ELSEVIER

Journal of Financial Economics 47 (1998) 55–82

JOURNAL OF Financial ECONOMICS

Risk management, capital budgeting, and capital structure policy for financial institutions: an integrated approach

Kenneth A. Froot[a], Jeremy C. Stein[b,*]

[a] *Harvard Business School, Boston, MA 02163, USA*
[b] *Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142, USA*

Received 10 January 1996; received in revised form 21 February 1997

Abstract

We develop a framework for analyzing the capital allocation and capital structure decisions facing financial institutions. Our model incorporates two key features: (i) value-maximizing banks have a well-founded concern with risk management; and (ii) not all the risks they face can be frictionlessly hedged in the capital market. This approach allows us to show how bank-level risk management considerations should factor into the pricing of those risks that cannot be easily hedged. We examine several applications, including: the evaluation of proprietary trading operations, and the pricing of unhedgeable derivatives positions. We also compare our approach to the RAROC methodology that has been adopted by a number of banks. © 1998 Elsevier Science S.A. All rights reserved.

Keywords: Risk management; Financial institutions; Capital allocation

JEL classification: G20; G31; G32

1. Introduction

One of the fundamental roles of banks and other financial intermediaries is to invest in assets which, because of their information-intensive nature, cannot be frictionlessly traded in the capital markets. The standard example of such an

* Corresponding author. Tel.: 617/253-3763; fax: 617/258-6855; e-mail: jcstein@mit.edu.

0304-405X/98/$19.00 © 1998 Elsevier Science S.A. All rights reserved
PII S0304-405X(97)00037-8

illiquid asset is a loan to a small or medium-sized company. A more modern example is the credit-risk component of a foreign exchange swap. Even if the currency risk inherent in the swap can be easily laid off by the dealer bank, the same is not likely true of the credit risk.[1]

At the same time that they are investing in illiquid assets, most banks also appear to engage in active risk management programs. Given a fixed capital structure, there are two broad ways in which a bank can control its exposure to risk. First, some risks can be offset by hedging transactions in the capital market. Second, for those risks where direct hedging transactions are not feasible, another way for the bank to control its exposure is by altering its investment policies. Therefore, with illiquid risks, the bank's capital budgeting and risk management functions become linked.

To see this point more clearly, return to the example of the foreign exchange swap. If a dealer bank is considering such a transaction, its own aversion to currency risk should not enter into the decision of whether or not to proceed. After all, if it doesn't like the currency risk embodied in the swap, it can always unload this risk in the market on fair terms. Thus, with respect to the tradeable currency risk, the risk management and investment decisions are separable. The same is not true, however, with respect to the illiquid credit-risk component of the swap. If the bank is averse to this risk, the only way to avoid it is by not entering into the swap in the first place.

This reasoning suggests that if the bank is asked to bid on the swap, its pricing should have the following properties. First, the pricing of the swap should be independent of the bank's own attitudes toward currency risk. That is, the bank should evaluate currency risk just like any other market participant, based only on the risk's correlation with systematic factors that are priced in the capital market. Second, the swap's pricing should depend on the bank's own attitude toward the credit risk. Thus, if the bank already has a portfolio of very highly correlated credit risks, it might bid less aggressively for the swap than another institution with a different balance sheet, all other things being equal. This should hold true even if the credit risk is uncorrelated with factors that are priced in the capital market.

Although this sort of approach to the pricing of bank products may sound intuitively reasonable, it differs substantially from the dominant paradigm in the academic literature, which is based on the classical finance assumptions of frictionless trading and absence of arbitrage. In the specific case of pricing the credit risk on a swap, the classical method boils down to a contingent-claims

[1] Although we use the term bank throughout for shorthand, we have in mind not just commercial banks, but other types of intermediaries as well, e.g., investment banks, insurance and reinsurance companies, etc. Indeed, many of the applications that we discuss below are set in the context of these other institutions.

model of the sort pioneered by Merton (1974).[2] This type of model, like any classical pricing technique, has the implication that the correct price for the swap is the same for any dealer bank, independent of the bank's pre-existing portfolio. Of course, this is because the classical approach by its very nature assumes away exactly the sorts of imperfections that make the bank's problem challenging and relevant. Indeed, it is only appropriate if either: (i) the bank can frictionlessly hedge all risks, including credit risks, in the capital market; or (ii) the Modigliani-Miller (1958) theorem applies, so that the bank has no reason to care about risk management.

Perhaps because the classical finance approach does not speak to their concerns with risk management, practitioners have developed alternative techniques for capital budgeting. One leading approach is based on the concept of risk-adjusted return on capital, or RAROC. The RAROC method effectively assesses a risk premium on an investment that is proportional to a measure of the investment's capital at risk, multiplied by a cost of capital. However, although the RAROC approach has some intuitive appeal, it is not clear that it is the optimal technique for dealing with the sorts of capital budgeting problems facing financial institutions. That is, RAROC, as currently applied, is not derived from first principles to address the objective of shareholder value maximization. Consequently, it has some features that might be considered troublesome, and it leaves other issues potentially unresolved. For example, should capital at risk be calculated based on an investment's total volatility, or on some sort of covariance measure? And once one has determined the amount of capital at risk, what is the right cost to assign to this capital? Should the cost of capital at risk depend on the strength of the bank's balance sheet, or other related variables?

Our primary goal in this paper is to develop a conceptual framework for capital budgeting that blends some of the most desirable features of both the classical approach and the RAROC bank-practitioner approach. To accomplish this goal, we build a model that is rooted in the objective of maximizing shareholder value in an efficient market, similar to the classical approach. However, the model also incorporates two other key features: (i) there is a well-founded concern with risk management, and (ii) not all risks can be hedged in the capital market. This allows us to capture the important intuition that bank-level risk-management considerations should enter into the pricing of those risks that cannot be hedged.

In principle, this framework can be applied to any company attempting to manage the risks associated with illiquid assets. Nonetheless, we think it is particularly well-motivated by the sorts of financial-industry problems

[2] For recent applications to the pricing of credit risk on derivatives, see, e.g., Cooper and Mello (1991), and Jarrow and Turnbull (1995).

discussed above. This is because getting the cost of capital right for any given instrument is very likely to be a first-order consideration for a financial institution. In contrast, for many industrial companies doing capital budgeting, the uncertainties associated with projecting cashflows on their physical assets may be so large as to swamp any modifications in discount rates that our method might suggest.

In order to make the bank's concern with risk management endogenous, we assume that there are increasing costs to raising new external funds. Thus, if a bank were to be hit with a negative shock that depleted its capital, it would incur some costs in rebuilding its balance sheet. In an effort to avoid these costs, the bank will behave in a risk-averse fashion. This basic rationale for risk management is essentially the same as that presented by Froot et al. (1993) in the context of nonfinancial corporations. It is also closely related to the banking models of Kashyap and Stein (1995), Stein (1996), and Greenwald et al. (1991), all of which emphasize the costs that banks face in raising non-deposit external finance. In order to keep the analysis simple, we ignore any potentially offsetting risk-taking incentives that might arise with government-provided deposit insurance. Thus, our model is most literally applicable either to financial institutions that are not insured commercial banks (e.g. investment banks) or to commercial banks that are sufficiently well-capitalized that one can for practical purposes safely ignore the incentive effects of deposit insurance.

