A Structural Model of Bank Default Linkages

Yvonne Kreis*
Gutenberg School of Management and Economics

Dietmar P.J. Leisen
Gutenberg School of Management and Economics

July 17, 2015

Abstract

We study default across the banking sector in a structural model of individual bank defaults. We show that the so-called asset correlation impacts, in a highly non-linear way, risk measures of the default frequency in the banking sector; this motivates capital ratio supplements in macro-prudential regulation. We assess the typical correlation value empirically and relate larger values to stress (crisis) periods in the banking sector. A theoretical application trades off efficiency gains from more concentrated banking sectors against increased systemic losses and allows us to derive an optimal level of asset concentration.

Keywords

default, structural model, systemic risk, macro-prudential regulation

JEL Classification

G01, G21

*Please address correspondence to: Yvonne Kreis, Gutenberg School of Management and Economics, University of Mainz, 55099 Mainz, Germany; Phone: ++49-6131-392-3992; Fax: ++49-6131-392-3782; Email: yvonne.kreis@uni-mainz.de.
1 Introduction

The recent financial crisis highlighted the interconnectedness of financial institutions and invigorated a discussion of contagion risks within the financial sector. The proposed models focus on correlations in market prices of financial institutions, e.g. Acharya et al. (2010), Adrian and Brunnermeier (2011), but default originates either in the balance sheet or changes therein. This leads us to study a structural model of default for individual banks (Merton (1974)), where default linkages are due to correlated (changes in) asset values across banks. Our goal is to analyze aggregate default risk across the banking sector.

Our model of aggregate default risk across the banking sector is closely related to models of aggregate default risk across a credit portfolio. However, there are at least two crucial differences: First of all, the number of entities is very different: typically, the banking sector is composed of a small number of financial institutions, but the analysis of credit portfolios assumes an infinite number of debtors; we will argue in the paper that this difference is akin to the role of macro-prudential regulation in addition to micro-prudential regulation. Second, the analysis of credit portfolios focuses on “small” correlations (between 0 and 24%), but our empirical analysis shows that that correlations in the banking are much larger on average (65% over the time period 1980-2015) and may even become very large (up to 88%).

Over the course of the last 35 years, we estimate rolling (average) correlations of banks in the banking sector both in relation to market indexes and in a principal component analysis. We then discuss the temporal evolution and argue that structural changes in the banking system lead to an increase in overall correlation, and that stress periods relate to periods with particularly large asset correlations in the banking sector. A theoretical discussion focuses on the frequency of bank defaults in the banking sector; it shows that the correlation value is a critical parameter in our analysis of the banking system that has been largely ignored so far: it affects both the default frequency and capital supplements for systemic risk in a highly non-linear way. Our results also show that the overall default frequency decreases as the number of banks become larger, i.e. concentration decreases; this analysis of default linkages, however, ignores that a more concentrated banking system delivers banking services more efficiently; we incorporate such efficiency gains, study the resulting trade-off
and characterize the optimal bank concentration.

Our paper is related to the literature on systemic risk. The current network literature on systemic risk studies banks’ interconnectedness on their asset/liability side in network models (see e.g. Poledna and Thurner (2014), Delpini et al. (2013)); however, an implementation is difficult, see, e.g. Gai and Kapadia (2010). We contribute to this literature on systemic risk by introducing a simplified conceptual framework that takes a “condensed” view on network models and bank’s interconnectedness; there are several motivations for such a framework. First of all, it is a straightforward extension of the well-known structural models of individual defaults to the banking system. Second, within a continuous-time setup, Duan and Zhang (2013) start modeling interbank relationships\footnote{The impact of a bank’s default on some other bank can be expressed through a form of sensitivity; second round effects can then be interpreted as an adjustment of that sensitivity and adding them all up provides the “total” sensitivity to a (default) shock.} and end up with a common factor describing default. Finally, there is historical support for such a condensed view; for example, regarding the Lehman Brothers’ default, Chakrabarty and Zhang (2012) report that default events do not only affect banks with direct exposure, but have an impact on all banks (with the strength of said impact determined by the respective factor loadings).

Our paper contributes to measurement of default linkages (systemic risk). Differently to previous approaches that center on market values, we relate default to the balance sheets. In addition, our approach to measuring risk in the banking system has practical advantages since it is easy and straightforward to implement. We estimate the changing correlations over time and relate them to stress period; this may eventually lead to a (forward-looking) indicator of stress periods for financial regulators. In addition, we relate small correlation values (close to zero) to micro-prudential regulation; based on the observed much larger values we then argue for macro-prudential regulation. We derive the associated capital supplements to cover systemic risk: they appear reasonable in “normal” times; however, they should be set in relation to correlations in periods of stress and with that premise the current capital supplement appear too small.

A final contribution is to the theory of bank organization. Following the financial crisis, researchers and central banks have suggested the implementation of bank size restrictions
(see e.g. Haldane (2012)) to enhance overall financial stability and to eliminate any too-big-to-fail issues. However, such limits come at the cost of efficiency losses. Since there has been no agreement on a specific bank size limit, we address the related research question of optimal design of a banking sector. Assuming completely homogeneous banks, we discuss costs and benefits of different banking sector designs, and find that at typical correlation levels the banking system should be less concentrated than it currently is.

The remainder of this paper proceeds as follows: We describe the basic setup in section 2. The next section discusses empirical asset correlations. Section 4 address the role of micro- and macro-prudential regulation, in particular we derive expected default frequency and the capital ratio supplement for systemic risk. In the following section 5, we study the trade-offs in bank concentration. Section 6 concludes.

2 The setup

We study a banking sector composed of $N$ financial institutions (henceforth banks) and index these by $i = 1, \ldots, N$. Our main interest lies in default linkages between financial institutions and their impact on the entire financial system; to keep our exposition simple we focus on a single time period with dates $t = 0$ and $t = 1$. At times it will be necessary to adjust for the physical time between both dates; as usual we express that difference in years and denote it by $\Delta t$.

Within the banking system, each bank can default on its obligations (without further specification of causes) at date $t = 1$. For simplicity, we assume that all banks have identical default probabilities $p_i = p$ and identical (fractional) loss-given-default $L_i = L$. We denote for every bank $i$ by $E_i, \hat{E}_i$ the date $t = 0, 1$ book values of equity and by $A_i, \hat{A}_i$ the date $t = 0, 1$ book values of assets, respectively; also, we denote by $D_i$ the date $t = 1$ book value of debt and assume it is known at date $t = 0$. The date $t = 1$ is the result of economic activity that materializes over the time period and therefore we assume that $\hat{A}_i$ is a (strictly positive) random variable. For future analysis we define

$$
\sigma^2_i = \text{Var}\left(\ln \frac{\hat{A}_i}{A_i}\right), \quad V_i = \frac{\ln \frac{\hat{A}_i}{A_i}}{\sigma_i}, \quad K_i = \frac{\ln \frac{D_i}{A_i}}{\sigma_i},
$$

and refer to $V_i$ as the risk-adjusted return on assets and to $K_i$ as the risk-adjusted capital
ratio. (Usually, $K_i$ is called the distance-to-default but we refer to it differently for reasons that will become apparent after equation (9) in subsection 4.3.) Note that $V_i$ is a random variable with variance equal to 1. Throughout, we assume that $V_i$ is a standard normal random variable.

