Capital regulation of banks:
Where do we stand and where are we going?

Robert Bichsel and Jürg Blum, Systemic Stability,
Swiss National Bank, Zürich
1 Introduction

For almost two decades now, capital adequacy requirements have been the focus of international banking regulation. After a period of worldwide liberalisation and deregulation, the Basel Capital Accord of 1988 ("Basel I") marked the beginning of a new phase of re-regulation with an attempt to bring about an international harmonisation of banking regulations. In June 2004, several years of revising and renegotiating Basel I have led to the endorsement of a new capital adequacy framework ("Basel II"), which is planned to come into effect by 2007. This important milestone is an opportunity for us to take stock of the current state of capital regulation.

The purpose of this paper is threefold. First, we take a step back and address the question of the economic rationale for capital regulation in the banking sector. Second, we give an overview of instruments currently available for determining the minimal level of required capital. And finally we identify three principles which we believe should guide the evolution of capital adequacy requirements in the future.

Our main message is that capital adequacy rules and, more specifically, "risk-weighted capital requirements" are socially desirable. However, risk-measurement and information-asymmetry issues, which are inherent to banking activities, prevent the implementation of first-best capital adequacy rules, i.e., capital requirements that fully and exactly reflect banks' underlying risks. In particular, we stress the fact that the hopes raised by the so-called "full-model approach", according to which risk-measurement issues could be addressed by delegating risk assessment to the banks themselves, are misplaced. A consequence of the inability of capital requirements to fully and exactly reflect banks' underlying risks is that any realistic capital adequacy scheme will leave banks some room for regulatory arbitrage. Acknowledging this, we claim that, as a safeguard against banks exploiting misaligned risk weights and choosing extremely risky portfolios, risk-sensitive capital requirements should be complemented by a capital floor which is independent of banks' risks.

1 We would like to thank Niklaus Blattner, Bertrand Rime, and an anonymous referee for helpful comments and discussions.


2 Rationale for capital regulation

This section summarises the main motivation for the regulation of bank capital. After outlining the factors that determine banks' optimal capital structure, we discuss two reasons why banks' voluntary choices of capital structure may differ from a socially efficient structure. In particular, we argue that banks have a tendency to hold too little capital and too much debt. In this context, capital adequacy requirements should be viewed as an instrument to align banks' capital ratios with the socially optimal ones.

The optimal capital structure of banks

Discussions about capital structure usually start with the seminal result by Modigliani and Miller (1958). They show that under the assumption of perfect markets, the financing choice between debt and equity does not affect the value of a firm, i.e., the financing mix is irrelevant and capital structure is not uniquely determined. This result provides a convenient starting point for further discussions about optimal capital structure. Not because it is realistic, as in reality capital structure clearly can matter, but because it highlights the factors that influence firms' financing choices.

Crucial for the result by Modigliani and Miller is the assumption of perfect, frictionless markets. This assumption implies, in particular, that there are no asymmetries of information and everybody is able to obtain credit at the market rate. In reality, however, there are two important deviations from this frictionless world: taxes and bankruptcy costs. On the one hand, debt provides a "tax shield". As opposed to dividends paid to equity holders, interest paid to lenders is a tax-deductible expense. Therefore, increasing the level of debt lowers the taxes paid by a firm and increases its after-tax pay-out to debt and equity holders. On the other hand, increasing the level of debt raises the probability of default and thereby the expected bankruptcy costs. In this simple tradeoff theory, a firm will borrow up to the point where the marginal benefit of tax shields is equal to the marginal expected cost of financial distress.

3 These classic deviations were already analyzed by Modigliani and Miller (1958) themselves. In addition, a large body of literature has developed which explores other deviations from the perfect markets paradigm. See Myers (2001) for a survey of the various competing theories of capital structure.
Whatever determines the optimal debt-equity choice of firms in general, there are two main reasons why debt is more attractive to banks than to other firms. First, a considerable share of banks’ debt consists of demand deposits and other very short-term liabilities. Since depositors value the high liquidity of these claims, they are willing to accept a lower interest rate than they could receive by investing in less liquid assets. Due to this “liquidity premium”, the marginal costs of (short-term) debt are lower for banks than other firms. Second, banks’ debt holders are protected by an extensive safety net. Thanks to (explicit) deposit insurance and (implicit) government guarantees, banks’ debt is perceived to be relatively safe – independent of the banks’ actual risk. This “subsidy” also contributes to a higher preference for debt by banks. Indeed, while additional factors influence the capital structure decision, banks typically have very low capital ratios (see Box).

