René M. Stulz

Risk Management, Governance, Culture, and Risk Taking in Banks

1. Introduction

The Oxford Dictionary defines risk as a situation that involves exposure to danger. It also states that the word comes from the Italian word risco, which means danger. I call risks that are only danger bad risks. Banks—and any firm for that matter—also have opportunities to take risks that have an ex ante reward on a standalone basis. I call such risks good risks.¹

One might be tempted to conclude that good risk management reduces the exposure to danger. However, such a view of risk management ignores the fact that banks cannot succeed without taking risks that are ex ante profitable. Consequently, taking actions that reduce risk can be costly for shareholders when lower risk means avoiding valuable investments and activities that have higher risk. Therefore, from the perspective of shareholders, better risk management cannot mean mean risk management that is more effective at reducing risk in general because reducing risk in general would mean not taking valuable projects. If good risk management does not mean low risk, then what does it mean? How is it implemented? What are its limitations? What can be done to make it more effective? In this article, I provide a framework to understand the role, the organization, and the limitations of risk management in banks when it is designed from the perspective of increasing the value of the bank for shareholders.

In corporate finance, the well-known Modigliani-Miller theorem of leverage irrelevance implies that the value of a firm does not depend on its leverage. For the theorem to hold, markets have to be frictionless, so there cannot be transaction costs of any kind. As has been stressed by modern banking research, there is no reason for banks to exist if the conditions of the Modigliani-Miller theorem hold. With the Modigliani-Miller theorem, a bank has the same value whether it is mostly financed by debt or mostly financed by equity. Hence, the value of a bank is the same irrespective of its risk of default or distress. It follows that if the conditions for the Modigliani-Miller theorem apply, a bank has no reason to manage its risk of default or its risk of financial distress (see, for example, Stulz [2003]).

When the Modigliani-Miller theorem does not apply, the most compelling argument for managing risk is that adverse outcomes can lead to financial distress and financial distress is costly (Smith and Stulz 1985). When a firm is distressed, it loses its ability to implement its strategy effectively and finds it more difficult and expensive to conduct its business. As a result, the value of a firm’s equity is reduced by the present value of future costs of financial distress. When a

¹ For a related useful taxonomy, see Kaplan and Mikes (2012). The authors distinguish between preventable, strategic, and external risks and show that the role of risk management differs across these types of risk.
firm manages risk so that it reduces the present value of these future costs of distress by more than the cost of reducing risk, firm value increases. Banks differ from firms in general because they create value for shareholders through their liabilities as part of their business model. Banks produce liquid claims and the value of a bank depends on its success at producing such claims. For instance, the value of a bank depends on its deposit franchise. A bank’s ability to issue claims that are valued because of their liquidity depends on its risk, so that risk management is intrinsic to the business model of banks in a way that it is not for nonfinancial firms (DeAngelo and Stulz 2015).

Since an increase in risk can enable a bank to invest in assets and projects that are valuable but can also lead to a loss in value because of an adverse impact on the bank’s risk of financial distress and its ability to create value through liabilities, there is an optimal amount of risk for a bank from the perspective of its shareholders. A well-governed bank will have processes in place to identify this optimal amount of risk and make sure that its actual risk does not differ too much from this optimal amount. Theoretically, the bank’s problem is simple: it should take any project that increases its value, taking into account the costs associated with the impact of the project on the bank’s total risk. But in practice, the bank’s problem is difficult because risk-taking decisions are made all the time throughout the bank and each decision affects the bank’s probability of financial distress to some degree. As a result, risk-taking decisions cannot be evaluated in isolation but must be assessed in terms of their impact on the overall risk of the bank.

In principle, if there is an optimal level of risk for a bank, the cost of taking on a new risk that increases the bank’s total risk should be traded off against the potential gain from taking the risk. However, ignoring hedges, it would never make sense for a bank to take a risk that destroys value as a standalone risk. We call such risks bad risks. They correspond only to danger. An example is a trader who writes underpriced deep-out-of-the-money puts because he believes that, if the puts are exercised, he will not receive a bonus anyway, while if they are not exercised, his bonus will be higher. Such a purchase is a negative net present value project for shareholders as a standalone project since the firm sells an asset for less than it is worth. Writing an overpriced put would be a positive net present value project on a standalone basis. Hence, such a risk would be a good risk. However, writing this option creates risk for the bank that may or may not be worth it given its total risk and the costs associated with its total risk. With our examples, both the bad risk and the good risk increase the bank’s total risk. While it is clear that taking the bad risk makes no sense for the bank, we cannot determine whether it makes sense for it to take the good risk by considering the good risk on a standalone basis. This is because taking the good risk increases the total risk of the bank.

At a point in time, how the risk of a project contributes to the total risk of the bank depends on the other risks the bank is exposed to at that time. Consequently, when risk taking is decentralized, the trade-off between how a project’s risk contributes to the bank’s risk and its expected return cannot be made in real time for most risk-taking actions because the project’s contribution to the bank’s value and its risk depends on the bank’s total risk at that time. Instead, a shortcut is typically used, which is to focus on risk separately (ignoring return) and manage the overall amount of risk of the bank by imposing limits on the risk that can be taken by units of the bank and/or by charging units for the risks they are taking. The risk management function in a bank measures and monitors risk taking by a bank’s units to ensure that their risk remains within prescribed limits and that the bank has the right amount of risk. A bank’s risk management function is generally called a bank’s risk management, and I follow that language. Unfortunately, focusing separately on risk has the potential to destroy value if not done well when it leads the bank to reject projects that are valuable for the institution despite their risk.

There are two fundamentally different ways that a bank’s risk management can destroy value. First, risk management can fail to ensure that the bank has the right amount of risk. This failure can come about for a number of reasons. In particular, risk management can fail to uncover bad risks that should be eliminated, it can mismeasure good risks, and it can fail in its task to measure the firm’s total risk. Second, risk management can be inappropriately inflexible, so that increases in risk are prevented even when they would be valuable to the institution. When risk management becomes too inflexible, it destroys value because the institution no longer has the ability to invest in valuable opportunities when they become available, and it also becomes less effective in making sure that the firm has the right amount of risk. The reason is straightforward: as risk managers become policemen, they are viewed within the institution as an obstacle rather than as partners in creating value. Striking the right balance between helping the firm take risks efficiently and ensuring that employees within the firm do not take risks that destroy value is a critical challenge for risk management in any bank.