In so-called structural models of default, see, e.g. Merton (1974) it is common to assume that default occurs when the value of assets is no longer sufficient to cover the value of debt, i.e. $A_i < D_i$. Note that $\bar{E}_i = \bar{A}_i - \bar{D}_i$, i.e. default occurs when the book value of equity becomes negative. Our setup is in line with structural models of default, see, e.g., Moody’s KMV approach (Crosbie and Bohn (2002)), since they start directly with the assumption that default occurs, iff

$$V_i < K_i.$$ (2)

Banks may be interconnected via the interbank market or via investments in the same or comparable assets. We follow the spirit of asset-based models and assume that

$$V_i = a_i \cdot Y + \sqrt{1 - a^2} \cdot \varepsilon_i,$$

where $Y, \varepsilon_i$ are independent standard normal random variables. Here $Y$ is a single (common) factor and $-1 \leq a_i \leq 1$ describes the individual bank’s sensitivity to the common factor. Here and throughout the theory parts of this paper we assume for simplicity that the sensitivity $a_i = a$ is the same for all banks in the banking system; then $\rho = a^2$ describes the correlation of risk-adjusted asset values for any two banks $i, j$. In the remainder of this paper we study exclusively correlations $\rho$ instead of sensitivities $a$.

Throughout this paper we denote by $P_{\rho,K}[:]$ the probability measure describing the events in this banking system and by $E_{\rho,K}[:]$ the expectation operator. Since all banks have the same default probability (by assumption), they all have the same distance-to-default, $K_i = K$ for all $i = 1, \ldots, N$. Then, $p = P_{\rho,K}[V_i < K_i] = \mathcal{N}(K)$ describes the (identical) default probability of (any) bank $i$. Here and throughout this paper we denote by $\mathcal{N}$ the cumulative distribution function for standard normal random variables, and by $\mathcal{N}^{-1}$ its inverse.

Our setup follows the spirit of so-called structural models, but departs from it in several important ways. First, we study a banking system composed of a large but finite number of banks; this differs from the usual application that focuses on an infinite number of banks.
Second, and most important, asset sensitivities are usually held fixed and both, the actual size and the time-variation in these sensitivities, are ignored; we incorporate a (potentially stochastic) time dependence of crucial parameters. There are several reasons to study time-changing (stochastic) sensitivities. First of all, we expect that during systemic/crisis events, the asset dynamics change significantly. In fact, the literature documents that correlations of stocks and bonds across markets rise in times of stress, the so-called flight-to-quality; this means within our framework that sensitivities must change drastically during such periods.

3 Time-dependent asset correlations

This section studies asset correlations (sensitivities) empirically. The quantitative insights will guide us through the following sections.

3.1 Estimation approach and data

We are interested in estimation of time-dependent correlations, we would desire such information on a frequent basis, e.g. daily, but asset information is available on a quarterly basis. To circumvent this, we note that stock price information is available on a daily basis; to describe how we proceed, let us suppose for a moment that we are interested in values at a physical time between the two dates $t = 0, 1$. Then, recall that the starting point of our analysis is a structural model (of default); therein the market value of equity (stock price) on a given day is the value of the claim on asset values exceeding liabilities $(\bar{A}_t - D_t)^+$. While the value of this claim at any date is driven by the (unobserved) asset value, we recall the idea of $\Delta$-hedging and note that the functional dependence of the stock price on the underlying asset value is approximately linear, with the $\Delta$ being the sensitivity. This, however, means that, in a first approximation, we can estimate asset correlations by estimating stock return correlations.

---

2 While we focus on a single factor model, more elaborate models take into account multiple factor dependencies with time-invariant sensitivities, see, e.g., Crouhy et al. (2000), p. 384 ff. To introduce our ideas in a setup as simple and straightforward as possible, we focus on a single factor model.