As will become clear, one key feature of this modeling approach is that it highlights a trade-off between managing risk via ex ante capital structure policy versus managing risk via capital budgeting and hedging policies. Aside from engaging in hedging transactions, a bank has two other methods for controlling the risk of being caught short of funds. First, it can adopt a very conservative capital structure. If there are no costs to holding a lot of capital, this will be the preferred way of dealing with the problem. In the limit when the bank holds a very large capital buffer, risk-management concerns will no longer enter investment decisions, and the model will converge back to the classical paradigm. Alternatively, if holding capital is costly due to tax or agency effects, for example, the bank can control its risk exposure by investing less aggressively in (i.e., charging a higher price for bearing) non-hedgeable risks. In this case, risk-management concerns will have a meaningful impact on capital budgeting policies. The bottom line is that in our framework, optimal hedging, capital budgeting and capital structure policies are jointly and endogenously determined.[3]

The remainder of the paper is organized as follows. In Section 2, we lay out the basic timing and assumptions of our model. In Section 3, we analyze

[3] Other papers which investigate the linkages between capital budgeting and risk management include Merton and Perold (1993), and Merton (1995a,b).

the model's implications for hedging, capital budgeting and capital structure decisions. In Section 4, we discuss some evidence from the commercial banking and reinsurance industries which is particularly helpful in illustrating that the model's basic premise, namely, that intermediary capital constraints have a first-order effect on their pricing of unhedgeable risks, is on strong empirical ground. Section 5 gives several examples of how the basic framework can be applied to different sorts of capital budgeting problems facing financial institutions. Section 6 presents a detailed comparison of our approach with the RAROC method. Section 7 concludes.

2. Timing and assumptions

The model we propose has three time periods, 0, 1, and 2. In the first two periods, the bank chooses its capital structure, and then makes capital budgeting and hedging decisions. These two periods are at the heart of our analysis. The last period closes the model, by giving the bank a well-founded objective function that incorporates both shareholder-value maximization as well as a concern for risk management.

2.1. Time 0: Bank chooses its capital structure

The bank enters Time 0 with an initial portfolio of exposures. This portfolio will result in a Time 2 random payoff of Z_P. The random variable Z_P is normal, and can be written as $Z_P = \mu_P + \varepsilon_P$, where μ_p is the mean and ε_P is a mean-zero disturbance term. For simplicity, we assume that this portfolio of exposures requires no net financing. That is, it is a zero-net-wealth portfolio.

The only decision facing the bank at Time 0 is how much equity capital to hold. Specifically, the bank can raise an amount of capital K, and invest the proceeds in riskless Treasury bills. Holding financial slack in this manner involves direct deadweight costs. These costs might in principle arise from a number of sources. For concreteness, it is useful to think of these costs as being driven by taxes. Thus the deadweight costs of holding an amount of capital K are given by τK, where τ is the effective net tax on cash holdings. The tax cost of holding equity-financed slack is just the mirror image of the tax advantage of debt finance. Although it involves deadweight costs, we will show below that banks typically opt to hold nonzero levels of capital. This is because holding capital allows banks to tolerate risks better, and thereby price their products more aggressively.

One point here deserves further comment. Although we have assumed that there are costs associated with a given stock of capital K, there is no sense in which banks have trouble adjusting K at Time 0. This runs counter to the spirit of many models of financing under asymmetric information (see, e.g., Myers and

Majluf, 1984), where, loosely speaking, costs are incurred not by having equity on the balance sheet per se, but rather by having to raise new external funds.[4] Indeed, we will assume momentarily that there are exactly such flow costs of new external finance at Time 2. Moreover, these flow costs will be a convex function of the amount raised, and they will be the driving force behind the bank's desire to manage risk.

Thus it is clearly a shortcut to ignore the potential for convex costs of external financing at Time 0. We do so in order to better focus on the question we ultimately wish to address with the Time-0 analysis: What is the appropriate long-run 'target' capital structure for the bank? In other words, how should the bank be seeking to position itself over the long haul: as a AAA credit, a BBB credit, or something in between? We recognize that, if at any point in time, the bank is far away from its ideal target, it may face costs of adjustment in getting to the target quickly, but it is nonetheless interesting to ask the question of what the target should be.

2.2. *Time 1: Bank invests in new products and makes hedging decisions*

We begin by focusing on the case where there is only one new product at Time 1. However, it is easy to generalize the results to the case of multiple new products, as we do below. The new product offers a random payoff of Z_N at Time 2. This payoff is normal, and can be written as $Z_N = \mu_N + \varepsilon_N$, where μ_N is the mean and ε_N is a mean-zero disturbance term. The magnitude of the bank's exposure to the new product is a choice variable, and is given by α. As before, it is assumed for simplicity that the exposure to the new product does not require any cash to be put up at Time 1. For example, the exposure may represent the assumption of a forward position where no money changes hands at the time the position is put on.

Given that we are working with zero-net-wealth forward positions, it is more natural to specify the dollar payoff on these positions, as opposed to a percentage rate of return. However, it is easy to reinterpret our notation in terms of the implied percentage returns on the associated underlying assets. Think of the forward position as being on an underlying asset whose Time-1 price is normalized to one. Then, if z is the rate of return on the underlying asset, the dollar payoff on the forward Z that we are defining satisfies $Z = z - r$, where r is the riskless rate of interest between Time 1 and Time 2. Thus all the statements we make below about required dollar payoffs on various forward positions can be trivially reinterpreted as statements about required rates of return on the underlying assets, simply by adding r.

[4] In other words, our Time-0 analysis of bank capital structure is of the static-tradeoff variety, as opposed to the dynamic pecking-order behavior implied by models of asymmetric information.

The second decision to be made at Time 1 is how to hedge both the initial and new exposures. We will have much more to say about the hedging technology later on. For now, we establish a piece of notation by defining the aggregate hedging position as H, and the associated payoff on this hedging position as Z_H.

Given the assumptions that have been made so far, the bank's realized internal wealth at Time 2, w, is given by

$$w = Z_P + \alpha Z_N + Z_H + K(1 - \tau). \tag{1}$$

In words, the amount of cash the bank has on hand at Time 2 will depend on the realizations on its old exposures, on its new product, and on its hedging positions, as well as on the amount of capital raised at Time 0.

2.3. *Time 2: Bank reacts to its cashflow realization*

As noted above, we need to build into the model a reason for the bank to care about risk management. That is, there has to be an incentive for the bank to care about the distribution of w. To do so, we follow Froot et al. (1993). In particular, we assume that after w is realized, the bank has a further, non-stochastic investment opportunity. For example, the bank might be able to extend some new loans. This investment requires a cash input of I, and yields a gross return of $F(I)$, where $F(I)$ is an increasing, concave function. The investment can either be funded out of internal sources, or funds can be raised externally in an amount e. Thus $I = w + e$. The hitch is that there are convex costs to raising external finance. These costs are given by $C(e)$.[5]

The solution to the bank's Time-2 problem can be denoted by $P(w)$, as follows:

$$P(w) = \max F(I) - I - C(e), \quad \text{subject to } I = w + e. \tag{2}$$

Froot et al. (1993) demonstrate that $P(w)$ is, in general, an increasing concave function, so that $P_w > 0$, and $P_{ww} < 0$. It is the concavity of the $P(w)$ function that generates a rationale for risk management. This concavity in turn arises from the convexity of $C(e)$, interacting with the concavity of $F(I)$ (see Froot et al., 1993, for more explicit details). Loosely speaking, the more difficult it is for the bank to raise external funds on short notice at Time 2, the more risk averse it will be with respect to fluctuations in its internal wealth, w.

[5] In Froot et al. (1993), we give a number of microeconomic rationales, based on agency or information problems, to justify this specification for the $C(e)$ function. We also show explicitly how such a convex functional form arises in a specific optimal contracting setting, a variant of the costly-state verification model due to Townsend (1979), and Gale and Hellwig (1985). Moreover, Stein (1996) generates a similar formulation in a banking model where non-deposit liabilities are subject to adverse selection problems.

2.4. Structure of the model

If one were to leave Time 1 out of the model, and keep only the parts corresponding to Time 0 and Time 2, we would be left with a very standard pecking order story of corporate investment and financing. At Time 2, the presence of increasing costs of external finance can lead to underinvestment, or a level of I that is less than the first-best. This distortion can be partially alleviated to the extent that financial slack can be built up at Time 0. Thus, it may be desirable to hold such slack on the balance sheet, even if there are some costs to doing so. This is very much in the spirit of Myers (1984), Myers and Majluf (1984), and the large literature that has followed these papers.