In this article, I first discuss the determinants of a firm’s optimal risk level in general, and then I turn to banks. In Section 3, I examine the role of governance and risk management in helping a bank achieve its optimal risk level. I offer an analysis of the determinants of the organization of risk management in Section 4. I assess the tools used by risk
management to ensure that the bank does not take on an excessive amount of risk in Section 5. In Section 6, I show that the limitations of the tools used by risk management create an important role for incentives and for a firm’s culture. Section 7 presents my conclusions.

2. Determining the Risk Appetite

In a market economy, there are compelling reasons for corporations to be run to maximize shareholder wealth. These reasons apply to banks as well. However, no corporation maximizes shareholder wealth in a vacuum. In particular, corporations are constrained in their actions by laws and regulations. Laws and regulations play a special role with banks because bank failures and weaknesses can have damaging effects on the financial system and the economy. If a bank is managed to maximize shareholder wealth, it will choose a level of risk consistent with that objective. A bank with too much risk could not conduct its business even if regulators allowed it to do so. Such a bank would find it hard to fund itself. While deposit insurance guarantees depositors against losses, it does not guarantee that they have continuous access to their deposits. Further, many short-term liabilities of banks are not insured. To the extent that safe and liquid deposits are a source of value for banks, too much risk will limit a bank’s ability to supply safe and liquid deposits and hence will adversely affect the value of the bank.

Some borrowers may have no reason to care if the bank they borrow from is too risky, but others will care. Borrowers who rely on their relationship with the bank could see that relationship jeopardized or lost if the bank becomes distressed or fails. They might therefore seek to borrow elsewhere rather than deal with a risky bank. If the bank is in the derivatives business, counterparties will be leery of dealing with it if it is too risky. The bank might also find it difficult or expensive to hire employees because potential employees will be reluctant to make bank-specific human-capital investments in a bank that is too fragile.

These and other reasons can explain why a bank that is too risky is worth less. At the same time, however, a bank that has no risk whatsoever might not be worth much either. Of course, if a bank could find valuable projects whose value it could capture without having to bear the risks, perhaps because it could perfectly hedge all those risks, that bank would have considerable value already and might not be able to increase its value by taking risks. In practice, however, banks cannot eliminate all risks through hedging and diversification. Hence, they have to take some risks to create wealth for their shareholders.

There are many ways to define risk. Shareholders who hold diversified portfolios have no reason to care about the volatility of the return of a stock in their portfolio on a standalone basis. They only care about the volatility of their portfolios. If a stock’s volatility increases so that shareholders’ portfolios become more volatile, shareholders can change their asset allocations. Hence, the risk that shareholders care about when they consider a bank is risk that makes the bank worth less than it would otherwise be worth. For risk to affect shareholder wealth, it has to affect future cash flows or the rate at which these cash flows are discounted. The possibility of unexpectedly low cash flows in the future that would make the bank distressed will reduce the value of the bank now because the market will adjust its value for the possibility that the bank will incur distress costs. These costs arise because the bank is no longer able to execute its strategy. Hence, the loss to shareholders is the loss that arises when the bank cannot implement its strategy. Viewed from this perspective, the risk that has to be managed to maximize shareholder wealth is the risk of financial distress.

For now, I will assume that the risk of financial distress is appropriately captured by the bank’s credit rating. Given the previous discussion, the optimal rating of a bank is generally not the highest rating, AAA, but some other rating. This is because, typically, achieving a AAA rating requires the bank to give up too many valuable risky projects. Suppose that a specific bank’s value is at its highest when the bank is given an A rating. An A rating essentially corresponds to a very low probability of default. From 1981 to 2011, the annual average default rate for A-rated credits was 0.08 percent, according to Standard and Poor’s. Hence, by targeting a specific probability

2 See, for instance, Poloncheck, Slovin, and Sushka (1993) for evidence that corporate borrowers are affected adversely when their relationship bank becomes distressed.

of default, the bank achieves its desired level of risk. For that institution, a higher rating than A will necessarily limit its activities so that it would have to give up projects. A lower rating than A might make it impossible for the bank to keep engaging in value-creating activities. This might be the case, for instance, if potential counterparties are not willing to transact with it if it has such a rating.

A bank with more of a deposit franchise and with more relationship lending is likely to prefer a higher rating than an institution that is engaged in more transactional activities. Similarly, a bank that enters into long-term derivatives contracts might find a higher rating more valuable than one that does not. Consequently, the rating that maximizes bank value differs across banks. The exhibit above shows the relationship between ratings and bank value for two different banks, Bank Safe and Bank Risky. In both cases, the relationship is concave, so that there is a maximum value. However, in the case of Bank Safe, firm value falls steeply if the bank is riskier than its target rating and increases only moderately as it increases its risk toward the target rating. For Bank Risky, the relationship between bank value and rating is substantially different. Its target rating is BBB and its value rises significantly as it increases its risk toward its target and falls sharply if it exceeds it. For both banks, having too much risk is extremely costly in terms of their value. However, for one bank, having too little risk has little cost, while for the other it has a large cost.

The relationship between bank value and risk presented for Bank Safe and Bank Risky in Exhibit 1 is sharply different from the relationship that would prevail if the Modigliani-Miller leverage irrelevance theorem applied to banks. In the Modigliani-Miller case, bank value would be the same irrespective of the bank’s risk of default or of financial distress. In other words, the bank could achieve exactly the same value if its rating were AAA or CCC. The reason for this is straightforward. If the Modigliani-Miller theorem applies, the firm can always alter its leverage at zero cost and hence achieve a specific rating through changes in leverage—for instance, by issuing equity and investing the proceeds in fairly priced risk-free securities. Since changing leverage has no impact on value when the Modigliani-Miller theorem applies, it follows that there is no relationship between bank value and risk of default in that world.

