tend to take similar strategies, such as fire sales, during a crisis, which may result in more similar portfolio holdings across all banks. According to the theoretical model in section 4, that similarity leads to an increase of the SII measures on all banks. Hence, the timing of the sharp rising of the SII measures is still an open issue for further study.

There are a few other observations from the moving window analysis. Notice that the financial system we have constructed contains twenty-eight banks. An SII measure of 15 means that if a certain bank fails, there will be an average of fourteen extra bank failures simultaneously. This is half of the entire system, which must be considered as a severe risk. Hence, the observed SII measures from the end of 2008 to mid-2009 indicate that the banking system suffers severe systemic risk during the crisis. Moreover, it is remarkable that Community Bank, the least systemically important bank from the full sample analysis, also showed the least systemic importance during the crisis. Nevertheless, the absolute level of the SII measure reached a comparable level with the other large banks. This suggests that size may not be a good proxy of systemic importance, particularly during periods of crisis.

5.3 Test "Too Big to Fail"

We use the estimated systemic importance measures to check whether they are correlated with the size measures. The correlation test is across different banks; thus, we need to have a unified value for each individual bank on each size measure. Since the sample period ends in year 2009, we first consider the end-of-2009 values of each size measure. The second approach is to take an average of the size measures over the full sample period (from 1987 to 2009). Then, we calculate the Pearson correlation coefficients between each pair of size measure and systemic importance measure across

Table 4. Correlation Coefficients: Full Sample Analysis

		SII	PAO	VI
End of 2009	Total Asset	0.1790	0.3968**	0.3943**
		(0.3622)	(0.0366)	(0.0379)
	Total Equity	0.1892	0.4053**	0.4027**
		(0.3348)	(0.0324)	(0.0336)
	Total Debt	0.1399	0.3640*	0.3615*
		(0.4777)	(0.0569)	(0.0588)
Average	Total Asset	0.1733	0.3746**	0.3723*
		(0.3779)	(0.0495)	(0.0510)
	Total Equity	0.1980	0.4031**	0.4006**
		(0.3126)	(0.0334)	(0.0347)
	Total Debt	0.1542	0.3546*	0.3523*
		(0.4334)	(0.0641)	(0.0660)

Notes: SII is the systemic importance index, defined as the number of expected bank failures given a particular bank fails; see (2). PAO is the probability of having at least one extra bank failure when a particular bank fails, defined in (1). VI is the vulnerability index, defined as the probability of failure given there exists at least another bank failure in the system; see (3). The numbers in parentheses are the p-values for testing whether the correlation coefficient is significantly different from zero. The upper panel reports the results based on using end-of-2009 values of the size measure, while the lower panel reports the results based on using the average of the size measure across the full sample period (from 1987 to 2009). ***, **, and * denote significance at the 1 percent, 5 percent, and 10 percent level, respectively.

twenty-eight banks. Moreover, we test whether the correlation coefficient is significantly different from zero. The results are shown in table 4.

Generally, the PAO and VI measures are positively correlated with the size measures, with significance level 5 percent or 10 percent. Put differently, the SII measure is not correlated with any size measure. We have argued throughout the paper that the SII measure is a more informative measure of systemic importance, because it considers the systemic impact of the failure of a particular bank. We conclude that the systemic impact of a bank failure is not correlated with the size measures. Therefore, "too big to fail" is not valid on the impact level at least for the constructed banking system. Nevertheless, with the positive correlation between the other two systemic importance measures and the size measures, the big banks are more likely to cause extra crises in the system. The bottom line is that it is not proper to use the size measures as a proxy of systemic importance. Instead, it is necessary to consider all of the systemic importance measures when identifying systemically important financial institutions.

We have carried out extensive robustness checks for the observed results. First of all, with the full sample estimation on the systemic risk measure, we consider the size measure in other years (e.g., end of 2008, 2007, etc.). The results are comparable with those using the end-of-2009 value. We omit the details.

Secondly, instead of the Pearson correlation, we consider the Spearman correlation, which only emphasizes the correlation between ranking orders. The Spearman correlation coefficients between the SII measure and the size measures are significantly different from zero. Since the Pearson correlation considers the absolute level while the Spearman correlation considers only the ranking orders, we find support of the "too big to fail" argument within the constructed system, only in terms of ranking the order.

Thirdly, with the moving window results on the systemic importance measures, we can get the end-of-year estimates on the systemic importance measures from 1994 to 2009 (sixteen years). We pool all of the bank-year estimates together, which results in $28 \cdot 16 = 448$ estimates for each systemic importance measure, and also 448 observations for each size measure. We then calculate the Pearson correlation coefficient for each pair and repeat the test on the significancy. The results appear in the first panel of table 5. None of the three systemic importance measures are correlated with any of the size measures.

Moreover, since we have obtained sixteen-year data on systemic importance and size for twenty-eight banks, we also perform the Pearson correlation analysis at the level of each year. We observe that the significant positive correlation between size and the PAO measure (and the VI measure) is robust for the sub-period 1994–99. From 2000 to 2009, the significance disappears. Interestingly, for the first period, the SII measure is also positively correlated with the three size measures. The (in)significant results are robust within each sub-period. To further explore these phenomena, we divide the period 1994–2009 into two sub-periods: 1994–99 and 2000–09,

Table 5. Correlation Coefficients: Moving Window Analysis

		SII	PAO	VI
Full Sample	Total Asset	0.0263	0.2073	0.2069
		(0.8943)	(0.2900)	0.2908
	Total Equity	0.0474	0.2188	0.2183
		(0.8107)	(0.2633)	(0.2645)
	Total Debt	−0.0080	0.1665	0.1667
		(0.9678)	(0.3971)	(0.3966)
Period 1: 1994–99	Total Asset	0.4541**	0.4950***	0.4911***
		(0.0152)	(0.0074)	(0.0080)
	Total Equity	0.4727**	0.5111***	0.5072***
		(0.0111)	(0.0054)	(0.0059)
	Total Debt	0.4648**	0.5014***	0.4973***
		(0.0127)	(0.0066)	(0.0071)
Period 2: 2000–09	Total Asset	0.2192	0.2295	0.2297
		(0.2624)	(0.2402)	(0.2396)
	Total Equity	0.2261	0.2374	0.2375
		(0.2473)	(0.2238)	(0.2235)
	Total Debt	0.1915	0.1907	0.1914
		(0.3290)	(0.3310)	(0.3293)

Notes: SII is the systemic importance index, defined as the number of expected bank failures given a particular bank fails; see (2). PAO is the probability of having at least one extra bank failure when a particular bank fails, defined in (1). VI is the vulnerability index, defined as the probability of failure given there exists at least another bank failure in the system; see (3). The numbers in parentheses are the p-values for testing whether the correlation coefficient is significantly different from zero. The upper panel reports the results based on pooling all bank-year observations (448 observations). The middle and lower panels report the results based on pooling bank-year observations in two periods: 1994–2000 and 2001–09. ***, **, and * denote significance at the 1 percent, 5 percent, and 10 percent level, respectively.

according to the individual year results. Then we pool bank-year data in each sub-period and repeat the Pearson correlation analysis. The results are reported in the second and third panels of table 5. It confirms the results from the individual-year analysis: in the first period, all three systemic importance measures are highly correlated with the size measures, while in the second period, the correlations disappear. It suggests that using size as a proxy for systemic importance was proper in the 1990s, but that the situation has changed from the beginning of the new century. Therefore, it is particularly important to consider the measures on systemic importance within the current financial world.

Last but not least, using daily equity returns may present a potential problem due to the heteroskedasticity in high-frequency financial returns. Diebold, Schuermann, and Stroughair (2000) argue that the fundamental assumption of EVT—that the observations are independent and identically distributed (i.i.d.)—is usually violated when dealing with high-frequency financial returns. They provide two potential solutions: firstly, one may investigate low-frequency block maxima, which reduces the dependency across time; secondly, one may apply a conditional extreme value model. In line with the second type of solution, Chavez-Demoulin, Davison, and McNeil (2005) consider a point process approach to decluster the financial return data. Notice that the non-i.i.d. problem is mainly for evaluating the univariate tail events. Because our study is under a multivariate framework, which models cross-sectional dependence of tail events, the non-i.i.d. problem on the time dimension is of less concern. Nevertheless, we conduct a robustness check in line with the first type of solution proposed by Diebold, Schuermann, and Stroughair (2000). Instead of daily returns, we consider monthly returns. Similar to the block maxima approach, considering monthly returns reduces the interdependent problem on the time dimension. Due to the low frequency, we have to use the full sample (276 months) for the analysis. The results appear in table 6. The estimates of the three systemic risk measures are similar to those using daily returns—at least in terms of ranking the systemic importance. The absolute level is higher than that using daily returns, because it is more likely to observe simultaneous tail events in the same month. We repeat the Pearson correlation analysis as shown in table 7. Again, we confirm that the systemic importance measures are not correlated with the size measures.

6. Conclusion

This paper considers three measures of systemic importance of financial institutions in a financial system. Since we regard the system as

Table 6. Systemic Importance Measures: Monthly Data

Banks	SII	PAO	VI
BancorpSouth	10.15	75.00%	11.28%
Bank of America	11.70	85.00%	12.59%
Bank of Hawaii	11.60	100.00%	14.49%
BB&T	13.50	90.00%	13.24%
Central Pacific	8.65	90.00%	13.24%
Citigroup	12.00	95.00%	13.87%
City National	10.20	80.00%	11.94%
Community Bank	9.10	75.00%	11.28%
Comerica	13.40	100.00%	14.49%
Cullen/Frost Bankers	10.20	95.00%	13.87%
First Horizon National	9.85	75.00%	11.28%
FNB	9.45	75.00%	11.28%
JP Morgan	10.80	90.00%	13.24%
KeyCorp	12.45	90.00%	13.24%
M&T Bank	13.10	95.00%	13.87%
Marshall & Ilsley	13.15	100.00%	14.49%
Old National Bancorp	8.25	70.00%	10.61%
PNC	12.20	95.00%	13.87%
Regions Financial	11.85	95.00%	13.87%
Sterling	9.80	65.00%	9.92%
Suntrust Banks	12.85	100.00%	14.49%
Synovus Financial	11.25	95.00%	13.87%
TCF Financial	10.60	90.00%	13.24%
US Bancorp	12.80	100.00%	14.49%
Valley National Bancorp	10.80	85.00%	12.59%
Webster Financial	11.35	90.00%	13.24%
Wells Fargo	13.50	90.00%	13.24%
Wilmington Trust	11.35	95.00%	13.87%

Notes: SII is the systemic importance index, defined as the number of expected bank failures given a particular bank fails; see (2). PAO is the probability of having at least one extra bank failure when a particular bank fails, defined in (1). VI is the vulnerability index, defined as the probability of failure given there exists at least another bank failure in the system; see (3).

the combination of individual institutions, it is a multivariate, rather than bilateral, relation. We consider the PAO measure proposed by Segoviano and Goodhart (2009), as well as two new measures: the SII measure, which measures the size of the systemic impact if one

Table 7. Correlation Coefficients: Monthly Data

		SII	PAO	VI
End of 2009	Total Asset	0.2137	0.1028	0.1068
		(0.2749)	(0.6027)	(0.5884)
	Total Equity	0.2254	0.1034	0.1075
		(0.2488)	(0.6005)	(0.5862)
	Total Debt	0.1489	0.0673	0.0714
		(0.4497)	(0.7335)	(0.7182)
Average	Total Asset	0.1981	0.1460	0.1488
		(0.3121)	(0.4585)	(0.4498)
	Total Equity	0.2259	0.1530	0.1560
		(0.2478)	(0.4371)	(0.4279)
	Total Debt	0.1728	0.1272	0.1300
		(0.3792)	(0.5189)	(0.5098)

Notes: SII is the systemic importance index, defined as the number of expected bank failures given a particular bank fails; see (2). PAO is the probability of having at least one extra bank failure when a particular bank fails, defined in (1). VI is the vulnerability index, defined as the probability of failure given there exists at least another bank failure in the system; see (3). The numbers in parentheses are the p-values for testing whether the correlation coefficient is significantly different from zero. The upper panel reports the results based on using end-of-2009 values of the size measure, while the lower panel reports the results based on using the average of the size measure across the full sample period (from 1987 to 2009). There is no significance result under the 10 percent level.

bank fails, and the VI measure, which measures the impact on a particular bank when the other part of the system is in distress.

We use a theoretical model based on affine portfolio holdings to show that a large bank is not necessarily more systemically important in terms of the three proposed systemic importance measures. Only with diversified banking activities might a large bank become systemically important. In contrast, the crisis of an isolated large bank will not pose a threat to the system.

The discussion can be extended to regulation policy debate. With acknowledgment of the trade-off between micro-level risk management and systemic importance, we must conclude that there is a great need for macroprudential approaches on financial regulation and supervision. Moreover, measuring systemic importance is the key to identifying systemically important institutions when imposing macroprudential regulations.

Besides developing the theoretical model, we conduct an empirical analysis—using multivariate EVT—to estimate the systemic importance measures. The empirical observation confirms that the PAO measure is not as informative as the SII measure in terms of distinguishing the systemically important banks. A moving window analysis shows similar results. Moreover, the VI measure is shown to be as informative as the PAO measure in terms of identifying systemically important banks.

We use the estimated systemic importance measures to test whether they are correlated with the measures on bank size. Regarding the systemic impact of bank failure measured by the SII measure, there is no empirical evidence supporting the "too big to fail" argument in terms of the Pearson correlation. In contrast, the other two systemic importance measures, PAO and VI, are positively correlated to the size measures. When considering the Spearman correlation, we find support for "too big to fail." Moreover, we find that in the more recent period the correlations disappear, which suggests that particular attention should be given to the systemic importance measures in recent years.

The empirical analysis in this paper is based on an artificial bank system. Therefore, the evidence from the empirical analysis should not be regarded as either support or disproof of the "too big to fail" argument. The bottom line is that we show the possibility of having a banking system in which the size measures are not a good proxy of the systemic importance.

Although in the current empirical analysis our proposed SII measure is shown to be more informative than the PAO measure proposed by Segoviano and Goodhart (2009), we address one potential drawback of the SII measure: it is a simple counting measure that takes no account of the differences between potential losses when different financial institutions fail. In other words, when calculating the expected number of failures in the system, the SII measure does not distinguish whether it causes a failure of a big bank or a small bank. This could be improved by considering the expected total loss in the system if one bank fails—i.e., calculating the expected shortfall conditional on a certain bank failure, which incorporates the size of all banks. Acharya, Santos, and Yorulmazer (2010) have designed systemic importance measures in this manner, while taking the heavy-tailedness of individual returns into consideration. To model the dependence structure, they use dynamic conditional correlation (DCC) models, which have an EVT flavor but deviate from the multivariate EVT framework. A systemic importance measure addressing the conditional expected shortfall under the multivariate EVT framework may overcome the drawback of the proposed SII measure. This is left for future research.

Appendix 1. Proofs

Proof of Proposition 1

Recall the definition of the PAO measure in (1). We have that

$$\begin{aligned}
PAO_i(p) &= \frac{P(\{\exists j \neq i, \text{ s.t. } X_j > VaR_j(p)\} \cap \{X_i > VaR_i(p)\})}{P(X_i > VaR_i(p))} \\
&= \frac{1}{p} P(\{\exists j \neq i, \text{ s.t. } X_j > VaR_j(p)\}) + 1 \\
&\quad - \frac{1}{p} P(\{\exists j \neq i, \text{ s.t. } X_j > VaR_j(p)\} \cup \{X_i > VaR_i(p)\}) \\
&= \frac{1}{p} P(\{\exists j \neq i, \text{ s.t. } X_j > VaR_j(p)\}) + 1 \\
&\quad - \frac{1}{p} P(\{\exists j, \text{ s.t. } X_j > VaR_j(p)\}) \\
&=: I_1 + 1 - I_2.
\end{aligned}$$

From the definition of the L function in (4), as $p \to 0$, $I_1 \to L_{\neq i}(1, 1, \ldots, 1)$ and $I_2 \to L(1, 1, \ldots, 1)$, which implies (6).

Proof of Proposition 3

Recall the definition of the SII measure in (2). We have that

$$SII_i(p) = \sum_{j=1}^{d} E(1_{X_j > VaR_j(p)} | X_i > VaR_i(p))$$

$$
\begin{aligned}
&= \sum_{j=1}^{d} P(X_j > VaR_j(p) | X_i > VaR_i(p)) \\
&= \sum_{j=1}^{d} \frac{P(X_j > VaR_j(p), X_i > VaR_i(p))}{P(X_i > VaR_i(p))} \\
&= \sum_{j=1}^{d} \frac{\begin{array}{c} P(X_j > VaR_j(p)) + P(X_i > VaR_i(p)) \\ -P(X_j > VaR_j(p) \text{ or } X_i > VaR_i(p)) \end{array}}{p} \\
&= \sum_{j=1}^{d} 2 - \frac{P(X_j > VaR_j(p) \text{ or } X_i > VaR_i(p))}{p}.
\end{aligned}
$$

From the definition of the L function in (4), as $p \to 0$,

$$
\frac{P(X_j > VaR_j(p) \text{ or } X_i > VaR_i(p))}{p} \to L_{i,j}(1,1).
$$

The relation (8) is thus proved.

Proof of Corrollary 1

Since $L(1,1,\ldots,1) - 1 < 0$, the relation (7) implies that a higher value of the VI measure corresponds to a higher level of $L_{\neq i}(1,1,\ldots,1)$. Together with (6), the corollary follows.

Proof of Theorem 1

Firstly, since the heavy-tailed feature in (11) assumes that the right tail of Y_i dominates its left tail, it is sufficient to assume that Y_i are all positive random variables for $i = 1, 2, 3$, i.e., without the left tail. We adopt this assumption in the rest of the proof.

We use the Feller convolution theorem to deal with the sum of independently heavy-tailed distributed random variables as in the following lemma.

LEMMA 2. *Suppose positive random variables U and V are independent. Assume that they are both heavy-tailed distributed with the same tail index α. Then, as $s \to \infty$,*

$$P(U + V > s) \sim P(U > s) + P(V > s).$$

Notice that the heavy-tailed feature implies $P(U > s)P(V > s) = o(P(U > s) + P(V > s))$, as $s \to \infty$. Hence, the Feller convolution theorem is equivalent to

$$P(U + V > s) \sim P(\max(U, V) > s);$$

i.e., the sum and the maximum of two independently heavy-tailed distributed random variables are tail equivalent. A proof using sets manipulation can be found in Embrechts, Küppelberg, and Mikosch (1997). With an analogous proof under multivariate framework, a multivariate version of the Feller theorem can be obtained. In the multivariate case, the tail equivalence between two random vectors is defined as the combination of having tail equivalence for each marginal distribution and having the same L function for the tail dependence structure. We present the result in a two-dimensional context in the following lemma without providing the proof.

LEMMA 3. *Suppose positive random variables U_1 and U_2 are independent. Assume that they are both heavy-tailed distributed with the same tail index α. Then for any positive constants m_{ij}, $1 \leq i, j \leq 2$, we have that the distribution functions of $(m_{11}U_1 + m_{12}U_2, m_{21}U_1 + m_{22}U_2)$ and $(m_{11}U_1 \vee m_{12}U_2, m_{21}U_1 \vee m_{22}U_2)$ are tail equivalent.*

To prove theorem 1, it is necessary to calculate the L function of (X_1, X_2, X_3) on points $(1, 1, 0)$, $(1, 0, 1)$, and $(0, 1, 1)$. The main instrument in the calculation is lemma 3. We start by comparing the individual risks taken by the three banks.

PROPOSITION 4. *For the three-bank model on (X_1, X_2, X_3), as $p \to 0$,*

$$VaR_2(p) = VaR_3(p) \sim cVaR_1(p), \tag{13}$$

where

$$c := \left(\frac{(2 - 2\gamma)^\alpha + 2\gamma^\alpha}{\gamma^\alpha + \mu^\alpha + (1 - \gamma - \mu)^\alpha} \right)^{-1/\alpha}.$$

Proof of Proposition 4

From lemma 2, we get that, as $s \to \infty$,

$$P(X_1 > s) \sim P((2-2\gamma)Y_1 > s) + P(\gamma Y_2 > s) + P(\gamma Y_3 > s)$$
$$= ((2-2\gamma)^\alpha + 2\gamma^\alpha)s^{-\alpha}K(s). \tag{14}$$

Similarly, we get that

$$P(X_2 > s) = P(X_3 > s) \sim (\gamma^\alpha + \mu^\alpha + (1-\gamma-\mu)^\alpha)s^{-\alpha}K(s). \tag{15}$$

By comparing (14) and (15), the relation (13) is a direct consequence.

We remark that $c < 1$. This is in line with the fact that the large bank X_1 takes more risks than the small banks. Moreover, when $\gamma > 1/2$, we get $c > 1/2$. In this case, the comparison between bank risk taking is not proportional to the size. At a relative level, the small banks are taking more risk. In other words, the large bank X_1 benefits from diversification.

Next, we calculate $L(1,1,0)$. Denote $v(p) := VaR_1(p)$. From lemma 3, we get that, as $p \to 0$,

$$\begin{aligned}
&P(X_1 > VaR_1(p) \text{ or } X_2 > VaR_2(p)) \\
&\quad \sim P((2-2\gamma)Y_1 \vee \gamma Y_2 \vee \gamma Y_3 > v(p) \text{ or} \\
&\qquad \gamma Y_1 \vee (1-\gamma-\mu)Y_2 \vee \mu Y_3 > cv(p)) \\
&\quad = P\left(Y_1 > \frac{v(p)}{(2-2\gamma)\vee\frac{\gamma}{c}} \text{ or } Y_2 > \frac{v(p)}{\gamma\vee\frac{1-\gamma-\mu}{c}} \text{ or } Y_3 > \frac{v(p)}{\gamma\vee\frac{\mu}{c}}\right) \\
&\quad \sim P\left(Y_1 > \frac{v(p)}{(2-2\gamma)\vee\frac{\gamma}{c}}\right) + P\left(Y_2 > \frac{v(p)}{\gamma\vee\frac{1-\gamma-\mu}{c}}\right) \\
&\qquad + P\left(Y_3 > \frac{v(p)}{\gamma\vee\frac{\mu}{c}}\right) \\
&\quad \sim \left[\left((2-2\gamma)\vee\frac{\gamma}{c}\right)^\alpha + \left(\gamma\vee\frac{\mu}{c}\right)^\alpha + \left(\gamma\vee\frac{1-\gamma-\mu}{c}\right)^\alpha\right] \\
&\qquad \times P(Y_1 > v(p)).
\end{aligned} \tag{16}$$

From (14), we get that $p \sim ((2-2\gamma)^\alpha + 2\gamma^\alpha)P(Y_1 > v(p))$ as $p \to 0$. Together with (16) and (5), we have that

$$L(1,1,0) = \lim_{p\to 0} \frac{P(X_1 > VaR_1(p) \text{ or } X_2 > VaR_2(p))}{p}$$
$$= \frac{\left((2-2\gamma)\vee\frac{\gamma}{c}\right)^\alpha + \left(\gamma\vee\frac{\mu}{c}\right)^\alpha + \left(\gamma\vee\frac{1-\gamma-\mu}{c}\right)^\alpha}{(2-2\gamma)^\alpha + 2\gamma^\alpha}. \qquad (17)$$

Due to the symmetry, we have that $L(1,0,1) = L(1,1,0)$. Following a similar calculation, we obtain that

$$L(0,1,1) = \frac{\left(\frac{\gamma}{c}\right)^\alpha + 2\left(\frac{\mu\vee(1-\gamma-\mu)}{c}\right)^\alpha}{(2-2\gamma)^\alpha + 2\gamma^\alpha}. \qquad (18)$$

Proposition 3 implies that $SII_2 = SII_3$. Moreover, $SII_1 > SII_2$ if and only if $L(1,1,0) < L(0,1,1)$. Hence it is only necessary to compare the values of the L function at the two points.