3 Mathematically, we can always represent a multi-factor model as a single factor model as long as we permit a time-changing (stochastic) correlation structure; in this light modeling stochastic correlations is inevitable in a multi-factor world.
<table>
<thead>
<tr>
<th></th>
<th>ABC</th>
<th>BoA</th>
<th>BNY</th>
<th>Cap1</th>
<th>Citi</th>
<th>GS</th>
<th>HSBC</th>
<th>JPM</th>
<th>Key</th>
<th>MS</th>
<th>PNC</th>
<th>State</th>
<th>Sun</th>
<th>Bancorp</th>
<th>Wells</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>0.8875</td>
<td>0.2912</td>
<td>0.1149</td>
<td>0.6549</td>
<td>-0.3721</td>
<td>0.6113</td>
<td>0.4166</td>
<td>0.5504</td>
<td>0.3726</td>
<td>0.0207</td>
<td>0.2334</td>
<td>0.7871</td>
<td>0.3473</td>
<td>0.4738</td>
<td></td>
</tr>
<tr>
<td>BoA</td>
<td>0.8875</td>
<td>0.3269</td>
<td>0.0680</td>
<td>0.5825</td>
<td>-0.4350</td>
<td>0.5415</td>
<td>0.3551</td>
<td>0.5614</td>
<td>0.4618</td>
<td>-0.0945</td>
<td>0.2201</td>
<td>0.6993</td>
<td>0.3427</td>
<td>0.4247</td>
<td></td>
</tr>
<tr>
<td>BNY</td>
<td>0.2912</td>
<td>0.3269</td>
<td>0.1685</td>
<td>0.4633</td>
<td>0.2049</td>
<td>0.4729</td>
<td>0.4042</td>
<td>0.4894</td>
<td>0.4178</td>
<td>-0.0437</td>
<td>0.2697</td>
<td>0.2768</td>
<td>0.2953</td>
<td>0.0618</td>
<td></td>
</tr>
<tr>
<td>Cap1</td>
<td>0.1149</td>
<td>0.0680</td>
<td>0.1685</td>
<td>0.4409</td>
<td>0.2896</td>
<td>0.2527</td>
<td>0.1622</td>
<td>0.1363</td>
<td>0.3911</td>
<td>0.5858</td>
<td>0.4049</td>
<td>0.4067</td>
<td>0.2421</td>
<td>0.4414</td>
<td></td>
</tr>
<tr>
<td>Citi</td>
<td>0.6549</td>
<td>0.5825</td>
<td>0.4633</td>
<td>0.4409</td>
<td>-0.1001</td>
<td>0.4720</td>
<td>0.5615</td>
<td>0.3911</td>
<td>0.3749</td>
<td>0.1978</td>
<td>0.5321</td>
<td>0.6925</td>
<td>0.4056</td>
<td>0.5341</td>
<td></td>
</tr>
<tr>
<td>GS</td>
<td>-0.3721</td>
<td>-0.4350</td>
<td>0.2049</td>
<td>0.2896</td>
<td>-0.1001</td>
<td>0.2218</td>
<td>-0.1218</td>
<td>-0.1232</td>
<td>0.6360</td>
<td>0.2124</td>
<td>-0.0934</td>
<td>0.6213</td>
<td>-0.2161</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSBC</td>
<td>0.6113</td>
<td>0.5415</td>
<td>0.4729</td>
<td>0.2527</td>
<td>0.4720</td>
<td>-0.0161</td>
<td>0.7640</td>
<td>0.7298</td>
<td>0.4145</td>
<td>-0.3056</td>
<td>0.1788</td>
<td>0.4015</td>
<td>0.6729</td>
<td>0.0888</td>
<td></td>
</tr>
<tr>
<td>JPM</td>
<td>0.4166</td>
<td>0.3551</td>
<td>0.4042</td>
<td>0.1622</td>
<td>0.5645</td>
<td>0.2218</td>
<td>0.7640</td>
<td>0.7298</td>
<td>0.4145</td>
<td>-0.3056</td>
<td>0.1788</td>
<td>0.4015</td>
<td>0.6729</td>
<td>0.0888</td>
<td></td>
</tr>
<tr>
<td>Key</td>
<td>0.5504</td>
<td>0.5614</td>
<td>0.4894</td>
<td>0.1363</td>
<td>0.3911</td>
<td>-0.1218</td>
<td>0.7298</td>
<td>0.2250</td>
<td>0.5181</td>
<td>-0.2464</td>
<td>-0.0666</td>
<td>0.2489</td>
<td>0.3589</td>
<td>0.2046</td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>0.3726</td>
<td>0.4618</td>
<td>0.4178</td>
<td>0.3911</td>
<td>0.3749</td>
<td>-0.1232</td>
<td>0.4145</td>
<td>0.2246</td>
<td>0.5181</td>
<td>-0.0878</td>
<td>0.2698</td>
<td>0.3596</td>
<td>0.1546</td>
<td>0.1487</td>
<td></td>
</tr>
<tr>
<td>PNC</td>
<td>0.0207</td>
<td>-0.0945</td>
<td>-0.0437</td>
<td>0.5858</td>
<td>0.1978</td>
<td>0.6360</td>
<td>-0.3056</td>
<td>0.2093</td>
<td>-0.2464</td>
<td>-0.0878</td>
<td>0.5624</td>
<td>0.2343</td>
<td>-0.0142</td>
<td>0.5064</td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>0.2334</td>
<td>0.2201</td>
<td>0.2697</td>
<td>0.4049</td>
<td>0.5321</td>
<td>0.2124</td>
<td>0.1788</td>
<td>0.4710</td>
<td>-0.0666</td>
<td>0.2698</td>
<td>0.5624</td>
<td>0.4554</td>
<td>0.0802</td>
<td>0.3303</td>
<td></td>
</tr>
<tr>
<td>Sun</td>
<td>0.7871</td>
<td>0.6993</td>
<td>0.2768</td>
<td>0.4067</td>
<td>0.6925</td>
<td>-0.0934</td>
<td>0.4015</td>
<td>0.4005</td>
<td>0.2489</td>
<td>0.3596</td>
<td>0.2343</td>
<td>0.4554</td>
<td>0.2274</td>
<td>0.5396</td>
<td></td>
</tr>
<tr>
<td>Bancorp</td>
<td>0.3473</td>
<td>0.3427</td>
<td>0.2953</td>
<td>0.2421</td>
<td>0.4056</td>
<td>0.6213</td>
<td>0.6729</td>
<td>0.5468</td>
<td>0.3589</td>
<td>0.1546</td>
<td>-0.0142</td>
<td>0.0802</td>
<td>0.2274</td>
<td>0.1062</td>
<td></td>
</tr>
<tr>
<td>Wells</td>
<td>0.4738</td>
<td>0.4247</td>
<td>0.0618</td>
<td>0.4414</td>
<td>0.5341</td>
<td>-0.2161</td>
<td>0.0888</td>
<td>0.2534</td>
<td>0.2046</td>
<td>0.1487</td>
<td>0.5064</td>
<td>0.3303</td>
<td>0.5396</td>
<td>0.1062</td>
<td></td>
</tr>
<tr>
<td>S&amp;P-500</td>
<td>0.1946</td>
<td>0.0632</td>
<td>-0.1916</td>
<td>0.5536</td>
<td>0.3145</td>
<td>0.5783</td>
<td>-0.3531</td>
<td>0.2972</td>
<td>-0.3856</td>
<td>-0.2114</td>
<td>0.8129</td>
<td>0.5832</td>
<td>0.3512</td>
<td>-0.0095</td>
<td>0.5067</td>
</tr>
</tbody>
</table>

Table 1: Pairwise correlations of 15 major U.S. financial institutions and their correlation with the S&P 500; estimation based on time period January 1980 and May 2015.

Bank Sample consists of ABC: Associated Banc Corp; BoA: Bank of America; BNY: Bank of New York Mellon; Cap1: Capital One; Citi: Citigroup; GS: Goldman Sachs; HSBC: HSBC; JPM: JP Morgan; Key: KeyCorp; MS: Morgan Stanley; PNC: PNC; State: State Street; Sun: SunTrust Banks; Bancorp: U.S. Bancorp; Wells: Wells Fargo
At the end of 2014, there were 5,572 commercial banks in the United States. It is neither practical nor interesting to study all these banks at the same time: according to Federal Reserve Economic Data, banks are highly heterogeneous in the size of their balance sheets; in particular, 68 banks with more than US-$ 15 billion in total assets jointly represent more than 90% of the banking sector. For our analyses, we select a fixed sample of 15 U.S. financial institutions that were among the largest by assets over the past decade. At the end of 2014, the average balance sheet in our sample was US-$ 806.51 billion. (The smallest balance sheet was US-$ 25.65 billion, the largest US-$ 2,527.01 billion). Out of the 15 banks that we selected for our sample, 11 banks were among the largest 15 U.S. banks at the end of 2014. For these 15 U.S. financial institutions, we collect daily stock prices between 01 January 1980 (or the respective listing date) and 29 May 2015.

Table 1 shows the estimated correlations between these 15 stocks as well as their correlations with the S&P 500. The correlation values show an average of 0.2985 with a minimum of -0.4350 (Goldman Sachs and Bank of America) and a maximum of 0.8875 (Banc Corp and Bank of America).

During the last 35 years, financial markets have been through several crisis periods. For example, our sample period encompasses the Savings & Loans Crisis in the early 1980, the dotcom bubble in the late 1990s as well as the housing bubble in the mid 2000. Also, the regulation of financial institutions has undergone significant changes: the Financial Institutions Reforms, Recovery and Enforcement Act of 1989 changed the regulation of S&L institutions, the Riegle-Neil Interstate Banking and Branching Efficiency Act of 1994 allowed for nationwide branching and acquisitions, and the Gramm-Leach-Bliley Financial Services Modernization Act of 1999 repealed the Glass-Steagall Act of 1933. We expect that these structural changes in the banking sector over the last 35 years impact the level of asset correlations.

The remainder of this section estimates correlations at daily frequency over the course of the last 35 years. Our intention is to study both the level of correlations over time and their short-term variations perspective. We carry out a discussion of two choices for the factor \( Y \): the first subsection looks at market indexes as application for the common factor \( Y \), while the second subsection extracts the factor \( Y \) from a principal component analysis. In
all our estimations, we estimate sensitivities over a rolling 3-month window; our plots show a smoothed version that takes rolling averages of the estimated sensitivities over 6-month periods.