If the Modigliani-Miller theorem applies, decision making in a bank can be decentralized as long as new projects do not have an adverse impact on existing projects. If new projects do not affect the value of existing projects, it is optimal for the bank to take all projects that create value on a standalone basis. However, if there is an optimal level of risk for the bank as a whole, a new project necessarily has an impact on other projects because it changes the bank’s aggregate level of risk and hence changes its own value through its impact on the risk of the bank. Consequently, fully decentralized decision making cannot be optimal when the Modigliani-Miller theorem does not apply and there is an optimal level of risk for a bank.

With the approach presented so far, bank value is highest if the bank achieves a specific target rating that depends on characteristics of the bank, such as its strategy and business model. But in practice, not all banks are rated. I have focused on a rating as a measure of risk because it is intuitive. However, a rating corresponds to a probability of default, and a bank that does not have a rating can still figure out the probability of default that is optimal. Obviously, banks might choose to tailor their risk in a more complex way. They might want to specify how they are affected by specific shocks. For instance, a bank might choose to set a level of risk such that it can survive a major recession with only a one-notch downgrade. An obvious difficulty with multiple constraints on a bank’s risk is that these constraints might be inconsistent and their impact on bank value might be hard to assess. At the same time, however, multiple constraints can be advantageous in that they could make it more likely that a bank will be well positioned following adverse shocks.

A bank’s risk appetite is the result of an assessment of how taking on more risk affects the opportunities that the bank can capitalize on. This assessment can change as the bank’s opportunities change. Consequently, a bank’s risk appetite cannot be inflexible. At the same time, however, the risk appetite is not determined in such a precise way that a small shift in opportunities will affect it.
Banks differ from other firms because their failure can have systemic effects. If a producer of widgets fails, as long as there are other producers of widgets, the impact on society will be extremely limited and will be immaterial for most. The same is not true if a large bank or a group of smaller banks fails. While it is important for society to limit the systemic risk that a bank creates, there is no a priori reason that a bank that has less systemic risk is worth more for its shareholders. It follows that a bank that maximizes its value for its shareholders may have an amount of systemic risk that is excessive from the perspective of society.

Because of the role of banks and the consequences of bank failures, regulators impose restrictions on banks’ ability to take risks on the asset side and they require banks to satisfy minimum capital requirements. As a result, each bank's systemic risk is reduced. These restrictions and requirements also mean that a bank chooses its level of risk subject to constraints. However, these constraints do not change the bottom line, which is that there is an optimal level of risk for a bank and this optimal level of risk differs across banks depending on the nature of their business. Because the optimal level of risk differs across banks, the costs to shareholders of constraints imposed by regulators are not equal across banks. For instance, Boyson, Fahlenbrach, and Stulz (2014) show that banks with high franchise value have incentives to choose low-risk strategies, so that for such banks, capital requirements are unlikely to be constraining.

In the framework of Section 2, there is, for each bank, a level of risk such that the value of the bank is maximized for shareholders. This level of risk is not zero. Good governance should ensure that the firm chooses this level of risk. This means making sure that the firm has processes in place that enable it to measure its risk, understand how firm value is related to risk, and maintain the right level of risk.

An obvious concern for shareholders is that management might do a poor job managing the firm’s risk or might have incentives to take risks that are not in the interest of shareholders. To alleviate this concern, the board has to ensure that the firm has the capability to measure and manage risk so that it has the right level of risk given its risk appetite, and has to ensure that it uses this capability effectively so that it actually takes the right level of risk. This means that the bank should have a risk management organization in place capable of making sure that it has the right level of risk. I discuss risk management organizational issues in the next section.

An important governance issue is that the bank’s board of directors has to have enough expertise to assess management’s efforts in measuring and managing risks. Understanding whether a firm takes the right risks is a rather complex and technical task. Even if the board has the proper expertise, it may be difficult for it to develop such an understanding. While boards require an external assessment of a firm’s accounting, they do not typically require such an assessment of what is effectively a firm’s risk accounting (though auditors may comment on various aspects of risk management). It would seem that risk audits might be valuable tools in helping the board reach the proper level of comfort that management is handling a bank’s risk properly.

An important implication of this view of risk governance is that good risk governance does not mean less risk. In fact, it could well be that management, left to itself, would choose for the bank to have too little risk rather than what is best for shareholders. Good governance means that the bank has the right amount of risk for its shareholders. This amount of risk may not be the amount that is appropriate from the perspective of society as a whole because shareholders may not have the proper incentives to take into account the externalities

3. Governance and Risk Taking

In Section 2, I presented a risk appetite framework from the perspective of the bank’s shareholders. Good governance means that shareholders get the maximum benefit from their ownership of the firm (Shleifer and Vishny 1998). With banks, regulation is a constraint that shareholders have to meet. Given the constraint, shareholders still want to maximize their wealth, and hence a well-governed bank should have mechanisms in place so that the level of risk chosen by management maximizes shareholder wealth subject to the constraints imposed by regulation. In this section, I address key trade-offs that must be made when designing a firm’s risk governance. This section is not meant to address general governance issues in banking, since excellent reviews of those issues already exist (Mehran, Morrison, and Shapiro 2011; Mehran and Mollineaux 2012; de Haan and Vlauh 2013) and the topic goes beyond the risk issues I am focused on.
properly measured beforehand. Alternatively, it could be while for them ex ante.