Denote

$$Q := ((2-2\gamma)^\alpha + 2\gamma^\alpha)(L(1,1,0) - L(0,1,1)).$$

We study the sign of Q in order to compare the systemic importance measures SII_i, $i = 1,2,3$.

Case 1: $\frac{2}{3} \le \gamma < 1$

Since $\gamma > 1/2$, we have $c > 1/2$. Thus, $\frac{\gamma}{1-\gamma} > 2 > \frac{1}{c}$, which implies that

$$\gamma > \frac{1-\gamma}{c} > \max\left(\frac{1-\gamma-\mu}{c}, \frac{\mu}{c}\right).$$

Therefore

$$Q = \left(\left((2-2\gamma)\vee\frac{\gamma}{c}\right)^\alpha + 2\gamma^\alpha\right) - \left(\left(\frac{\gamma}{c}\right)^\alpha + 2\left(\frac{\mu\vee(1-\gamma-\mu)}{c}\right)^\alpha\right) > 0.$$

Hence, $SII_1 < SII_2 = SII_3$.

CASE 2: $\gamma = 1/2$

In this case, we still have $c \geq 1/2$. The equality holds if and only if $\mu = 1/4$. Due to the symmetric position between X_2 and X_3, without loss of generality, we assume that $\mu \leq 1/4$. Then we get $\mu \leq 1 - \gamma - \mu$. Moreover, the inequality $\mu/\gamma \leq 1/2 \leq c$ implies that $\gamma \geq \frac{\mu}{c}$ and it is not difficult to obtain that $\gamma \leq \frac{1-\gamma-\mu}{c}$. Hence,

$$
\begin{aligned}
Q &= \left(1 + 2^{-\alpha} + \left(\frac{1/2-\mu}{c}\right)^{\alpha}\right) - \left(\left(\frac{\gamma}{c}\right)^{\alpha} + 2\left(\frac{1/2-\mu}{c}\right)^{\alpha}\right) \\
&= 1 + 2^{-\alpha} - \left(\frac{1}{2^{\alpha}} + \left(\frac{1}{2} - \mu\right)^{\alpha}\right) c^{-\alpha} \\
&= 1 + \frac{1}{2^{\alpha}} - \frac{1 + (1-2\mu)^{\alpha}}{2^{\alpha}} \frac{2^{\alpha} + 2}{1 + (2\mu)^{\alpha} + (1-2\mu)^{\alpha}} \\
&= 1 + \frac{1}{2^{\alpha}} - \left(1 + \frac{2}{2^{\alpha}}\right)\left(1 - \frac{(2\mu)^{\alpha}}{1 + (2\mu)^{\alpha} + (1-2\mu)^{\alpha}}\right) \\
&\leq 1 + \frac{1}{2^{\alpha}} - \left(1 + \frac{2}{2^{\alpha}}\right)\left(1 - \frac{(1/2)^{\alpha}}{1 + (1/2)^{\alpha} + (1/2)^{\alpha}}\right) \\
&= 0.
\end{aligned}
$$

Here we used the facts that $\frac{(2\mu)^{\alpha}}{1+(2\mu)^{\alpha}+(1-2\mu)^{\alpha}}$ is an increasing function with respect to μ and $\mu \leq 1/4$. The equality holds if and only if $\mu = 1/4$. Hence, in case $\gamma = 1/2$, we conclude that $SII_1 \geq SII_2 = SII_3$, with the equality holding if and only if $\mu = 1/4$.

CASE 3: $0 < \gamma \leq 1/3$

In this case, it is not difficult to verify that $2 - 2\gamma > \frac{\gamma}{c}$ and $\gamma < \frac{1-\gamma}{2c}$. Due to the symmetry, we only consider the case $0 < \mu \leq \frac{1-\gamma}{2}$. Then we have that $\mu \leq 1 - \gamma - \mu$ and $\gamma < \frac{1-\gamma-\mu}{c}$. Therefore

$$
Q = (2 - 2\gamma)^{\alpha} + \left(\gamma \vee \frac{\mu}{c}\right)^{\alpha} - \frac{\gamma^{\alpha} + (1-\gamma-\mu)^{\alpha}}{c^{\alpha}}. \tag{19}
$$

For any fixed γ, c is a function of μ denoted by $c(\mu)$. For $0 < \mu < \frac{1-\gamma}{2}$, $c(\mu)$ is a strictly decreasing function. Thus $g(\mu) := \frac{\mu}{c(\mu)}$ is a continuous, strictly increasing function. Notice that $g(0) = 0 < \gamma$ and $g((1-\gamma)/2) > \gamma$. There must exist a unique $\mu^* \in (0, (1-\gamma)/2)$ such that $g(\mu^*) = \gamma$.

Denote $c^* := c(\mu^*)$. From $\frac{\mu^*}{c^*} = \gamma$, we get that

$$\frac{\gamma^\alpha}{\mu^{*\alpha}} = (c^*)^{-\alpha} = \frac{(2-2\gamma)^\alpha + 2\gamma^\alpha}{\gamma^\alpha + \mu^{*\alpha} + (1-\gamma-\mu^*)^\alpha}.$$

It implies that

$$\frac{\gamma^\alpha}{\mu^{*\alpha}} = (c^*)^{-\alpha} = \frac{(2-2\gamma)^\alpha + \gamma^\alpha}{\gamma^\alpha + (1-\gamma-\mu^*)^\alpha}.$$

Continuing from equation (19), we get that

$$Q(\mu^*) = (2-2\gamma)^\alpha + \gamma^\alpha - (\gamma^\alpha + (1-\gamma-\mu^*)^\alpha)(c^*)^{-\alpha} = 0.$$

Hence, we conclude that for $\mu = \mu^*$, $SII_1 = SII_2 = SII_3$.

For $0 < \mu < \mu^*$, it is clear that

$$\begin{aligned} Q(\mu) &= (2-2\gamma)^\alpha + \gamma^\alpha - (\gamma^\alpha + (1-\gamma-\mu)^\alpha)c^{-\alpha} \\ &= (2-2\gamma)^\alpha + \gamma^\alpha - (\gamma^\alpha + (1-\gamma-\mu)^\alpha)\frac{(2-2\gamma)^\alpha + 2\gamma^\alpha}{\gamma^\alpha + \mu^\alpha + (1-\gamma-\mu)^\alpha} \\ &= \frac{((2-2\gamma)^\alpha + \gamma^\alpha)\mu^\alpha - \gamma^\alpha(\gamma^\alpha + (1-\gamma-\mu)^\alpha)}{\gamma^\alpha + \mu^\alpha + (1-\gamma-\mu)^\alpha} \end{aligned}$$

is a strictly increasing function with respect to μ. Moreover, for $\mu^* < \mu < \frac{1-\gamma}{2}$, Q is calculated as

$$\begin{aligned} Q(\mu) &= (2-2\gamma)^\alpha + (\mu^\alpha - \gamma^\alpha - (1-\gamma-\mu)^\alpha)c^{-\alpha} \\ &= (2-2\gamma)^\alpha + (\mu^\alpha - \gamma^\alpha - (1-\gamma-\mu)^\alpha)\frac{(2-2\gamma)^\alpha + 2\gamma^\alpha}{\gamma^\alpha + \mu^\alpha + (1-\gamma-\mu)^\alpha} \\ &= \frac{2(2-2\gamma)^\alpha\mu^\alpha + 2\gamma^\alpha(\mu^\alpha - \gamma^\alpha - (1-\gamma-\mu)^\alpha)}{\gamma^\alpha + \mu^\alpha + (1-\gamma-\mu)^\alpha}, \end{aligned}$$

which is also a strictly increasing function with respect to μ. Therefore, for $0 < \mu < \mu^*$, $Q < 0$ while for $\mu^* < \mu < \frac{1-\gamma}{2}$, $Q > 0$. Correspondingly, we have $SII_1 > SII_2 = SII_3$ in the former case and $SII_1 < SII_2 = SII_3$ in the latter case.

Due to the symmetry, a similar result holds when $\frac{1-\gamma}{2} < \mu < 1-\gamma$, and the switch point is then $1-\gamma-\mu^*$. More specifically, for $\frac{1-\gamma}{2} < \mu < 1-\gamma-\mu^*$, $SII_1 < SII_2 = SII_3$; and for $\mu = 1-\gamma-\mu^*$, $SII_1 = SII_2 = SII_3$; for $1-\gamma-\mu^* < \mu < 1-\gamma$, $SII_1 > SII_2 = SII_3$. The theorem is thus proved for case 3.

Proof of Lemma 1

Along the lines of the proof of theorem 1, we obtained that in a three-bank model,

$$\begin{aligned} SII_1 - SII_2 &= \{1 + (2 - L(1,1,0)) + (2 - L(1,0,1))\} \\ &\quad - \{(2 - L(1,1,0)) + 1 + (2 - L(0,1,1))\} \\ &= L(0,1,1) - L(1,0,1). \end{aligned}$$

Therefore, $SII_i > SII_j$ if and only if $L_{\neq i}(1,1,\ldots,1) > L_{\neq j}(1,1,\ldots,1)$. Together with (6), the lemma is proved.

We remark that the above calculation is not based on the affine portfolio model. Hence, the result is valid in a general three-bank model. However, one cannot obtain similar relations when the number of banks exceeds three.

Appendix 2. Statistical Estimation on the L Function

Consider independently and identically distributed (i.i.d.) observations from the random vector $(X_1,\ldots,X_d)$ denoted by

$$\{(X_{1s}, X_{2s}, \ldots, X_{ds}) | 1 \leq s \leq n\}.$$

The sample size is n. The non-parametric approach of estimating the L function starts from the assumption (4). Roughly speaking, the estimation takes a certain p-value for which the VaR for each dimension can be estimated by the order statistics. Then the probability in the numerator of (4) is estimated by a counting measure. To ensure that $p \to 0$, theoretically we take a sequence $k := k(n)$, such that $k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$, and get an empirical estimation of the L function from replacing p with k/n and using the empirical estimation on the distribution function of $(X_1, X_2 \ldots, X_d)$. The explicit estimator is given as

$$\hat{L}(x_1,\ldots,x_d) := \frac{1}{k}\sum_{s=1}^{n} 1_{\exists 1 \leq i \leq d, \text{ s.t. } X_{is} > X_{i,n-[kx_i]}},$$

where $X_{i,1} \leq X_{i,2} \leq \cdots \leq X_{i,n}$ are the order statistics of the i-th dimension of the sample, $X_{i1}, \ldots, X_{in}$, for $1 \leq i \leq d$. Particularly, $L(1, 1, \ldots, 1)$ is estimated by

$$\hat{L}(1, 1, \ldots, 1) := \frac{1}{k} \sum_{s=1}^{n} 1_{\exists 1 \leq i \leq d, \text{ s.t. } X_{is} > X_{i,n-k}}.$$

For the estimator of the L function, usual statistical properties, such as consistency and asymptotic normality, have been proved. See, e.g., de Haan and Ferreira (2006).

Practically, the theoretical conditions on k are not relevant for a finite sample analysis. Thus, how to choose a proper k in the estimator is a major issue in estimation. Instead of taking an arbitrary k, a usual procedure is to calculate the estimator of $L(1, 1, \ldots, 1)$ under different k values and draw a line plot against the k values. With a low k value, the estimation exhibits a large variance, while for a high k value, since the estimation uses too many observations in the moderate level, it bears a potential bias. Therefore, k is usually chosen by picking the first stable part of the line plot starting from low k, which balances the trade-off between the variance and the bias. The estimates follow from the k choice. Because k is chosen from a stable part of the line plot, a small variation of the k value does not change the estimated value. Thus, the exact k value is not sensitive for the estimation on the L function. Such a procedure has been applied in univariate EVT for tail index estimation.

With the chosen k, we in fact consider a tail event as the loss return exceeds a VaR with tail probability level k/n. In our empirical application, the chosen k value differs according to the sample size. For the full sample analysis (sample size 6,000), we choose $k = 180$, which corresponds to a p level at 3 percent; for the moving window analysis (sample size 2,000), we choose $k = 100$, which corresponds to $p = 5\%$; for the monthly data analysis (sample size 276), we choose $k = 20$, which corresponds to $p = 7.2\%$. With a lower number of observations, we choose a slightly higher level of p, albeit in a low absolute level.

References

Acharya, V. V., J. A. C. Santos, and T. Yorulmazer. 2010. "Systemic Risk and Deposit Insurance Premiums." *Economic Policy Review* (Federal Reserve Bank of New York) 16 (1): 89–98.

Acharya, V. V., and T. Yorulmazer. 2007. "Too Many to Fail—An Analysis of Time-Inconsistency in Bank Closure Policies." *Journal of Financial Intermediation* 16 (1): 1–31.

Adrian, T., and M. Brunnermeier. 2008. "CoVaR." Federal Reserve Bank of New York Staff Report No. 348.

Allen, F., E. Carletti, and A. Babus. 2009. "Financial Crises: Theory and Evidence." *Annual Review of Financial Economics* 1 (1): 97–116.

Allen, F., and D. Gale. 2000. "Financial Contagion." *Journal of Political Economy* 108: 1–33.

Bernanke, B. 2009. "Financial Reform to Address Systemic Risk." Speech at the Council on Foreign Relations, Washington, DC, March 10.

Chavez-Demoulin, V., A. C. Davison, and A. J. McNeil. 2005. "Estimating Value-at-Risk: A Point Process Approach." *Quantitative Finance* 5 (2): 227–34.

Cifuentes, R., H. S. Shin, and G. Ferrucci. 2005. "Liquidity Risk and Contagion." *Journal of the European Economic Association* 3 (2–3): 556–66.

Dasgupta, A. 2004. "Financial Contagion through Capital Connections: A Model of the Origin and Spread of Bank Panics." *Journal of the European Economic Association* 2 (6): 1049–84.

De Bandt, O., and P. Hartmann. 2001. "Systemic Risk in Banking: A Survey." In *Financial Crisis, Contagion, and the Lender of Last Resort: A Reader*, ed. C. Goodhart and G. Illing, 249–98. Oxford University Press.

De Haan, L., and A. Ferreira. 2006. *Extreme Value Theory: An Introduction.* Springer.

De Vries, C. G. 2005. "The Simple Economics of Bank Fragility." *Journal of Banking and Finance* 29 (4): 803–25.

Diebold, F. X., T. Schuermann, and J. D. Stroughair. 2000. "Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management." *Journal of Risk Finance* 1 (2): 30–35.

Diebold, F. X., and K. Yilmaz. 2009. "Measuring Financial Asset Return and Volatility Spillovers, with Application to Global Equity Markets." *Economic Journal* 119 (534): 158–71.

Embrechts, P., C. Klüppelberg, and T. Mikosch. 1997. *Modelling Extremal Events: for Insurance and Finance*. Springer.

Freixas, X., B. M. Parigi, and J.-C. Rochet. 2000. "Systemic Risk, Interbank Relations, and Liquidity Provision by the Central Bank." *Journal of Money, Credit, and Banking* 32 (3): 611–38.

Hartmann, P., S. Straetmans, and C. G. de Vries. 2004. "Asset Market Linkages in Crisis Periods." *Review of Economics and Statistics* 86 (1): 313–26.

———. 2005. "Banking System Stability: A Cross-Atlantic Perspective." NBER Working Paper No. 11698.

International Monetary Fund. 2009. *Global Financial Stability Report* (April).

Lagunoff, R., and S. L. Schreft. 2001. "A Model of Financial Fragility." *Journal of Economic Theory* 99 (1): 220–64.

Poon, S., M. Rockinger, and J. Tawn. 2004. "Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications." *Review of Financial Studies* 17 (2): 581–610.

Rajan, R. G. 2009. "Too Systemic to Fail: Consequences, Causes and Potential Remedies." Written statement to the Senate Banking Committee Hearings, May 6.

Segoviano, M., and C. Goodhart. 2009. "Banking Stability Measures." IMF Working Paper No. 09/04.

Stern, G. H., and R. J. Feldman. 2004. *Too Big to Fail: The Hazards of Bank Bailouts*. Brookings Institution Press.

Zhou, C. 2008. *On Extreme Value Statistics*. PhD Thesis, Tinbergen Institute.

Discussion of "Are Banks Too Big to Fail?"

Stefan Straetmans
Maastricht University School of Business and Economics

Chen Zhou's paper is quite ambitious. Not only is it focused on developing and comparing measures of systemic risk (abbreviated by SR in the rest of this discussion); it also tries to clarify whether and to what extent bank size triggers SR. The practical relevance of this research is obvious, as there is an urgent need for simple and economically intuitive SR indicators for policymakers (governments, central banks, regulatory authorities). During non-crisis times, it may be desirable for policymakers to monitor whether SR remains below some level that is deemed to be critical; SR may also be used as a criterion to judge whether a distressed bank is sufficiently important from a systemic point of view to justify its bailout. Although one can imagine that structural reforms in the financial sector can reduce part of the existing SR potential, it will probably never be completely wiped out in the future, which implies that SR monitoring and measuring will remain relevant.

The size-SR relation that constitutes the second theme of the paper seems to be taken for granted nowadays by governments or central banks that often made liquidity provisions conditional on breaking up financial institutions into smaller pieces.[1] However, are bigger banks necessarily fueling SR? This is not obvious. According to finance theory, SR is either induced by "direct" channels via interbank market linkages or "indirect" channels such as similar portfolio holdings in bank balance sheets. However, bigger banks are not necessarily more interconnected with the rest of the financial system, nor do they necessarily exhibit more diversified portfolio holdings. The paper—although mainly empirically inclined—actually contains a parsimonious model of bank interdependence that shows that there *may* be a relation between size and SR. In the end, however, whether the size-SR relation exists or not remains an empirical issue.

[1] As to date, however, not much has been done yet in terms of "breaking up" the banking system (especially in the United States).

The problem one faces if one wants to model or measure SR (or, alternatively, the related concept of bank contagion) is that there is no generally accepted definition available. The paper therefore proposes to measure three competing probabilistic measures of bank stock returns' extreme co-movements. The already existing PAO index of Segoviano and Goodhart (2009) is taken as a benchmark for comparison with two newly proposed measures. Loosely speaking, the PAO index is the conditional probability of having at least one extra bank failure in the event a particular bank fails. Turning around the conditioning event of the PAO index, one gets the conditional probability measure that a particular bank fails in the event that at least one other bank fails—i.e., the "vulnerability" index (VI). Finally, the systemic impact index (SII) is defined as the expected number of bank failures in the event that one particular bank fails. All three measures summarize specific information on the risk spillover in the banking system. However, whereas the PAO and VI measures only provide probabilities of banking failures conditional on other banking failures, the SII indicator is more a truly multivariate measure in that it reflects how many banks are potentially influenced when one particular bank fails.

After defining the SR indicators, statistical Multivariate Extreme Value Theory (MEVT) is put to work to estimate the three measures by means of so-called tail dependence functions, or tail copulas. To my knowledge, the only two papers that previously applied MEVT techniques to assess banking system stability are Hartmann, Straetmans, and de Vries (2006) and de Jonghe (2010). In other words, the approach is pretty novel to the area of banking and SR. The statistical concept of tail copula deserves some further clarification because it constitutes the crucial device for calculating the three considered SR indexes. It is a tail version of the ordinary copula that holds in the center of a joint distribution. The tail copula exists under fairly general conditions when the joint distribution function of sets of random variables is well defined. More specifically, the tail copula $l(u, v)$ (assume for simplicity a bivariate pair of random variables) is defined as

$$l(u, v) = \lim_{t \to 0} t^{-1} P\{X > VaR_1(tu) \text{ or } Y > VaR_2(tv)\};$$

see, e.g., Hartmann, Straetmans, and de Vries (2004). The quantiles VaR_1 and VaR_2 are the Value-at-Risk levels that correspond to marginal exceedance probabilities of tu and tv, respectively. In contrast to correlation analysis, the curvature of $l(u, v)$ completely determines the dependence structure of joint risks in their tails. The tail copula relates marginal and joint probabilities in the following way. First define the exceedance probabilities $p_1 = P\{X > VaR_1(p_1)\}$, $p_2 = P\{Y > VaR_2(p_2)\}$, and $p_{12} = 1 - P\{X \leq VaR_1(p_1), Y \leq VaR_2(p_2)\}$ for the sake of notational convenience. One can easily show that the bivariate excess probability p_{12} and the marginal probabilities p_1 and p_2 are related via the tail copula. For sufficiently small $t > 0$,

$$l(u, v) \approx t^{-1}[1 - P\{X \leq VaR_1(tu), Y \leq VaR_2(tv)\}].$$

Choose $tu = p_1$ and $tv = p_2$, so that $l(u, v) = l(t^{-1}p_1, t^{-1}p_2)$. Moreover, the linear homogeneity property of the l-function implies that $tl(t^{-1}p_1, t^{-1}p_2) = l(p_1, p_2)$. Hence for small values of p_1 and p_2, approximately,

$$p_{12} \approx l(p_1, p_2).$$

Thus the joint probability p_{12} only depends on the marginal probabilities p_1 and p_2 and tail copula $l(., .)$. Non-parametric estimates of the tail copula are used to calculate time-constant values of the SR indicators for a given sample. A second part of the empirical analysis provides a correlation analysis between the systemic contribution measures and some size proxies. The analysis is performed for twenty-eight banks and moving window samples. The correlation outcomes do not reveal a robust relation between SR estimates and size proxies. The statistical and economic significance of the correlations seems to fluctuate across sub-samples, SR measures, and size proxies.