3.2 Market indexes

This subsection studies two potential choices for the common factor: the S&P-500 and the Dow Jones U.S. Financials, a bank index. For both, we calculate separately the sensitivity of each financial institution in our sample towards the respective index. Since we extract this information from pairwise correlations, we use the Fisher z-transformation to calculate an average sensitivity. (For details we refer to Fisher (1915), and Gayen (1951)).

Panel (a) of figure 1 shows the average correlation for the entire time-period, while panel (b) takes a deeper look at the last 10 years; both panels are organized similarly, i.e. they show the average correlation of our sample with the stock price index over a rolling 3-month window. In addition, the linear trend in average correlation is depicted over the respective sample period.

Let us first discuss the average correlation based on the S&P 500. For the time period up to 1995, the average correlation shows significant variation. For example, slightly negative correlation values in early 1994 are followed by average correlations of well over 0.75 in early 1995. Thereafter, average correlation remains at a fairly high level up to mid 1999 when a strong decrease in average correlations occurs. From 2001 to 2007, average correlations seem to stabilize once more around 62%. The subsequent financial crisis is evident in our results. There is a first increase in 2008 (Lehman default), a further increase in average correlation in mid-2010 and mid-2012. During crisis periods and times of increased uncertainty, the average correlation increases (sharply).

It may be argued whether the S&P-500 is a reasonable choice for common factor Y. As the S&P-500 encompasses a large sample of U.S. stocks, we consider it well appropriate to represent the financial market development. As an additional example, we consider for that purpose return data of the Dow Jones U.S. Financials from July 2005 to May 2015. Results for the respective average correlation of U.S. financial institutions with this benchmark are also presented in panel (a) of figure 1. While the average correlation with the Dow Jones
U.S. Financials is generally larger than with the S&P-500, there is only a difference in level, but not in temporal development.

This similarity in results regarding the S&P-500 and the Dow Jones U.S. Financials is also present in panel (b) of figure 1 which focuses on the time period of 2005 to 2015. Here, individual crisis events like the Lehman default in fall 2008, the outbreak of the Greek debt crisis in 2010 are evident. This focus on a shorter time period structurally represents the current situation and the impact of individual crisis events on average sensitivity.

In conclusion, our results show that over the course of these 35 years of data, average correlation of financial institutions’ stock prices and the market index seem to have strongly increased. This indicates a higher co-movement among financial institutions’ stock prices, i.e. they are apparently more dependent on (influenced by) a single common factor. Moreover, stock price correlations seem to increase in crisis periods, but the identification of individual crisis events seems limited.

3.3 Principal component analysis

While asset correlations derived in the previous subsection aim at identifying the common market sensitivity as an average, a principal component analysis (PCA) provides a different perspective. This approach aims at identifying principal components explaining variance in
Figure 2: Principal Component Analysis: Average Correlation of U.S. Financial Institutions

observations, but it neither requires nor aims at an identification of explanatory factors.\footnote{A PCA uses orthogonal linear transformation to convert observations of correlated variables into values of uncorrelated variables. The resulting principal components explain the variance in observations (see e.g. Johnson and Wichern (2007), chapter 8).}

Figure 2 presents the average factor loadings w.r.t. the first principal component which we interpret to represent average asset sensitivity among financial institutions.\footnote{As an alternative, we studied the time series of the percentage of variation explained by the first principal component (its explanatory power on total variation). It is analogous to the one depicted in figure 2. To save space we only present results for averages, as this is analogous to our earlier analysis, see figure 1.} In panel (a) of figure 2, the average is 65.12\%, but the time series shows significant variation. Similar to results of the correlation analysis, average asset sensitivity is fairly volatile from 1980 to 1994. It then appears to stabilize at values of about 62.5\%. From fall 2004 up to spring 2008, the average correlation then falls to approx. 60\%, and only starts to increase in spring 2008 shortly before the collapse of Lehman Bros. In subsequent months, a drastic increase in average correlation is evident, reaching a maximum value of over 90\% in summer 2012. Over the complete time period, we once more find evidence for an increase in common asset sensitivity among financial institutions. The average in figure 2 shows an increase from approximately 57\% in early 1980 to close to 73\% in 2015. We consider these results to represent strong evidence in favor of a time-dependent modeling of asset correlations.

Regarding the development of asset sensitivity leading up to and during the financial crisis, we refer to panel (b) of figure 2. For the sample period 2005 - 2015, the average asset
sensitivity is 70.99% (minimum: 38.31%; maximum: 87.70%). This average correlation is significantly larger than the average of 65.12% for the sample period 1980 - 2015. Regarding the temporal developments over the past decade, panel (b) of figure 2 shows the following: Up to the default of Lehman Bros. (September 2008), average asset correlations were generally between 55% and 65%. Only in the time period directly leading up to the Lehman bankruptcy and in subsequent months, average asset correlations drastically increased. Hence, at that time, the asset value of financial institutions in the U.S. was significantly determined by a common shock, a systemic shock caused by the Lehman default.

While the Lehman bankruptcy was a severe crisis event (systemic shock), our analysis also documents a strong decrease in asset sensitivity in late 2009/10, only to be followed by another strong increase starting in spring 2010. This second increase in asset correlations coincides with the outbreak of the Greek debt crisis and subsequent negotiations on a Greek rescue package.

In conclusion, results of these analyses confirm that asset correlations do behave differently during normal periods than during crisis periods. In addition, our results have shown how systemic events over the past decade have influenced the US financial market, and how individual events have influenced the perception of systemic risk. Jointly, this empirical evidence warrants a time-dependent modeling of asset correlations.

### 3.4 Linking times of banking sector stress to large correlations

Our previous results documented both that asset correlations change dramatically over time and that correlations can become very large (close to the theoretical maximum 100%). This subsection studies whether periods with large correlations can be linked to times of banking sector stress.

Over the time period 2005 - 2015, figure 3 shows that the increased level of uncertainty following the Lehman crisis in fall 2008 is associated with a high level of correlations. Moreover, the impact of the initial news of the Greek debt crisis is evident in 2010. The high level of average asset correlations in 2012 indicate an extended period of increased uncertainty in the market. This could on the one hand be explained by the numerous negotiations of Greek rescue packages, on the other hand, this might also be caused by the US stock market
decline in fall 2011. Moreover, the significant correlation spike in 2013/14 may represent the increased level of perceived uncertainty caused by U.S. government getting close to the debt ceiling. These results are first evidence for periods of high asset correlations coinciding with systemic crisis events. It may be interesting to use periods of large correlations as an early warning system for periods of systemic crisis stress. The easiest way for this is to set a threshold and take crossing that threshold as a warning signal. Choosing an appropriate threshold is a difficult topic that warrants thorough discussion but is beyond the scope of this paper. Our interest here is merely in pointing out what could potentially be done and so we look at a threshold level of 80.62%. This threshold represents the 80% quantile of asset correlations in the U.S. market over the past decade. We depict the threshold as a dashed horizontal line in

---

In addition to the time period 2005-2015 we studied the 1980 - 2015 time period. There, a peak in average correlation in late 1987 represents the impact of the Black Monday on 19 October 1987. The high level of average correlation in 1990 can be explained by a recession in the U.S. and a higher level of uncertainty due to Iraq’s invasion of Kuwait causing an increase in oil prices. Furthermore, the events of 1994 are when the bond market suffered significant losses of more than US-$ 1 trillion. Finally, we see a double spike in (and larger) correlations in late 2002 and mid-2003; they coincide with a stock market crash October 2002 and the subsequent freezing of IPO markets for most of 2003. These results further confirm our analysis but the extended plot does not provide additional insights.
figure 3. It shows that asset correlations may present an adequate warning for immediate systemic events. However, the warning signal appears to be lagging. We attribute this to two shortcomings: first, the threshold may not be chosen appropriately; second and more important, we plot six-months lagging averages and this may have the correlations dragging behind. In addition, it appears that it is not the level itself that should be used as a warning signal but rather (strong) increases in correlations. Further analysis of these issues would be warranted but is beyond the scope of this paper.