What we have added to this basic pecking order story are the ingredients that come into play at Time 1. Specifically, the bank now has two other tools at its disposal beyond simply holding financial slack that it can use to offset the underinvestment distortions caused by costly external finance. First, the bank can adjust its risk exposure by undertaking hedging transactions, H. Second, the bank can also adjust its risk exposure by varying the amount α it invests in the new product. Thus overall, the bank optimizes by picking the right combination of three policy variables: K, H and α. It is in this sense that capital structure, hedging and capital budgeting decisions are linked to one another.

3. Analysis of the model

To understand the properties of the model, it is easiest to work backwards. We have already seen that any given realization of w at Time 2 can be mapped into a non-stochastic payoff $P(w)$, once we specify $F(I)$ and $C(e)$. The next step is to ask, from the perspective of Time 1, when w is uncertain, what is the market value of the bank? And, given this valuation function, how should a bank that seeks to maximize its value set its hedging and capital budgeting policies at Time 1? That is, how should the bank choose H and α?

3.1. Valuing the bank at Time 1

From the perspective of Time 1, the ultimate payoff $P(w)$ to a bank shareholder is a random variable. In order to value this random variable, we need a pricing model. Without any real loss of generality, we can assume that asset prices are determined by a simple one-factor model. In this setting, the value of the bank's shares V, will be

$$V = \{EP(w) - \gamma \operatorname{cov}(P(w), M)\}/(1 + r), \tag{3}$$

where M, for market, is the one priced factor, γ is the equilibrium excess return per unit of variance for bearing M risk, and r is the riskless rate of interest between Time 1 and Time 2.

3.2. Optimal hedging policy at Time 1

Suppose the bank designs its risk management policy so as to maximize shareholder value V. What should it do? To build intuition, let us first consider a simple case where all the bank's risks are perfectly tradeable. In other words, these risks can be frictionlessly unloaded in the market, on terms that are just fair, given the correlation of the risks with the priced factor.In Appendix A, we prove the following:

Proposition 1: *In the case where all risks are perfectly tradeable, the bank maximizes value by hedging completely. That is, it picks its hedging position H so that the payoff on the hedging position* $Z_H = -\varepsilon_P - \alpha\varepsilon_N + k$, *where k is a constant.*

This result is fairly intuitive. It is a generalization of a result presented in Froot, Scharfstein, and Stein (1993), where we considered the more restrictive case in which investors in the capital market were risk neutral and hence where there were no priced factors. The one subtlety is that while it is tempting to conclude from the proposition that the bank simply behaves like a risk-averse individual, this is not quite correct. In general, a risk-averse individual will not wish to completely shun systematic risk, as this involves a reduction in expected return. Rather, an individual will typically opt for an interior solution in which he bears some systematic risk. However, this is not true for a publicly traded bank in our setup. A bank does not reduce shareholder value by sacrificing return in exchange for a reduction in risk, so long as the terms of trade are set in an efficient market, that is, so long as the hedging transactions have a net present value (NPV) of zero. Since there is no cost to reducing risk, we are left with the pure effect that, because of the concavity of $P(w)$, risk reduction on fair terms is always desirable.

Of course, by assuming that all risks can be sold without friction in the capital market, we have trivialized the bank's problem. As was stressed in the introduction, the very existence of intermediaries, such as banks, is testimony to the fact that certain risks are somewhat information intensive, and hence cannot be traded with perfect liquidity. To take a first cut at capturing this notion, we make the following decomposition. We assume that a bank's exposures can be classified into two categories: (i) perfectly tradeable exposures, which, as above, can be unloaded frictionlessly on fair-market terms, and (ii) completely non-tradeable exposures, which must be retained by the bank no matter what.

In terms of our previous notation, this amounts to decomposing both the pre-existing and new risks, ε_P and ε_N, as follows:

$$\varepsilon_P = \varepsilon_P^T + \varepsilon_P^N, \tag{4}$$

$$\varepsilon_N = \varepsilon_N^T + \varepsilon_N^N, \tag{5}$$

where ε_P^T is the tradeable component of ε_P, ε_P^N is the non-tradeable component, and so forth. For now, assume that the priced factor M is fully tradeable, so $\text{cov}(\varepsilon_P^N, M) = \text{cov}(\varepsilon_N^N, M) = 0$.

There are a number of different examples that help illustrate what we have in mind with this decomposition, and we will develop several of these in more detail shortly. For the time being, it may be helpful for concreteness to think of the new product as being an investment in a company whose stock is not publicly traded.[6] Clearly, some of the risk associated with such an investment may be tradeable, to the extent that it is correlated with, say, macroeconomic conditions, and hence can be hedged with some sort of contract on an aggregate variable. However, some of the idiosyncratic exposure associated with the investment cannot be laid off, at least not frictionlessly. This is what we are trying to capture.

In this environment, a simple extension of Proposition 1 can be proven (see the Appendix A):

Proposition 2. A bank will always wish to fully hedge its exposure to any tradeable risks. That is, it picks its hedging position H so that the payoff on the hedging position $Z_H = -\varepsilon_P^T - \alpha\varepsilon_N^T + k$, *where k is a constant.*

3.3. Capital budgeting policy at Time 1

3.3.1. The case of a single investment decision

Now that we have established how the bank sets its hedging policy, we can turn to the capital budgeting question, namely, what should the bank's desired investment α in the new product be at Time 1, and how does this investment depend on the new product's expected return and its risk characteristics?

To attack this question, we begin by noting that, given the results of Proposition 2 on optimal hedging policy, we can rewrite Time 2 wealth w as

$$w = \mu_P + \varepsilon_P^N + \alpha(\mu_N + \varepsilon_N^N) + k + K(1 - \tau). \tag{6}$$

[6] To be absolutely literal in terms of our notation, the investment should be thought of as a forward position in the non-traded company. However, as mentioned above, it is trivial to reinterpret our results in terms of required returns on the underlying assets.

Eq. (6) reflects the fact that all the tradeable risk components have been hedged out of w, leaving only the non-tradeable components. More structure can be put on this expression by pinning down k, which is the expected return on the hedging transactions. Given that these transactions have a net present value of zero, the expected return just offsets the systematic risk, so

$$k = -\gamma(\mathrm{cov}(\varepsilon_P^T, M) + \alpha\,\mathrm{cov}(\varepsilon_N^T, M)). \tag{7}$$

Substituting Eq. (7) into Eq. (6), we obtain

$$w = \mu_P + \varepsilon_P^N + \alpha(\mu_N + \varepsilon_N^N) - \gamma(\mathrm{cov}(\varepsilon_P^T, M) + \alpha\,\mathrm{cov}(\varepsilon_N^T, M)) + K(1-\tau). \tag{8}$$

Thus the bank's objective function is to maximize its value V as given by Eq. (3), subject to the constraint that w satisfy Eq. (8). To keep things simple, we assume that the bank views itself as a price-taker, that is, it takes μ_N as a fixed parameter, and chooses the quantity invested α accordingly. Thus the first order condition is simply that $dV/d\alpha = 0$. In the Appendix A, we show that this condition reduces to

$$\alpha^* = \{(\mu_N - \gamma\,\mathrm{cov}(\varepsilon_N^T, M)) - G\,\mathrm{cov}(\varepsilon_N^N, \varepsilon_P^N)\}/G\,\mathrm{var}(\varepsilon_N^N), \tag{9}$$

where $G \equiv -EP_{ww}/EP_w$ is a measure of the bank's effective risk aversion.

Unlike the case of an individual decisionmaker, the bank's risk aversion G is an endogenous variable. In particular, G will depend on the amount of capital K that the bank holds. It is easy to see that in the limiting case where K becomes infinitely large, G converges to zero, for any arbitrary specification of the underlying $C(e)$ function. In other words, with infinite capital, the bank becomes risk neutral, as in the classical setting. This is because the probability of it ever having to seek costly external finance falls to zero. Moreover, it can be shown that in the sort of optimal contracting setup considered in Froot, Scharfstein, and Stein (1993), this convergence is always monotonic, in other words, $dG/dK \leqslant 0$ everywhere. We will assume that this declining risk aversion property holds in what follows.