To make the issues clearer, consider the situation where it is optimal in terms of shareholder value to increase the risk of a bank. This greater risk may make the bank more fragile but also more valuable. If an adverse realization of the increased risk taken by the bank leads it to become distressed, this can have an adverse impact on other banks that are counterparties of the bank. For instance, a default by the bank could mean that other banks sustain losses on unsecured obligations from the defaulting bank. As these other banks sustain losses, they become financially weaker and potentially endanger the

Because the optimal amount of risk from the perspective of shareholders need not be the optimal amount for society, it would be wrong to believe that somehow better governance makes banks safer. It can make them more valuable but also riskier.

stability of the financial system. A bank maximizing shareholder wealth will take into account the potential impact of its actions on the financial system only to the extent that they affect its value. This means that the bank is likely to take too much risk from the perspective of society because it will ignore the impact of that risk on society beyond what is reflected in its value. For instance, the fact that a failure of the bank could lead counterparties of its counterparties to fail will be a cost that has little impact on the value of the bank but may have considerable impact on the safety of the financial system. Hence, to make sure that banks take proper account of the impact of their actions on the financial system, constraints have to be put on the actions they can take and/or taxes have to be imposed on actions that are costly to the financial system.

Existing empirical research does not seem to support the proposition that better governance in banks leads to less risk. The credit crisis provides a natural experiment for testing this proposition. If it were correct, we would expect better-governed banks to be less affected by the crisis since they would have been less exposed to risks that manifested themselves during the crisis, assuming these risks were properly measured beforehand. Alternatively, it could be that the risks were not or could not be properly assessed in advance. In any case, there is no evidence suggesting that better-governed banks performed better during the crisis.

Specifically, research examines four dimensions of governance. First, evidence shows that banks with boards that were more shareholder friendly performed worse than other banks, not better (Beltratti and Stulz 2012; Erkens, Hung, and Matos 2012). Anginer et al. (2013, 2014) provide a more general exploration of the relationship between governance, performance, and capitalization using an international data set. They find that banks with better governance have less capital, and, strikingly, that better governance is associated with more insolvent risk for banks and that the effect is larger in countries with better fiscal health. The authors attribute this stronger effect to the fact that there is more value for banks in exploiting the financial safety net. Laeven and Levine (2009), using a cross-country data set, show that when ownership is more concentrated, so that shareholders have more power, banks take more risk.

Second, the governance literature emphasizes that more stock ownership by top management leads to better alignment of incentives between management and shareholders. However, existing evidence shows that banks whose management had more of a stake performed worse during the crisis, not better (Fahlenbrach and Stulz 2011).

Third, there is a considerable literature that focuses on CEOs’ ability to entrench themselves so that they can pursue their own objectives rather than maximize shareholder wealth. Such entrenched CEOs are likely to take less risk than shareholders would like them to because they could lose their jobs if their banks experience distress. Ferreira et al. (2013) show that managers of banks that were more entrenched were less likely to be bailed out during the crisis. Relatedly, Chen, Hong, and Scheinkman (2010) show that institutional investors had a preference for banks that were taking more risk before the crisis.

Finally, there is no evidence that banks whose boards had more financial expertise performed better (Minton, Taillard, and Williamson 2014). All this evidence, at the very least, implies that better governance did not lead banks to perform better during the crisis. Of course, the implication is not that better governance is bad for shareholders; rather, the correct implication is that better governance does not mean less risk. Better governance meant taking risks that would have been rewarding for shareholders had there not been a crisis. Because a crisis like the one that transpired, if it was contemplated at all, was viewed as an exceedingly low-probability event, the evidence supports the view that shareholders saw the taking of these risks as worthwhile for them ex ante.
4. The Organization of Risk Management

In this section, I discuss the trade-offs that affect how risk management should be organized in a bank. Consider a bank where employees throughout the organization can take risks. Suppose that the top management could know exactly what the bank's risk is at each point in time, and suppose further that it could instantly hedge risk at zero cost. In this case, risk management would be straightforward. Having determined its risk appetite, the bank could control its risk through hedging by top management. As long as risk takers in the bank only took projects that create value regardless of their risk, top management would have no reason to monitor the risk in the sense of assessing risk decisions made by employees. All the bank would have to do is measure the risk taken within the bank and control it through hedging.

Real-world banks cannot control risk this way for at least three important reasons:

1. Limitations in risk-measurement technology: While real-time risk measures exist for a number of activities within banks, such measures do not exist for banks as a whole. Further, risk measurement is imperfect and can be quite imprecise. Finally, risk measurement can be affected by behavioral biases. For instance, over-optimism and groupthink can lead to important issues being ignored or underappreciated (Greenbaum 2014).

2. Limitations on hedging: Even if a bank had a highly precise measure of its overall risk, it does not follow that it could safely manage its overall risk through hedging by top management. Some risks cannot be hedged and hedges may not work out as planned.

3. Limitations regarding risk-taker incentives: Risk takers do not take only those risks that increase the value of the bank. Some risk takers turn out to be rogue traders. More importantly, however, risk takers often are rewarded in ways that give them incentives to take risks that are not as valuable to the bank as they are to the risk takers. It is even possible that risk takers can gain from taking risks that destroy value for the bank. This problem is made worse by the limitations in risk measurement tools.

These three limitations mean that risk has to be monitored and managed throughout the organization. To help with this task, large banks have risk management organizations that employ risk managers and are headed by a chief risk officer (CRO). Despite their title, risk managers, for the most part, do not manage risk. They primarily measure it, monitor it, and help those who do manage risk. To see this more concretely, consider the interactions between the head of a trading desk at a bank and the bank's risk managers. The head of the trading desk manages the risk taken by the desk, taking into account the opportunities that are available and their risk. He does so within constraints set by senior management and possibly the board. Risk management will help in setting these constraints and may have a more direct role because of delegation from senior management and possibly the board. Risk management will monitor the risk of the desk and make sure that that risk stays within the limits that have been set by the bank. Similarly, at the firm level, risk management also has a monitoring and advising role, but the top risk manager in a company is the CEO, not the CRO.

Section 2 presented a framework for understanding the type of risk management an organization should select to maximize shareholder wealth. If the relationship between bank value and risk is close to flat, risk management cannot create much value by making sure that the bank's aggregate risk is at its optimal level. In contrast, if too much risk results in a sharp drop in bank value, risk management that keeps the bank from taking on too much risk creates significant value in that the bank would be worth much less if the market lacked confidence in its ability to manage risk. It therefore follows that the extent of a bank's investment in risk management depends on how its value is related to its risk. The size of the investment in risk management is an investment decision like any other for a bank. Therefore, it has to compare costs and benefits. Excessive investment in risk management can destroy value just as much as insufficient investment in risk management can.