The case for capital regulation
As discussed in the previous section, banks typically choose a positive capital ratio, even on a voluntary basis. Nevertheless, from an overall economic point of view, unregulated banks tend to hold too little capital. Two effects contribute to the divergence of a bank’s privately optimal capital ratio from the socially optimal capital ratio.

First, banks fail to give attention to the negative externalities and costs to third parties that would occur if they went bankrupt. While this is true for all firms, it can be argued that the externalities in banking are especially severe. These negative externalities include disruptions to the payments system and a general loss of confidence in the banking system (with possible contagious bank runs on other banks). The reduction of credit due to a banking crisis can slow economic growth and lead to costs in terms of reduced GDP. It is true that banks themselves have an interest in holding capi-

Box: Banks’ capital ratios in perspective

The ratio of capital to total assets for banks worldwide is typically well below 10%. For instance, the two Swiss big banks – which are well-capitalised according to their risk-weighted capital ratios – have (unweighted) capital ratios below 3%. The following comparisons help to put these numbers into perspective:

- Typical capital ratios of firms that are listed on a stock exchange range between 30% and 40%. Large, internationally successful companies such as Nestlé (40%) or Novartis (63%) also have high capital ratios.
- Historically, banks’ (unweighted) capital ratios used to be much higher than nowadays. Around 1900, for instance, the capital to total assets ratios of the Swiss big banks were higher than 20% and those of the cantonal banks well exceeded 10%.

- Banks themselves usually consider their own borrowers creditworthy if these borrowers have minimal capital ratios in the order of magnitude of 30%. Depending on other characteristics of borrowers, this limit may be higher or lower.

These comparative figures illustrate that, as a percentage of their assets, banks’ capital buffers are extraordinarily thin both from a historical perspective and when compared to other industries. However, this does not necessarily mean that banks’ capital ratios are currently too low, i.e., below what would be desirable from the perspective of the social optimum.

4 For an extensive survey of the literature on capital structure, see also Harris and Raviv (1991).
5 For an overview of the costs of banking crises, see Hoggarth and Saporta (2001).

6 On a consolidated basis, end of 2004. See section 3.2 for a definition of risk-weighted capital ratios.
tal in order to avoid bankruptcy and ensure their continued existence. Due to limited liability, however, they neglect the consequences of their insolvency as described above, and therefore hold too little capital relative to the socially optimal amount that would take these costs into account.7

Second, as described in the previous section, part of banks’ preference for debt stems from the subsidy they enjoy on their debt. Due to generally underpriced deposit insurance and government guarantees, not all bank debt fully reflects the underlying risks. The costs of this safety net are borne by the deposit insurance, the government, or ultimately the taxpayer. Again, since banks do not take the true total cost of debt into account, they have a tendency to borrow more than socially optimal.8

Given the tendency of banks to choose capital ratios that are too low relative to first best, capital regulation can be viewed as an obvious attempt to correct this market failure.9 By counteracting the described distortions, capital requirements aim to improve overall economic efficiency.

3 Capital adequacy rules

Capital requirements can take different forms. We distinguish three broad classes of capital adequacy schemes: The simple gearing-ratio limit, the traditional risk-weighted capital requirements, and the so-called full-model approach.10

3.1 Gearing-ratio limit

Imposing a “gearing-ratio limit” (GRL), i.e., setting an upper bound to a bank’s debt-to-capital ratio, is probably the most obvious way to prevent a bank from holding excessive amounts of debt. From a prudential regulatory perspective, this instrument has three attractive features. First, a GRL guarantees a minimum buffer that protects a bank against the consequences of losses. The lower the gearing ratio, the bigger the shocks a bank can withstand without failing. Second, a GRL sets the minimum loss that has to be borne by shareholders. This affects the bank’s risk-taking behaviour: the lower the gearing ratio, the smaller the moral-hazard effect due to limited liability and hence the smaller the risk-taking incentive – comparable to deductibles in insurance contracts. Third, the GRL is a simple and transparent rule. It is simple for banks to apply, while compliance is easily verified by supervisory authorities and market participants.