Note that in the polar case where $G = 0$, the bank will invest an infinite amount in anything with a return that exceeds the market risk premium. However, when $G > 0$, and investment requires the assumption of non-tradeable risk, the bank will be more conservative. The greater the contribution of the new non-tradeable risk to the variance of the bank's overall portfolio of non-tradeable risk, the more pronounced the conservatism.

When $G > 0$, the investment behavior embodied in Eq. (9) cannot in general be summarized in terms of a single hurdle rate, as is the common convention in most corporate capital budgeting applications. This is because the bank's risk aversion makes the required return an increasing function of the amount invested in the new product. However, in the limiting case where α goes to zero, there is a simple hurdle rate representation. We can interpret this case as one

where the bank must make a decision to either accept or reject an investment opportunity of small fixed size. For such an investment, the hurdle rate is given by

$$\mu_N^* = \gamma \operatorname{cov}(\varepsilon_N^T, M) + G \operatorname{cov}(\varepsilon_N^N, \varepsilon_P^N). \tag{10}$$

Eq. (10) can be thought of as a simple two-factor pricing model. The first factor is the standard market-risk factor. The novel twist lies in the addition of the second factor. To the extent that a bank takes on a non-tradeable risk, this risk should be priced based on its correlation with the bank's pre-existing portfolio of non-tradeable risks ε_P^N. Moreover, the unit price of non-tradeable risk is given by G. Ultimately, the extent to which non-tradeable risk gets priced depends on the factors which shape the curvature of the $P(w)$ function, namely, the $F(I)$ and $C(e)$ functions, as well as the initial amount of capital K held by the bank.

3.3.2. Multiple investments, interdependencies, and decentralization

Thus far, we have assumed that the bank only has the opportunity to invest in one new product at Time 0. However, it is straightforward to extend the results to the case where the bank can invest in multiple new products simultaneously. Suppose there are n new products, indexed by i. To streamline the notation slightly, it will be helpful to work with the expected payoff on each product in excess of the market risk premium. Thus, we define $\pi_i \equiv \mu_{Ni} - \gamma \operatorname{cov}(\varepsilon_{Ni}^T, M)$. The first-order conditions now become

$$\alpha_i^* = \{\pi_i - G \operatorname{cov}(\varepsilon_{N_i}^N, \varepsilon_P^N + \textstyle\sum \alpha_j^* \varepsilon_{N_j}^N)\}/G \operatorname{var}(\varepsilon_{N_i}^N), \tag{11}$$

for all i and $j \neq i$. Eq. (11) is of exactly the same form as Eq. (9). The only difference is that the appeal of the ith new product now depends not only on its covariance with the bank's pre-existing non-tradeable exposures, but also on its covariance with any other new non-tradeable exposures that are taken on at Time 1. There is one condition for each new product, and these conditions must be solved simultaneously to yield the optimal portfolio mix of new products.

The solution for this set of equations is

$$\alpha^* = \Omega^{-1}\{\pi - GC_{NP}\}/G, \tag{12}$$

where α^* and π are now both $n \times 1$ vectors, Ω is the $n \times n$ variance-covariance matrix for the non-tradeable exposures on the new products (i.e., $\Omega_{ij} = \operatorname{cov}(\varepsilon_{N_i}^N, \varepsilon_{N_j}^N)$), and C_{NP} is an $n \times 1$ vector whose ith element is $\operatorname{cov}(\varepsilon_{Ni}^N, \varepsilon_P^N)$.

Eq. (12) can be given a simple interpretation. The term $(\pi - GC_{NP})$ can be thought of as a vector of net returns for the i products, where the netting takes into account both the market risk of these products and their covariance with the bank's pre-existing portfolio. Once this netting has been done, the desired

mix of the new products follows from a standard mean-variance optimization, using the net returns as the means.

Eqs. (11) and (12) make it clear that things are more complicated in this multi-investment setting than in the usual corporate capital budgeting framework. In the usual framework, investment decisions are independent of one another. If their cash flows are held fixed, the appeal of project i does not depend on whether or not project j is undertaken. Here, this no longer holds true. In fact, there are two distinct sources of interdependence.

First, there is what might be termed a covariance spillover effect. Holding fixed G, investment in any product i will be less (more) attractive to the extent that there is also a significant investment made in another product j with positively (negatively) correlated non-tradeable risk. Thus, with non-zero covariances across the new products, α_i typically depends not only on own-project return π_i, but also on the π's of all the other products.

Second, and somewhat more subtly, there is what might be termed a bank-wide cost of capital effect. Even if all the covariances across the new products are zero, in other words, Ω is a diagonal matrix, investment decisions are in general interdependent because they can all influence the value of G. For example, if the bank takes a large, very risky position in one product, even if this position is completely orthogonal to all others, this might raise G and thereby make the bank less willing to take on any other risks.

These interdependencies imply that in order for the bank to make optimal investment decisions, these decisions must be centralized. If one thinks of individual product managers as observing the π's of their own products but not of others, one cannot simply delegate the investment decisions to these managers, even under the strong assumption that there are no agency problems and that the managers would therefore act in the bank's best interests with the information that they have. Rather, the information of the individual managers must be pooled.

As a practical matter, however, such centralized decisionmaking may present its own set of difficulties. This is likely to be especially true when decisions are made at very high frequencies, in which case the costs and delays associated with transmitting new information to headquarters each time an investment is made may be prohibitive. To take an extreme example, consider a bank with several hundred different traders who adjust their positions on an almost continuous basis. Clearly, in such a polar situation, complete centralization of decisionmaking is impossible.

This observation raises the question of whether one can approximate the full information centralized solution in a decentralized setting where individual product managers cannot condition on the contemporaneous π's of other new products. Analogous to the case with one new product, things become simpler if one is willing to entertain the limiting situation where the α's approach zero. In this case, Eq. (12) reduces to $\pi^* = GC_{NP}$, which is just a vector version of

Eq. (10). In other words, in the limiting case where all the investments in the new products are small, one can use the same two-factor hurdle rate approach independently for multiple investments that one would use if there was only a single investment. Of course, to the extent that the investments in the new products are not small, this decentralized approximation will be an imperfect one.

Overall, this line of reasoning suggests that the right question is not whether or not the bank should centralize its decisionmaking, but rather how often headquarters should gather information and use this pooled information to help guide investment decisions. Loosely speaking, what we have in mind is a dynamic version of the model wherein each time headquarters gathers information, it can update its estimate of both G, and the stochastic characteristics of the pre-existing portfolio. These updated estimates can then be passed back to individual product managers, who will use them to form hurdle rates and thereby do their best to approximate optimal incremental investment decisions on a decentralized basis over the interval of time before the next round of information pooling.

Although we have not analyzed such a dynamic model formally, we suspect that the following basic trade-off would emerge. On the one hand, shortening the interval between rounds of information pooling should lead to smaller deviations from the full-information centralized solution. On the other hand, this will also increase the costs of information transmission. The task is to properly balance these two competing considerations.

Again, we should emphasize that this informal story ignores any agency issues associated with delegating investment decisions to individual product managers. In reality, these considerations are likely to be important. For example, for a given information set, a product manager may have a tendency to take what the bank would view as excessive risks, because his reward structure is inherently a convex function of outcomes. In this case, decentralization may involve not only setting appropriate hurdle rates, but also imposing position limits or capital constraints on individual managers.

3.4. Optimal capital structure at Time 0

We are now in a position to fold back to Time 0 and solve for the optimal capital level K. There is a simple trade-off at work. On one hand, as noted above, a higher K reduces the bank's effective risk aversion G. From an ex ante perspective, this allows the bank to invest more aggressively in products that promise an above-market return at Time 1. On the other hand, a higher K also involves deadweight costs of τK.

To illustrate the first part of the tradeoff most transparently, consider a simple one-product case where the investment in question is a small one. In this setting, a natural question to ask is how the bank's hurdle rate, as given by Eq. (10),

changes with K:

$$d\mu_N^*/dK = \text{cov}(\varepsilon_N^N, \varepsilon_P^N)dG/dK = (\pi_N^*/G)dG/dK. \quad (13)$$

Since dG/dK is negative, Eq. (13) says that if the hurdle rate is initially above the market-required return, that is, if $\pi_N^* > 0$, the hurdle rate will fall smoothly as the amount of capital K is increased.