4 Macro-prudential regulation

The goal of so-called micro-prudential regulation is to ensure that every bank is safe and sound individually. In addition to that, macro-prudential regulation is concerned about the banking system as a whole. This section introduces risk measures to quantify the overall impact of default linkages on the banking system and then studies their implications for prudential regulation.

4.1 Expected default frequency

We define the default frequency $M_N$ by setting

$$M_N = \frac{\sum_{i=1}^{N} X_i}{N}, \text{ where } X_i = 1_{V_i < K} = \begin{cases} 1, & \text{default} \\ 0, & \text{otherwise} \end{cases}. \quad (4)$$

This is a random variable that describes the fraction of banks that default. (This allows us also to describe easily the (total) loss across the banking system, since all banks share the same LGD; we refrain from studying this term in this paper, however.) The expectation fulfills:

$$E[M_N] = \frac{1}{N} \sum_{i=1}^{N} E[X_i] = \frac{1}{N} \sum_{i=1}^{N} p_i = p. \quad (5)$$

There are two important special cases for which it is straightforward to characterize (approximately) the distribution of $M_N$. Let us consider these two cases now to gain insight into risk across the banking system. Let us first assume for a moment that $N = \infty$ and $p \neq 0$;
then we interpret the limit \( N \to \infty \) as \( M_\infty \); it takes values on \((0,1)\) with the well-known density, see Vasicek (2002):

\[
\sqrt{\frac{1-\rho}{\rho}} \exp \left( -\frac{1}{2\rho} \left( \sqrt{1-\rho} N^{-1}(x) - K \right)^2 + \frac{1}{2} (N^{-1}(x))^2 \right) \text{ for } 0 < x < 1.
\]

Note that this density is the distribution underlying the measurement of large diversified credit portfolios, e.g. credit retail portfolios, and that it has entered as such into the Basel capital rules. These rules consider a correlation according to a specified formula with a maximum of 24% and derive a so-called Worst-Case-Default-Rate (WCDR) that is defined as the default rate that will not be exceeded with a pre-specified confidence level \( c = 99.9\% \). Based on this density Vasicek (2002) derives a closed-form expression for \( WCDR_{N=\infty} \) that has entered into the Basel regulation.

While we might want to apply this result directly to studying the banking system we want to point out two important observations that let us question such a straightforward application: first, Vasicek (2002) notes that the density is unimodal for \( \rho < 0.5 \), monotone for \( \rho = 0.5 \) and U-shaped for \( \rho > 0.5 \); this means that the density changes shape qualitatively and suggests that it is important to determine the actual size of the sensitivity. Our second observation, however, notes that the Basel rules take a maximum correlation of 24% but our empirical analysis in the previous section shows that it is at a minimum of 38% and that typical values are in the 70% range. While the Basel formula is likely a reasonable approximation for very large portfolios with small correlations, it is questionable that it is appropriate for a banking system of reasonable size (much smaller \( N \)) and much larger correlations.

To illustrate further the issues, let us now assume for a moment that the correlation \( \rho = 0 \) but \( N \) is finite. There are no default linkages and the risk-adjusted asset values are driven by idiosyncratic risk only. Under this assumption, the Law of Large Numbers applies and tells us that \( M_N \) converges to \( p \ P_{\rho,K} \)-almost surely. For reasonably large \( N \) this suggests that \( M_N \approx p \), i.e. the average default frequency is approximately the individual default probability. If we ignore the actual dependence on \( N \), we may view this as saying that the default frequency is characterized by the individual bank default probability \( p \). This in turn then suggests that prudential regulation should focus exclusively on individual banks.
micro-prudential regulation).

However, the reader may have noticed that our argument for micro-prudential regulation hinges crucially on a sufficiently large \( N \). From a quantitative perspective, we are interested in the result that for given size \( N \), the difference between \( M_N \) and \( p \) is “small.” This, however, is not the case. To illustrate this we continue to assume the special case \( \rho = 0 \) and invoke the Central Limit Theorem:

\[
M_N \sim \mathcal{N}(p, \sigma_N), \quad \text{where} \quad \sigma_N = \sqrt{\frac{p(1-p)}{N}}. \tag{6}
\]

Surprisingly, this is quite large in typical applications. For example, suppose that \( p = 1\% \), then with a fairly large number of banks \( N = 1000 \) we find \( \sigma_N = 0.3146\% \); this means that we can be confident with 99% probability that the default frequency is within the interval \([0.189\%, 1.811\%] = [1\% \pm 2.5758 \cdot 0.3146\%]\). A typical banking system is composed of a small number of very large banks and a large number of comparatively small banks; while the actual total number of bank entities in the US is larger than 5,000 from a legal perspective, from an economic perspective we believe that a much smaller number is relevant. With that in mind our calculation suggests that even with uncorrelated banks we can only be 99% confident that the default frequency is not larger than 1.8%, i.e. we can only be 99% confident that it does not become larger than twice the average (the individual) default probability. While our analysis puts a lower bound on this, we already find these numbers to be strikingly large from a regulatory perspective.

So far we have considered separately two special cases to gain intuition. It is now time to get to our initial problem with a finite size \( N \) and potentially non-zero correlation. To determine the density of the default frequency we proceed analogous Vasicek (2002) and denote for all \(-\infty < y < \infty\) by

\[
p(y) = P[X_i = 1|Y = y] = \mathcal{N} \left( \frac{N^{-1}(p) - \sqrt{\rho} y}{\sqrt{1-\rho}} \right) = \mathcal{N} \left( \frac{K - \sqrt{\rho} y}{\sqrt{1-\rho}} \right)
\]

the default probability conditional on the realization of the factor \( Y = y \). Note that this is independent of \( i \). For given \( N < \infty \), the default frequency \( M_N \) takes values in \( \left\{ \frac{i}{N} \right\}_{i = 0}^{N} \}. We have

\[
P \left[ M_N = \frac{i}{N} \right] = \binom{N}{i} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}(p(y))^i(1-p(y))^{N-i}dy. \tag{7}
\]
Figure 4: Density of default frequency $M_{N=1000}$; individual default probability $p = 1\%$.

Different to Vasicek (2002) we do not intend to study the limit $N \to \infty$ and, therefore, we cannot further simplify this. (The reader may be tempted to approximate the binomial distribution by a suitable normal distribution through a conditional application of the Central Limit Theorem; however, this leads to poor results for two reasons: (1) the typical values of $N$ that we study are much smaller than the $N = 1,000$ that we looked at above; (2) for correlated asset values it is inevitable to condition on the realization of $Y = y$ but the quality of the approximation is not uniform in the realization of $Y = y$.) Throughout, we calculate this out numerically.