The risk-taking framework also helps in assessing how independent the risk management function should be. One often-held view is that risk management is the equivalent of the audit function, but for risk. From this perspective, since the audit function in a firm is independent, the risk manage
The occurrence of an undesirable outcome is not evidence of excessive risk taking or bad management. It could simply be the realization of an extremely low-probability event that was fully contemplated by the bank when it chose its strategy.
values for the index had higher returns during the crisis. Further, they find that bank holding companies with a higher value of the index have less tail risk, measured by the average return on the five worst daily stock returns during a year. The studies investigate how risk management affects tail risk and stock returns. Risk management does not target these measures, and the relationship between metrics that risk management does focus on and these measures does not appear straightforward. Therefore, one would want to know through which channels risk management affects stock returns and stock tail risk measures because an understanding of these channels would give reassurance that the relationships documented in these studies are not spurious. An interesting paper by Berg (2014) provides some evidence on this issue. He shows that, in a bank where loan officers are rewarded according to loan volume, having risk management monitor loan decisions reduces the probability of default of loans in the bank’s loan portfolio. Another issue with these studies is that a financial institution could have good risk governance because it is costly for that institution to have too much risk and so it wants low risk. Hence, the institution sets up its risk management organization to ensure that it will have low risk. Viewed from this perspective, the empirical evidence shows that a financial institution that wants to have low risk can achieve low risk. Simply paying a CRO a higher salary relative to the CEO will not ensure that a financial institution has low risk.

5. Tools and Challenges in Achieving the Optimal Level of Risk

If all the risks of a firm could be captured by a reliable value-at-risk (VaR) measure, the risk framework presented in Section 2 could be implemented in a conceptually straightforward way. I show this in the first part of this section. I then turn to the limitations of using VaR to manage firm-wide risk.

5.1 Using VaR to Target Risk

The risk framework of Section 2 implies that a firm wants to target the probability of making a loss that could put it in financial distress or in default. In other words, it wants the probability of a loss that exceeds a threshold amount to be its target probability. Hence, if the firm wants its probability of default within a year to be, for the sake of illustration, 0.06 percent, it wants the loss that has only a 0.06 percent probability of being exceeded to be the largest loss it could incur without being forced into default. A loss that is exceeded only with a probability \(p\) over one year is the value at risk (VaR) over one year at the probability level \(p\). It follows that the risk framework leads directly to the use of VaR as a firm-wide risk measure (Nocco and Stulz 2006). The use of VaR is ubiquitous in risk management, which gives rise to a constant debate about the merits of VaR. However, despite its weaknesses, VaR is the right risk measure in a wide range of circumstances. Consider a bank that has chosen a risk appetite that implies that its probability of failure is 0.06 percent over one year. This means that the bank is expected to fail less than once in a thousand years. Suppose that the bank has $100 billion of assets and $10 billion of equity. If all the risks that the bank faces could be measured through a bank-wide VaR, the bank should have an equity cushion such that there is a 0.06 percent probability that it will make a loss that would be larger than its equity cushion. If this bank has a bank-wide VaR of $15 billion, it has taken too much risk given its risk appetite because its probability of default is higher than 0.06 percent. Hence, this bank should either reduce the risk of its assets or raise additional equity.

Within a bank, a VaR can be estimated for any risk-taking unit (see, for instance, Litterman [1996]). For instance, a VaR can be estimated for the book of a trader as well as for the unit that the trader belongs to. Starting from the smallest units for which VaR is estimated, the VaRs can be aggregated so that the bank-wide VaR is a function of the VaRs of these units as well as of the correlations in risks across these units. Further, using the VaRs of the smallest units and the correlations, it is possible to assess how each unit contributes to the risk of the bank. For instance, a bank could estimate how much of its risk as measured by VaR is accounted for by a specific trader.

The fact that the bank-wide VaR results from the aggregation of VaRs of units of the bank means that risk management can target the bank’s VaR by setting limits on the VaRs of units of the bank. With such an approach, if all units are within
their limits, the VaR of the bank should not exceed the VaR that corresponds to its risk appetite.

5.2 Setting Limits

The risk framework provides guidelines for how VaR limits should be set. First, the firm’s risk appetite specifies the firm-level VaR limit. Second, within the firm, VaR limits should depend on the profitability of the risk-taking unit in relation to its VaR. Ideally, the marginal unit of risk should have the same expected profit across all risk-taking units of the bank. It would make little sense for a bank to allow a unit to take up large amounts of risk if that unit cannot use that risk to create value for the bank. Because profit opportunities change, it follows that limits cannot be unchangeable. When profit opportunities appear for a sector of the bank, it makes sense for limits to be adjusted. However, if the bank’s risk appetite has not changed, VaR limits cannot be increased in one sector of the bank without being decreased elsewhere. Of course, if profit opportunities change for the bank as a whole, so that the expected return from risk taking increases, it can be optimal for the bank to change its risk appetite and, as a consequence, its firm-wide VaR limit as well.

With the risk framework of Section 2, a bank targets its probability of default over a year. To properly target this probability of default, it has to make sure that its risk does not depart from its target over the year. This means that it must monitor and set limits at a higher frequency during the year. For instance, the bank can monitor and control the risk of trading activities in liquid markets using a one-day VaR. Within the year, the bank can change limits in response to unexpected losses. This flexibility means that it has the ability to take more risks if it expects that it can adjust its risk easily.

An obvious problem with setting limits is that the bank’s units might not make full use of their ability to take risk. Consider a unit with a daily VaR limit of $10 million. If that unit can alter its VaR through trades quickly and at low cost, it will operate close to its limit as long as it has opportunities to trade. However, if a unit cannot alter its VaR quickly and at low cost, it will want to keep some risk capacity in reserve so that it can take advantage of opportunities if circumstances change.