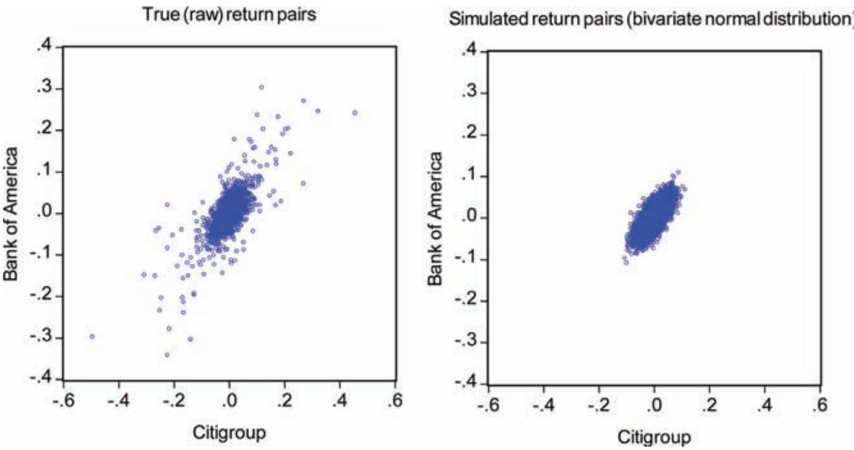
The main advantage of the statistical MEVT methodology is that it enables one to tackle the very low frequency nature of systemic events. Previous empirical approaches to modeling systemic events including the use of multinomial logit models (see, e.g., Gropp, Lo Duca, and Vesala 2009) or quantile regression methodology (see, e.g., Adrian and Brunnermeier 2008) typically analyze collapses in bank stocks corresponding to marginal exceedance probabilities that do not fall below the 1 percent level. This does not really correspond with very infrequent tail events. Upon assuming the use of daily return data, a 1 percent marginal exceedance probability corresponds to a quantile or crisis level that is expected to be exceeded once every 100 days. Of course, the true frequency of systemic events remains a subjective matter, but most people would probably agree that joint stock price co-movements that are expected to happen more than once a year can hardly be dubbed "extreme" or "systemic."

Another appealing feature of the MEVT approach constitutes the fact that the SR indexes can be estimated without knowing the marginal distributions—i.e., one does not need to estimate the Value-at-Risk levels VaR_1 and VaR_2 that characterize the banks' tail risk; it suffices to choose the marginal exceedance probabilities p_1 and p_2.[2] Thus, the considered SR indicators purely reflect information on the dependence between extreme bank stock returns and are not "contaminated" by, e.g., asymmetries or inequalities in the marginal distributions.

Yet another advantage is that the MEVT methodology takes account of "non-linear dependence" and "tail dependence"—provided these empirical stylized facts are present. It is often suggested that contagion phenomena or systemic risk spillovers may be non-linear dependence phenomena that cannot be captured by simple linear approaches like, e.g., regression/correlation analysis. Also, it is by now generally accepted that financial returns exhibit "tail dependence" (see, e.g., de Vries 2005 for the case of bank stock returns). Capturing the tail dependence feature is essential for accurately assessing SR. Pairs of random variables are tail dependent if the joint conditional exceedance probability $P\{X > s|Y > s\}$ does not vanish when the exceedance level s grows large, i.e., $\lim_{s\to\infty} P\{X > s|Y > s\} > 0$. To illustrate the difference between tail dependence and tail independence, I pick a representative pair of bank stock returns (Citigroup; Bank of America) from Chen Zhou's data set and assume they are jointly normally distributed. Upon estimating the means, standard deviations, and correlation, the joint distribution is completely determined. Next, I use the bivariate normal model as a simulation vehicle to draw a sample of the same size as the raw data sample (n = 6,109 return pairs). Figure 1 shows both data clouds.

[2]Without loss of generality, Chen Zhou's paper assumes that $p_1 = p_2 = p$.

Figure 1. Extreme Bank Stock Return Co-Movements: Real vs. Simulated Data



For sake of comparison, the axes are identically scaled. One can see that the normal data cloud is not able to reproduce the joint downward crashes in the bank stocks that are clearly visible in the true data cloud (tail dependence). That there is so little extremal dependence in the right-hand-side graph may seem surprising because the Pearson correlation between the two bank stock return series is found to equal 0.6915. However, the bivariate normal distribution is characterized by tail independence, which implies that a non-zero correlation in the center of the distribution vanishes in the tails. The above example is a popular reminder that one should be careful with choosing parametric models for modeling extremal spillovers and SR: either one opts for parametric models that nest both data features of heavy tails and tail dependence or one decides to work with purely non-parametric techniques that pick up these stylized facts in the data automatically. Chen Zhou's paper chooses the latter approach.

The MEVT approach is also characterized by certain limitations. First of all, an empirical approach that uses bank stock returns to evaluate SR is by definition unable to evaluate the potential systemic impact of large non-quoted banks. Also, the market efficiency assumption that bank stocks fully reflect a bank's liquidity and solvency situation at each time instance may be considered by some as too restrictive. Moreover, the fact that MEVT estimation techniques only use tail observations implies that one needs sufficiently large samples to start with, as one typically uses only 1 to 5 percent of the full sample for estimating the tail features.[3] Finally, the tail copula methodology as employed in this paper does not allow SR indicators to be truly time varying but reflects the SR over a given sample period (typically several years). This problem is partly remedied in the paper by estimating the SR indicators over a rolling window size.

The correlation analysis of the size-SR relationship also raises some questions. The non-robustness of the correlation outcomes is puzzling. Maybe the true underlying relation—if it exists—is nonlinear in nature, which would invalidate the use of linear correlation analysis. Also, the correlation analysis is based on SR indicators and size variables of twenty-eight banks, which represents only a small cross-section of the U.S. banking sector. One should also be careful with correlating truncated variables like the SR indicators to a size variable without some specific transformation of the latter. The computation of rank correlations between size and SR is probably the simplest way to circumvent this problem. Finally, one should be aware that the SR variable itself is an estimate with a standard deviation. To what extent the cross-sectional variation in SR point estimates is statistically and economically significant can only be judged if one includes confidence intervals.

This paper is by far not an endpoint in SR evaluation, and one could think of several potential extensions. First, the fuzziness of the SR definition opens the door for yet other measures. For example, one might be interested in assessing the impact—in money terms—on the financial sector as a whole of one failing financial institution by means of an aggregate expected loss in the event of some individual bank failure. Sharp losses on a bank market index or portfolio can be used as a basis for calculating this expected loss. Let B stand for the return series of the individual bank and M stand for a

[3] Experience shows that full samples of 500 return observations may already be sufficient in order to obtain reasonably accurate estimates of the tail features. The sample in Chen Zhou's paper largely fulfills this lower-bound restriction for the sample size.

banking market index return series such that $B > 0$ and $M > 0$ represent a loss (or negative return). The expected loss of the banking index conditional upon an individual bank failure (or, alternatively, expected shortfall) boils down to

$$E(M > VaR_M(p)|X > VaR_X(p)),$$

with $P\{M > VaR_M(p)\} = P\{X > VaR_X(p)\} \equiv p$.

It would also be of interest to calculate the pre-bankruptcy values of SR for those banks that actually went bust and to see to what extent the indicators for those banks are outliers as compared with liquid and solvent banks, i.e., can the SR indicators be considered as early-warning indicators? Moreover, for sake of sensitivity analysis, one would like to quantify the SR indicators using alternatives for bank stock returns such as the distance-to-default (dd) measure or credit default swap (CDS) spreads. Finally, size is not the only potential determinant of a bank's contribution to overall systemic instability. Future research should also focus on identifying other variables that may influence the level of SR. If one manages to identify a stable relation between these determinants and a generally accepted SR definition, this "transmission mechanism" may be exploitable by policymakers for the sake of controlling SR just like, e.g., central bankers manipulate interest rates or monetary aggregates to fight inflation. The modeling of the interaction between SR indicators and its potential triggers in a truly time-varying way constitutes one important future research challenge.

References

Adrian, T., and M. Brunnermeier. 2008. "CoVaR." Federal Reserve Bank of New York Staff Report No. 348.

De Jonghe, O. 2010. "Back to the Basics of Banking? A Micro-Analysis of Banking System Stability." *Journal of Financial Intermediation* 19 (3): 387–417.

De Vries, C. G. 2005. "The Simple Economics of Bank Fragility." *Journal of Banking and Finance* 29 (4): 803–25.

Gropp, R. E., M. Lo Duca, and J. Vesala. 2009. "Cross-Border Bank Contagion in Europe." *International Journal of Central Banking* 5 (1): 97–140.

Hartmann, P., S. Straetmans, and C. G. de Vries. 2004. "Asset Market Linkages in Crisis Periods." *Review of Economics and Statistics* 86 (1): 313–26.

———. 2006. "Banking System Stability: A Cross-Atlantic Perspective." In *The Risk of Financial Institutions*, ed. M. Carey and R. M. Stulz, 133–93. Chicago and London: The University of Chicago Press.

Segoviano, M., and C. Goodhart. 2009. "Banking Stability Measures." IMF Working Paper No. 09/04.

Systemic Risk: Changing the Regulatory Perspective*

Jean-Charles Rochet
University of Zürich and Toulouse School of Economics

The article puts forward the view that the regulatory perspective on systemic risk should be changed drastically. The sub-prime crisis has indeed revealed many loopholes in the supervisory/regulatory framework for banks—in particular, the inability to deal with the too-big-to-fail syndrome and also the lack of resiliency of interbank and money markets. To a large extent, the contagion phenomena that took place in these markets were the necessary outcomes of the passive attitude of banking supervisors, who have let large banks develop a complex and opaque nexus of bilateral obligations. We propose two reforms: adopting a platform-based (instead of institution-based) regulatory perspective on systemic risk and encouraging a generalized move to central counterparty clearing.

JEL Codes: G21, L51.

1. Introduction

This article puts forward the view that the regulatory perspective on systemic risk should be changed drastically. Indeed, the sub-prime crisis has revealed many loopholes in the supervisory/regulatory system. But the main lesson that can be drawn from the actions taken (and statements made) by public authorities during this crisis is

*This paper was prepared as a commentary on the third session of the Second Financial Stability Conference of the International Journal of Central Banking, Madrid, Bank of Spain, June 17–18, 2010. It draws on my keynote address in the Annual Meeting of the Finnish Economic Association, Turku (February 5–6, 2009), as well as my presentations in the 54th Economic Conference of the Federal Reserve Bank of Boston, Cape Cod (October 21–23, 2009) and in the 4th annual meeting of the Swiss Finance Institute, Zürich (November 30, 2009). I thank the participants in these meetings for their useful comments.

that, in the future, any large financial institution that encounters financial problems can expect to be bailed out by public authorities on the grounds that it is too big to fail (TBTF) (alternative terms are too interconnected to fail, large and complex banking organization, or systemically important financial institution). The turmoil that followed the failure of Lehman Brothers in September 2008 has indeed led politicians to believe they had to commit to an unconditional support of any troubled financial institution whose failure might create major disruptions. Of course, this commitment is a disaster in terms of moral hazard and market discipline. From a forward-looking perspective, public authorities could not convey a worse message to market participants and bank managers.

A similar pattern emerged after the Continental Illinois bailout in 1984,[1] and at the time, it took more than five years for market discipline to be somewhat restored.[2] But this bailout was a single event, and the Comptroller of the Currency at the time tried to maintain, as much as he could, some ambiguity on which banks were really TBTF. Today there is no more ambiguity: all large financial institutions will always be rescued. Public authorities of G-20 countries have even agreed to publicly commit to a systematical bailout. Unless resolute reforms are undertaken, it will probably take a very long time to restore market discipline again. The situation is even aggravated by the fact that an indirect outcome of the crisis is an increased concentration of the banking systems of many countries, the surviving banks becoming even bigger than before and, in some countries at least, close to being too big to be bailed out.

Another major source of concern for public authorities is the complete lack of resiliency of interbank and money markets during the recent crisis. It is amazing to see how some shocks to the relatively small sub-prime market could lead to the complete dry-up of

[1]In May of 1984, Continental Illinois was bailed out by the U.S. federal government. It was only the seventh largest bank in the United States, but it was a money-center bank holding large deposits of hundreds of smaller banks. U.S. supervisors feared that its failure could propagate toward many of these smaller banks.

[2]Flannery and Sorescu (1996) show that banks' debt spreads only started reflecting default risks around 1989, after a regulatory transition toward letting market participants share the losses when a banking firm fails.

liquidity markets for more than a year. This paper argues that this lack of resiliency is due to a fundamental mistake in the way these markets were conceived. To a large extent, the contagion phenomena that took place on these markets were the necessary outcomes of the passive attitude of banking supervisors, who have let large banks develop an enormous and opaque nexus of bilateral obligations. In Rochet and Tirole (1996), Jean Tirole and I explored the theoretical justifications of such a decentralized organization of the interbank markets and found only one possible answer: market discipline. More precisely, we found that the only possible explanation why prudential authorities could have let banks organize the trade of their reserves vis-à-vis the central bank in an over-the-counter (OTC) fashion was the hope to promote what we called peer monitoring, i.e., the mutual surveillance of banks by their competitors. However, this hope was misplaced: the price to pay for this mutual surveillance—namely, the risk of contagion—was too heavy. Market discipline could only have worked if public authorities could have convinced market participants that they would not intervene if a systemic crisis occurred, which revealed not credible.

A logical consequence of this result, which we did not put forward clearly enough in Rochet and Tirole (1996), is that the current, decentralized, organization of interbank markets has a huge cost (contagion risk) but no benefit. Market discipline does not work for the interbank market, not only because of the strong likelihood of a public bailout in the event of a crisis but also because of the faulty conception of its industrial organization. Decentralized trading of bank reserves has a major drawback: it bundles liquidity risk with counterparty risk, which makes price discovery almost impossible.

The rest of this article is structured as follows. Section 2 reviews traditional justifications for prudential regulation of banks. Section 3 examines what is left of these justifications after the crisis. Section 4 illustrates the main issues with a discussion of the main concerns for the repo market. Section 5 shows that the traditional paradigm of banking theory has become obsolete. Section 6 advocates in favor of a generalization of centralized trading and systemic risk. Section 7 suggests a way to change the regulatory perspective on systemic risk. Section 8 concludes.

2. Banking Regulation: The Classical Doctrine

The traditional doctrine holds that prudential regulation of banks is essentially justified by two reasons:

- It protects depositors and limits the liability of deposit insurance funds.

  This is microprudential regulation, analyzed in detail by Dewatripont and Tirole (1994). Without going into detail, the important ingredients of an efficient microprudential regulation are independence and accountability of supervisors, use of market discipline, a lender-of-last-resort policy governed by the Bagehot principles, no injection of public money, and cost-minimizing resolution of failures. This doctrine is best illustrated by the U.S. Federal Deposit Insurance Corporation Improvement Act (FDICIA), which articulated very clearly the notion of prompt corrective action.
- It protects the banking and financial "systems."

  This is macroprudential regulation. It aims at avoiding contagious failures, spillovers, and major disruptions to the banking and financial system. It justifies renouncing all the principles stated above, i.e., introducing exceptions to FDICIA, possible intervention of the Treasury, liquidity injections by the central bank, and (temporarily?) abandoning the recourse to market discipline.

This duality between micro- and macroprudential regulations is well illustrated by the doctrine employed by the Bank of Canada (1995) in its lender-of-last-resort policies:

> For solvent financial institutions requiring . . . credit, the Bank can provide Emergency Lending Assistance (ELA). ELA is intended to overcome a market failure associated with financial institutions that have a significant share of their liabilities as "deposits" (fixed-value promises to pay, redeemable at very short notice) and whose assets are generally highly illiquid. The Bank of Canada Act requires that such lending be secured by collateral pledged by the borrowing institution. . . . The collateral eligible to secure credit from the SLF is the same as that eligible for intraday credit in the Large Value Transfer System. . . .

> It is the policy of the Bank to lend only to institutions that are judged to be solvent in order to mitigate moral hazard that can arise from such potential intervention, and to avoid damaging the interests of unsecured creditors.
>
> In conditions of severe and unusual stress on the financial system more generally, the Bank has authority to provide liquidity through outright purchases of a wide variety of securities issued by any Canadian or foreign entities, including non-financial firms. . . . In other words, the Bank has the authority to provide liquidity to a broad range of financial and non-financial institutions when the Governor of the Bank judges that such transactions are justified to safeguard the safety and soundness of Canada's financial system.

This is all very fine. Alas, the sub-prime crisis has shown that these doctrines were largely insufficient.

3. What's Left of the Classical Doctrine After the Crisis

One of the striking features of the sub-prime crisis was that shocks to the relatively small sub-prime market could provoke the distress of vital parts of the financial infrastructure, especially interbank and money markets. This overreaction was largely due to the uncertainty of market participants about the impact that a decline in real estate prices and the beginning of a recession might have on a sizable fraction of the assets held by large banks. These large banks are the main players in these liquidity markets, which are vital to modern economies.

Confronted with this freezing of money markets, central banks did what they could to substitute these failing markets. They organized several kinds of lending facilities and de facto provided the intermediation of a large part of liquidity flows among banks and also between banks and some non-banks. In parallel, public authorities all over the world injected large amounts of capital and provided a whole spectrum of guarantees to financial institutions, in the hope of restarting these vital liquidity markets. These (largely improvised) interventions were very costly and only partially succeeded in restarting liquidity markets and restoring confidence. But the important message is that the justification for public intervention was not so much avoiding contagious failures but rather maintaining the integrity of some parts of financial infrastructure that are deemed "vital" to the economy.

As for the future, envisaging less costly ways to maintain financial stability should be on the top of the reform agenda. In particular, it would be disastrous to let market participants consider that all large financial institutions will always be rescued (and their creditors insured) if they are again in a situation of financial distress. Taxpayers of most countries will not be willing to accept a second dose of the sort of blanket guarantees that governments have committed to provide to large financial institutions in the hope of maintaining financial stability.

As I already argued, the main issue is how to improve supervision of systemically critical firms and to strengthen the resilience of the financial system to the unwinding of such a large firm. This implies that any "systemically important" firm must receive especially close supervisory oversight of its risk taking, risk management, and financial condition, and be held to high capital and liquidity standards.

This poses the major difficulty of identifying these TBTF firms. What criteria should be used to determine when a firm (not necessarily a bank) is TBTF and when it is not? The paper by Chen Zhou (2010), presented in this session, proposes new statistical measures of systemic risk building on previous contributions by Adrian and Brunnermeier (2009) and Goodhart and Segoviano (2009). Another interesting approach is the methodology proposed by Tarashev, Borio, and Tsatsaronis (2009) for the allocation of systemwide risk to each individual institution, in line with its systemic importance. This methodology combines statistical risk measures with the Shapley value, a widely used solution concept in cooperative game theory. This approach could be used to provide guidelines for defining which firms should be subject to an alternative regime as systemically important, and the process for invoking that regime. A more pragmatic solution could be to adapt the procedures used for invoking the so-called systemic risk exception under FDICIA.

Another important question is which agency should decide which institutions are systemic. Many central banks around the world already have an explicit statutory basis for their oversight of critical payment and settlement systems. As I argue in more detail below, a natural corollary is that these central banks should also be in charge of systemic risk supervision and, in particular, should decide which institutions are systemic and which are not. This is not the current situation in the United States, where the Federal Reserve does not have explicit oversight authority for systemically important payment and settlement systems. Reforming this might be reasonable.

The main lesson that can be drawn from the behavior of public authorities during the sub-prime crisis is that protecting financial infrastructure—i.e., the institutions that support trading, payments, clearing, and settlement—has become the fundamental reason behind macroprudential regulation and supervision. The aim here is not only to make the financial system as a whole more resilient but also to reduce the need for future government intervention. I claim that this requires a drastic change in the regulatory/supervisory perspective.

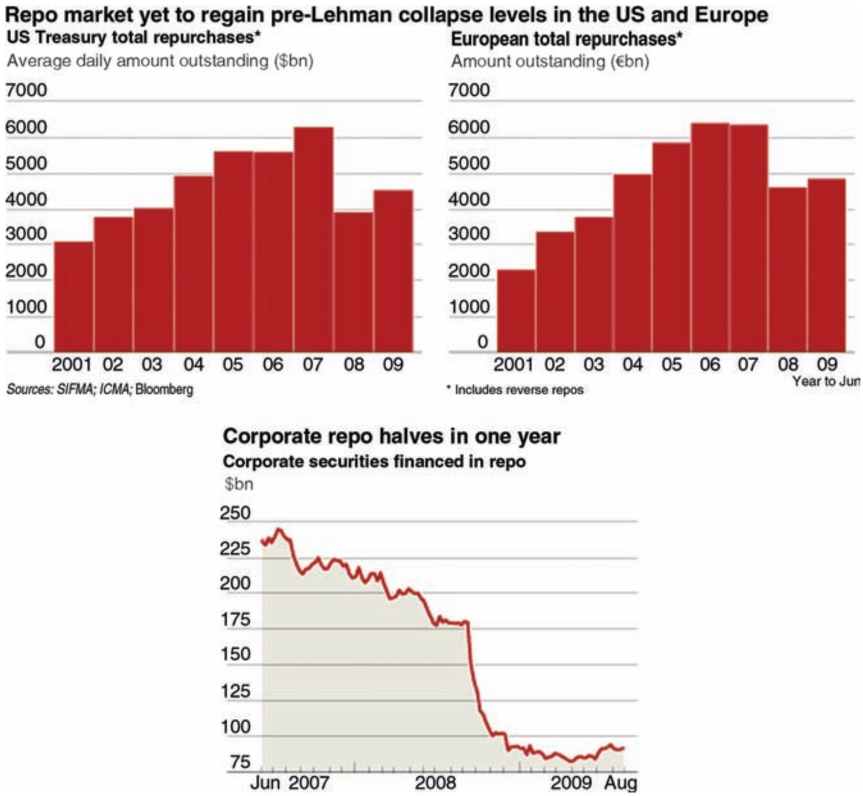
4. An Illustration: The Repo Market

A repo is a sale of securities coupled with an agreement to repurchase the securities at a specific price on a later date. Repo markets perform essential functions: first, they provide secured investments to cash investors on the money market; second, they allow the borrowing and lending of securities; and third, they indirectly boost liquidity on crucial financial markets such as debt markets (especially Treasuries) and derivatives markets.

To avoid the collapse of these markets during the crisis, central banks have taken extraordinary actions. For example, the Federal Reserve has established temporary facilities such as the Primary Dealer Credit Facility (PDCF) and the Term Securities Lending Facility (TSLF) to provide liquidity to market participants. Other central banks like the European Central Bank and the Bank of England have agreed to lend to more counterparties and have enlarged the scope of eligible collateral. However, these facilities have not prevented a sharp reduction in the activity of these repo markets, as illustrated by figure 1, taken from *Financial Times* (2009).