Figure 4 plots the density of the default frequency for $N = 1,000$ and three different choices of $\rho = 0, 0.1, 0.2$. For $\rho = 0$ we see that the density is approximately normal and that it has approximately the standard deviation that we calculated above ($0.3\%$). Figure 4 also displays the default frequency for $\rho = 0.1$ and $\rho = 0.2$. We see that for these correlations, the density becomes skewed and has a considerable right tail, i.e. the probability density for large values of $M_N$ increases with larger correlations; put differently, such events are more likely as we increase correlations. Above we considered the 99% confidence level with a correlation of $\rho = 0$ already as “large”; given the larger tails for non-zero correlations, the level will likely increase further. Also, we recall from our discussion there that the typical size is much smaller than 1,000 from an economic perspective.
4.2 Macro-prudential risk measures

From the perspective of macro-prudential regulation, we are interested in the departures of $M_N$ from the micro-prudential reference value $p$, in particular in those values $M_N \geq p$, where the default frequency exceeds our micro-prudential expectations. For that purpose, we define the so-called *Conditional Expected Default Frequency (CEDF)* by setting

$$\text{CEDF} = E_{p,K}[M_N|M_N \geq p].$$

This measure describes the expected default frequency conditional on the default frequency exceeding the micro-prudential “reference” level. This is a coherent risk measure in the sense of Artzner et al. (1999) and we will use it throughout this paper as a measure of risk in the banking system as a whole.

To gain some understanding on this risk measure we fix again for a moment $\rho = 0$. Then we know from equation (6) that $M_N$ is approximately normal distributed such that the conditional distribution $P_{\rho=0,K}[M_N \leq m|M_N \geq p] - p$ is that of a so-called half-normal; this implies that $\text{CEDF} = E_{p=0,K}[M_N|M_N \geq p] \approx \sigma_N \sqrt{2/\pi} = \sqrt{\frac{p(1-p)}{N}} \sqrt{2/\pi}$. For example for $p = 1\%$, $N = 1000$, this gives the approximation $\sigma_N \cdot 79.8\% = 0.2511\%$ for the $\text{CEDF}$. This matches well our discussion in the previous subsection for $\rho = 0$.

In general, however, there is no closed-form approximation for our risk measure $\text{CEDF}$ and so we calculate it numerically using the description in equation (7). Figure 5 plots the conditional expected default frequency as a function of the correlation $\rho$; we plot this for three different choices for the size of the banking system: $N = 10$, $N = 20$ and $N = 1000$. Whatever the size $N$ of the banking system, we see that increasing the correlation increases the conditional expected default frequency, as expected. The dependency is highly non-linear: for $N = 10$ this can be seen clearly for $\rho$ starting at 0.7; for $N = 20$ we see this clearly (at least) for $\rho$ greater than 0.8; for $N = 1,000$ this shifts further to the right. Note also that increasing the size $N$, the conditional expected default frequency decreases, i.e. a more diversified banking system is more “stable”.

---

7Our measure is related to measures of systemic risk that have been considered, e.g. SRISK, Systemic/Marginal Expected Shortfall (Acharya et al. (2010)), and CoVaR (Adrian and Brunnermeier (2011)). For an overview on systemic risk measures we refer the reader to e.g. Upper (2011) and for a comparison to Benoit et al. (2013). A comparison with our measure is not straightforward.
In addition to conditional expected default frequency we study the usual Value-at-Risk (VaR) measure and the associated Conditional Value-at-Risk (VaR) measure (CVaR, a.k.a. expected tail loss). Vasicek (2002) studied the former measure for an infinitely large credit portfolio which is mathematically similar to an infinitely large banking system ($N = \infty$). For that purpose we refer to the former as WCDR$_N$ and to the latter as EWCDR$_N$. Formally, they are defined using a confidence level $0 < c < 1$ by setting

$$WCDR_N = \min \{\theta | \text{Prob}[M_N \leq \theta] \leq 1 - c\} \text{ and } EWCDR_N = E[M_N | M_N \geq WCDR_N].$$

Figure 5 plots the WCDR in Panel (a) and the EWCDR in Panel (b) with a confidence level of 99%; both plots show these as a function of the sensitivity $\rho$ for three different choices for the size of the banking system ($N = 10, N = 20$ and $N = 1000$). We see that for “small” numbers $N = 10, 20$, WCDR jumps while for $N = 1000$ it becomes a smooth line. This is essentially as expected but is an inconvenient feature in such plots. In addition, these jumps show up in the EWCDR plot, although they tend to be much smaller.

Note that EWCDR is a coherent risk measure in the sense of Artzner et al. (1999), since it corresponds to a form of expected tail loss. However, WCDR is not a coherent measure since it corresponds to the VaR concept. Therefore, we focus on EWCDR. When we compare EWCDR with CEDF in figure 5 we see that both show results that are quantitatively
Figure 6: Macro-prudential risk measures based on VaR concept; confidence level \( c = 99\% \) comparable. Since the CEDF measure is easier to understand, easier to implement and numerically more stable, we adopt the CEDF risk measure for further analysis in this paper.

### 4.3 Systemic capital buffer

Our goal is to study the role of capital rules in prudential regulation. In prudential regulation it is common to look at risk-adjusted assets; the advanced measurement approach permits this to be based on the Value-at-Risk (VaR): for market risks, calculations are based on the 99% quantile at the two-week horizon; for credit risks, calculations are based on a 99.9% quantile at the one-year horizon. Minimum equity to be held (as required by the regulator) is VaR multiplied by a factor between 3 and 4, depending on the performance of risk management system.

For simplicity, we focus our discussion on the variance-covariance approach (normal distribution approximation) and a multiplicative factor of 3. This means that (at a minimum) \( E_i = 3 \cdot \text{VaR}_i \), where \( \text{VaR}_i = 2.326 \cdot \sqrt{\frac{2}{52\Delta t}} \text{Var}(\tilde{A}_i) \) describes the Value-at-Risk of bank \( i = 1, \ldots, N \). (Recall that \( \Delta t \) is the (physical) time-difference in years between both dates. Here, the division by \( \Delta t \) is used to annualize variance and multiplication by \( 2/52 \) serves to scale down to a horizon of two weeks.) Based on equation (1) we know that \( \tilde{A}_i = A_i \exp(\sigma_i V_i) \), where \( V_i \) is a standard normal random variable. This implies that
\[ \text{Var}(A_i) = A_i^2 \exp(\sigma_i^2) (\exp(\sigma_i^2) - 1) \approx A_i^2 \sigma_i^2. \] (The approximation uses the property that \( \sigma_i \) is a small number.) Then we have

\[ E_i = 3 \cdot 2.326 \sqrt{\frac{2}{52\Delta t}} \sigma_i A_i. \] (8)

Moreover, we know that in our structural model default of bank \( i \) occurs if, and only if, \( \bar{A}_i < D_i \). Using once more the property \( \bar{A}_i = A_i \exp(\sigma_i V_i) \) we then find that default occurs if, and only if, \( \sigma_i V_i < \ln(D_i/A_i) \). Typically, the equity portion in banks is small and so we use the common approximation \( \ln x \approx x - 1 \) for \( x \approx 1 \). For simplicity, we treat this as an approximation and find that default occurs if, and only if, \( \sigma_i V_i < \ln(D_i/A_i) \). (Here we used the balance sheet \( D_{it} + E_{it} = A_{i,t} \).) Comparing this with equation (2) tells us that

\[ K_i = -3 \cdot 2.326 \sqrt{\frac{2}{52\Delta t}} \text{ and that } E_i = -K_i \sigma_i A_i. \] (9)

This motivates us to call the term \( K_i \) the risk-adjusted capital ratio right from the start in section 2.