An important issue in setting limits is determining the level of aggregation for which limits are set. In practice, this is often described as the issue of selecting the level of granularity of limits. Consider the case where a limit is set for a department that trades in mortgage-backed securities. More granular limits would be limits at the trader level. Even more granular limits would be for maturity bins at the trader level. More granular limits make it much harder, and sometimes impossible, for risk-taking units to accumulate large unmonitored pockets of risk. However, more granular limits also make it much more difficult for risk-taking units to aggressively take advantage of good opportunities without negotiating a relaxation of limits. As limits become less granular, the discretion of the risk-taking units increases. More discretion makes it easier for these units to take advantage of opportunities quickly, but it also makes it easier for them to end up with large losses.

5.3 The Limits of Risk Measurement

Measuring risk at the firm level presents obvious difficulties. First, aggregating VaR measures to obtain a firm-wide risk measure is fraught with problems. Second, VaR does not capture all risks. Third, VaR has substantial model risk. I assess these issues in turn.

To organize the analysis, I will continue using the risk framework of Section 2. Hence, the bank targets a probability of default. I will assume that it targets that probability over a one-year horizon. The firm defaults or fails if it makes a loss large enough that it exhausts its equity buffer. So, to properly target a probability of default, the firm has to correctly measure the risk of a loss that exceeds the size of the equity buffer. This means that all risks that could lead to losses have to be modeled. If the firm targets a probability of default of 0.06 percent but models only some of the risks, it will have a higher probability of default if its equity buffer corresponds to the one-year VaR obtained from the modeled risks.

A typical approach for a bank is to divide risks into market, credit, and operational risks. Basel II introduced this division and requires banks to hold capital for each of these types of risk. Unfortunately, a firm-wide VaR that is obtained by aggregating market, credit, and operational risks will typically not reflect all risks. Such an approach misses business risks if these risks are not modeled as part of operational risk. For many banks, noninterest income is a large component of revenue. This income is variable and it tends to be low when the bank makes losses on loans. Such income has to be modeled when assessing the amount of equity necessary to support the targeted probability of default. Second, credit VaRs do not necessarily model the risk arising from unexpected changes in interest rates and credit spreads. More generally, interest rate risks in the banking book and interest rate risks arising from liabilities are typically not included in firm-wide VaRs.
The firm-wide measurement apparatus used by banks is focused on risks arising from the asset side. In practice, however, banks can fail because their funding vanishes (see, for example, Duffie [2010]). Before the crisis, funding liquidity risk was often not even part of risk management in banks but instead was the focus of the treasury department. Now, funding liquidity risk is an issue that is given more attention by risk management. However, it is still not the case that funding risk is integrated in the firm-wide VaR analysis. A shock to funding can force the bank to sell assets at a loss. Further, shocks to funding are more likely to happen in periods when markets for securities are themselves less liquid, so that selling assets quickly will be costly because they are sold at a discount.

If a bank divides risks between market, credit, and operational risks, it has to aggregate these risks to obtain a firm-wide measure of risk (Rosenberg and Schuermann 2006). This aggregation requires estimates of the correlations between these types of risks. It turns out that aggregate risk is very sensitive to these estimates. To see this, suppose that a bank has a VaR of $1 for each type of risk. If the correlations are 1 among the risks, the bank-wide VaR is $3. If the correlations are 0, the bank-wide VaR falls to $1.73. Unfortunately, data to estimate such correlations are sparse. Yet, these correlation estimates make an enormous difference in the amount of equity that is required to target a given default probability. Mistakes in correlation estimates could lead a bank to have too little capital and to have a risk of default much larger than its targeted risk of default.

Another important problem in aggregating risk is that different types of risk have different statistical distributions. While market risk generally has a fat-tailed symmetric distribution but can often be well-approximated by the normal distribution, the distributions for credit risk and operational risk are both fat-tailed and highly skewed. Risks that are normally distributed can be added up in a straightforward way because the sum of normally distributed variables is a normally distributed variable. However, it is not straightforward to add risks that follow different distributions. One approach that the literature has focused on is the use of copulas. Implementing this approach in practice has proven challenging, especially in the context of yearly frequencies, where there is only limited data available for estimation.

A VaR is a forecast. When it is estimated for the firm as a whole, it is a forecast for the firm as a whole. One can assess whether a VaR is properly estimated by examining the VaR exceedances (see, for example, Christoffersen [2011]). If a bank estimates a one-day VaR at the 5 percent level for its trading book, it expects the VaR to be exceeded roughly thirteen times in a year. If the VaR is exceeded fewer than thirteen times, it is a potential indication that the bank's VaR estimates are biased upward. Alternatively, if the VaR is exceeded more than thirteen times, the VaR may be biased downward or random variation may be such that the unbiased VaR was exceeded more than thirteen times. Statistical tests have been developed that can be used to assess whether a VaR is biased given sampling variation. The problem with an annual VaR estimated at the 0.06 percent probability level is that there cannot be a sufficient history to reliably assess whether the VaR is unbiased. The fact that a 0.06 percent VaR is not exceeded over a period of five years tells us almost nothing. Consequently, risk measures used to assess the appropriate size of a capital buffer cannot be back-tested satisfactorily. The only way to assess whether such risk measures are reliable is to assess the process that is used to produce them. However, such an approach does not resolve the key issue that the one-year VaR estimated for extremely low probability levels (such as the 0.06 percent in my example) is very sensitive to assumptions made about the extreme tail of the distribution of the value of the bank. These assumptions cannot be tested robustly in the way that assumptions for a 5 percent daily VaR can be tested.