In the United States, a very popular form of repo is the triparty repo, in which an intermediary (a clearing bank) facilitates transactions by providing operational services (custody of securities, settlement, valuation of collateral) and, more importantly, extending

Figure 1. Impact of the Crisis



intraday credit to market participants. On average, more than $1.7 trillion was exchanged in this way in the United States every working day of the first quarter of 2010.

There are several reasons why this market poses several threats to financial stability (Federal Reserve Bank of New York 2010). For one, the U.S. market is completely dominated by only two clearing banks: Bank of New York Mellon and JP Morgan Chase, which are thus perfect examples of systemic institutions. Moreover, participation in this market is also very concentrated: the ten biggest dealer banks represent 85 percent of cash borrowing and 65 percent of security lending (with a typical position, for a large dealer, of around $200 billion). Moreover, the main actors are also big in other markets. For example, JP Morgan Chase is also the largest

OTC derivatives dealer in the United States (with a $79 trillion notional position at the end of 2009) and the largest manager of hedge funds (with $53.5 billion in assets under management at the end of 2009). The risk of contagion to and from the repo markets is thus enormous.

There are also other, equally important concerns (Federal Reserve Bank of New York 2010), such as the fact that this market relies on huge amounts of intraday credit. This is because all tri-party repos are unwound each morning, even if most of them are renewed later in the day. Moreover, there is no regulatory oversight of this market. It is quite opaque, and some of its big actors are foreign and thus not directly supervised by U.S. authorities. This has led to many forms of inadequate risk-management practices by cash lenders and clearing banks, such as huge maturity mismatch, high leverage, and very loose collateral policies. Finally, there are no clear contingency plans for sharing losses and managing collateral in the event a large participant defaults.

In 2009, the Federal Reserve Bank of New York asked a group of senior U.S. bankers to make recommendations for addressing these concerns, and subsequently published a white paper on the issue (Federal Reserve Bank of New York 2010). The most important of these recommendations were as follows: to perform an automatic substitution of securities when a repo is renewed, in order to reduce the need for intraday credit; to improve risk-management practices of dealers and clearing banks; and to improve transparency, notably by mandating the disclosure of aggregate statistics on collateral and haircuts.

But even if these measures are implemented, they will eliminate the main issues: first, the two clearing banks are too big to fail; second, the risks taken by large banks in other markets can spill over to the repo market (and vice versa); and finally, there will still be a risk of a run on a large dealer at the slightest suspicion of its solvency.

5. The Need for a New Paradigm in the Theory of Banking

The classical model of banking (Diamond and Dybvig 1983) is inadequate for describing the activities of large modern banks. Let us briefly recall the main features of this model, which was very influential in developing our understanding of commercial banking:

- Banks transform short-term retail deposits into long-term opaque loans.
- The socially optimal degree of maturity transformation depends on the preferences of households (which determine the liquidity insurance needs of depositors) and on the technology of firms (which determines the investment needs of borrowers).
- Maturity transformation generates an intrinsic instability of banks, which calls for some form of regulation.
- Bank runs can be prevented by an adequate combination of deposit insurance, solvency regulation, and microprudential supervision.

As shown very clearly in Duffie's (2010) book (*How Big Banks Fail and What to Do about It*), this is not how big banks live and die nowadays. We need to build a different model, along the following lines:

- Dealer banks intermediate the "backbone" markets for securities and derivatives.
- They invest in marketable securities (as opposed to opaque loans) and also provide a whole bundle of services to investors (collateralized lending, asset management, brokerage services, etc.).
- Transformation is used to provide more liquidity to investors (but how much is too much?).
- The sources of fragility of these banks are different: ill-designed market infrastructures and excess transformation.

Indeed, modern runs on a bank take different routes (Duffie 2010): wholesale deposits (as in the Northern Rock case), novation (see a definition of this term below) demands by the counterparties of the bank on OTC contracts (as in the case of Bear Stearns), and flight of prime brokerage clients (as in the case of Morgan Stanley). New sources of fragility have appeared: collateral triggers after downgrades (as in the case of AIG Financial Products) and ultimately loss of clearing/settlement privileges (as in the case of Lehman Brothers).

In fact, the risk of contagion in interbank markets is largely due to two difficulties: the default externalities generated by the existing complex nexus of OTC transactions, and the legal uncertainty about loss-sharing rules in the event a large participant defaults, which provides an incentive to run at the slightest suspicion of problems.

The proposed remedies are too radical (like the Volker rule) or difficult to implement, such as the "living wills" (Herring 2010), additional regulatory requirements for firms identified as "systemic," the international harmonization of resolution procedures (Avgouleas, Goodhart, and Schoenmaker 2010), and finally the international cooperation between supervisors (which is unlikely to be effective when needed, as illustrated by Dewatripont, Rochet, and Tirole 2010).

We propose simpler (but more radical) ways to solve these difficulties:

- Adopt the central counterparty clearing (CCP) model for all "vital" market infrastructures.
- Change the regulatory perspective (platform based instead of institution based).

These two proposals are now discussed in turn.

6. Generalizing Central Counterparty Clearing

Many commentators have argued that the lack of transparency of interbank exposures on money markets and derivatives has played a major role in the propagation of the crisis. OTC transactions are typically very opaque and can be a major source of systemic risk. U.S. Secretary Geithner has fostered the development of central clearing platforms for credit derivatives. Along the same lines, Pennacchi (2009) discusses deposit-insurance-related reforms that would improve the efficiency of the financial system. The first reform he identifies is "to mitigate TBTF by reducing counterparty risk via centralized clearing (and possibly exchange-trading) of derivatives." See also Bernanke (2009): "To help alleviate counterparty credit concerns, regulators are also encouraging the development of well-regulated and prudently managed central clearing counterparties for OTC trades."

Bernanke (2009) puts forward a similar proposal for repo markets:

> The Federal Reserve and other authorities also are focusing on enhancing the resilience of the tri-party repurchase agreement (repo) market, in which the primary dealers and other major banks and broker-dealers obtain very large amounts of secured financing from money-market mutual funds and other short-term, risk-averse sources of funding. For some time, market participants have been working to develop a contingency plan for handling a loss of confidence in either of the two clearing banks that facilitate the settlement of tri-party repos. Recent experience demonstrates the need for additional measures to enhance the resilience of these markets, particularly as large borrowers have experienced acute stress. The Federal Reserve's Primary Dealer Credit Facility, launched in the wake of the Bear Stearns collapse and expanded in the aftermath of the Lehman Brothers bankruptcy, has stabilized this critical market, and market confidence has been maintained. However, this program was adopted under our emergency powers to address unusual and exigent circumstances. Therefore, more-permanent reforms are needed. For example, it may be worthwhile considering the costs and benefits of a central clearing system for this market, given the magnitude of exposures generated and the vital importance of the market to both dealers and investors.

The guiding principle of central counterparty clearing is that after two parties have agreed on a trade, the clearing platform steps into each trade by acting as counterparty to each side. This is called novation, a mechanism by which the platform essentially becomes "the buyer to every seller and the seller to every buyer." This mechanism allows the netting of multilateral (not only bilateral) exposures but also the centralization of collateral, which introduces diversification effects, especially if there is some degree of cross-pledging between different types of markets.

To reduce the risk and possible consequences of a default by a clearing member or one of its customers, CCPs have developed several risk-management procedures. The primary protection is provided by *initial margin*, a deposit which clearing members are required to place in an account with the CCP. CCPs typically also make margin calls to ensure that they remain protected over time as prices change. They usually also have access to additional default resources, such as mutual guarantee funds or insurance cover, and require clearing members to fulfill financial requirements to reduce the likelihood of default.

To protect themselves and the clearing house against client defaults, members are generally required to set a minimum level of margin for their clients according to rules set down by the clearing house. De facto, CCP failures have been extremely rare. Knott and Mills (2002) find only three cases: Paris in 1973, Kuala Lumpur in 1983, and Hong Kong in 1987.

In principle, CCPs mark to market positions daily. Thus they should be exposed only to the extent that a one-day price movement exhausts the entire margin of a clearing member. In practice, CCPs may be exposed over a longer period, as it may take time to decide whether a member should be declared in default and then to close out positions. Several studies have attempted to quantify the potential exposure of clearing houses over one or more days. Some of these models are purely statistical and pre-specify acceptable coverage levels in a purely exogenous fashion. By contrast, Fenn and Kupiec (1993) develop a model that aims at minimizing the total sum of margin, settlement costs, and the cost of settlement failure. Clearing houses need to trade off several objectives when they set their margins. Requiring high margins and good-quality collateral is costly to members. Marking positions to market and settling gains or losses, on either a daily or more frequent basis, also entails costs. To arrive at an optimal margin level, the clearing house must balance these costs against the potential losses resulting from a default of contracts.

By helping to manage counterparty risk and by providing netting services, CCPs allow market participants to economize on collateral, compared with what they would otherwise need to hold to ensure equivalent protection in bilaterally cleared markets. Regulators also often recognize the reduction in counterparty risk by allowing clearing members to hold less capital than if they were exposed directly to other market participants. Clearing members may also reduce the resources spent on monitoring individual counterparties, insofar as their actual counterparty is the CCP. Through the design of clearing members' margining and collateral requirements, CCPs reduce the probability of immediate propagation to solvent members of losses incurred by the insolvent one.

Moreover, a CCP clearly improves transparency, which explains why reforms are often resisted by those currently enjoying an information advantage (i.e., the major OTC derivatives dealers). As exemplified by the Lehman failure, when a major player in bilaterally cleared derivatives markets fails, it is not immediately apparent to the remaining market participants (who are absorbing the losses) how big the losses are and how the failed firm's counterparties are affected. The effects of this uncertainty can be devastating on market confidence, as illustrated by Bear Stearns, Lehman, and AIG. This uncertainty is mitigated by a CCP that has effective means of allocating losses and no incentive to use the information it holds for its own profits. This neutrality alleviates the information concerns of market participants. A CCP also increases operational efficiency, by centralizing the monitoring of trades and reducing potential for disputes.

CCPs have proven to be resilient even under stressed market conditions such as the ones we are facing today and have showed their ability to ensure normal market functioning in the event of failure of a major market player. A case in point is the successful unwinding of the interest rate swap positions left open following the default of Lehman Brothers.

7. A Change in Regulatory Perspective: Protecting Platforms, Not Banks

The main objective of macroprudential regulation should be to protect platforms (i.e., vital parts of financial infrastructure), not individual banks! Many central banks are given the rather vague objective of "maintaining financial stability," which gives them too much discretion and opens the door to lobbying by large institutions and political pressure. This could be limited if central banks were given a more precise mandate. The one I propose here is to guarantee the integrity of a precise list of financial markets and infrastructures that are deemed "vital": interbank (both secured and repo) markets, money markets, and some derivative markets and large-value payment systems (LVPSs). To do so, it would be useful to learn from the experience of private clearing houses, which have developed sophisticated policies for protecting themselves against the failure of their participants.

Typically, private clearing houses distinguish between their members, who have a privileged status, and ordinary participants. In counterpart to their privileged status, the clearing members are supposed to implement a set of risk-mitigation policies, such as collateral and capital requirements and bilateral credit limits. For example, members are typically required to make an upfront deposit to a default fund intended to cover losses that exceed the defaulting member's margins. I believe central banks could adopt a similar policy and condition the direct participation of financial institutions to the "vital" part of the financial infrastructure on special requirements (such as solvency and liquidity requirements) that would go beyond the standard requirements imposed on deposit-taking institutions by microprudential regulators.

In effect, my proposal would aim at replacing the notion of "systemically important institution" with that of "systemically important platform." Such platforms would only be directly accessible to a group of "officially recognized financial institutions" that would have to comply with special regulatory requirements and would be directly supervised by the central bank. The status of "officially recognized financial institution" could be revoked by the central bank if these special regulatory requirements are not satisfied. A special resolution procedure would be created for these institutions, so that the central bank has the legal powers to close it down, or at least restrict its activities before it is too late. Again, this is line with the position recently expressed by Bernanke (2009):

> The United States also needs improved tools to allow the orderly resolution of a systemically important non-bank financial firm, including a mechanism to cover the costs of the resolution. In most cases, federal bankruptcy laws provide an appropriate framework for the resolution of non-bank financial institutions. However, this framework does not sufficiently protect the public's strong interest in ensuring the orderly resolution of non-depository financial institutions when a failure would pose substantial systemic risks. Improved resolution procedures for these firms would help reduce the too-big-to-fail problem by narrowing the range of circumstances that might be expected to prompt government intervention to keep the firm operating.

These "officially recognized financial institutions" would be the equivalent of existing "systemically important institutions," which have access to special liquidity assistance facilities and possible government guarantees in case of distress. But there would be an important difference: the central bank would choose who belongs to the club and who does not! If the advantages associated with membership far exceeded the costs, the threat of revoking the status would work as an important disciplining device. OTC markets would still be active, but since they would be penalized by regulation, it is likely that they would become small and therefore not in a position to jeopardize the entire system.

Traditional prudential regulation is targeted at financial firms (institutions based). The new regulation/supervision of systemic risk would instead be targeted at infrastructures (platform based). Each systemic authority would be mandated to guarantee (separately) the safety of the small number of infrastructures (exchanges, CCPs, LVPSs) that are deemed "vital" within its jurisdiction (political decision). This would have significant advantages:

- There would be no need to reinvent the wheel for regulatory requirements: it would be enough to adopt market best practice put in place by private CCPs.
- There would be less need for "living wills": loss-sharing procedures would be specified ex ante at the level of each platform.
- There would be less need for international cooperation of supervisors: each "platform" would have to be sound independently of what is going on elsewhere.
- This would eliminate the rationale for splitting or downsizing banks (thus preserving scale and scope economies).

Note that this perspective differs from "functional supervision": several activities can be performed on the same platform, while the same activity can be performed on several platforms.

8. Conclusion

This paper puts forward a change of regulatory perspective on systemic risk and suggests a reversal of the balance of power between large banks and supervisors. Instead of letting some banks grow big and opaque enough to constitute a threat to the financial system, my proposal is to let the central bank, as the systemic risk supervisor, decide which banks are safe enough to be allowed as members of the financial "platforms" that are deemed vital for the economy: large-value payment systems, unsecured and collateralized interbank markets, and some derivative markets. The central bank would receive an explicit mandate for guaranteeing the continuity of these platforms and for regulating membership.

If the advantages associated with membership to these platforms far exceeded the costs, the threat of revoking the member status would work as an important disciplining device. OTC markets would still be active, but since they would be penalized by regulation, it is likely that they would become small and therefore not in a position to jeopardize the entire system.

References

Adrian, T., and M. Brunnermeier. 2009. "CoVaR." Staff Report No. 348, Federal Reserve Bank of New York.

Avgouleas, E., C. Goodhart, and D. Schoenmaker. 2010. "Living Wills as a Catalyst for Action." DSF Policy Paper No. 4.

Bank of Canada. 1995. "Bank of Canada Lender-of-Last-Resort Policies." *Financial System Review* 49–55.

Bernanke, B. S. 2009. "Financial Reform to Address Systemic Risk." Speech at the Council on Foreign Relations, Washington DC, March 10.

Dewatripont, M., J.-C. Rochet, and J. Tirole. 2010. *Balancing the Banks*. Princeton, NJ: Princeton University Press.

Dewatripont, M., and J. Tirole. 1994. *The Prudential Regulation of Banks*. Cambridge, MA: MIT Press.

Diamond, D., and P. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91 (3): 401–19.

Duffie, D. 2010. *How Big Banks Fail and What to Do about It*. Princeton, NJ: Princeton University Press. Forthcoming.

Federal Reserve Bank of New York. 2010. "Tri-Party Repo Infrastructure Reform." White Paper.

Fenn, G. W., and P. Kupiek. 1993. "Prudential Margining Policy in a Futures-Style Settlement System." *Journal of Futures Markets* 13 (4): 389–408.

Financial Times. 2009. "Bank Runs Left Repo Sector Exposed." September 10.

Flannery, M., and S. Sorescu. 1996. "Evidence of Bank Market Discipline in Subordinated Debenture Yields: 1983–1991." *Journal of Finance* 51 (4): 1347–77.

Goodhart, C., and M. Segoviano. 2009. "Banking Stability Measures." IMF Working Paper No. 09/04.

Herring, R. 2010. "Wind-down Plans as an Alternative to Bailouts: The Cross-Border Challenges." In *Ending Government Bailouts as We Know Them*, ed. K. Scott, G. Shultz, and J. Taylor, chapter 7. Stanford, CA: Hoover Institution Press.

Knott, R., and A. Mills. 2002. "Modelling Risk in Central Counterparty Clearing Houses: A Review." *Financial Stability Review* (Bank of England) (December): 162–74.

Pennacchi, G. 2009. "Deposit Insurance." Paper prepared for the AEI Conference on Private Markets and Public Insurance Programs, Washington, DC, January 15.

Rochet, J.-C., and J. Tirole. 1996. "Interbank Lending and Systemic Risk." *Journal of Money, Credit, and Banking* 28 (4): 733–62.

Tarashev, N., C. Borio, and K. Tsatsaronis. 2009. "The Systemic Importance of Financial Institutions." *Quarterly Review* (Bank for International Settlements) (September): 75–87.

Zhou, C. 2010. "Are Banks Too Big to Fail?" DNB Discussion Paper No. 232.

It's Broke, Let's Fix It: Rethinking Financial Regulation*

Alan S. Blinder
Princeton University

This paper considers a wide range of financial reform issues, focusing on macroprudential regulation and on the recent reforms in the United States—although the principles apply globally. It emphasizes such issues as the needs for a systemic risk regulator and an orderly resolution mechanism for systemically important financial institutions, the risks posed by proprietary trading and skewed compensation incentives, and the design of new capital and liquidity requirements for banks. Throughout, an attempt is made to relate concrete regulatory proposals to the abstract principles that should govern financial regulation.

JEL Codes: G01, G18, G28, L51.

1. Why Regulate Finance?

The whole world is now grappling with how to redesign the financial regulatory system, which is widely perceived to have failed on a massive scale and with catastrophic consequences.[1] In June 2009, the U.S. Treasury (2009a) released a lengthy white paper with a long, comprehensive list of reform proposals. Naturally, that document became the focal point of the discussion in the United States. The

*I am indebted to Olivia Hamilton for excellent research assistance and to Princeton's Center for Economic Policy Studies for financial support. For helpful suggestions, I thank Martin Baily, Jean-Pierre Danthine, William Dudley, Peter Fisher, Mark Flannery, John Gieve, Charles Goodhart, Douglas Harris, Anil Kashyap, Peter Kenen, Jean-Pierre Landau, Burton Malkiel, several members of the Pew Financial Reform Task Force, of which I am a member, and conference participants at the Federal Reserve Bank of Boston conference in October 2009 and the IJCB/Bank of Spain conference in June 2010.

[1]This is not to imply that regulatory failures were the *only* cause of the financial crisis of 2007–09. But they were certainly among the causes.

Treasury subsequently supplemented it with a sixteen-part draft bill and a variety of other documents. In December 2009, the House of Representatives passed a bill which was similar but not identical to the Treasury's proposal, and after many months in the congressional sausage grinder the Senate followed suit in July 2010. The Dodd-Frank Act is now the law of the land in the United States, though thousands of pages of detailed regulations are yet to be written.

Before you set out to do something—such as revamping the entire financial system—it is always a good idea to figure out *why* you are doing it and *what* you are trying to accomplish. According to the standard economic canon, government intervention into private business affairs normally is justified under one of the following five rubrics:

1. to create and enforce rules of the game and keep the system honest;
2. to guard against undue concentration, thus keeping markets competitive;
3. to redistribute income, e.g., through the tax-and-transfer system;
4. to correct externalities or other market failures, e.g., those due to asymmetric information;
5. to protect the interests of taxpayers, e.g., in cases in which public money is being spent or put at risk.

Income redistribution is of only marginal concern in the context of the financial crisis.[2] But the other four are all quite apposite and will be given substantial attention here. Rules were inadequately enforced or, in some cases, never made. Emergency rescue operations have increased concentration in the banking industry, and the too-big-to-fail (TBTF) doctrine may have been abused by many once-illustrious financial companies. A variety of miscreants imposed enormous costs on innocent bystanders by dragging the world's economies down. And taxpayers around the world have been forced to shoulder a variety of huge actual and potential bills. All

[2]The main issue was that many low-income people were victimized in the sub-prime lending markets.

this suggests the need for some fundamental rethinking of the rules and regulations that govern the financial system.

One other generic distinction is relevant before getting down to specifics. Economists typically draw a sharp distinction between regulation of prices, quantities, and entry, on the one hand, and regulations designed to promote health and safety and to protect consumers, on the other. Call these "price" and "safety" regulations for short. The prototypical attitude among many economists is deep skepticism about price regulations, which are normally designed to fleece someone, but open-mindedness about safety regulations. Pertinent questions must be asked on a case-by-case basis, of course. Does a regulatory proposal really advance the cause it purports to advance? Does it do so fairly and efficiently? Is a better alternative available? But if the answers are yes, yes, and no, the proposed regulation may be worth supporting. My discussion will focus on financial safety regulation. If you are looking for someone who wants to bring back Glass-Steagall or Regulation Q, you'll have to look elsewhere.

Within that general framework, I suggest the following four main reasons for (different kinds of) financial regulations, all of which play major roles in this paper:

1. *Consumer protection*: To protect customers from anti-competitive behavior (and hence from excessively high prices), from fraud, from deceptive practices, and perhaps even—though this is far more controversial—from their own foolishness and gullibility.
2. *Taxpayer protection*: To limit the costs to taxpayers of the government's safety net for financial institutions. The huge bailout costs that taxpayers in many countries are now bearing are spectacular examples. Ex ante taxpayer protection often involves guarding against or limiting moral hazard. Ex post taxpayer protection involves, inter alia, such things as least-cost resolution.
3. *Financial stability*: To protect the financial system against various sorts of systemic risks that might be triggered by contagious runs, breakdowns of the "financial plumbing," or failures of large institutions that are either too big or too interconnected with others to fail—or, rather, to fail messily.

4. *Macroeconomic stability*: To limit the adverse spillover effects of financial shocks on the real economy and/or to limit the financial propagation and magnification of shocks that originate outside the financial sector—in short, to mitigate booms and busts.

Notice that safe and sound operation of banks and other financial institutions contributes to each of these objectives. Sound institutions do not fleece their customers. Safe institutions do not hand taxpayers large bills. Safe and sound institutions contribute to, rather than undermine, financial and macroeconomic stability. It's no wonder that when you wake a bank supervisor up from a deep sleep, the first words out of his or her mouth are liable to be "safety and soundness." That attitude pervades this paper.