As of Time 0, the bank's objective function is to pick K so as to maximize $V - K$, recognizing that $V = V(w(\alpha^*(K), K))$. In words, K affects w directly through the amount of financial slack that will be available at Time 2, as well as indirectly through its influence on the optimal investment strategy α^*. One can use the envelope theorem to show that the solution to this problem can be written as

$$EP_w = 1/(1 - \tau). \quad (14)$$

Eq. (14) has an intuitive interpretation. The bank has two choices. It can hold more slack K as a buffer at Time 0, or it can be forced, in an expected sense, to seek more financing later on, with the attendant costs. It should optimally set K so that the expected shadow value of external funds at Time 2 just balances the cost of holding more capital at Time 0. In the limiting case where $\tau = 0$, so that there are no deadweight costs of holding capital, the bank holds an arbitrarily large amount. This drives the expected shadow value of external funds EP_w to one, which in turn implies that G converges to zero. Thus the bank behaves in a classical manner, doing capital budgeting according to a purely market-based model of risk and return. In contrast, as τ increases above zero, the bank holds less capital, thereby raising its effective risk aversion G, and amplifying the deviations from textbook capital budgeting principles.

4. Does capital really matter for intermediary pricing?

The fundamental premise of our model is that an intermediary's capital position affects its pricing of unhedgeable risks, even those risks which are completely idiosyncratic, and hence would be unpriced in a classical setting. While there is likely to be some degree of truth to this proposition, it is less obvious that capital constraints have enough of a first-order impact on intermediary investment decisions to warrant the sort of detailed capital budgeting analysis that we have been undertaking.

In general, it can be tricky to marshall unambiguous evidence in favor of the hypothesis that intermediary capital matters for capital budgeting. Empirical work typically faces two obstacles. First, hurdle rates cannot be directly observed, and second it is hard to find exogenous shocks to capital. Consider for example, the large recent literature on the so-called capital crunch effect in the commercial banking industry (see Sharpe, 1995). This literature has produced

a great deal of evidence that reductions in bank capital lead to substantial declines in lending volume, which is certainly consistent with the notion that capital matters for loan supply decisions. But there is also in many cases an alternative interpretation, namely that reductions in bank capital are symptomatic of a deterioration in the lending environment, so that falling lending volume may reflect a scarcity of good lending opportunities, rather than changes in banks' hurdle rates.

One recent paper that makes progress on this identification problem is Houston et al. (1996). Using data on the individual subsidiaries of bank holding companies, they find that the lending of any one bank in the holding company is strongly impacted by shocks to holding company capital, even when these shocks come from the holding company's non-bank subsidiaries. Arguably, shocks to the non-bank subsidiaries are unrelated to the lending opportunities of the banking subsidiaries, so the inferences are less ambiguous in this experiment.[7]

Perhaps even more clear-cut evidence of the importance of intermediary capital comes from work by Froot and O'Connell (1997) on the catastrophe reinsurance business. Catastrophe reinsurance is unique in at least two ways. First, to the extent that one can estimate the actuarial value of policies, the empirical analog of a reinsurer's hurdle rate can readily be calculated, as the market price charged by the reinsurer for a policy, less its actuarial value. Second, reinsurer capital is largely driven by shocks, such as earthquakes and hurricanes, that are exogenous, in the sense of not affecting the appeal of writing future reinsurance, holding fixed actuarial value.

Froot and O'Connell (1997) document the following key facts about the catastrophe reinsurance business. First, the average hurdle rate in this industry is very high. On average, over the period 1980–1994, price is on the order of four times actuarial value. Such a markup is especially striking given that the risks being insured are essentially uncorrelated with the market portfolio, so that a classical model would imply prices roughly equal to actuarial values. Second, prices rise sharply relative to actuarial values in the wake of large disasters that deplete reinsurer capital. For example, after Hurricane Andrew in 1992, the price index constructed by Froot and O'Connell approximately doubled. This type of pattern suggests that capital plays an extremely important role in intermediaries' pricing of unhedgeable risks.

A subtle caveat here is that a disaster like Hurricane Andrew may reveal information about the actuarial value of future hurricane policies. For example, it may show that severe storms are more likely, or construction in Florida is shoddier, than previously thought. If this is true, one has to be careful in making inferences based on how the price of hurricane reinsurance responds to a bad

[7] See also Peek and Rosengren (1997) for related evidence.

hurricane. However, Froot and O'Connell (1997) provide evidence that the prices of all catastrophic reinsurance contracts, even those which contain no Florida exposure or no hurricane exposure, such as California earthquake-based policies, respond in a similar way to a bad hurricane. This would seem to be definitive evidence of the importance of reinsurer capital.

5. Applications

5.1. A private investment or lending group

As noted above, the most literal interpretation of our tradeable versus non-tradeable risk decomposition would be to think of a group in a bank that invests in private companies, or slightly more generally, companies whose stock is sufficiently illiquid as to preclude easy hedging of the bank's position. To take a concrete illustration, a merchant banking group may have the opportunity to invest in a portion of the equity of a private company. How much should the group be willing to pay for a stake of a given size? Or equivalently, for a given price, how large a stake should the group be willing to take?

As in any valuation exercise of this sort, the group will have to do a discounted cash flow analysis of the company in question. The cashflow projections will be done in the usual manner. However, our results suggest that if risk management is a serious concern for the group's parent bank, the discount rates should differ from those used in a classical setting. Specifically, to calculate an appropriate discount rate, the group should start with the classical rate (e.g., a weighted average cost of capital based on the capital asset pricing model (CAPM)), and then add a premium π given by the appropriate variant of Eqs. (9)–(12).

For example, suppose the purchase price for the company in question has already been set at \$1000, and the group is deciding how many shares it wants to bid for. Its internal projections suggest that the company will have level cashflows of \$200 per year forever. In other words, the group perceives the internal rate of return on the investment as being 20%. Moreover, the group's best estimate of the classical discount rate for this company is 17%. Thus from the group's perspective, the deal offers a return premium π of 3%. If this is the only decision facing the bank at this time, one can now determine the optimal stake α in the company from Eq. (9).

5.2. A proprietary trading desk

A slightly less literal, but nonetheless useful, application of our framework is to a proprietary trading operation located inside a larger financial institution. For the time being, suppose we are thinking of a desk that trades actively in

simple linear instruments such as futures and forwards. At first glance, it might appear that our approach would be of little use in thinking about such a desk. To the extent that all the instruments it deals in a relatively liquid, it can in principle hedge any risk it faces, and therefore our tradeable versus non-tradeable risk decomposition would seem to have little bite.

However, one needs to be a bit careful with the interpretation of the words "non-tradeable". Even if all the risks facing the desk are hedgeable in principle, this obviously cannot be what the desk does in practice. If it did hedge out all of its risks, the desk would have no business. In other words, being a trading desk by definition requires intentionally assuming certain exposures. Ostensibly, these exposures are justified by the desk's ability to earn a positive return on average, even after adjusting for market-wide risk factors. The presence of such positive, albeit subjective, risk-adjusted returns makes such exposures non-tradeable in our sense.

Seen in this light, our framework can be helpful in thinking about two closely related questions facing the managers of a trading desk. First, there is the ex ante capital budgeting question: given a particular directional prediction concerning an asset, how aggressively should the desk invest in that asset? Second, there is the ex post performance measurement question: how can one evaluate whether the desk made enough of a profit to compensate for the risks it imposed on the bank as a whole?

In the case where the trading desk hedges out all the risk associated with the priced market factor M, our previous results apply directly, so that one can again use the appropriate variant of Eqs. (9)–(12) to answer both of these questions. For example, if the desk is making a go/no-go decision on a trade ε_N of relatively small size that is uncorrelated with M, Eq. (10) says that the decision should be to go ahead only if the trade offers a subjective expected return that exceeds $G\ \mathrm{cov}(\varepsilon_N, \varepsilon_P^N)$.