In the Basel accords, a comparable systemic capital supplement has been implemented in the so-called systemic risk charge. This additional capital buffer – currently set at up to 3.5%-points in the worst bucket, see Basel Committee on Banking Supervision (2011) of the Basel Committee on Banking regulation – has to be held by banks considered systemically important. Let us denote \( q_i^* = E_i^* / A_i \) the additional capital charge related to the additional equity \( E_i^* \) to hold; similarly, let us denote \( q_i = E_i / A_i \) the common (micro-prudential) capital charge which relates to the common (micro-prudential) equity \( E_i \) to hold. For discussions, it is more appealing to think of the systemic risk charge as a relative add-on, i.e. the relative systemic capital supplement is defined as \( q_i^* / q_i \). Assuming the equity to hold corresponds to the 8% set forth under the Basel accords, the respective relative systemic capital supplement is currently set at a maximum of 3.5%/8% = 0.4375.

Equation (9) tells us that we

\[ \frac{q^*}{q} = \frac{E^*}{E} = \frac{K^*}{K_i}. \] (10)

This tells us that we need to determine the ratio \( \frac{K^*}{K_i} \). Note that we do not intend to justify the so-called Cooke-ratio of 8% in the Basel accord; instead we derive the capital to hold directly from a VaR calculation.
Figure 7: Relative capital supplement $q_t^\rho/q_t$ for macro-prudential regulation, as a function of correlation $\rho$.

From the perspective of micro-prudential regulation the Basel accord translates into a default probability

$$p = p_i = P_{\rho,K}[V_i < K_i] = \mathcal{N}\left(-3 \cdot 2.326 \sqrt{\frac{2}{52\Delta t}}\right)$$  \hspace{1cm} (11)

at the time-horizon $\Delta t$ between both dates. For macro-prudential regulation, we fix a (macro-prudential) target probability $p^*$ and look for a capital supplement $q^*$ that ensures

$$E_{\rho_{o,K^*+K}}[M_N|M_N \geq p] = p^*.$$  \hspace{1cm} (12)

The distribution of $M_N$ is here driven by an additional parameter $K^*$ in a way that $K^* + K$ that replaces $K$. The probability $p$ is set by equation (11) and unaffected by the choice for $K^*$.

In implementations we do not set the time-horizon $\Delta t$; instead we note a one-to-one relationship with $p$ in equation (11) and set first the individual default probability $p$ and then derive $K$ accordingly. Figure 7 studies the relative systemic capital supplement for $p^* = 1\%$ when we set $p = 1\%$; it plots this as a function of the correlation for three different sizes of the banking system: $N = 10, 20, 100$. We see that an increase in the number of banks leads to a smaller relative capital supplement; this is essentially some form of diversification.
For a banking system with a fairly large number \( N = 100 \) we see for the typical correlation value \( \rho = 75\% \) that the relative capital supplement is approximately \( 1/4 \); this means that the absolute capital supplement in relation to the current 8\% for micro-prudential risk should be an add-on of roughly 2\%-points. In crisis periods where the correlation runs up to 90\% our calculations suggest a capital supplement of about \( 1/3 \) relative to the 8\% Cooke ratio, which translates into an absolute add-on of roughly 2.5\%-points. Currently there is no bank in the worst bucket in terms of the systemic risk buffer, i.e. 2.5\%-points is currently the highest systemic risk buffer imposed to banks, in practice. Our add-on thus appears quite in line with the systemic risk capital buffer under the new Basel accords.

However, the new Basel accord only considers a number of SIFIs that is much smaller than \( N = 100 \). Here, the appropriate size is \( N = 20 \) or even \( N = 10 \); then we expect the capital charges to be much larger than those suggested so far. For \( N = 20 \) we find that with \( \rho = 75\% \) (\( \rho = 90\% \)) the relative capital supplement is approximately 0.5 (0.6), translating into an add-on of 4\%-points (4.8\%-points). For \( N = 10 \) we find that with \( \rho = 75\% \) (\( \rho = 90\% \)) the relative capital supplement is approximately 0.7 (0.85), translating into an add-on of 4.8\%-points (6.6\%-points). In both these cases the additional capital charge for macro-prudential risk is much larger than that under the new Basel accord, not only with the typical correlation value but even for important with the correlation value in crisis periods (that we are particularly concerned about).

5 What is the Optimal Number of Banks?

The previous section studied risk measures based on the default frequency for a banking sector and showed that they are decreasing in the number of banks considered, i.e. systemic risk of a banking sector consisting of a larger number of small banks is lower. In fact, a banking sector with a large (from a mathematical perspective, an infinite) number of (infinitely small) banks is most favorable. Yet, a smaller banking sector (i.e. a banking sector consisting of a small number of comparatively large banks) may also prove beneficial for individual banks’ health. For example, a smaller banking sector is generally characterized by a lower level of competition, and banks in this market are often able to gain higher overall
profits similar to a monopoly rent. This section studies the trade-offs involved to determine the optimal number of banks\(^8\).

Throughout, we assume that whenever a bank enters/leaves the market, total assets are redistributed to once more ensure homogeneity in asset size, i.e. throughout, we consider a banking sector with equal sized banks. These assumptions enable us to quantify the impact of changes in the total number of banks and thereby bank size for given total assets.

### 5.1 Efficiency Gains

Regarding the relation of profits and different levels of bank sector size, we focus on the net interest margin (NIM) as one of the main asset-based income sources for credit institutions. For example, in a small banking sector with only a few banks offering their services, each individual bank will be able to set higher (lower) interest rates on loans (deposits), and thereby increase their NIM. A higher NIM in combination with increased profits in a monopoly / oligopoly like market strengthens individual banks’ earnings, in turn strengthens their health, and jointly increases overall bank sector stability (see e.g. Keeley (1990), and Beck et al. (2006))\(^9\).

Akhavein et al. (1997) analyze the impact of mega-mergers on efficiency and prices in the banking sector. They focus on U.S. banking organizations with at least $1$ billion of assets in at least one year between 1980 and 1990, and define mega-mergers as mergers between at least two banks having at least $1$ billion in assets each. To determine the efficiency gains in these mega-mergers, Akhavein et al. (1997) calculate a frontier profit function for each bank (see also Berger et al. (1993)) and then determine the total profit efficiency ratio as actual profits over maximal profits. For mega-mergers, results show an increase in the asset-weighted average from 44% pre-merger to 71% post-merger (increase of 27 percentage points). The respective peer groups of banks experience an increase in average total profit efficiency ratio of 10 percentage points over the same sample period (pre- and post-merger).

---

\(^8\)Following the Lehman default, questions relating to optimal bank size as well as optimal bank sector design / size have been raised (see e.g. Haldane (2012)). Our analysis is closely related to these considerations.