2. Principles of Sound Regulation

I have already enunciated and emphasized two principles of sound regulation, whether financial or not:

1. Regulation should be designed to mitigate some well-articulated problem.
2. Regulations should concentrate on "safety" issues rather than on "price" issues.

But these are not the only relevant principles. There is a third:

3. Regulation should not stifle valuable innovation.

Stated thus, as an abstract principle, no one would object. But, perhaps especially in the financial sphere, the application of this principle to concrete cases is rife with ambiguities and judgment calls. Which innovations are "valuable"? In most areas of human endeavor, progress is unidirectional: Technology gets unambiguously better, albeit at variable rates; it never deteriorates. In finance, as in academia, I am less sure. For example, simple interest rate swaps and plain-vanilla asset-backed securities (ABS) appear to be clear technological advances. So are credit default swaps (CDS) used to hedge credit risk. But I am loath to defend the social utility of CDO^2s or "naked" CDS used for gambling (or to assist in bear raids)[3]—not to mention exotic mortgages or credit cards with terms that seem designed to victimize naïve consumers. Others, however, hold different views.[4]

Next, I have already mentioned the following:

4. Regulation should be efficient.

To me, this means two main things. One is that regulations should not impose *unnecessary* costs on businesses or consumers. The necessary costs are, well, necessary. This simple principle is, of course, familiar from business applications, and the rationale for it is exactly the same. Just as firms strive to produce any given level of output at the lowest possible cost, regulators should strive to achieve their goals at the lowest possible cost—including, especially, the costs imposed on others. The second important efficiency principle is that regulations should be designed to work *with* incentives rather than to fight them. Wherever possible, we want the invisible hand to assist the visible hand—and vice versa.

But in the public domain, the efficiency principle has an important companion that is far less salient in, and far less understood by, the private sector:

5. Regulation should be, and be seen to be, fair.

This principle includes due process, of course. Good regulation follows the law and is not arbitrary or capricious. To the maximum extent possible (which will never be 100 percent), regulators' actions should be predictable and understood by the regulated. Notice also the phrase "and be seen to be." In democratic governments, where officials serve and are accountable to the public, the *appearance* of fairness is almost as important as the reality. Without it, political legitimacy is thrown into question.

[3]A CDO is a collateralized debt obligation, which pools a variety of debt instruments and tranches them. A CDO^2 is a CDO of CDOs.

[4]For example, many people argue—with some validity—that naked CDS help provide liquidity to the CDS market. But maybe this market got too liquid! In any case, these positions should always be properly collateralized.

6. The regulatory system should not leave large gaps, whereby important activities that should be regulated escape regulation.

The financial industry is teeming with highly innovative, even ingenious people. That ought to benefit society. But when their prodigious talents are turned toward escaping regulation, the commonweal is put in jeopardy. For example, in the United States, the absence of a federal mortgage regulator and *any* effective regulator of either derivatives or asset-backed securities played major roles in allowing a house-price bubble to turn into one of the biggest and most pervasive financial crashes in history.

That said:

7. Regulatory overlap should not leave firms confused or needing to satisfy one regulator at the expense of another.

Anyone who has ever been involved in financial regulation knows that some degree of overlap is inevitable. For example, within a U.S. bank holding company regulated by the Federal Reserve, the bank itself may be regulated by the Office of the Controller of the Currency (OCC) while the broker-dealer subsidiary is regulated by the Securities and Exchange Commission (SEC). It is a good principle to try to minimize, rather than maximize, such regulatory overlap. But perhaps even more important, a company should not get conflicting guidance or instructions from regulator A and regulator B. Anyone who has ever lived in the regulatory world knows that this sometimes happens, too.

There is an eighth principle, which is much in dispute. So I will offer both versions:

8a. Regulation should be *by function or instrument*; that is, the same activity should be regulated by the same regulator, regardless of the type of institution that performs it.

8b. Regulation should be *by institution*; that is, all of the potentially disparate activities of a single financial institution should be regulated by the same regulator.

Both principles are appealing. But obviously, in a world of multifunction firms, no regulatory system can satisfy both. So which version should take precedence?

I must admit to a certain ambivalence on this question, since neither pure model quite fits.[5] Strict functional regulation can founder on close inter-relationships across functions (e.g., sound mortgage underwriting standards and consumer protection) and may leave significant gaps in the regulatory system (e.g., who was supposed to regulate credit default swaps?), especially since executives of complex financial giants try to manage them as single enterprises.[6] But strict regulation by institution requires the regulator of a large, complex institution to have a tremendous range of in-house expertise, lest it find itself overwhelmed (e.g., the Office of Thrift Supervision trying to regulate AIG Financial Products). It is presumably this conundrum that led several countries to create a single, all-purpose financial regulator with authority over substantially all financial institutions and markets—e.g., the UK's Financial Services Authority (FSA). But that innovation is not noteworthy for uniform success across countries, and bank regulation in Britain, in particular, is now being returned to the Bank of England.

3. When the Tide Goes Out, You See the Rocks

The rolling financial crisis of 2007–09 revealed a number of weaknesses in the financial regulatory structure, some of which were not apparent before—though perhaps they should have been—and some of which were. There seems to be no natural way to order the items on what is by now a long list of needed regulatory reforms, and many of the items overlap in complicated ways. So I will simply proceed in a top-down fashion, starting with systemic risk and working my way down to the supervision of individual companies, and finally to

[5]Reflecting its usual lack of ambivalence, the *Financial Times* entitled an editorial (May 26, 2009) "Regulation Time: Authority Must Be Assigned by Function, Not Product." But as Saunders, Smith, and Walter (2009, p. 154), who generally favor regulation by function, note, "regulation by function is not enough in the case of LCFIs" (large, complex financial institutions). I agree.

[6]Cf. Tarullo (2009, p. 8), who says that the Gramm-Leach-Bliley Act of 1999 "elevated the concept of 'functional regulation' to the potential detriment of a more effective consolidated supervision."

consumer protection.[7] In each case, my focus will be on what failed, why, and what might be done to fix it. As I proceed down the list, I will discuss many, but certainly not all, of the recently enacted U.S. reforms.

3.1 The Need for a Systemic Risk Regulator

The term "systemic risk" connotes risks that threaten the entire (or at least a large portion of the) financial *system* and thus, by inference, threaten the entire economy. Even cataclysmic events, such as the bursting of the tech stock bubble after 2000, may not pose systemic risks if they can be confined to a single market[8]—although a big enough crash may undermine many institutions, and perhaps other markets. Similarly, the failure of a single large institution need not, but could, pose a systemic risk. That worry, for example, is the origin of the old "too big to fail" (TBTF) doctrine, its younger cousin the "too interconnected to fail" (TITF) doctrine, and shining the regulatory spotlight on systemically important financial institutions.

Between the summer of 2007 and the spring of 2009, the United States and other nations experienced a frightening cascade of financial ructions the likes of which the world had not seen since the Great Depression—when finance was far simpler. This searing experience seems to have created, among other things, a widespread agreement that we need a macroprudential or "systemic" risk regulator (really a systemic risk *supervisor*, but I'll stick to standard nomenclature) with a broad view over the entire financial landscape—an agreement reflected in the Treasury's proposal, the law enacted by the U.S. Congress, and reform efforts in many other nations. But what would such a regulator do—and how?

Presumably, its main task would be to serve as an early warning-and-prevention system, constantly on the prowl for looming risks

---

[7]While lengthy, my list is not encyclopedic. Several important issues are not taken up here—for example, accounting standards, money market mutual funds, insurance regulation, and what to do with Fannie Mae and Freddie Mac. For the most part, the Dodd-Frank Act punted on these, too.

[8]This will normally require, among other things, that it's an equity market with little debt involved.

that extend across markets and/or across different classes of institutions and that are growing large enough to have systemic consequences if troubles arise. Easier said than done. Recognizing systemic risk ex ante calls to mind former U.S. Supreme Court Justice Potter Stewart's famous test for recognizing pornography: "You know it when you see it." But the question, just as with pornography, is: Do you? Different people see things differently.

That said, the task of identifying systemic risks while they are still bubbling up may not be impossible, as long as our performance standards are appropriately modest. For example, Borio and Drehmann (2009) have developed a statistical indicator of bubbles based on the simultaneous occurrence of rapid increases in asset prices (of stocks or houses) *and* rapid growth of credit. Huang, Zhou, and Zhu (2009) propose a way to estimate the systemic risk of specific institutions by looking at their CDS spreads and stock prices. Barrell et al. (2009) predict banking crises in OECD countries with data on capital adequacy, liquidity ratios, and property prices. There are others. But all this research is new, and probably none of it is ready for prime time. That said, the ideas hold promise and merit further study.

Here is a counterfactual case study meant to illustrate what a systemic risk regulator in the United States might have seen, had one been in place in 2005–06. It might have noticed that a variety of banks *and* non-banks were granting a very large volume of dubious sub-prime mortgage loans, *and* that a correspondingly large volume of fixed-income securities (some of them quite opaque) were being built on these shaky foundations, *and* that many systemically important institutions were acquiring extremely large positions in these very loans and securities, *and* that a truly colossal volume of derivatives (e.g., CDS) were being written on these securities without much capital behind them.

That was a long sentence, but its most important words are the four uses of the conjunction "and." The essence of the systemic risk regulator's job is to look *across* markets and *across* types of businesses. The lending practices of U.S. banks were, in fact, supervised by four federal banking agencies and fifty state agencies—albeit poorly, as it turned out. However, no one had a good window on the lending activities of non-bank mortgage lenders, which accounted for most of the sub-prime mortgages, including an inordinate share of the worst ones.[9] Thus no regulator kept a watchful eye on the mortgage market *as a whole* even as the volume of subprime lending skyrocketed.[10] While many government bureaus (e.g., the SEC and the banking agencies) had some peripheral involvement with asset-backed securities, no regulator was responsible for overseeing the gigantic ABS markets—which, in many cases, buried their products in off-balance-sheet structured investment vehicles (SIVs) and conduits.[11] In addition, regulators either did not know how many of these ill-fated assets were on (and off) their institutions' balance sheets or, what would have been even worse, allowed unconscionable risk concentrations to build up anyway.[12] Finally, the U.S. government made fateful decisions in 1998 (when Commodity Futures Trading Commission [CTFC] head Brooksley Born was overruled) and again in 2000 (when the Commodity Futures Modernization Act protected derivatives from regulation) to turn a blind eye toward derivatives. This deliberate neglect left, for example, the oversight of AIG's credit default swaps business in the hands of the badly overmatched Office of Thrift Supervision (OTS).[13]

How might things have been different if a systemic risk regulator had been in place then? The residential mortgage market, plus the mortgage-backed securities (MBS), CDOs, CDS, and other instruments built on these mortgages, constituted the largest financial market in the world. So you might have thought the systemic risk regulator would have kept a watchful eye on it. If it had, it would

[9] According to Chomsisengphet and Pennington-Cross (2006, p. 38), "subprime loans were originated mostly by non-depository and monoline finance companies."

[10] According to Morgenson (2008), the value of newly issued sub-prime mortgages rose 159 percent from 2002 to 2006.

[11] This practice was, in turn, largely motivated by avoiding capital charges. More on that below.

[12] For example, the SEC's Inspector General later reported (U.S. SEC 2008) that the Commission knew Bear Stearns was dangerously overconcentrated in sub-prime securities and yet did nothing about it. Bank managements and boards failed, too. For example, the *New York Times* reported that Citigroup's CEO first learned of the company's $43 billion in mortgage-related assets in 2006. See Dash and Creswell (2008).

[13] AIG purchased a small savings bank in 1999, presumably to qualify as a thrift holding company under the Gramm-Leach-Bliley Act of 1999, and thus to be regulated by the OTS, which has now been abolished. I have heard it claimed that AIG and others fashioned these contracts as swaps, rather than as insurance policies, to avoid insurance regulation and capital requirements.

have seen what the banking agencies apparently missed: a lot of dodgy mortgages being granted by non-bank lenders with no federal regulator.[14] It might then have been natural for the systemic risk regulator to look into the solidity of the securities that were being manufactured from these mortgages. That investigation *might have* turned up the questionable AAA ratings that the rating agencies were showering on these securities, but it certainly *would have* uncovered the huge risk concentrations both on and off banks' balance sheets. And, unless it was totally incompetent, it would have been alarmed to learn that a single insurance company (AIG) was on the sell side of an inordinate share of all the CDS that had been issued—and that the company did not have nearly enough capital to back them. Hmm. That counterfactual really does suggest that history might have turned out much better.

Some people would end the role of the systemic risk regulator right there—as an investigative body and whistle blower whose job is to alert other agencies to mounting hazards. Maybe such a limited role is appropriate. But if systemic problems and weaknesses are discovered, *someone* should presumably take steps to remedy them—or at least to safeguard the rest of the system. Whom?

Under one model, the systemic risk regulator would act like the family doctor, taking a holistic view of the patient, making a general diagnosis, and then referring patients to appropriate specialists for treatment: to the SEC for securities problems, to the banking agencies for safety and soundness issues, to the ____ for potential problems with derivatives, etc. (We really do need to fill in that blank, don't we?[15])

There is a problem here, of course. If several different agencies are involved in the solution, the process will work much better if their actions are well coordinated. Anyone who has been involved in a bureaucratic turf war knows that achieving coordination is harder

---

[14] Journalists were calling attention to the dangers inherent in the rapid growth of sub-prime mortgages as early as June 2004. For example, the *New York Times*' Edmund Andrews (2004) wrote that "problems may be just over the horizon, especially in markets where housing prices have risen far faster than personal income." In that same year, Federal Reserve Governor Edward Gramlich (2004) gave a prescient speech calling attention to some of the "challenges" posed by sub-prime mortgage lending.

[15] The Dodd-Frank Act shares this authority between the SEC and the CFTC.

than it sounds. Hence one key question is: Would the systemic risk regulator turn out to be the field marshal of a well-coordinated army, or find itself herding cats? I guess the soon-to-be-established Financial Stability Oversight Council (FSOC)—consisting of (count 'em) the heads of the Treasury, the Federal Reserve, the OCC, the Federal Deposit Insurance Corporation (FDIC), the SEC, the CFTC, the Federal Housing Finance Agency (FHFA), the National Credit Union Administration, the new Consumer Financial Protection Bureau, and an independent presidential appointee "having insurance expertise"—is supposed to solve that problem. I wonder.

An alternative model would work more like a full-service HMO, where the internist/gatekeeper refers patients to in-house specialists for treatment as necessary. Establishing a comprehensive systemic risk regulator like that would mean giving it a much broader grant of authority—not only to *diagnose* problems but to *fix* them. It would also require a tremendous range of in-house financial expertise, in both depth and breadth. This is, roughly speaking, the FSA model, which failed in practice in the United Kingdom but succeeded in Canada.

But that leaves one big unresolved problem. When problems of truly systemic proportions blow up, or threaten to, a lender of last resort is likely to be crucial to the solution—at least to the immediate rescue phase thereof. And only the central bank can serve this function. So this more expansive view of the role of the systemic risk regulator points straight to the central bank. In fact, it is the main reason why I see this question *of principle* as being open and shut *in practice*. At least in the United States, if there is to be a systemic risk regulator, it must be the Federal Reserve. Like Secretary Timothy Geithner, "I do not believe there is a plausible alternative,"[16] for at least five reasons.

First, if the systemic risk regulator is to take strong actions (as I think it should), rather than just flag problems, then it really must have—among other things—lender-of-last-resort powers.

Second, systemic risk regulation is a natural first cousin (if not a sister) of monetary policy. Both are *macroprudential* functions. Every central bank sees itself as the nation's principal guardian of

[16] Secretary Geithner was quoted on a variety of financial wires on June 17, 2009.

financial stability, whether or not its explicit legal mandate says so. Since synergies between monetary policy and financial stability abound, it is unnatural, inefficient, and probably even dysfunctional to separate them.[17]

Third, and implicit in what I just said, the central banks of most countries—unlike any other agency—would not be starting from scratch in monitoring and thinking about systemic risk. Most of them already have the eyes and ears (though not enough of them) to do this job, and the broad view (though, again, not broad enough) of the entire financial landscape.[18] It must have these things in order to do monetary policy properly. So the Federal Reserve and many other central banks, unlike de novo agencies, would hit the ground running.

Fourth, the nature of the job requires an effective systemic risk regulator to be fiercely independent of politics. Probably no other agency of government has as much independence as the central bank—even where, as in the United States, that independence is more a matter of tradition than law.[19] Perhaps an equivalent degree of independence might be developed, eventually, in a systemic risk regulator other than the central bank. But doing so would certainly take time, probably a lot of it.

Fifth, I am deeply skeptical that the job can be done well by a consortium or committee of regulators, such as the above-mentioned FSOC. Creating a hydra-headed systemic risk regulator, as some have proposed, invites delays, disagreements, and turf wars. It also dilutes accountability. So the U.S. financial reform sensibly made the Federal Reserve the operating arm of the systemic risk regulation apparatus, with the FSOC serving more as a policymaking body and board of directors. That said, if the systemic risk

[17] See, for example, Peek, Rosengren, and Tootell (2009) or Blinder (2010).

[18] Part of the "not enough" in the United States is that the Gramm-Leach-Bliley Act (1999) made the Federal Reserve only the "umbrella" supervisor of financial holding companies. In practice, that meant that the Federal Reserve did not pay enough attention to what went on *inside* subsidiaries whose primary regulator was some other agency—indeed, it was not authorized to do so. See Bernanke (2009) and Tarullo (2009). The Dodd-Frank Act is supposed to end this practice for large financial holding companies.

[19] Congress can override any Federal Reserve decision by a simple majority vote. But it never does.

regulator is to have authority to shut down dangerous institutions, it probably should require consent from *someone*. Under the new U.S. system, that someone is the Secretary of the Treasury—subject to the approval of the courts. That seems like a sensible choice to me. After all, life-or-death decisions on individual companies are inherently political and may commit taxpayer money. So it is appropriate for someone with political legitimacy to make such decisions. The Secretary of the Treasury, the country's CFO, is the President's designated agent.[20]

In fact, the absence of such a political check was one major source of unease about the Federal Reserve's extensive use of section 13(3) powers during the crisis. That once-dormant but now-controversial provision of the Federal Reserve Act allows the Board, "in unusual and exigent circumstances," to "discount for any individual, partnership, or corporation, notes, drafts, and bills of exchange" as long as they are "secured to the satisfaction of the Federal Reserve bank" and the Federal Reserve has "evidence that such individual, partnership, or corporation is unable to secure adequate credit accommodations from other banking institutions." In short, in an emergency, the Federal Reserve was authorized to lend to *virtually anyone* against *virtually any collateral*.

That broad grant of authority, obviously, was neither checked nor balanced. Perhaps the Federal Reserve should have been required to seek permission ex ante from the Secretary of the Treasury, at least in cases involving individual companies,[21] and then to report ex post to Congress on why its use of section 13(3) powers was justified. The new Dodd-Frank Act requires such reporting (and auditing) and also *prohibits* section 13(3) lending to individual companies, which may prove to be a step too far.

Returning to the choice between the family doctor and HMO models, it is, of course, possible to envision hybrids. For example, even if the central bank takes on the more-limited family doctor role, it will probably be among the specialists that get called in to

[20]Once the company is designated to be shut down, the actual job of receivership is done by the FDIC.

[21]Section 13(3) cases that deal with *markets*, rather than with *companies*, sit much closer to monetary policy proper; so perhaps the Secretary's permission should not be required.

affect a cure. Alternatively, if the central bank takes on the more-powerful HMO role, the government might want to limit some of its other powers, lest the central bank become too monolithically powerful. The U.S. Treasury and Congress debated this issue at length in 2009–10, but in the end took away only the Federal Reserve's authority over consumer protection.[22]

3.2 The Need for an Orderly Resolution Mechanism for Financial Giants

The too-big-to-fail and too-interconnected-to-fail doctrines have been roundly criticized in recent years, largely because of the moral hazards they create, but also on fairness grounds and because they expose taxpayers to potentially huge liabilities. All three criticisms are valid. Living under the government's protective umbrella gives financial giants an unfair competitive advantage (e.g., access to cheaper funding), puts taxpayers on the hook, and may encourage excessive risk taking.

But let's not forget the (valid) rationale for having a TBTF doctrine in the first place: It is very risky to let a giant financial institution—which will, of course, have thousands of counterparties and probably a global reach—go bankrupt. Lehman Brothers demonstrated that fact painfully in September 2008. Within days, the entire global financial system was melting down at a frightening pace. Within weeks, officials in numerous countries were all but declaring that no other large financial firm would be allowed to fail, and backing that pledge with piles of taxpayer money.

Yet the aforementioned objections to the status quo ante are both powerful and valid. So it's no wonder that recent experience with TBTF and TITF, though "successful" (e.g., the financial meltdown was halted), seems to have left no one happy—not regulators, not legislators, and certainly not ordinary citizens. What to do? There are several options.

First, we could try to jettison the two doctrines, replacing them by "sink or swim," presumably coupled with a heavy dose of caveat

[22] As a political compromise, Congress housed the new Consumer Financial Protection Bureau at the Federal Reserve but made it independent of the Federal Reserve. If that sounds strange, it is.

emptor. In my view, this option is attractive mainly in *laissez-fairey* tales. In the real world, it is probably neither possible nor advisable. Not possible because genies are not easily put back into bottles; the precedents already set would undermine the credibility of any future proclamation of "never again." Not advisable because, as just noted, there really is a good reason to have some sort of TBTF doctrine, even though business failures are inherent in capitalism. Yes, it feels right to punish miscreants. But it doesn't feel right to see millions of innocent bystanders dragged down with them. Recent experience also suggests that market discipline was highly overrated, some might even say oxymoronic. Where was the market discipline—until the very last minute—of Bear Stearns, Lehman Brothers, AIG, and others?

Second, we could adjust our regulatory and anti-trust policies to make it difficult or impossible for any financial institution to be too big to fail in the first place, as suggested by Bank of England Governor Mervyn King (2009), among others. Frankly, I think it is probably a good idea to impart some greater anti-bigness tilt to regulatory policies once the crisis is over.[23] But our expectations should be realistic, which in this case means modest. Specifically, we are not going to create an environment in which no U.S. financial institution is TBTF or TITF—and I think the same is true in many other countries. Think about it. The United States is a huge country. Its biggest firms are extremely large in just about every industry. Why should finance be different? And modern finance is all about interconnections. Furthermore, the United States is and likely will remain the global financial leader. Globalization itself involves some minimum scale and scope that depends, inter alia, on developments in other countries. So international competitive considerations suggest that both the size and scope of America's largest financial institutions will at least approximate those of other major nations. The same holds, ipso facto, for other major countries.[24]

[23] For example, crisis-induced mergers and acquisitions have left the United States with a banking system that is too concentrated according to the Riegle-Neal Interstate Banking and Branching Efficiency Act (1994).