No discussion of risk management can be complete without addressing the issue of risks that are not known—the famous black swans of Nassim Taleb or the “unknown unknowns” of Donald Rumsfeld. These rare risks are not relevant for VaR when the VaR is estimated at probability levels that are not extremely low. Hence, they do not create a bias in such VaR forecasts. However, the role of these risks becomes much more consequential when assessing an annual VaR at extremely low probability levels, such as the 0.06 percent level. The losses corresponding to such a VaR are caused by extremely rare events, so that one's understanding of what such rare events could be becomes important. A focus on historical data and the use of established statistical techniques cannot by itself be sufficient to estimate a VaR at the 0.06 percent level because the historical data generally encompasses a period that is too short to develop an accurate representation of extreme losses that have an annual probability of less than 0.06 percent of occurring.
A 0.06 percent VaR is one that should be exceeded less than once every thousand years. In other words, a bank that targets a 0.06 percent probability of default should be able to survive just about any crisis. This suggests another approach to investigating whether the VaR is correctly estimated. Since the bank should survive almost all crises, a simple way to assess whether the bank’s targeting of the probability of default is done correctly is to simulate what the performance of the bank would be if historical crises were to repeat. This approach amounts to performing stress tests. If such tests show that the bank would be unable to survive past crises, it is likely that its VaR is biased. More generally, however, stress tests can help us understand the risks that a bank is exposed to and whether it has enough equity to withstand adverse realizations of these risks.

6. INCENTIVES, CULTURE, AND RISK MANAGEMENT

Risk measurement is never perfect. Even if it were, there would still be the problem that firm value does not depend on risk alone. Risk management that is structured so that it rigidly keeps a bank’s risk below some pre-specified level and does so through a large set of inflexible limits may well succeed in controlling risk, but in the process it may prevent the institution from creating wealth for its shareholders. In a bank, risk management is part of the production technology. If risk management works well, the institution creates more value because it can issue more liquid claims and because it has more capacity to take profitable risks.

An unfortunate tendency among some board members and regulators is to think of the risk management function as a compliance function in the same way that auditing is a compliance function. Assuredly, there is an important compliance element to risk management. If a limit is set for a specific risk, the risk function must ensure that the limit is respected and understand why it is exceeded if it is. However, auditors are never in a position to conclude that departures from generally accepted accounting principles (GAAP) can create shareholder wealth. In contrast, risk managers who have some discretion over limits have to know when limit exceedances should be allowed and when a business line should be forced to respect a limit. Risk managers also have to determine, or help determine, when limits have to be changed and when it is appropriate for the institution to adjust its risk appetite.

Banks always face trade-offs between risk and expected return. To complicate matters, risk and expected return are measured imperfectly. If the costs to an institution of having more risk than is optimal are extremely high, that institution may benefit from having a risk management organization that operates as a police department that enforces rules. In this case, it would also make sense for the organization to account for limitations in risk measurement by imposing a substantial risk buffer—in other words, set a limit for the risk measure that is lower than the objective to account for the fact that the risk measure might understate risk. However, this is not typically the situation that an institution faces. In general, an institution can lose a lot from not being able to take advantage of opportunities that might be precluded by an inflexible risk organization. Further, difficulties in assessing risk mean that a risk management organization might make incorrect risk assessments without having a dialogue with business units. Unfortunately, such a dialogue is often impossible if the risk management function is viewed as a compliance unit rather than an essential part of the firm that seeks to implement policies that increase firm value.

Hall, Mikes, and Millo (2013) and Mikes, Hall, and Millo (2013) conducted a clinical study of two banks, which they denote as Saxon Bank and Anglo Bank. Their study shows vividly the issues involved in the positioning of risk management within the organization. In Saxon Bank, risk managers succeeded in being part of the important decisions. They helped shape these decisions and could make sure that risk considerations would be taken into account. In contrast, in Anglo Bank, risk management was divided between a group more focused on formal measures and a group more focused on intuition and interpersonal relationships. The group more focused on formal measures became dominant, but the risk management function failed in that it had no influence on the main decisions of the bank. Importantly, employing the formal measures of the role of risk management used in the literature discussed earlier, it is not clear that these two banks could be distinguished, yet risk management played a fundamentally distinct role in the two. This indicates that new measures for the role of risk management are needed.

If everyone in an organization is focused on ensuring that the institution takes risks that increase firm value and not
risks that decrease it, risk management becomes a resource in making this goal possible. Lines of business cannot know by themselves the extent to which the risks they take increase firm value because the amount of risk the bank can take at a given point in time depends on other risks taken by other lines of business. Hence, risk management has to bring to these risk-taking decisions the perspective of the firm as a whole to make sure that the firm itself does not have a suboptimal amount of risk. By bringing in this perspective, risk managers face potential conflicts with managers who are concerned about their unit only. Hence, for risk management to work well, it has to be that executives within the firm have reasons to care about the firm as a whole. This outcome requires incentives that reward executives if they create value for the firm as a whole and makes them bear adverse consequences from taking risks that destroy value.

Setting correct incentives for risk taking is complex. However, as Rajan (2006) discusses, poor incentives can impose large costs, both on shareholders and on society at large. Many banks have developed a bank-wide mechanism that can properly assess the cost of taking specific risks. Such a mechanism is called risk capital (see, for example, Matten [2000]). For a bank, risk capital is the amount of capital the bank requires to support the risks it takes so that, as a whole, its level of risk meets its risk appetite. As a unit of the bank takes a risk, the bank can keep its aggregate level of risk by acquiring more equity capital to support its risk taking.

If a bank does not force executives to take into account the cost of their risk taking for the bank as a whole, all of the burden of limiting risk will be borne by risk management.

This greater equity capital has a cost and this cost should be taken into account when evaluating the risk. Taking this equity capital cost into account may mean that it is no longer worthwhile to take the risk. If a bank does not force executives to take into account the cost of their risk taking for the bank as a whole, all of the burden of limiting risk will be borne by risk management. Such an approach is problematic for two reasons. First, it means that risk limits end up running the lines of business because the lines of business have no reason to internalize the cost of risk. Second, when risk is managed mostly through limits, the risk capacity of the bank is used less efficiently—risk-bearing capacity becomes allocated more through rationing than through the price mechanism.

Incentives should be set right, but incentives have limits. It is not possible to set up an incentive plan so precisely calibrated that it leads executives to take the right actions in every situation. Executives have to deal with situations that nobody thought possible. Employment contracts are by their very nature incomplete. A further issue is that not all risks can be quantified or defined. When a bank focuses on specific risks that it quantifies and can account for in employee reviews and incentive plans, there is an incentive for employees to take risks that are not quantified and monitored.