\(^9\)In addition to these profit considerations, Mishkin (1999) argues strongly in favor of bank consolidation focusing on potential economies of scale and cost-effective structures. A similar argument is made in the G10 report on the consolidation of the financial sector highlighting potential diversification gains (see Group of 10 (2001)).
Akhavein et al. (1997) interpret these results as an average 17% efficiency increase for merged institutions.

We interpret this result as saying that efficiency gains express themselves by a multiplicative factor \((1 + 17%)\) in the total profit efficiency ratio of merged institutions. The total profit efficiency ratio is defined as actual profits over maximal profits. Since we do not have further information on maximal profits, we assume this 17% increase to occur in actual profits. Moreover, we then suggest - in first application - to replace total profits by the NIM. This leads us to translate the efficiency increase as the NIM of merged institutions by a multiplicative factor \((1 + 17%)\).

There are currently 68 U.S. banks with more than US-$ 15 billion in total assets representing more than 90% of all assets in the sector. These 68 large banks have a NIM of 2.94\% (according to the Federal Reserve System). If 2 of these 68 financial institutions were to merge, the respective entity from this mega-merger might benefit from additional efficiency gains of 17\% compared to all other banks as documented by Akhavein et al. (1997). Following the above lines of thought, we assume this merged entity to increase the NIM to 2.94\% \cdot 1.17 = 3.44\%. In the banking sector now consisting of 67 financial institutions, the average NIM could then be determined as a weighted average of the NIM of 2.94\% for 66 financial institutions and the NIM of 3.44\% for one financial institution. The average NIM for the whole banking sector would then equal 2.95\%. For different bank sector concentration, the respective NIM under a first consideration of potential efficiency gains could be calculated as

\[
\text{NIM}_{N-1} = \frac{N-2}{N-1} \cdot \text{NIM}_N + \frac{1}{N-1} \cdot \text{NIM}_N \cdot (1 + 17\%).
\]

This presentation of NIM as a function of the number of banks can be approximated by

\[
\text{NIM}_N = 6.17\% \cdot N^{-17.66\%}.
\]

The relation between potential NIM and bank sector size is then as follows: In a smaller bank sector, competition between banks decreases allowing NIM to increase. In contrast, in a large bank sector, banks face higher competition which lowers their rent and thus allows for lower NIM.
5.2 Net impact of efficiency gains and systemic costs

Both the NIM as well as the overall losses associated with systemic events are based on total assets (in the banking sector); we now analyze the net effect of these opposite effects of bank sector size, i.e.

\[ \Delta_N = \text{NIM}_N - \text{CEDF}, \]

where the conditional expected default frequency \( \text{CEDF} \) of the previous section is based on a probability \( p = 1\% \). Figure 8 shows the net effect \( \Delta_N \) for different number of banks \( N = 1, \ldots, 60 \). Since the conditional expected default frequency \( \text{CEDF} \) depends on the level of asset correlation, we plot results for three different levels for the U.S. banking sector (as derived in section 3.3): \( \rho = 38.31\%, 70.99\%, 87.70\% \). These three choices are motivated by our analysis in section 3; they correspond to the minimum, average and maximum level of the correlation in our PCA analysis of subsection 3.3.

Whatever the correlation in figure 8, the net impact initially increases as we increase the number of banks \( N \) but ultimately decreases from thereon as we further increase \( N \). Overall, this figure shows that there is a maximum, an optimal number \( N^* \) of the banking system. In addition, for small numbers \( N \), the net impact \( \Delta_N \) is negative, but becomes positive and at the optimal number \( N^* \) the net impact is positive.
Figure 9 shows the optimal number of banks $N^*$ as a function of the correlation $\rho$. As the correlation $\rho$ increases, the optimum number increases. For asset correlations up to 78%, results indicate an optimal bank sector size of 20 or less. Yet, once asset correlations are larger than 78%, the optimal number of banks increases significantly. At 85%, the optimal number is 46 banks, close to correlations of 90%, a number $N = 100$ banks in the bank sector is optimal. (At the individual default level, this is the maximum as we cannot distinguish discrete numbers of default further.)

We can interpret figure 9 further in light of our discussion in section 3. For the “minimum” asset correlation of 38.31%, that we see in that section, the optimum is at $N^* = 4$ banks; for the “average” asset correlation of 70.99%, figure 9 shows that an optimum is reached for a banking system consisting of 11 banks. For the “maximum” asset correlation of 87.70%, the optimum is for a banking system of 84 banks.

At the average value of asset correlation (70.99%) and for a banking sector with 11 banks, the profits through the NIM prevail the associated expected losses most. Thus, during the past decade and from a net profit point of view, a banking system with only 11 individual banks would have been most profitable for the overall banking system. However, the past decade has also seen significant systemic events, and during such times periods, the level
of asset correlation may change dramatically. During such periods, asset correlations are fairly large (in the sample considered in section 3, the maximum lies at 87.70%) increasing the probability for large losses in the banking system. Since it is advantageous to diversify these potential losses, a banking sector with a large number of banks is optimal. Here, the potential profits through NIM have a lower impact as a smaller bank sector (in our example of 11 or fewer banks) would actually lead to a negative net effect.

The opposite reasoning holds during time periods of low uncertainty and low systemic risk levels. Then, the correlation among banks is comparatively small (at 38.31%). This low correlation leads to lower expected losses since the default linkage is weaker in this banking system. Hence, the incentive of profit maximization in NIM can be applied and a smaller bank sector is optimal. The optimum of 4 can even be found in the current distribution of assets in the U.S. bank sector: The four largest banks in the U.S. have total asset of more than US-$1,000 billion each. While a high level of concentration (like \( N = 4 \)) can be found in the market, it appears unreasonable from our net impact analysis: based on the typical correlation, the systemic risk component is too large to justify such a concentrated banking system.

These results highlight once more that a calibration at the average level of asset correlation may be misleading by underestimating the impact of systemic events during crisis periods. This implies that, on average, a bank sector with a larger number of banks (less concentrated) may be considered advantageous as it diversifies potential losses over several banks while still allowing for a positive net effect even in times of crisis.

6 Conclusion

This paper studied a structural model of default; default linkages arise from correlation in the underlying asset values. Our approach convinces through its easy and straightforward implementation. Previously, an analogous setup has been studied to describe risk in large credit portfolios (typically retail credit), but our application differs in two important directions from a quantitative perspective: First, the number of banks is small and we showed that this is quantitatively very different from an application of the Central Limit Theorem.
(infinite number of banks). Second, the typical correlation value is empirically closer to 1 than to 0 and thus quantitatively very different from the typical value in the analogous credit portfolio application. We further pointed out that this is highly relevant by showing that the correlation affects the density, risk measures and associate capital in a highly non-linear way. We motivated macro-prudential regulation in addition to micro-prudential regulation in that way. Our empirical applications also documented the time-dependent nature in asset correlation and linked time periods of particularly large correlation to stress periods; ultimately, this approach may lead to a (forward-looking) stress indicator for bank regulators. Finally, we applied our setup to the question of the optimal number of banks (optimal asset concentration) trading off diversification arguments (systemic risk considerations) against efficiency considerations and characterized the optimal bank concentration.

References


Duan, J.-C., Zhang, C., 2013. Cascading defaults and systemic risk of a banking network, National University of Singapore.