[24] That said, two-trillion-dollar balance sheets are probably not necessary to, and may be inimical to, business success. Here, the "too big to manage" doctrine may hold sway.

Third, we could recognize the inevitability of having some TBTF institutions, and charge them for the privilege. I think that is, in Sherlock Holmes's words, a capital idea (pun intended) to which I will return.

Fourth, we could develop a new resolution mechanism, perhaps patterned on what the FDIC now does with small banks (often *before* the bank's net worth goes negative), that would enable the authorities to wind down a systemically important financial institution (including a non-bank) in an orderly fashion—rather than just throwing it to the Chapter 11 wolves. This last idea is the one Congress adopted, after much debate. Among its presumed virtues are the following:

- The TBTF doctrine would morph into "too big to be put into Chapter 11," but not "too big to be seized and its management thrown out." That change alone should go a long way toward reducing moral hazard.
- Taxpayers would be mostly relieved of the burdens of costly bailouts. I say "mostly" because even least-cost resolution often imposes costs on taxpayers, and because the "systemic risk exception" (discussed below) explicitly recognizes that costs are justified if they mitigate systemic damage.
- Regulators would no longer have to keep large "zombie banks" (and non-banks) on life support for fear of the systemic consequences of shutting them down.

As in many aspects of financial reform, the details matter. Many of the most important issues are lawyers' questions, regarding which I have no comparative advantage. But let me make a few points about good design—beginning with this question: What do we want the new resolution authority to accomplish that present arrangements do not? I would list six objectives. To the extent possible, it should:

1. maintain the continuity of the payments and settlements systems, which might mean continuing the core operations of some firms;
2. create resolution procedures that are both well defined (though legal certainty may be unattainable) and expeditious;

3. reduce, or in the limit eliminate, contagion to other institutions;
4. minimize the threat and/or the severity of disruptive runs prior to resolution;
5. minimize costs to taxpayers; and
6. as part of objective 5, *not* make all liability holders whole (in contrast to recent bailout practice).

Notice, however, that some of these objectives conflict with others. For example, if the resolution procedure imposes (even well-defined and predictable) losses on debt holders (objective 6), as a way to minimize taxpayer burdens (objective 5), debt holders will have incentives to run whenever they smell trouble (thus contradicting objective 4 and maybe even objectives 1 and 3). Perfection is not achievable—and we should keep that in mind when appraising concrete proposals.

To me, there are two basic approaches, each with advantages and disadvantages.

The first would be to establish a special resolution authority to do for systemically important financial institutions something like what the FDIC does so well for failing (small) banks: make the resolution process orderly, predictable, and fast. This is roughly what the U.S. Congress just did. But notice at least three critical differences. First, resolving a multi-function financial giant goes way beyond what the FDIC is accustomed to doing over a weekend. The job is so different, both quantitatively and qualitatively, that closing on Friday afternoon and reopening on Monday morning is out of the question. So "fast" may be an unattainable goal. Second, such a resolution authority must keep systemic risk and financial stability in mind, rather than be narrowly legalistic like a bankruptcy court. Doing so would no doubt push decisions out of the legal realm and into the realm of political economy—which has both costs (possible political interference) and benefits (political legitimacy and attentiveness to broader economic concerns). But it is necessary, if systemic risk is to be contained. One critical focus of a special resolution authority would be to limit externalities, which bankruptcy is *not* designed to do. Third, a special resolution authority would presumably be allowed to impose something akin to prompt corrective action *before* insolvency, again in stark contrast to what a bankruptcy proceeding does.[25]

Since other countries need to grapple with this issue too, it is worth replaying parts of the U.S. debate. The Treasury's original white paper modeled its proposed resolution procedure "on the 'systemic risk exception' contained within the existing FDIC resolution regime" and reserved it "only for extraordinary times" in which an actual or impending default by a systemically important institution "would have serious adverse effects on the financial system or the economy."[26] The Secretary of the Treasury, after consulting with the President and obtaining approvals from the relevant regulators, would "generally" appoint the FDIC as the conservator or receiver with "broad powers to take action."[27]

As Congress debated the idea, two of the biggest issues were the role of the courts and whether "bailouts" should be allowed. The compromises in the House and Senate did a little of each. As mentioned earlier, if the company objects, the Secretary of the Treasury needs court approval in order to proceed with special resolution under Dodd-Frank. More important, the Act eliminates the conservatorship option, which I would have preferred to keep, mandating receivership instead. It also requires that management be dismissed and that unsecured creditors bear some losses, two features that, I think, will help reduce moral hazard.

The second approach to special resolution would be to enact a new "Chapter 16" of the bankruptcy code,[28] designed to remedy the defects of Chapter 11 in cases of systemic risk—defects that were painfully obvious when Lehman collapsed. For example, Chapter 16 would maintain the continuity of payment and settlements

[25] On these points and others, see Cohen and Goldstein (2009).

[26] These words almost precisely mimic the words in FDICIA Section 13(c)(4)(G): "serious adverse effects on economic conditions or financial stability." It is noteworthy that the systemic risk *exception* is an exception to the principle of least-cost resolution. It thus recognizes that avoiding systemic risk can justify the expenditure of public money, as noted above.

[27] All the quotations in this paragraph come from pages 77–78 of the white paper.

[28] The present U.S. bankruptcy code has fifteen chapters.

systems,[29] provide for uniform treatment across different types of businesses (e.g., banks, insurance companies, securities firms, etc.), and work relatively smoothly across national borders. As compared with the Treasury's recommended approach, Chapter 16 would presumably be less political/economic and more legalistic. It would thus engender greater legal certainty, and perhaps harsher penalties for failure, at the cost of less flexibility in dealing with "unusual and exigent circumstances." It would almost certainly be far less attentive to macroprudential aspects and externalities on other firms, perhaps excluding them altogether. And it would probably also be slower. For example, it is hard to imagine how a Chapter 16 proceeding could be brought *before* the institution failed to make payments. These last two drawbacks strike me as particularly significant. Perhaps they struck Congress the same way.

In either case, prompt resolution will be assisted by "living wills," a novel regulatory requirement first proposed by Brunnermeier et al. (2009) that Treasury also proposed in its white paper. Under Dodd-Frank, each systemically important firm would be required to report "resolution plans" to the Federal Reserve, the FDIC, and the FSOC, and to keep them up to date. Like seemingly everyone, I find this idea attractive. But I wonder how practical it would be. For example, how often would the plan have to be updated? Every time a new division was created, spun off, or amalgamated? Every time the firm's balance sheet changed in a major way? If living wills are both detailed and updated frequently, which might be necessary if they are to serve their regulatory purpose, they might impose large burdens on firms. More likely, banks will find ways to make compliance a cookie-cutter, check-the-box exercise that does not consume much time or energy.

Whichever approach is selected, there is a serious human resources concern that seems to have been ignored. Special resolutions of systemically important financial institutions will presumably be rare events, perhaps occurring only once every several decades,

[29] Chapter 11 goes some way in this direction already, by exempting repos and derivatives from the usual automatic stay in bankruptcy, that is, by allowing payments to continue to be made. While these provisions do maintain continuity, questions have been raised about why such contracts (after netting, which clearly makes sense) should be put at the head of the payments queue in case of bankruptcy.

and requiring the skills of possibly hundreds of attorneys, accountants, bank examiners, etc. How could the FDIC, or anyone else, keep such a large, highly skilled resolution force at the ready, prepared to descend en masse on a failing institution on short notice, without even knowing in advance whether the main problems would lie in banking, securities, trading, or insurance? Presumably, personnel would have to be seconded from several federal agencies. Analogously, where on the federal bench would we find a judge prepared to take on a massive Chapter 16 case, when he or she had probably never seen one in his entire judicial career? In this respect, at least, Chapter 16 seems to have one advantage: In a bankruptcy proceeding, the burden of working out the details of the financial restructuring falls on the company and its creditors, not on the court.

3.3 The Roster of Financial Regulators

America's crazy-quilt of banking and financial regulators, which has been substantially unchanged for a long time, has come in for a great deal of criticism of late—much of it well justified. I think it is safe to say that, starting from scratch, no one would ever design the current U.S. system. And, indeed, it is hard to find anything like it elsewhere in the world. The organization chart is a management consultant's nightmare, with overlaps and gaps galore. I'll divide this broad issue into parts, starting with the one that may be the most delicate politically.

3.3.1 Rearranging the Regulatory Deck Chairs

Unlike many other aspects of its proposal, where it was bold, the U.S. Treasury proposal was timid when it came to trimming the number of regulators and rationalizing their functions. And after a flirtation with major deck chair rearrangements in the Senate, Congress settled on essentially the same approach. Only one of the four current federal banking agencies will be merged out of existence (the OTS). Thus, many banks will still be able to shop around for a friendlier federal regulator; their choices will just be trimmed from four to three.[30] Furthermore, despite the similarities in their

[30]But large banks will not be able to avoid the Federal Reserve.

functions (how different are futures and options, really?), and the potential dysfunction inherent in their traditional (and sometimes bitter) rivalry, the SEC and the CFTC are not slated to be merged. A twenty-five-year-old management consultant could have done better.

Why such timidity? One reason might be the presumed (by some) virtues of competition among regulators. But I believe the main reasons are grounded in *realpolitik*. Tidying up the organizational chart by eliminating the CFTC and collapsing all four banking agencies into one makes good sense in the abstract. It's what I'd want my students to recommend in a classroom exercise. But either proposed reorganization would likely trigger fierce bureaucratic and congressional turf wars that reformers might lose. So I'm willing to forgive Treasury and the two banking committees for their decision to husband their political capital for more important, and more winnable, fights. Furthermore, tidying up the organizational chart is almost certainly *less* important than getting the regulators—all of them—to improve their performance. After all, regulatory failure on a grand scale was one major cause of the mess.

Still, since this paper is about *principles* for regulatory reform, I feel obliged to state the obvious: Having two regulators rather than six would be better, as most of the rest of the world recognizes.

3.3.2 Filling Regulatory Gaps

The unhappy history that started to unfold in the United States in the summer of 2007 highlighted at least five major regulatory gaps. First, the United States had no federal mortgage regulator. Second, we had no effective regulation of derivatives. Third, we didn't even have much knowledge about, much less regulatory authority over, the activities of hedge funds. Fourth, regulators, and perhaps even top executives, had inadequate understanding of what was going on in off-balance-sheet entities such as conduits and SIVs. Fifth, no one was effectively regulating the gigantic ABS markets. The last four of these five failings were not limited to the United States. They were and remain common across the globe.

Remember, one of the main regulatory principles mentioned earlier was that large gaps in the regulatory structure should be avoided. The status quo ante failed miserably on that criterion. And while nature may abhor a vacuum, financial market participants will rush to fill one—as they did. Banks certainly played a key role in the sub-prime lending debacle, but the problems really burgeoned out of control in America's unregulated, non-bank sector. Securitization ran wild. No regulator understood what AIG was up to in the CDS market. To this day, I think no one really understands the role hedge funds played in the drama. And a vast profusion of conduits and SIVs were used to hide assets from prying regulatory eyes. The new U.S. law proposes to fill all five gaps. Let's hope it succeeds.

Mortgages. All residential mortgages (and many other consumer products) will fall under the jurisdiction of the new Consumer Financial Protection Bureau (CFPB). This agency could, among other things, mandate that mortgage lenders (whether banks or not) provide simple, clear disclosures (a one-pager would be nice). An even better idea—to offer consumers "plain vanilla" mortgage options whose standardized terms would be intelligible and facilitate comparison shopping—was dropped in the House. I would have liked to have had that. I also preferred a true federal mortgage regulator with authority over both banks and non-banks.[31] But the CFPB is clearly a step in the right direction. Similarly, to protect consumers, I had hoped that all mortgage lenders would be forced to abide by a suitability standard, like the one that now applies to financial advisers. Instead, the new law imposes only a "duty of care" on all mortgage originators and bans "unfair, deceptive, or abusive" lending practices.[32]

Since a free-standing consumer protection agency proved to be a major bone of contention between liberals and conservatives, the Act placed the new agency—physically and budgetarily—at the Federal Reserve Board, even though its head is to be an independent presidential appointee who does *not* report to the Chairman of the Federal Reserve. I have no idea how this will work in practice.

Derivatives. While the regulation of derivatives is fraught with peril, it is not hard to improve upon what the world has now—which is practically nothing. I have argued for years that the most

[31]Brunnermeier et al. (2009) seem also to take this view.

[32]In tort law, a *duty of care* requires a party to an agreement to adhere to a standard of reasonable care in exercising its responsibilities. A suitability standard goes further. It makes it a violation (of ethics or law) to sell someone a financial product that requires greater sophistication or ability to absorb losses than he or she possesses.

important step the world's governments could take would be to push as much derivatives trading as possible onto organized exchanges—whether by cajoling, regulatory incentives, or regulatory coercion. Cajoling might mean, for example, letting banks know that their regulators view over-the-counter (OTC) derivatives as far riskier than exchange-traded derivatives. Arched eyebrows often work. Incentives might mean, for example, higher capital charges on OTC derivatives than on exchange-traded derivatives. Under such a regime, regulatory arbitrage might actually enhance rather than undermine safety and soundness. Coercion might mean, for example, banning certain types of institutions (e.g., insured depositories) from trading in OTC derivatives, except perhaps for unambiguous hedging and/or market-making purposes.[33] I have long advocated the middle approach: providing strong regulatory incentives to move trading onto organized exchanges.

Once again, purity should not be the goal. There are cases in which customized derivatives really are appropriate, and such arrangements will never be sufficiently standardized to trade on exchanges. Furthermore, any financial contract whose payoffs depend on any other financial price can be considered a "derivative," and the government certainly does not want to prevent two companies from making a one-off financial deal that is in both parties' interests—as long as the deal poses no systemic risk. Moreover, genuinely customized derivative contracts, by their very nature, should not be of systemic importance. If and as they outgrow that status, they should become standardized and forced onto exchanges where, among other things, sufficient capital and collateral would be required.

The regulation of derivatives proved to be a major battle ground in the U.S. financial reform process. The Treasury white paper (p. 48) proposed to subject OTC derivatives to a "robust regime" of regulation that included "conservative capital requirements," margins, reporting requirements, and "business conduct standards." The subsequent draft legislation allowed for either clearinghouses or exchanges to accomplish that. What's the difference? An exchange will presumably clear its trades through a clearinghouse. So either mechanism offers the advantages of central clearing, multilateral

[33] This would be a piece of the so-called Volcker Rule.

netting, greater transparency, and the interposition of a third (and very creditworthy) counterparty between buyer and seller—which would substantially reduce (and in most cases eliminate) counterparty risk. The main additional step taken by an exchange, as opposed to just a clearinghouse, seems to be the public reporting of price and volume data. Since such reporting seems desirable both for public policy and market efficiency reasons, I have always favored exchanges. But the major gains from a systemic-risk standpoint come from moving from the status quo to central clearinghouses. The further step from clearinghouses to exchanges would be icing on the cake.

The bill that emerged from Congress favors standardization and exchange trading. Its main problem is the large number of exceptions it allows. How this will work in practice remains to be seen.

Hedge Funds. The supervision and regulation of hedge funds is the dark continent of financial regulation; we have virtually no experience with it. Complexities abound in this domain for many reasons, not the least of which is that the term "hedge fund" connotes a form of legal organization, not a type of financial activity. The truth is that hedge funds do almost everything. Some are risky, others are safe. Some use a great deal of leverage, others use little. Some specialize in large directional bets, others shun them. And so on. Furthermore, aside from some presumed involvement in bear raids, hedge funds do not appear to have played major roles in the crisis.

Recognizing that maintaining the secrecy of their investment strategies is often a key component of a hedge fund's business model, I have long thought that most of the answer is to require "regulatory transparency," meaning that funds should keep regulators—but not the public—informed about their portfolios and exposures, almost in real time. That seemed, more or less, to be the Treasury's attitude as well.[34] The white paper (p. 12) proposed that all funds above "some modest [size] threshold" register with the SEC and disclose information that is "sufficient to assess whether" it "poses a threat to financial stability." It goes on to explain that regulators need to know "how such funds are changing over time and whether any such funds have become so large, leveraged, or interconnected that

[34] And also that of Calomiris (2009, p. 87) and the Committee on Capital Markets Reform (2009, p. 93).

they require regulation for financial stability purposes" (p. 37)—e.g., by being declared systemically important. While Treasury originally proposed to treat private equity and venture capital funds in essentially the same way, Congress exempted them on the grounds that it seems unlikely that a private equity or venture fund would ever grow large enough to pose systemic risks.

However, Brunnermeier et al. (2009) raise a point that, while reinforcing the need for regulatory transparency, suggests that it may not be enough. Suppose many hedge funds, each of them too small to pose systemic risk, exhibit strong herding behavior; that is, they all (or most of them) move at once and in the same direction. Then the "herd" might pose systemic risks that no single hedge fund, acting alone, would. In such a case, which sounds all too plausible, the systemic risk regulator will certainly need the transparency just discussed. But it may also need some authority to alter behavior in unusual circumstances for macroprudential reasons. Exactly how is a tricky question, to say the least.

Off-Balance-Sheet Entities. Thinly capitalized conduits, SIVs, and other off-balance-sheet entities were among the most viral transmission mechanisms for the crisis. Small losses on, e.g., ABS were sufficient to render them insolvent. The slightest "runs," which generally took the form of inability to roll over commercial paper, were sufficient to render them illiquid. The size and prevalence of SIVs and conduits were also among the biggest surprises to regulators, to the investing public, and sometimes, it seemed, even to the top executives of the parent companies.

While there may be legitimate business reasons for taking certain activities off balance sheet, many such arrangements were designed primarily to avoid or minimize regulatory capital charges. That should always have been considered a bad reason. The Treasury and Congressional proposals require financial institutions to report off-balance-sheet exposures and require regulators to take off-balance-sheet entities into account in computing capital requirements. I'm not sure why they didn't take the next step and require that such entities be consolidated onto the parents' balance sheets.[35]

[35]Sufficient capital is not the only issue here; sufficient liquidity is another. When the SIVs became illiquid, they called upon the "liquidity puts" their parents had given them.

Asset-Backed Securities. Problems with asset-backed securities were, of course, pervasive during this financial crisis, and the ABS markets are still pretty moribund. In addition to a general lack of regulation and supervision, one alleged problem was that securitization enabled—perhaps even encouraged—a dangerous game of "hot potato." Mortgage originators quickly passed credit risks along to securitizers who, after a short delay, passed them on to investors.[36] Neither the originators nor the securitizers, it seemed, paid enough attention to the quality of the credits they were passing along. After all, underwriting standards really do not matter if you sell the mortgage before the first payment comes due. Thus, *homo economicus* notwithstanding, the process seemed to hide risks rather than extinguish them.

This perceived weakness of securitization has spawned suggestions to require originators and/or securitizers to retain a fraction of their handiwork in their own portfolios. As one example, the Treasury white paper called upon bank regulators to require *originators* of securitized loans to hold on to 5 percent of them—adding, however, that the regulators might prefer "to apply the requirements to securitization sponsors rather than loan originators" (p. 44). Indeed, the bill that passed Congress moved the 5 percent minimum requirement from the originators to the *securitizers*; it also authorized regulators to set it higher.

I applaud the "skin in the game" approach, but would have liked it strengthened in two ways. First, the 5 percent requirement seems rather puny—less than the sales tax, so to speak. Second, I would like to motivate *both* originators *and* securitizers to conduct more serious due diligence. Mandating 5 percent retention by *each* of these two parties would wound both birds with a single stone.

But "wound" is not "kill." One thing we witnessed as this crisis unfolded was that many securitizers retained very large CDO positions on their own books—often concentrated in the lowest-rated tranches. Thus, they had more than just "skin" in the game—they had muscle and bone. Such flagrant disregard of elementary risk-management standards was shocking. But most of it was

[36]Sometimes there were even more links in the chain.

voluntary,[37] which suggests that "skin in the game" requirements are no panacea.

3.3.3 The Glass-Steagall Issue

A number of vocal critics have argued that tearing down the walls that formerly separated commercial banking, investment banking, and insurance (under the Glass-Steagall Act) was among the fundamental causes of the financial crisis in the United States.[38] I hear this claim almost every day.

First, a personal confession: As Vice Chairman of the Federal Reserve Board in the mid-1990s, I was always lukewarm toward repeal of Glass-Steagall. But my main reservation was amorphous and improvable: the worry that Wall Street's "burn 'em and churn 'em" culture might infect commercial banking. So I reluctantly endorsed the Board's pro-repeal position for two main reasons: because markets had torn huge holes in the alleged walls anyway, and because the government should not ban activities unless it has good reasons to do so. That said, I think the Gramm-Leach-Bliley (GLB) Act of 1999 has gotten a bad rap in this episode.

I often pose the following question to critics who claim that repealing Glass-Steagall was a major cause of the financial crisis: What disasters would have been averted if Glass-Steagall was still on the books? I've yet to hear a good answer.[39] While mortgage underwriting standards were disgraceful, they were promulgated by banks and mortgage finance companies and did not rely on any new GLB powers. The dodgy MBS were put together and marketed mainly by free-standing investment banks, not by newly created banking-securities conglomerates. All five of the giant investment banks (Goldman Sachs, Merrill Lynch, Morgan Stanley, Lehman

[37] Tett (2009) offers a good discussion of this practice—and of objections to it that went unheeded.

[38] For one example, see Kuttner (2007). Paul Volcker also aligned himself with this view.

[39] The closest is that Citicorp would not have been allowed to merge with Travelers, which had previously acquired both Smith Barney and Salomon Brothers. But Citi could probably have done everything it did in the mortgage markets anyway.

Brothers, and Bear Stearns) got themselves into severe trouble without help from banking subsidiaries,[40] and their problems certainly did not stem from conventional investment banking activities—the historic target of Glass-Steagall. Similarly, Wachovia and Washington Mutual died (and Bank of America and Citigroup nearly did) of banking diseases, not from entanglements with or losses imposed on them by related investment banks. In short, I don't see how this crisis would have been any milder if GLB had never passed.[41] Furthermore, most of the rest of the world never had GLB.

3.4 Dysfunctional Compensation Incentives

I come now to a dangerously jagged rock that was plainly visible for many years before the tide went out, but was nonetheless ignored: the perverse incentives for excessive risk taking that are built into many compensation schemes.

Consider the prototypical compensation plan for a trader, whether employed by a commercial bank, an investment bank, or a hedge fund. If, through skill or luck, he earns copious profits for his firm, he will receive a non-trivial share of that profit in his year-end bonus. In the halcyon days, it was not uncommon for such bonuses to run into multiple millions of dollars; some traders amassed dynastic wealth. If, on the other hand, a trader loses tens of millions of dollars of the firm's money, his bonus (which is the vast majority of his compensation) will be lost. Other than that, he will not share in the firm's pain. In particular, profits in one year are often *not* clawed back when there are losses in subsequent years.[42] He may not even lose his job. A similar problem inheres in the compensation plans of many other income generators—such as mortgage salespeople, who get paid for originating loans and not docked when loans go bad.