An added wrinkle arises in the case where the trading desk also has a subjective view on the priced factor M, and therefore chooses not to hedge out the risk associated with M, even though this is in principle feasible. (Recall that in deriving our capital budgeting results, we relied on Proposition 2, which said that without such a subjective view, the bank would always hedge M-risk completely.) In Appendix A, we show that the relevant analogue to the hurdle-rate result in Eq. (10) is now given by

$$\mu_N^* = \gamma\ \mathrm{cov}(\varepsilon_N, M) + G^M\ \mathrm{cov}(\varepsilon_N, \varepsilon_P^N), \tag{15}$$

where G^M is a slightly modified version of the risk-aversion parameter G that takes account of the complication that the bank is now bearing priced M-risk. We define G^M formally in terms of primitive variables in Appendix A. Thus, the basic logic is exactly the same as before and the only change is that the risk aversion parameter is calculated differently.

5.3. *Pricing non-hedgeable derivative positions*

Finally, our approach may be useful in helping to price derivatives positions that a bank cannot cost-effectively hedge. For concreteness, suppose the bank is acting as a dealer and has been asked to write a put option on the equity of another firm. If the option can be effectively delta-hedged by trading in the underlying equity, we have the case where all the risk is tradeable. Thus the option should be priced using the standard Black-Scholes (1973) approach.

However, suppose instead that the firm in question is either privately held, or only very thinly traded. More precisely, the current market value of the firm is observable, but because of either trading costs or short-selling constraints, it is infeasible to hedge the option. Thus, if the bank writes the option, it must bear the associated exposure. What price should the bank now charge for the option?

To attack this problem, we need to make a few further assumptions. First, the market value of the firm's stock, S, follows a lognormal diffusion process with drift θ (for simplicity, there are no dividends) and instantaneous variance σ:

$$dS = \theta S\, dt + \sigma S\, dz. \tag{16}$$

Second, we also need to assume that we are in the limiting case where the investment under consideration is very small. This allows us to express the bank's required return on the stock μ_S^* as a fixed constant, in the form of Eq. (10).[8] Note that, if the stock is positively correlated with the rest of the bank's portfolio, then $\mu_S^* > \theta$. That is, the bank's required return exceeds the market's required return. This result arises from the bank's risk aversion. If the bank were forced to hold the underlying stock directly, it would value it at a discount relative to the market. As before, we can define the premium that the bank requires on the stock, above and beyond the market required return, as $\pi_S^* \equiv \mu_S^* - \theta = G\,\mathrm{cov}(dS/S, \varepsilon_P^N)$.[9]

With these definitions in hand, Appendix A shows that the value of the option from the bank's perspective F, satisfies the following partial differential equation:

$$(r - \pi_S^*)SF_S + F_t + \tfrac{1}{2}\sigma^2 S^2 F_{SS} - rF = 0. \tag{17}$$

This is exactly the same valuation equation that one would obtain in a classical pricing setting where the underlying stock paid a proportional dividend of π_S^*. Thus, to adjust the option's value for non-hedgeability, all one needs to do is

[8] The only distinction is that, because we are now talking about the required return on a stock, instead of a forward position, one must add the riskless rate r to the expression in Eq. (10) to obtain μ_S^*. See Section 2.2.

[9] In this continuous time setting, ε_P^N should now be interpreted as the instantaneous innovation in the rate of return on the non-tradeable component of the pre-existing portfolio.

take a standard pricing model and augment the dividend yield by π_S^*. The one slight complication is related to early exercise. In the example above, the holder of the put option will generally attach a different value to the option than the bank. And the holder's decision of when to exercise the option will be determined so as to optimize value from his perspective. This exercise strategy must then be incorporated when valuing the option from the bank's perspective.

In the context of our example of a bank writing a put option, Eq. (17) implies that the bank will charge a higher price than suggested by a standard model. The intuition is as follows: the put option is written on a stock which the bank discounts at a rate of π_S^* more than the market. Thus, from the bank's perspective, the underlying stock is worth less than it would be in the open market, and accordingly, the put option is worth more. The upward effect on the option price is the same as would occur if the option was fully hedgeable, but was written on a stock that paid a dividend of π_S^*.

More generally, the same method can be used to value a wide range of illiquid derivatives positions. For example, one might wish to value illiquid credit risks of the sort discussed in the Introduction. Following Merton (1974), one could take the approach of modeling a bank's credit exposure to a firm as being equivalent to a short put position in the firm's market value. Of course, this basic type of model can be tailored along a number of dimensions, according to how one wants to treat issues of priority in bankruptcy, for example. But whatever the specific variant of the perfect markets pricing model is chosen, our results suggest that it can be readily adjusted for illiquidity, once the value of π_S^* has been established.

6. Comparison with RAROC-based capital budgeting

As noted in the Introduction, an increasingly popular approach to capital budgeting among banks is based on the concept of RAROC, or risk-adjusted return on capital.[10] Now that we have worked through several examples using our own framework, it is useful to see how our method compares with RAROC. For simplicity, we will consider the case of a small, normally distributed new investment opportunity. In this case, our model says that the appropriate hurdle rate that will maximize value is given by the two-factor model of Eq. (10).

[10] The RAROC concept was developed at Bankers Trust in the late 1970's. For recent descriptions of how it can be used for capital budgeting, see Zaik et al. (1996), James (1996), and Uyemura et al. (1996).

6.1. The RAROC approach

Different banks implement RAROC in different ways. We will discuss some of the variations momentarily. However, the general approach can be described as follows. Each investment under consideration is allocated a certain amount of capital. Multiplying the allocated amount of capital by a cost of capital yields a capital charge. The hurdle rate for the investment is then the relevant riskless rate, plus the capital charge. Translated into our notation, we have:

$$\mu_N^R = E_N^R(k_N^{ER} - r), \tag{18}$$

where μ_N^R is the RAROC required return, in excess of the riskless rate, for the new investment, E_N^R is the amount of capital the RAROC model allocates to this investment, and k_N^{ER} is the cost of capital as computed by the RAROC model. In all variants of RAROC of which we are aware, the capital allocation E_N^R is related to a measure of the investment's risk. Thus, RAROC can be thought of as a one-factor risk-pricing model.

Comparing Eq. (18) with Eq. (10), we can see that our model will coincide with the RAROC approach only if the following three conditions are satisfied:

1. The investment in question must contain no priced tradeable risk. In other words, it must be the case that $\text{cov}(\varepsilon_N^T, M) = 0$. Said differently, if the bank uses RAROC to evaluate an investment, this investment must be considered on a post-hedged basis, where all the priced tradeable risk has already been hedged out.
2. One must be able to express the capital charge E_N^R for the new investment as a linear function of that investment's covariance with the bank's existing portfolio. In other words, one must be able to write $E_N^R = \zeta\,\text{cov}(\varepsilon_N^N, \varepsilon_P^N)$, for some parameter ζ.
3. Finally, it must be that $\zeta(k_N^{ER} - r) = G$, where G is our risk aversion parameter.

6.2. Potential pitfalls in the implementation of RAROC

The three conditions above highlight different ways that one can go wrong – at least relative to our model's implications for value-maximizing behavior – in applying a RAROC approach to capital budgeting. We now discuss each of these in turn, and where possible, comment on the current state of practice as we understand it.

6.2.1. Inadequate separation of priced and non-priced risks.

One of the key implications of our two-factor model is that a bank should evaluate investments according to *both* their correlation with any priced market

factors and their correlation with the bank's existing portfolio. Taken literally, the one-factor RAROC approach does not allow for these two degrees of freedom. Of course, if the RAROC model is only used for investments that contain no priced market risks, perhaps because these priced risks have already been sold off before the model is applied, then there will be no problem. But to see where things can easily go wrong, think back to our example of the proprietary trading desk in Section 5.2 above. One can easily imagine situations where the desk increases its exposure to the priced market factor M, without changing either its total volatility or its correlation with the rest of the bank's portfolio, (perhaps because the rest of the bank is M-neutral). In this case, the required return for the desk should increase, but a direct application of the standard RAROC methodology is unlikely to capture this effect.