Because of the limits of risk management and incentives, the ability of a firm to manage risk properly depends on its corporate culture as well. There is a large organizational behavior literature on corporate culture and a smaller economics literature on the topic (for a recent review, see Bouwman [2013]). An often-used definition of corporate culture from the organizational behavior literature is that an organization’s culture is “a system of shared values (that define what is important) and norms that define appropriate attitudes and behaviors for organizational members (how to feel and behave)” (O’Reilly and Chatman 1996). An important aspect of corporate culture is that it is the result of learning over time. This aspect of culture is emphasized by the following definition: “Culture is what a group learns over a period of time as that group solves its problems of survival in an external environment and its problems of internal integration” (Schein 1990). As a result, a culture is hard to change. It also has to be transmitted to new hires and it may leave with key employees. Hence, a firm’s culture is not permanent.

Within the economics literature, culture is a mechanism that makes the corporation more efficient because it simplifies communication and facilitates decisions. From this perspective, having a strong culture has important fixed costs but it decreases marginal cost (Hermalin 2001). The organizational behavior literature is more focused on characterizing a firm’s culture, so it has various typologies of corporate cultures. With the organizational behavior approach, different organizations have different cultures and an organization may not necessarily have the culture that maximizes shareholder wealth or ensures the success of the organization. For instance, Cartwright and Cooper (1993) distinguish between a role-oriented culture which is very bureaucratic and centralized; a task/achievement-oriented culture, which emphasizes teamwork and execution; a power-oriented culture, which is highly centralized and focuses on respect of authority; and a person/support-oriented culture, which is egalitarian and nurtures personal growth.

Limited empirical work exists on the relationship between culture and corporate outcomes, in part because it is difficult to measure the dimensions of culture. As one author put it
more than twenty years ago, “Organizational culture is a complex phenomenon, and we should not rush to measure things until we understand better what we are measuring” (Schein 1990). Two recent studies have used data from surveys of employees on how attractive their companies are as a place of work. Guiso, Sapienza, and Zingales (2015) show that companies whose managers are viewed as trustworthy and ethical have higher valuations and higher profitability. Popadak (2013) finds that improvements in shareholder governance change a firm’s culture, in that the firm becomes more results-oriented but less customer-oriented, and employee integrity falls. In that study, shareholders gain initially from the better governance, but these gains are partly offset over time because of the change in culture.

The literature on culture does not focus on risk taking or, for that matter, on the issues that are unique to the financial industry. An exception is Fahlenbrach, Prilmeier, and Stulz (2012). The authors do not use a direct measure of culture. Instead, they show that latent characteristics of banks, which could be explained by culture, are helpful to understanding how crises affect banks. Specifically, they show that a bank’s performance in the crisis of 1998 helps predict its performance in the recent crisis. This effect is of the same magnitude as bank leverage in helping to understand bank performance.

Firms in the financial industry differ from other firms in the extent to which employees typically make decisions regarding risk. A loan officer who can decide whether a loan is granted makes a decision to take a risk. She may have information about that risk that nobody else in the organization has. No one may ever know whether the decision was right from the perspective of the firm, for a number of reasons. First, it may not be possible for the loan officer to credibly communicate the information that she has. Second, the loan officer may have incentives to grant loans that she knows should not be granted. Third, loan outcomes are of limited use since expected defaults are not zero. A solution for the bank is to minimize the discretion of the loan officer by relying on statistical models for the decision. However, such a solution can be costly because it reduces flexibility and eliminates the use of soft information that can be valuable. A bank’s culture can constrain loan officer discretion in a way that leads to better outcomes for the bank. A bank with an underwriting culture that is highly focused on the interests of the bank will make it harder for a loan officer to deviate from the social norms within the bank because employees who are in contact with the loan officer might be able to assess that the officer is deviating from the bank’s norms and the extent to which she is doing so in a way that neither risk managers nor executives could.

Another example where corporate culture can make risk management more effective is with respect to acceptable interactions with risk managers. If the social norm is for traders to be confrontational when questioned, it is much harder for risk managers to correctly assess the risk of positions and how to mitigate this risk. In this case, the risk managers’ energies have to be devoted to fighting with traders and figuring out what they might be hiding.

A final example involves how employees use information about risk that they discover through routine interactions. Consider a situation in which an executive observes a trader on a desk that the executive is not responsible for take a position that cannot be expected to be profitable for the firm but might be very valuable to the trader if it pays off. For some reason, the trader’s own supervisor either does not understand the position or does not care. The position breaches no limits, so risk management has not investigated it. Depending on the firm’s culture, the executive could act very differently. In some firms, he would say nothing. In other firms, he would start a dialogue with the supervisor or the trader. In the latter firms, one would expect risk taking to be more likely to increase value since risk taking that destroys firm value is less likely to take place.

As far as I know, only Sorensen (2002) has examined the implications of corporate culture for risk outcomes. Sorensen predicts that a strong culture, by which he means strong agreement within a firm on shared values and norms, leads to more consistency. In other words, culture is a control mechanism. With a stronger control mechanism, there should be less variability in outcomes. His study examines the volatility of unexpected performance on measures of culture strength. He finds a strong negative relationship between the volatility of unexpected performance and culture strength. Unfortunately, his sample includes no firms from the financial industry.
7. Conclusion

The success of banks and the health of the financial system depend critically on how banks take risks. A bank’s ability to measure and manage risks creates value for shareholders. There is no simple recipe that enables a bank to measure and manage risks better. For risk taking to maximize shareholder wealth, a bank has to have the right risk management, but also the right governance, the right incentives, and the right culture. A risk management structure that is optimal for one bank may be suboptimal for another. Ultimately, the success of risk management in performing its functions depends on the corporate environment and on risk management’s ability to shape that environment. However, while better risk management should lead to better risk taking, there is no reason for a bank with good risk management to have low risk.


References (Continued)