[40] Merrill Lynch had a big bank, but it was not the source of Merrill's problems.

[41] One possibility is that AIG Financial Products would have been regulated better by someone other than the OTS. But, given the more-or-less blanket exemption of derivatives from regulation by the Commodity Futures Modernization Act, this seems unlikely.

[42] Even worse, traders sometimes collect profits on a mark-to-market basis on trades that are still open. That said, some companies *do* claw back profits when there are subsequent losses.

Pay plans that are structured in such a "heads I win, tails I don't lose" way create powerful incentives for traders to go for broke gambling with OPM ("other people's money"), for salespeople to originate too many dodgy mortgages, and so on. Add to that the typical demographic profile of a trader—young, smart, brash, single, and risk loving—and you have a dangerous brew.

Amazingly, financial executives were aware of these perverse incentives for decades and did nothing to correct them.[43] Why? One brilliant and famous hedge fund executive explained to me about fifteen years ago that it was because everyone else paid their traders the same way; any firm that deviated from the industry norm would expose its top talent to competitive raids. That answer is a variant of what should be called forevermore *The Chuck Prince Principle*: "As long as the music is playing, you've got to get up and dance."[44] It seemed a terrible answer to me then, as it does now. Besides, a company can offer its traders a compensation plan with the same *average* pay as its competitors but quite different incentives for risk taking.

A second possibility is that senior executives face analogously skewed incentives themselves and therefore *want* their traders to go for broke. Just like the traders and sales personnel who work for them, top executives of corporate financial institutions derive the lion's share of their compensation from bonuses that are normally linked to the firm's *annual* profits. They, too, get rich in the fat years and pass the losses on to shareholders in the lean years. So they, too, rationally want to go for broke. In fact, the situation may be even worse for CEOs, for they typically are blessed with "golden parachutes" that pay off handsomely if their contract is terminated. So they may not even lose when the coin comes up tails.

Notice that I used the adjective "corporate" to modify the noun "financial institutions" in the preceding paragraph. Outside the corporate sector—e.g., in the hedge fund and private equity worlds—the risk-taking incentives for the CEO and other top executives (but not

[43] When Raghuram Rajan (2005) made this point at the Federal Reserve's August 2005 Jackson Hole conference, he encountered such vociferous criticism that I rose to defend him "against the unremitting attack he is getting" (p. 394).

[44] Quoted in the *Financial Times*, July 10, 2007. Note the date. The music stopped just one month later.

for traders) are quite different. While the general partner will, of course, pass on most of the firm's losses to the limited partners, he will be stuck with a pro rata share himself. Indeed, it is typical for the heads of hedge funds to have much of their personal net worth tied up in the funds. This fact transforms OPM into what I (Blinder 2009) once called MOM ("my own money"), which people have a way of treating with greater respect. I don't think it was a coincidence that the big Wall Street firms became more adventurous when they switched from being partnerships to being corporations.[45]

If this analysis of compensation incentives is anywhere near correct, it points toward several possible remedies. One, which I will explore later, might be to move most trading activities out of the corporate sector, and especially out of the systemically important firms that potentially share their losses with taxpayers. This would be a variant of the "Volcker Rule."

Another might be to regulate compensation practices, e.g., by treating them as an integral part of the firm's risk-management system. The U.S. Treasury advocated this idea in its June 2009 white paper (p. 11): "Federal regulators should issue standards and guidelines to better align executive compensation practices of financial firms with long-term shareholder value and to prevent compensation practices from providing incentives that could threaten the safety and soundness of supervised institutions." In October 2009, the Federal Reserve did precisely that via some proposed supervisory guidance that it posted for comment on the Federal Register. After receiving extensive comments, the Federal Reserve issued its final guidance in June 2010. I see this as a potentially important step forward, but only if the guidance has teeth. We'll see.[46]

The Federal Reserve's guidance gave particular attention, by way of example, to clawbacks, to risk-adjusting incentive pay, and to reducing the sensitivity of pay to short-term performance. The Squam Lake Group added a nice twist to the clawback—or, better yet, holdback—idea: that a portion of the deferred compensation be

[45]That said, the leaders of some of the worst-offending *corporate* firms, such as Bear Stearns, Lehman Brothers, and AIG, had much of their personal net worth tied up in company stock. Alas, there are no easy answers here.

[46]Media reports in June 2010 suggested that banks were not paying much attention to the guidance.

forfeited if the firm becomes the beneficiary of any "extraordinary government assistance."[47]

A third option might be to enact (or voluntarily adopt) "say on pay" provisions, which give shareholders a voice and a vote on CEO pay—which was also enacted as part of Dodd-Frank.

While I support both "say on pay" and the Federal Reserve's regulatory initiative, I am a bit skeptical that government can legislate compensation practices effectively. State-imposed rules will be avoided, evaded, and twisted into knots by private-sector participants who have vested interests in doing so, and who are much deeper in the weeds. The job of fixing skewed compensation practices must presumably fall mainly on the (heretofore narrow) shoulders of corporate boards of directors and, more specifically, on their risk and compensation committees.[48] In saying that, I realize that relying on better board performance is standing on a thin reed—which, of course, raises a far bigger issue that is beyond the scope of this paper: How do we make corporate boards take their responsibilities to shareholders (as opposed to management) more seriously? One possibility might be to pay board members 100 percent in restricted stock. That, in turn, might help directors see the virtues of also compensating traders and CEOs that way.[49]

3.5 Reforming Regulatory Capital Standards[50]

I have heard it said that only three things matter in bank regulation: capital, capital, and capital. While I don't agree, the capital cushion is extremely important. At the conceptual level, our current bank capital standards—whether under Basel I, Basel II, or whatever comes next—have come in for two sorts of well-deserved criticisms: (i) the minimum capital they prescribe is inadequate, and (ii) capital requirements are procyclical.[51] Let me take up each in turn.

[47] French et al. (2010), p. 81.

[48] The Federal Reserve's guidance places responsibility squarely there, even recommending that banks that lack a compensation committee establish one—with a majority of non-executive directors. I agree.

[49] I owe this suggestion to Burt Malkiel.

[50] This paper was written long before the Basel III agreement in September 2010.

[51] In addition, there are many practical criticisms—such as that capital requirements are too easily avoided by using off-balance-sheet entities such as SIVs.

3.5.1 How Much Capital

Were banks holding enough capital when the crisis hit? The question seems to answer itself. Indeed, the U.S. Treasury declared that "higher capital requirements for banking firms are absolutely essential."[52] That is probably true. But I do not believe the worst leverage problems derived from the conventional Basel I capital-adequacy standards (4 percent tier 1 capital, 8 percent tier 1 and 2, etc.). Yes, I think 4 percent was too little. But the bigger leverage problems arose from (i) investment banks that operated (under a different regulatory regime) with thirty to forty times leverage, (ii) gimmicks such as thinly capitalized SIVs and conduits that (legally) avoided capital requirements, making actual leverage far greater than putative leverage, and (iii) derivatives that manufactured synthetic leverage that was vastly higher than literal leverage (the ratio of assets to equity).

That said, Basel II, which was never officially adopted in the United States, seemed to take several major steps backward. First, it generally reduced capital requirements relative to Basel I, which was almost certainly a mistake—especially for systemically important financial institutions. Second, its "advanced internal ratings based approach" placed far too much weight on banks' own risk models—which were found wanting, to put it charitably. Third, well-known quantitative tests by regulators a few years ago discovered that different banks, each using their own internal models, calculated alarmingly *different* capital charges for the *same* portfolio of assets. Fourth, the Basel II "standardized approach" relied far too heavily on ratings assigned by the credit rating agencies, which performed miserably.[53] (More on the rating agencies shortly.)

I am not a Neanderthal. I realize that the simple risk-buckets of Basel I were crude, misclassified many assets, and set up perverse incentives for regulatory arbitrage. It was not for naught that the world's bank supervisors labored for a decade trying to improve upon Basel I. Unfortunately, it is not clear that they succeeded. If you believe that Basel II is deeply flawed, as I do, two huge questions arise: First, what capital standards would be better? Second,

---

[52] U.S. Treasury (2009b, p. 1).

[53] Among the many places where these and other issues about Basel II are discussed, see Tarullo (2008).

given that leveling the international playing field has always been a central (and reasonable) objective of the Basel Accords, and that Europe and Japan have already adopted Basel II, what should the United States do now?

I prefer to leave such vexing questions to people more deeply knowledgeable on these matters than I. But here is one quick answer, at least as a stop-gap: Instead of letting banks use their own internal risk models to assess capital adequacy, create an admittedly imperfect standardized model that every major bank in the world must use for regulatory purposes. That's not a perfect solution. But it would at least eliminate the four problems listed above while regulators try to agree on Basel III. According to several insiders, good progress toward Basel III is being made. But the target for the Basel Committee is the end of this year, with adoption by member countries presumably coming long after that. How long can individual countries afford to wait?

The stop-gap approach I have just sketched is *not* favored by the U.S. Treasury, which last year issued a statement of "high-level principles" that should govern any new Basel-type regulatory framework.[54] These principles (in my phrasing, not Treasury's words, except where quoted) are:

1. Capital requirements should be higher.
2. Capital requirements should be even higher for systemically important firms.
3. Most regulatory capital should be common equity.
4. Capital requirements should take macroprudential considerations into account.
5. The "relative risk weights [must] be appropriately calibrated" and also "reflect the systemic importance" of the various exposures.
6. Procyclicality should be reduced.
7. An overall leverage constraint should be introduced, *in addition to* the risk-based capital rules.
8. A liquidity standard should be added to the capital standard.
9. The regulatory regime must be extended to non-banks that potentially pose systemic risks.

[54]U.S. Treasury (2009b). The quoted phrase comes from page 2 of that document.

I have just dealt with principle 1, and I will take up principles 2, 6, and 8 below. But first a few brief comments on some of the others.

Principles 4 and 5 on Treasury's list seem sensible, important, and devilishly difficult to implement, especially if international agreement is required. That said, one important sub-point under principle 5 is that banks should hold more capital against trading positions—another capital idea. The overall leverage constraint (principle 7) is also appealing, but far trickier than it sounds. Treasury's draft (p. 10) correctly notes *both* that "a simple leverage constraint would make the regulatory system more robust" (which is good) *and* that it "is a blunt instrument that... can create its own set of regulatory arbitrage opportunities and perverse incentive structures" (which is bad). The latter, of course, is why Basel I and II applied risk weights in the first place.

What about special, higher capital requirements on systemically important institutions that have the potential to impose large costs on taxpayers (Treasury's principle 2)? It seems axiomatic to me that, because TBTF status gives those institutions a competitive advantage in the capital markets, they should be required to pay for the privilege. But how? One natural answer is to impose higher capital charges on them. Another, discussed below, is to force them to hold some sort of contingent capital that would become real capital during a crisis. A third is to assess higher FDIC insurance premiums on TBTF institutions with insured deposits. It might be appropriate to do all three.

Notice, however, that imposing any such extra charges on systemically important institutions requires that regulators name them and thus "officially" confer TBTF status on them—as opposed to the current regime, which is sometimes characterized as constructive ambiguity. (Try guessing: Was Bear Stearns TBTF in February 2008? Was Lehman TBTF in August 2008?) Because some people prefer maintaining constructive ambiguity, they oppose naming the TBTF institutions.[55] I disagree for at least four reasons. First, except around the edges, people will always know which

[55] One way out of this dilemma, advocated by the Pew Task Force (2009, pp. 9–10), is to make capital and liquidity requirements higher for larger institutions, either via brackets or by a continuous formula. Then there is no need to designate TBTF institutions.

companies are TBTF anyway. Second, surprises in this regard (e.g., letting Lehman fail) can have catastrophic consequences. Third, the government cannot collect extra “fees” from TBTF institutions—which I consider crucial—unless it names them. And fourth, TBTF institutions should be subjected to a tougher supervisor regime.

3.5.2 Procyclicality

The second main accusation that has been leveled against the existing capital standards is that they are procyclical.[56] They aggravate credit cycles by allowing too little provisioning for losses when times are good (and loan-loss experience is favorable) and then stiffening requirements when times turn bad (and loan losses mount)—thus forcing banks to set aside more loan-loss reserves and scramble for more capital exactly when capital is hardest to find. The Treasury white paper explicitly recognized this shortcoming and urged the Basel Committee to address it, which the Committee is now doing. But it proposed nothing concrete for the United States.

Maybe it should have. In the United States, the tendency toward procyclicality in loan-loss provisioning is exacerbated by the practices of both the SEC and the Internal Revenue Service (IRS). The SEC requires banks, like all companies, to report GAAP earnings as a guide to investors—which is certainly a legitimate part of its job. However, one unfortunate side effect is that the Commission sometimes accuses banks of using loss provisions to smooth earnings by *over*-provisioning in good times and *under*-provisioning in bad times. That, of course, is precisely what we want in order to reduce procyclicality. The IRS, for its part, seeks to collect the taxes that companies owe according to the tax code. Again, that is their job. But in doing so, the tax agency has been known to prevent banks from setting aside “excessive” loan-loss reserves that would reduce their taxable income.

What to do? It’s debatable. But in my view, the legitimate concerns of the SEC and the IRS must be subordinated to the needs of macroprudential supervision in this case. However, the SEC and

[56] Among other studies of this issue, see Bikker and Metzemakers (2005), Bouvatier and Lepetit (2008), and Repullo and Suarez (2009). The later takes specific aim at Basel II.

the IRS probably see things differently. In any case, effectuating such changes would probably require changes in law—not to mention regulatory turf wars.

While removing SEC and IRS impediments to prudent provisioning would help, we must still address the *inherent* procyclicality problem. After all, loan losses *really are* smaller in booms and larger in busts. Spain seems to have handled this issue better than most countries with its system of "dynamic" or "statistical" provisioning. The name derives from that fact that provisions for loan losses are increased when actual losses run *below* statistically expected losses, e.g., in boom periods—and are analogously reduced in busts. The Spanish system is designed to be neutral over the business cycle but to stock up on loss provisions during the upswing so as to be ready for the downswing. The Basel Committee is currently considering whether "Basel III" should have this feature and, if so, how.

I myself am attracted to a particular idea for "contingent capital" suggested by the Squam Lake Working Group on Financial Regulation, an ad hoc panel of academic experts.[57] Their idea, which derives from Mark Flannery's (2005) clever earlier proposal for "reverse convertible debentures," is to *require* certain banks to issue a novel type of convertible bond.[58] Conventional convertible debt gets exchanged for equity at the option of the bondholder; and because this option has value, convertible debt bears lower interest rates than ordinary debt. The proposed new form of convertible debt would give the option to the regulators instead and/or have an automatic trigger.

Under the proposal, regulators would have the power, by declaring a systemic crisis, to *force* holders of these special convertibles (but not holders of other debt instruments) to convert to equity against their will—thus giving banks more equity capital (and less debt) just when they need it most.[59] Naturally, the existence of such an option would *detract* from the value of the convertible bond and therefore make its interest rate *higher* than that on ordinary

[57]See French et al. (2010, ch. 7).

[58]See also Rochet (2008) and Flannery (2009).

[59]The actual Squam Lake proposal has both a systemic (macro) trigger and a bank-specific (micro) trigger. It has been suggested that investors might not like the idea that regulators can trigger the debt-to-equity conversion.

debt—unless the debt-to-equity conversion was priced below market.[60] Indeed, if the requirement is limited to TBTF institutions, as seems appropriate, a higher interest rate on a fraction of their debt would constitute a natural part of the penalty cost for being TBTF. Furthermore, the spread on this new type of debt over regular debt could become a useful market indicator of the likelihood of a systemic crisis.

As in many cases, one key question is price—specifically, how large an interest rate premium would investors demand to cover the risk that their bonds could be converted into equity at adverse moments? If this premium proved to be very large, these new convertibles would be a very expensive form of capital that banks might shun, preferring ordinary equity instead. On the other hand, perhaps after an adjustment period, the premium might prove to be modest. Only experience will tell.

3.5.3 The Rating Agencies

The discussion of Basel II's shortcomings brings up another problem: what to do about the rating agencies—which were supposed to be among the built-in safeguards of our financial system but failed us badly.

The Treasury white paper proposed to solve, or perhaps I should say to mitigate, this problem with tougher regulatory oversight: "The SEC should continue its efforts to strengthen the regulation of credit rating agencies, including measures to promote robust policies and procedures that manage and disclose conflicts of interest, differentiate between structured and other products, and otherwise strengthen the integrity of the ratings process" (p. 14). I agree, and as a matter of fact, the rating agencies have been working hard on these "integrity" issues right now.[61]

But many observers think the fundamental problem lies deeper: with the issuer-pays model. As long as rating agencies are for-profit companies, paid by the issuers of the securities they rate, the agencies will have a natural tendency to try to please their

[60] I owe this last point to Mark Flannery.

[61] I happen to have personal knowledge of the strenuous efforts being made at one of the rating agencies to erect Chinese walls, to empower an ombudsman, etc.

customers—just as any business does. Unfortunately, the most obvious alternative, switching to an investors-pay model, is infeasible except in markets with very few investors. Otherwise, information flows too readily, and everyone wants to free ride. What to do? The way out of this dilemma, it seems to me, is to arrange for some sort of third-party payment.[62] The government (e.g., the SEC) or an organized exchange or clearinghouse seem to be the natural alternative payers. In either case, they could raise the necessary funds by levying a user fee on all issuers.

3.6 Liquidity Regulation

Earlier, I expressed skepticism about the notion that capital is the only thing that matters for safety and soundness. One reason is that banks and other financial institutions can also founder on insufficient *liquidity*. Indeed, this crisis demonstrated many times over that there is no bright line between illiquidity and insolvency in a maelstrom. It is at least arguable, for example, that institutions such as Bear Stearns, Lehman Brothers, Merrill Lynch, and others crumbled, or nearly did, from lack of liquidity rather than from lack of assets—just as in an old-fashioned bank run.[63] The point is that, where there is a severe maturity mismatch between assets and liabilities, a liquidity crisis can morph quickly into a solvency crisis. For example, firms that cannot roll over their short-term debt may be forced to sell illiquid assets at distressed prices.

The business models of what were once the Big Five securities firms in the United States—Merrill Lynch, Goldman Sachs, Morgan Stanley, Lehman Brothers, and Bear Stearns—were based on extreme leverage, as has already been mentioned. But they were also based on incredible reliance on extremely short-term financing, e.g., from the overnight repo market.[64] These liabilities were massive in volume and could run in a day. In a somewhat different way, many large bank holding companies (e.g., Bank of America and Citigroup) also put themselves in perilous liquidity positions by providing explicit or implicit liquidity guarantees to their SIVs

[62] This idea was suggested by Rochet (2008).
[63] See, among many others, Gorton and Metrick (2009).
[64] Among many sources that could be cited, see Morris and Shin (2008).

and conduits. When the crisis hit and the SIVs could no longer roll over their commercial paper, they turned to their parent banks for massive liquidity support.

Thus one regulatory lesson learned in this crisis is that minimum capital requirements need to be supplemented by minimum *liquidity* requirements, and the Basel Committee has taken this lesson to heart. But it is not an easy task.

Since liquidity issues result largely from maturity mismatch between the asset and liability sides of a firm's balance sheet, there are two basic approaches. Regulators could insist on, e.g., less reliance on very short-term non-deposit funding,[65] or they could constrain the composition of assets, insisting on more liquid ones.[66] In either case, problems of assigning the right "liquidity weights" arise. This is perfectly analogous to the Basel II problem of finding the right capital weights, and may be just as difficult.

To illustrate how hard it may be to develop liquidity standards, consider the schizophrenia evidenced when the U.S. Treasury tried to elaborate on its liquidity ideas. Its "principles" (2009b, pp. 11–12) state that "the liquidity regime should be independent from the regulatory capital regime" even though "capital regulation and liquidity regulation are highly complementary" so that there may be merit in "making regulatory capital requirements a function of the liquidity risk." Furthermore, liquidity regulations should be "strict but flexible." Got that?

3.7 Risk Management

The "M" in U.S. regulators' CAMELS ratings for banks stands for *management*. One essential aspect of bank management that failed badly was *risk* management—and not just at commercial banks. Some of the most elementary precepts of sound risk management, such as limiting asset concentrations, were violated massively and repeatedly by allegedly smart financial institutions. Incompetence is one thing; and it can be mitigated, though never eliminated, by

[65] I say "non-deposit" because even demand deposits are very sticky; they don't "run."

[66] See again Morris and Shin (2008).

more rigorous supervision. But I think there also was—and still is—a structural flaw.

Think of the typical position of a risk manager within the hierarchy of a large financial company that is making handsome trading profits while the good times roll.[67] She becomes worried that a certain business line is taking excessive risks with the firm's capital. The line manager disagrees; after all, his traders are making millions for the bank's bottom line. So they take their dispute to the CEO, where the argument may go something like this:

RISK MANAGER: Joe's traders are taking huge risks with the company's money. This could all blow up in our face.

BUSINESS LINE MANAGER: No way. We're beautifully hedged. Besides, my guys made $500 million for the bank last quarter. How much did your folks make, Jane?

RISK MANAGER: Well, actually, my risk controllers didn't earn the bank a dime. We're a cost center. You know that, Joe.

Soon the meeting is over, with the CEO siding with the business line manager,[68] who probably outranks her in the corporate hierarchy anyway.

Can this built-in bias against risk managers be overcome? I fear the answer may be no. One way to try is for the board of directors—those alleged guardians of shareholder interests—to hold management's feet to the fire by elevating the importance of the risk-management function, including giving the chief risk officer a direct reporting line to the board's audit or risk committee. But history suggests that this degree of board activism may require—at least—an external prod. To provide this, bank supervisors could insist on such changes. Since money talks in a financial institution, another, complementary approach might be to direct top management to design a compensation and promotion system that treats risk controllers on a par with the traders they allegedly control—the

---

[67] The best eyewitness account of this problem may be the anonymous "Confessions of a Risk Manager" in *The Economist* (2008).

[68] Remember the previous discussion of CEO compensation incentives.

objective being to get as much (or more) talent into risk management as in trading.[69] Maybe there are better ideas—such as the idea mentioned earlier to pay traders, CEOs, and directors all in restricted stock. But in the interim, bank supervisors should keep a sharp eye on companies' risk-management practices.

3.8 Consumer Protection

I come, finally, down to ground level—to the consumers of financial products.