6.2.2. Capital allocations based on measures of variance rather than covarience

As noted above, the amount of capital E_N^R allocated to an investment is typically related to some measure of that investment's risk. But in some applications of RAROC, the risk measure used is the investment's total volatility (see, e.g., Uyemura et al., 1996). This approach is at odds with value maximization. As we have shown, it makes more sense for the capital allocation to be driven by the investment's covariance with the rest of the bank's portfolio. Fortunately, the preferred covariance-based approach seems to be gaining some favor. For example, James (1996, 11) in discussing the implementation of RAROC at Bank of America, writes, "the amount of capital allocated varies with the contribution of the project to the overall volatility of earnings at B of A (the project's so called internal beta)". This statement is clearly in the spirit of our model.

6.2.3. Using an incorrect cost of capital

Even if the RAROC capital allocation is based on the appropriate covariance measure, one still has to come up with the right values of ζ and k_N^{ER}. From the perspective of our model, the individual values chosen do not really matter, so long as the two together satisfy $\zeta(k_N^{ER} - r) = G$.

The most common practice for selecting these values seems to be as follows. First, ζ is chosen to ensure that the probability of the bank defaulting is less than some threshhold level. For example, according to Zaik et al. (1996), at Bank of America, a default probability of 0.03% is tolerated. Second, the cost of capital, k_N^{ER}, is typically set to equal the required return on equity for the bank's shareholders, which could be calculated from the CAPM (see Zaik et al., 1996). This latter calculation appears in the context of our model to be illogical. The fallacy can be most easily seen by considering the polar case where the bank, as suggested by our model, hedges all priced risks. In this case, the bank's shareholders are left only holding non-priced risks, so their required return on equity is simply the riskless rate r. But if this is true, then the RAROC method says that the capital charge should be zero for any values of ζ and E_N^R!

Simply put, shareholders' required return on the bank's equity bears no relationship to what we would ideally like to capture, which is the parameter G. This parameter is in principle influenced by the deadweight costs of holding equity on the balance sheet, which is a very different concept than the required return on equity. To consider another polar case, as the tax rate τ goes to zero, so does G. In this case, our model would imply that it is costless, in a deadweight sense, to hold equity, so there should be no markup at all for non-priced risks. Yet a RAROC model driven by the required return on equity might incorrectly continue to apply a markup.

In fairness to the RAROC method, our risk aversion parameter G, while theoretically more appropriate, is harder to estimate from readily available data. To construct G, one ultimately needs to either measure or simulate the shadow value of internal funds P_w, and to understand how it varies as a function of the level of internal funds w.

7. Conclusions

This paper has stressed the relation between risk management, capital budgeting, and capital structure policies for banks. In our model, all three of these policies are shaped by two related primitive frictions. First, it is costly for banks to raise new external funds on short notice. Second, it is also costly for banks to hold a buffer stock of equity capital on the balance sheet, even if this equity is accumulated over time through retained earnings.

Given these frictions, bank-value maximization implies the following conclusions. First banks should hedge any risks that can be offloaded on fair-market terms. Second, banks should also hold some capital as a device for absorbing those illiquid risks which cannot be hedged, but the optimal amount of capital is limited. Finally, given limited capital, banks should value illiquid risks much as an individual investor would – according to their impact on overall portfolio risk and return – with the degree of risk aversion being a decreasing function of the amount of capital held.

Because capital budgeting in our model is driven by an exposure's impact on portfolio risk and return, decentralized implementation becomes a problem if the pre-existing portfolio cannot be treated as fixed for all time. Thus, for example, if several new investment opportunities arrive at once, the optimal allocation to each must be jointly determined. This requires a well-informed central planner operating out of bank headquarters.

The whole issue of centralization versus decentralization strikes us as very important in the context of financial institutions, where asset portfolios can in principle be adjusted quite rapidly, and where fully centralized decision making could become extremely unwieldy. The fact that our model has had little of substance to say on this front suggests that it is missing certain key ingredients.

As discussed above, the two most glaring omissions in this regard are: (i) the lack of an explicitly dynamic element, whereby information about investment opportunities evolves continuously over time, but where headquarters might choose to collect data only at discrete intervals, and (ii) the absence of any agency problems at the level of individual product managers. Incorporating one or both of these considerations into our general framework would be a natural next step, and might shed some light on a variety of practices, such as the widespread use of capital or position limits for individuals or small units within a bank, that seem to be important in the real world but which cannot be understood in the context of the current version of our model.

Acknowledgements

This research is supported by the Global Financial System Project at Harvard Business School (Froot), the Finance Research Center at the Massachusetts Institute of Technology (Stein), and the National Science Foundation (Stein). Thanks to Maureen O'Donnell and Melissa Cunniffe for help in preparing the manuscript, and to Fischer Black, Chris James, Bob Merton, Andre Perold, Richard Ruback, René Stulz, an anonymous referee and seminar participants at New York University and the National Bureau of Economic Research for helpful comments.

Appendix A.

A.1. Proof of Propositions 1 and 2

Since the bank may face numerous tradeable exposures, we prove Propositions 1 and 2 allowing for multiple exposures. We denote the bank's hedging policy as a set of coefficients h_i, which weight a set of L tradeable factors, x_i, $i = 1, \ldots, L$. This implies that $Z_H = \sum h_i x_i$. We assume that the last of these x_i's is the market factor, i.e., $x_L = M$. We show below that the value-maximizing set of h_i's is that which emerges from a simple regression of pre-hedged wealth, $Z_P + \alpha Z_N + K(1 - \tau)$, on the x_i's. This hedging policy is complete, in that it minimizes the variance of wealth by making post-hedged wealth at Time 2 uncorrelated with each tradeable factor.

To derive this, note that the bank chooses its hedging policy to maximize value, V.

This implies that

$$\frac{dV}{dh_i} = \frac{dE(P(w))}{dh_i} - \gamma \frac{d\,\mathrm{cov}(P(w), M)}{dh_i} = 0, \tag{A.1}$$

where dh_i represents a change in the hedge ratio for the ith exposure. Using the definition of covariance, we can write

$$\frac{dV}{dh_i} = E(P_w)\,E\left(\frac{dw}{dh_i}\right) + \operatorname{cov}\left(P_w, \frac{dw}{dh_i}\right) - \gamma\frac{d\operatorname{cov}(P(w), M)}{dh_i}. \tag{A.2}$$

Next, given that the components of w are normally distributed, we can use the fact that, for normally distributed x and y, $\operatorname{cov}(f(x),y) = E(f_x)\operatorname{cov}(x,y)$. Applying this yields

$$\frac{dV}{dh_i} = E(P_w)\,E\left(\frac{dw}{dh_i}\right) + E(P_{ww})\operatorname{cov}\left(w, \frac{dw}{dh_i}\right) - \gamma\frac{d(E(P_w)\operatorname{cov}(w, M))}{dh_i}. \tag{A.3}$$

We then perform the differentiation indicated in the last term, and observe that, if the bank performs its hedging in an efficient market, changes in risk and return must be related by

$$E\left(\frac{dw}{dh_i}\right) = \gamma\frac{d\operatorname{cov}(w, M)}{dh_i}. \tag{A.4}$$

Thus, noting that $dw/dh_i = x_i$, we are left with

$$\frac{dV}{dh_i} = E(P_{ww})\operatorname{cov}(w, x_i) - \gamma E(P_{ww}x_i)\operatorname{cov}(w, x_L) = 0. \tag{A.5}$$

Given the concavity of $P(w)$, the solution to these L equations involves setting each h_i so as to make its associated factor uncorrelated with internal wealth, i.e., setting $\operatorname{cov}(w, x_i) = 0$ for all i. This can only be accomplished by stripping all of the tradeable exposures from internal wealth. Note that this logic applies regardless of whether or not all the risks are tradeable. Thus, we have proven both Propositions 1 and 2.