Perhaps the greatest financial crisis the world has ever known had its roots in risky sub-prime mortgages in the United States, most of which should never have been written. Of course, some of the mortgagees put themselves in the financial line of fire knowingly and deliberately by gambling that ever-rising house prices would bail them out of otherwise-untenable mortgages. It is hard to stop consenting adults from engaging in reckless behavior. But other people, especially poor and unsophisticated ones, either had no idea what they were signing up for or were duped into dangerous mortgage contracts they did not understand. It is this group that commands our attention and needs our help.

Some would seek the answer in greater financial education, but I am skeptical—not because I doubt that financial literacy is desirable, but because I fear it may be unattainable. Surveys of consumers regularly turn up evidence of mind-blowing levels of financial illiteracy. Even the simplest ideas, such as compound interest or the notion that diversification reduces risk, are foreign to most people.[70] Maybe I'm wrong, but the situation looks hopeless to me. It's like teaching adherence to speed limits to people who don't understand the difference between miles per hour and kilometers per hour. So I fall back on the second line of defense—state intervention to protect people from deceptive practices.

What might that mean in practice? Banning certain financial products? Maybe some, but I'd hope not many. There really are

---

[69] According to Tett (2009, ch. 7), JP Morgan Chase raised the status and pay of risk managers when Jamie Dimon became COO. I have also heard it claimed that Goldman Sachs has followed this practice for years. Is it just a coincidence that these two firms weathered the storm better than most of their peers?

[70] Among many sources that could be cited, see Lusardi (2009).

people for whom an adjustable-rate mortgage (ARM) with a teaser rate makes perfect sense, maybe even an option ARM. And we don't want to squelch useful financial innovations. But people who do not understand complicated financial products should be warned—or if necessary, guided—away from such products. What we need, instead, are:

- suitability standards for consumer financial products, so that sellers who lure unsophisticated customers into inappropriate (for them) products face legal peril, just as stockbrokers do today;
- simple, plain-English disclosures, perhaps modeled on food labels, which tell people "what's inside" at a glance and thereby facilitate comparison shopping;
- requirements that financial institutions offer "plain vanilla" options that are straightforward, easy to understand, and easy to comparison shop;
- good default options so that people who simply cannot or do not make up their minds are put into reasonable options.

With the exception of the first, the Treasury white paper called for each of these. But Congress, bowing to industry pressure, dropped most of them. It remains to be seen what the new Consumer Financial Protection Bureau will (or will not) do along these lines.

One final point about consumer protection, which I make with some regret as a former Federal Reserve Vice Chairman: Congress, in its wisdom, gave much of the responsibility for enforcing the various consumer financial protection laws to the Federal Reserve. No one would call this task an integral part of central banking (Blinder 2010). Furthermore, I believe consumer protection will always be a "weak sister" inside a central bank, which attaches much greater importance to its other responsibilities, such as monetary policy. In practice, the Federal Reserve performed its consumer protection duties poorly and deserved to lose that authority—not as a punishment, but because another agency, focused on the mission, will probably do the job better. That judgment, I think, is reflected in the recent U.S. legislation.

4. The Propriety of Proprietary Trading

The crisis raised numerous questions about the proprietary trading (a.k.a. gambling) activities of large financial firms, both in the United States and elsewhere. In the United States, it led the Obama administration to advocate (belatedly) the so-called Volcker Rule, which would proscribe proprietary trading by depository institutions.[71]

Before diving into these perilous waters, I want to make it clear that I have no puritanical aversion to gambling, that I realize that risk taking is an integral part of capitalism, and that I understand that even the simplest financial businesses (e.g., commercial lending) are inherently risky. All that said, I am among those who wondered, even before the crisis, whether our leading financial institutions hadn't taken proprietary trading too far. And now that so many big gambles have gone awry, I'm pretty much convinced that they did.

Let me approach the conceptual issues by analogy. Start with the case of an isolated *homo economicus*, meaning a completely rational person who confers (and imposes) no externalities on anyone else. Call him Harry. Among other things, Harry has no family to support, no employees, and no access to the social safety net. (If this sounds unrealistic, keep reading.) If Harry chooses to wager most of his net worth on something, society should presumably let him. If he fails, he will bear the entire cost.

Now suppose Harry supports a family, or runs a business whose employees depend on him for their livelihoods. Since going broke now has serious consequences for others, *someone* might want to place limits on Harry's ability to gamble—as a way to protect these innocent bystanders. Precisely what those limits might be is a hard question, however. Since risk taking is such an important part of capitalism, we certainly do not want to overdo it. And if Harry is a sole proprietor, rather than a corporation, it's far from clear who might constrain his behavior anyway. Besides, Harry should recognize his responsibilities to others and impose sensible limits on himself. Many people will stop right there, leaving the government

[71] This "rule," which is due to Paul Volcker, was not part of the U.S. Treasury's original proposals. The administration embraced the idea in January 2010.

no role. Certainly, not much state intervention, if any, can be justified on this count.

But things get murkier when there is any sort of social safety net, for then taxpayers are obligated to pick up part of the bill for failure. If Harry, his family, or his employees are entitled to state-provided benefits in the event he goes broke, then the state—acting in the taxpayers' interest—seems to have reason to limit the risk that Harry can take, either by restraining his behavior (example: requiring seat belts in cars), or by requiring him to purchase insurance (example: automobile collision insurance), or by levying taxes or fees to finance state-provided insurance (example: unemployment insurance). Except on the libertarian fringe, these ideas are widely accepted in all countries. But practical applications are often contentious.

One direct application in the financial context is to deposit insurance, whereby the government socializes some of the risks of bank failures by collecting insurance premiums and paying off (insured) depositors when necessary. Recent events raise a parallel question in the case of proprietary trading losses: If the government will, in practice, shoulder some of the losses in the event of failure, should it either limit or "tax" risky trading activities?

Before attempting to answer that question, let me consider some possibly pertinent deviations from the hypothetical norm we have just discussed. First, suppose Harry is not a rational person. Then, especially if innocent bystanders are involved, but perhaps even if not, most people would consider it reasonable to impose some limits on Harry's gambling—on the paternalistic grounds that he must be "protected from himself."

Next, imagine that Harry is so rich, and his business empire so far-flung, that his financial ruin could have serious systemic consequences. Harry might then be considered "too big too fail." We could try telling Harry that he is on his own, but the words would not be credible. In such a case, society might want to charge Harry for the special safety net that sits under him. The financial-sector parallels here are obvious—and germane to proposals, such as those mentioned earlier, to charge institutions for the privilege of being TBTF. Think Citigroup.

Third, suppose Harry is not rich enough to be systemically important per se, but is nonetheless heavily entangled in counterparty relationships with dozens, if not hundreds, of other institutions, some of which are systemically important. His failure might then undermine the viability of these counterparties, and perhaps the entire financial system. One possible reaction, in this case, would hold that Harry's counterparties should limit their exposures to him so that his failure would *not* bring them down too. Fair enough. But what if they don't? What if, instead, Harry's failure threatens to bring down a significant portion of the house of cards? Then, again, we have a plausible case for state intervention. In finance, this thought is the origin of the "too interconnected to fail" doctrine. Think Bear Stearns or, for that matter, Long-Term Capital Management in 1998.

The morals of this story are two. First, when "no man is an island," various (but not all) forms of state intervention into risk-taking behavior become at least *potentially justifiable*. Second, when, in particular, a taxpayer-financed safety net will pick up some of the bills, state intervention of some kind becomes not merely justifiable but perhaps even *obligatory*—as a way to protect taxpayers.

This reasoning leads straight to a pretty radical-sounding question, evocative of the now-famous Volcker Rule: Does proprietary trading belong in *any* financial institution that has access to the financial safety net—including, in the United States, FDIC insurance, the Federal Reserve's discount window, and other safeguards? Answering "no" leads to some radical departures from current practice because the list of such institutions obviously includes *all* insured depositories and *all* TBTF (or TITF) institutions.[72] But I must admit that I have a hard time getting to "yes"—unless capital charges against trading activities are raised enormously. So the radical solution seems to merit serious consideration. For example, the Group of Thirty (2009, p. 28)—hardly a bunch of populist renegades—concluded that, "Large, systemically important banking institutions should be restricted in undertaking proprietary activities that present particularly high risks... and large proprietary trading should be limited by strict capital and liquidity requirements." That's not quite a ban, but it's getting close.[73]

[72]Thus the list of possibly eligible institutions extends beyond Volcker's depository institutions. It does *not*, however, include modest-sized hedge funds. More on this shortly.

[73]It is no coincidence that Paul Volcker was instrumental in drafting this report.

But there is a downside: Roping off "proprietary trading" from other, closely related activities of banks is not as easy as it sounds. For example, banks buy and sell securities, foreign exchange, and other assets for their clients all the time. Often, such buying and selling is imperfectly synchronized or leaves banks with open positions for other reasons. Does that constitute "proprietary trading"? Furthermore, market making has obvious synergies with dealing on behalf of clients. Do we want to label all such activities as "proprietary trading"? And what about times when investment banks bring out securities, can't sell all of them, and get stuck with some? The point is: There is no bright line. That is why Adair Turner (2009), the chairman of Britain's FSA, concluded that "we could not proceed by a binary legal distinction—banks can do this but not that—but had to focus on the scale of position-taking and the capital held against position-taking."

Next, consider explicit trading in various sorts of assets. Suppose a firm owns a corporate bond and buys a CDS to insure it. It has hedged, not gambled, because the CDS extinguishes the default risk of the bond. But buying or selling a "naked" CDS creates risk rather than destroying it. The trouble is, we cannot always tell the two actions apart. And what if the truly hedged bank subsequently sells the bond, leaving itself an open position in the CDS? It seems hopeless to expect regulators to monitor banks' trading activities with that degree of precision.

Is there a way out? Certainly not a perfect one. One idea might be to require banks to segregate all their trading activities (including dealing for customers and market making) into a separately—but not thinly—capitalized affiliate, and then to make it clear that the affiliate has *no* claim on the safety net and *no* claim on the parent.[74] That's a step. But, of course, the parent company will still suffer a reputational hit if its trading affiliate goes under. So the Chinese wall can never be impermeable.

Furthermore, this suggestion evokes the notorious SIVs that were the source of so much trouble during this crisis. There would be two important differences, however. First, if the parent has access

[74] A controversial proposal that was half-heartedly embodied in the Dodd-Frank Act segregates derivative trading in approximately this way. But I call it "half-hearted" because most derivatives are excepted.

to the safety net, the law should make it *illegal* for the parent to downstream capital to the affiliate.[75] Creditors of the trading companies need to know—*with legal certainty*—that they are on their own. Second, and presumably as a consequence of the first, trading affiliates would need sizable capital and liquidity cushions of their own. To make this idea work, of course, no trading affiliate can ever be permitted to grow large enough to be TBTF in its own right.

Finally, let's go back to Harry and now suppose he represents the leadership of a corporation. In that case, his putative risk controller is supposed to be the board of directors. But it's equally clear that too many boards were far too passive and permissive, if not somnolent, to safeguard shareholder interests effectively. Their failures raise a host of questions about corporate governance that extend well beyond the scope of this paper. But I want to close with one last issue: the difference between proprietary trading inside and outside the corporate form.

As I mentioned earlier, skewed compensation systems give both corporate executives and traders incentives to gamble too much with stockholders' money. For them, it's often, "Heads, I win; tails, the stockholders lose." But there is an alternative to the corporation, mentioned earlier: the partnership. Partnerships internalize most of these externalities because decisionmakers are "playing" with their own money. It is also true that, e.g., hedge funds rarely grow large enough to be systemically important. Both of these considerations suggest that the financial system might be sturdier if more proprietary trading were to migrate out of giant corporations and into, e.g., hedge funds—as long as no hedge fund is allowed to grow too big to fail.

5. Brief Summary

The efficient markets hypothesis notwithstanding, the case for laissez-faire in financial markets now looks extremely weak. Finance does not appear to be self-regulating—except at enormous cost.

[75]An analogous provision exists in bank supervision. The Federal Reserve's "source of strength" doctrine allows the holding company to transfer funds to help the bank, but not vice versa. But Congress has never enshrined this principle into law.

Regulation of financial businesses is readily justified by the needs to protect consumers from being duped, to protect taxpayers from shouldering large bills, and to protect citizens from the dangers of financial and macroeconomic instability.

But that macroprudential rationale for regulation is not a license for government to do anything it pleases. Regulation should be well targeted at specific problems. It should be efficiently designed and run. It should rarely (and then only with very good reasons) regulate prices, quantities, or limit entry.[76] And it should strive both to be fair and to avoid stifling valuable innovation. A sound regulatory structure will minimize both regulatory gaps (activities without a regulator) and regulatory overlaps (activities with multiple regulators). It will harness incentives rather than fight them. And its success will not require outguessing the markets, which is a hopeless task in any case.

America's current regulatory system falls short on many of these criteria, and it failed miserably when put to the test a few years ago—with appalling consequences. The same was true of the regulatory systems of most countries.[77] The systems therefore cry out for change. The recently enacted U.S. law should make the system sturdier and more effective, but certainly not foolproof.

While recognizing that no one can do the job perfectly, I believe that virtually every country would benefit from having a systemic risk regulator—and that that regulator should be the central bank. In these respects, my views align with those of the U.S. Treasury and Congress. I also believe that having some financial companies that are "too big (and/or interconnected) to fail" is inevitable, and so we need a new mechanism to resolve such institutions in an orderly way. While the most appropriate form of such a mechanism (e.g., a special resolution authority versus a new chapter of the bankruptcy code) is not entirely clear, it appears that special resolution holds some significant advantages over "Chapter 16." Again, that is what Congress also decided.

Simplifying the existing maze of federal regulators in the United States would be highly desirable. And going down from six agencies (the Federal Reserve, OCC, FDIC, OTS, SEC, and CFTC) to

[76] This last point is germane to the debate over Glass-Steagall.
[77] One notable exception was Canada, which fared quite well.

five hardly qualifies as a great leap forward. But I understand the political imperative to tread lightly here, and the new Consumer Financial Protection Bureau—though bringing us back up to six, if counted as a free-standing agency—should help fill an important gap.

What to do about regulating derivatives, which caused a great deal of trouble, and hedge funds, which probably did not, is far from clear. But pushing the trading of as many derivatives as possible onto organized exchanges, and insisting on regulatory transparency from all sizable hedge funds strike me as valuable first steps that the new U.S. law takes.

The Basel II capital standards need work, to put it mildly. They probably require too little capital and are procyclical to boot. They place far too much faith in both ratings agencies and banks' internal risk models, and they are too permissive about off-balance-sheet entities. They also need to be supplemented by liquidity requirements for, as we have seen, illiquidity can turn quickly into insolvency. And that's just for starters. The good news is that the talks that (hopefully) will lead to Basel III are dealing with all of these issues and others right now, purportedly effectively. The bad news is that Basel III will require international agreement, which will come slowly (if at all). Since we cannot wait another decade, this poses a difficult question for national authorities: Should they wait for Basel or go it alone?

Last, and certainly not least, the recent serial financial catastrophes highlighted three huge jagged rocks that were concealed by the high tide: dysfunctional compensation systems (for both traders and CEOs), sleepy if not irresponsible corporate boards, and the question of how much proprietary trading should be allowed in institutions that have (tacit or explicit) government backing. None of these problems are easy to solve, though I have made some suggestions for each. What is abundantly clear is that the status quo will no longer do. "If it ain't broke, don't fix it" is a wise, old adage that I quote frequently and try to live by. But it does not apply here.

References

Andrews, E. 2004. "The Ever More Graspable, and Risky, American Dream." *The New York Times*, June 24, 2004, C1.

Barrell, R. P., E. Davis, D. Karim, and I. Liadze. 2009. "Bank Regulation, Property Prices and Early Warning Systems for Banking Crises in OECD Countries." National Institute of Economic and Social Research Discussion Paper No. 330 (March).

Bernanke, B. 2009. "Lessons of the Financial Crisis for Banking Supervision." Remarks to the Federal Reserve Bank of Chicago Financial Structure Conference, May 7.

Bikker, J., and P. Metzemakers. 2005. "Bank Provisioning Behaviour and Procyclicality." *Journal of International Financial Markets, Institutions and Money* 15 (2): 141–57.

Blinder, A. 2009. "Crazy Compensation and the Crisis." *The Wall Street Journal*, May 28, 2009.

———. 2010. "How Central Should the Central Bank Be?" *Journal of Economic Literature* 48 (1): 124–34.

Borio, C., and M. Drehmann. 2009. "Towards an Operational Framework for Financial Stability: 'Fuzzy' Measurement and Its Consequences." BIS Working Paper No. 284 (June).

Bouvatier, V., and L. Lepetit. 2008. "Banks' Procyclical Behavior: Does Provisioning Matter?" *Journal of International Financial Markets, Institutions and Money* 18 (5): 513–26.

Brunnermeier, M., A. Crockett, C. Goodhart, A. Persaud, and H. S. Shin. 2009. *The Fundamental Principles of Financial Regulation.* Geneva Report on the World Economy. London: International Centre for Monetary and Banking Studies and the Centre for Economic Policy Research.

Calomiris, C. 2009. "Financial Innovation, Regulation, and Reform." *Cato Journal* 29 (1): 65–91.

Chomsisengphet, S., and A. Pennington-Cross. 2006. "The Evolution of the Subprime Mortgage Market." *Review* (Federal Reserve Bank of St. Louis) 88 (1): 31–56.

Cohen, H. R., and M. Goldstein. 2009. "The Case for an Orderly Resolution Regime for Systemically-Important Financial Institutions." Report commissioned by the Pew Task Force on Financial Reform (October 21).

Committee on Capital Markets Regulation. 2009. "The Global Financial Crisis: A Plan for Regulatory Reform." Report. (May 26).

Dash, E., and J. Creswell. 2008. "Citigroup Pays for a Rush to Risk." *The New York Times*, November 23, 2008.

The Economist. 2008. "A Personal View of the Crisis: Confessions of a Risk Manager." August 7, 2008.

Flannery, M. 2005. "No Pain, No Gain? Effecting Market Discipline via 'Reverse Convertible Debentures'." In *Capital Adequacy beyond Basel: Banking, Securities, and Insurance*, ed. H. Scott, 171–96. Oxford, UK: Oxford University Press.

———. 2009. "Stabilizing Large Financial Institutions with Contingent Capital Certificates." Working Paper (October). Available at http://papers.ssrn.com/so13/papers.dfm?abstract_id=1485689.

French, K. R., M. N. Baily, J. Y. Campbell, J. H. Cochrane, D. W. Diamond, D. Duffie, A. K. Kashyap et al. 2010. *The Squam Lake Report: Fixing the Financial System*. Princeton, NJ: Princeton University Press.

Gorton, G., and A. Metrick. 2009. "Securitized Banking and the Run on Repo." International Center for Finance at Yale School of Management Research Paper No. 14 (July).

Gramlich, E. 2004. "Subprime Mortgage Lending: Benefits, Costs, and Challenges." Remarks at the Financial Services Roundtable Annual Housing Policy Meeting, Chicago, Illinois, May 21.

Group of Thirty. 2009. *Financial Reform: A Framework for Financial Stability*. Washington, DC.

Huang, X., H. Zhou, and H. Zhu. 2009. "A Framework for Assessing the Systematic Risk of Major Financial Institutions." *Journal of Banking and Finance* 33 (11): 2036–49.

King, M. 2009. Speech at the Lord Mayor's Banquet for Bankers and Merchants, Mansion House, London, June 17.

Kuttner, R. 2007. "The Bubble Economy." *The American Prospect*, September 24, 2007.

Lusardi, A. 2009. "Household Savings Behavior in the United States: The Role of Literacy, Information, and Financial Education Programs." In *Policymaking Insights from Behavioral Economics*, ed. C. Foote, L. Goette, and S. Meier, 109–149. Boston, MA: Federal Reserve Bank of Boston.

Morgenson, G. 2008. "How the Thundering Herd Faltered and Fell." *The New York Times*, November 9, 2008.

Morris, S., and H. S. Shin. 2008. "Financial Regulation in a System Context." In *Brookings Papers on Economic Activity: Fall 2008*, ed. D. Elmendorf, N. G. Mankiw, and L. Summers, 229–74. Washington, DC: Brookings Institution Press.

Peek, J., E. Rosengren, and G. Tootell. 2009. "Is Financial Stability Central to Central Banking?" Paper presented at the Federal Reserve Bank of Boston's annual economic conference, Chatham, Massachusetts, October 22.

Pew Charitable Trusts Task Force on Financial Reform. 2009. "Principles on Financial Reform: A Bipartisan Policy Statement." (December). Available at http://www.pewfr.org/admin/section_introductions/files/FRP-POLICY-STATEMENT-FINAL.pdf.

Rajan, R. 2005. "Has Financial Development Made the World Riskier?" In *The Greenspan Era: Lessons for the Future*, 313–69. Proceedings of the Federal Reserve Bank of Kansas City Annual Economic Symposium, Jackson Hole, Wyoming, August 25–27.

Repullo, R., and J. Suarez. 2009. "The Procyclical Effects of Bank Capital Regulation." CentER Discussion Paper No. 2010-29S (October).

Rochet, J.-C. 2008. "Commentary: Rethinking Capital Regulation." In *Maintaining Stability in a Changing Financial System*, 473–81. Proceedings of the Federal Reserve Bank of Kansas City Annual Economic Symposium, Jackson Hole, Wyoming, August 21–23.

Saunders, A., R. Smith, and I. Walter. 2009. "Enhanced Regulation of Large, Complex Financial Institutions." In *Restoring Financial Stability: How to Repair a Failed System*, ed. V. Acharya and M. Richardson, 139–156. Hoboken, NJ: John Wiley & Sons, Inc.

Tarullo, D. 2008. *Banking on Basel: The Future of International Financial Regulation*. Washington, DC: Peterson Institute for International Economics.

———. 2009. "Global Financial Regulation in the Wake of the Crisis." Remarks to the Peterson Institute for International Economics, Washington, DC, June 8.

Tett, G. 2009. *Fool's Gold*. New York, NY: Simon & Schuster, Inc.

Turner, A. 2009. "Financial Regulatory Reform: A New Foundation." Speech at the British Bankers' Association Annual International Banking Conference, June 30.

U.S. Securities and Exchange Commission, Office of Inspector General, Office of Audits. 2008. "SEC's Oversight of Bear

Stearns and Related Entities: The Consolidated Supervised Entity Program." Report No. 446-A (September 25).

U.S. Treasury. 2009a. "Financial Regulatory Reform: A New Foundation—Rebuilding Financial Supervision and Regulation." (June 17). Available at http://www.house.gov/apps/list/hearing/financialsvcs_dem/testimony_-_treasury.pdf.

———. 2009b. "Principles for Reforming the U.S. and International Regulatory Capital Framework for Banking Firms." (September 3). Available at http://www.ustreas.gov/press/releases/docs/capital-statement_090309.pdf.