A.2. Derivation of Eq. (9)

The bank's first-order condition is

$$\frac{dV}{d\alpha} = \frac{dE(P(w))}{d\alpha} - \gamma\frac{d\operatorname{cov}(\mathrm{P(w)}, M)}{d\alpha} = 0. \tag{A.6}$$

This expression can be rewritten as

$$\operatorname{cov}\left(P_w, \frac{dw}{d\alpha}\right) + E(P_w)E\left(\frac{dw}{d\alpha}\right) - \gamma\operatorname{cov}(w, M)E\left(P_{ww}\frac{dw}{d\alpha}\right)$$
$$- \gamma E(P_w)\frac{d\operatorname{cov}(w, M)}{d\alpha} = 0. \tag{A.7}$$

Using the fact that for normally distributed x and y, $\text{cov}(f(x), y) = E(f_x)\text{cov}(x, y)$, and rearranging the terms, this expression can be rewritten as

$$\frac{dV}{d\alpha} = E(P_w)\left(E\left(\frac{dw}{d\alpha}\right) - \gamma\frac{d\,\text{cov}\,(w, M)}{d\alpha}\right) + E(P_{ww})\left(\text{cov}\left(w, \frac{dw}{d\alpha}\right)\right.$$
$$\left. - E\left(\frac{dw}{d\alpha}\right)\gamma\,\text{cov}(w, M)\right) - E(P_{www})\left(\gamma\,\text{cov}(w, M)\,\text{cov}\left(w, \frac{dw}{d\alpha}\right)\right) = 0. \tag{A.8}$$

Next, note that

$$\frac{dw}{d\alpha} = \mu_N + \varepsilon_N^N - \gamma\,\text{cov}(\varepsilon_N^T, M)$$
$$E\left(\frac{dw}{d\alpha}\right) = \mu_N - \gamma\,\text{cov}(\varepsilon_N^T, M)$$
$$\text{cov}\left(w, \frac{dw}{d\alpha}\right) = \text{cov}(\varepsilon_N^N, \varepsilon_P^N) + \alpha\,\text{var}(\varepsilon_N^N)$$
$$\frac{d\,\text{cov}(w, M)}{d\alpha} = \text{cov}(\varepsilon_N^N, M). \tag{A.9}$$

Substituting these expressions into Eq. (A.8), and rearranging, yields a generalization of Eq. (9) in the text:

$$\mu_N = \gamma\,\text{cov}(\varepsilon_N^T, M) + \gamma\,\text{cov}(\varepsilon_N^N, M) + G^M(\text{cov}(\varepsilon_N^N, \varepsilon_P^N) + \alpha\,\text{var}(\varepsilon_N^N)), \tag{A.10}$$

where

$$G^M = \frac{-(E(P_{ww}) - \gamma E(P_{www})\,\text{cov}(w, M))}{E(P_w) - \gamma E(P_{ww})\,\text{cov}(w, M)}. \tag{A.11}$$

One last step is required to generate Eq. (9) as it appears in the text. Eq. (A.10), above, is more general, in that it explicitly allows for the possibility that $\text{cov}(\varepsilon_N^N, M) \neq 0$ and hence $\text{cov}(w, M) \neq 0$, even after the bank has hedged. As we discuss in Section 5.2 of the text, a trading desk may wish to leave itself exposed to the market factor M in this way, if it has a subjective view on the expected return to this factor. However, in Eq. (9), we consider the simpler case, in which hedging completely removes the market factor, so that $\text{cov}(\varepsilon_N^N, M) = \text{cov}(w, M) = 0$, and G^M reduces to $G = -E(P_{ww})/E(P_w)$.

A.3. Derivation of Eq. (17)

By Ito's lemma, the instantaneous expected change in the value of the option is given by

$$E(dF) = \theta SF_S + F_t + \tfrac{1}{2}\sigma^2 S^2 F_{SS}. \tag{A.12}$$

This must be equal to $\mu_F^* F$, where μ_F^* is the bank's required return on the option. By Eq. (10), we have

$$\mu_F^* = r + \gamma \mathrm{cov}\left(\frac{dF}{F}, M\right) + G\mathrm{cov}\left(\frac{dF}{F}, \varepsilon_P^N\right). \tag{A.13}$$

By the linearity of the covariance operator, this can be rewritten as

$$\mu_F^* = r + \frac{F_S S}{F}\left(\gamma \mathrm{cov}\left(\frac{dS}{S}, M\right) + G\mathrm{cov}\left(\frac{dS}{S}, \varepsilon_P^N\right)\right) = r + \frac{F_S S}{F}(\mu_S^* - r) \tag{A.14}$$

Thus, the valuation equation is given by

$$\theta \mathrm{SF}_S + F_t + \tfrac{1}{2}\sigma^2 S^2 F_{SS} = F\left(r + \frac{F_S S}{F}(\mu_S^* - r)\right) \tag{A.15}$$

Substituting $\pi_S^* = \mu_S^* - \theta$ and rearranging gives Eq. (17), as it appears in the text.

References

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Cooper, I., Mello, A., 1991. The default risk of swaps. Journal of Finance 46, 597–620.

Froot, K., O'Connell, P., 1997. On the pricing of intermediated risks: theory and application to catastrophe reinsurance. Unpublished working paper, Harvard Business School, Boston, MA.

Froot, K., Scharfstein, D., Stein, J., 1993. Risk management: coordinating corporate investment and financing policies. Journal of Finance 48, 1629–1658.

Gale, D., Hellwig, M., 1985. Incentive-compatible debt contracts I: the one-period problem. Review of Economic Studies 52, 647–664.

Greenwald, B., Levinson, A., Stiglitz, J., 1991. Capital market imperfections and regional economic development. In: Giovannini, A. (Ed.), Finance and Development: Lessons and Experience. The Center for Economic Policy Research, pp. 65–93.

Houston, J., James, C., Marcus, D., 1996. Capital market frictions and the role of internal capital markets in banking. Unpublished working paper. University of Florida, Gainesville.

James, C., 1996. RAROC based capital budgeting and performance evaluation: a case study of bank capital allocation. Unpublished working paper. University of Florida, Gainesville.

Jarrow, R., Turnbull, S., 1995. Pricing derivatives on financial securities subject to credit risk. Journal of Finance 50, 53–85.

Kashyap, A., Stein, J., 1995. The impact of monetary policy on bank balance sheets. Carnegie-Rochester Conference Series on Public Policy 42, 151–195.

Merton, R., 1974. On the pricing of corporate debt: the risk structure of interest rates. Journal of Finance 29, 449–470.

Merton, R., 1995a. A model of contract guarantees for credit-sensitive opaque financial intermediaries. Unpublished working paper. Harvard Business School, Boston, MA.

Merton, R., 1995b. A functional perspective on financial intermediation. Financial Management 24, 23–41.

Merton, R., Perold, A., 1993. The theory of risk capital in financial firms. Journal of Applied Corporate Finance 5, 16–32.

Modigliani, F., Miller, M., 1958. The cost of capital, corporation finance and the theory of investment. American Economic Review 48, 261–297.

Myers, S., 1984. The capital structure puzzle. Journal of Finance 39, 575–592.

Myers, S., Majluf, N., 1984. Corporate financing and investment decisions when firms have information that investors do not have. Journal of Financial Economics 13, 187–221.

Peek, J., Rosengren, E., 1997. The international transmission of financial shocks: the case of Japan. American Economic Review 87, forthcoming.

Sharpe, S., 1995. Bank capitalization, regulation, and the credit crunch: a critical review of the research findings. Unpublished working paper # 95-20. Federal Reserve Board, Washington, D.C.

Stein, J., 1996. An adverse selection model of bank asset and liability management with implications for the transmission of monetary policy. Unpublished working paper. The Massachusetts Institute of Technology, Cambridge, MA. Forthcoming in RAND Journal of Economics.

Townsend, R., 1979. Optimal contracts and competitive markets and costly state verification. Journal of Economic Theory 21, 265–293.

Uyemura, D., Kantor, C., Pettit, J., 1996. EVA for banks: value creation, risk management and profitability measurement. Journal of Applied Corporate Finance 9, 94–133.

Zaik, E., Walter, J., Kelling, J., James, C., 1996. RAROC at Bank of America: From theory to practice. Journal of Applied Corporate Finance 9, 83–93.