At the same time, however, the simplicity of the GRL is also its main drawback. A GRL is a capital adequacy rule which takes only one component of a bank’s overall risk profile into account – its leverage. The second component – the riskiness of its assets – is ignored. As a consequence, imposing a GRL may provide an incentive to proceed to risk shifting, potentially leading to a higher risk of default. The underlying mechanism is as follows. From the bank’s point of view, the uniform capital requirement imposed by a GRL is too high for very safe assets and too low for high-risk assets. This means that, in the absence of any regulation, a bank would hold less capital against low-risk assets and more capital against high-risk assets. Hence, the bank may be tempted to substitute high-risk assets, which seem relatively cheap in terms of required capital, for low-risk assets, which seem too “expensive”.11

9 Of course, there are other ways to influence banks’ behaviour and risk. They include investment restrictions, liquidity requirements, interest rate ceilings, and measures to ensure greater market discipline (e.g., by reducing the financial safety net). See Freixas and Rochet (1997) for a textbook treatment. Here we choose to ignore these instruments and to focus exclusively on capital adequacy requirements.
10 A number of alternative approaches have been proposed, but have only played a limited role in the regulatory discussion. They include the pre-commitment approach (Kupiec and O’Brien, 1995), the supervisory approach (Estrella, 1998), and the base-plus approach (Shepherd-Walwyn and Litterman, 1998). For further reading on capital requirements, see Dewatripont and Tirole (1994) or the survey by Santos (2001).
11 For a formal treatment of this argument, see Kim and Santomero (1988) or Rochet (1992).
As a consequence, the effect of a GRL on a bank’s overall risk profile, i.e., its likelihood of failure, will depend on two effects. First, there is the buffer effect, which ensures that – for constant levels of risk – the higher a bank’s capital the less likely it is to fail. Second, there is the indirect effect of a GRL on banks’ risk-taking incentives. The sign of this second effect is ambiguous. Risk may decrease or increase, depending on the relative magnitude of the limited liability effect (risk reduction) and the asset substitution effect (risk increase). In theory, therefore, the net effect of a GRL on banks’ probability of failure is also ambiguous. In practice, however, it is still an open question whether the possibility of a GRL increasing banks’ overall risk profile is a plausible scenario or whether it is an unrealistically extreme case.12

3.2 Risk-weighted capital requirements

An obvious way to improve the GRL is to take into account the riskiness of banks’ activities when determining the level of required capital. Banks holding more risky assets should be subject to higher capital requirements. This approach is generally referred to as “risk-weighted capital requirements” (RWCR), reflecting the fact that under this scheme, a lower bound to the ratio between a bank's capital and a weighted sum of its assets is defined, whereby a higher weight is attached to riskier assets. While RWCR are conceptually simple and intuitively appealing, their practical implementation is hampered by a fundamental problem: the measurement of the riskiness of banks’ assets. This difficulty stems mainly from the inherent opaqueness of banking activities. Banks are specialised in the financing of projects for which direct access to the capital market is limited as a result of information asymmetries. Outsiders, who do not possess the same information as the banks, are at a disadvantage and are not able to adequately measure the assets’ riskiness.13 As a consequence, the supervisory authority, as an outsider, faces two related difficulties: It has to define risk weights for assets or classes of assets whose risk it cannot measure precisely, and it has to verify whether banks correctly assign their assets to the pre-defined risk classes. This will generally lead to risk weights which do not accurately reflect the underlying risks and hence give rise to “regulatory arbitrage” opportunities. As under a GRL, banks will have an incentive to substitute assets whose risk weights are considered to be too high for assets whose risk weights are too low.

The Basel Capital Accord of 1988 (Basel I), which established internationally harmonised capital requirements, provides a good example of a RWCR scheme which provides room for regulatory arbitrage opportunities. Basel I is characterised by a small number of broad and heterogeneous risk classes. For example, corporate loans are all given the same risk weight of 100% – regardless of borrowers’ credit standing and hence regardless of the underlying risk. More generally, under Basel I, risk weights that accurately reflect the underlying risks are the exception rather than the rule. Over time, it has become clear that banks have been finding (and actually exploiting) more and more ways to perform regulatory arbitrage.14

12 See Bichsel and Blum (2004) for an empirical study of the relationship between risk and capital in banks.

13 On the opaqueness of banks, see, e.g., Morgan (2002).

The concern about regulatory arbitrage was the main motivation for revising Basel I. The new capital adequacy framework (Basel II) aims to reduce the scope of regulatory arbitrage by improving the measurement of risk and better aligning risk weights with the underlying risks. The success of the revision appears somewhat mixed. On the one hand, substantial improvements have been made. Under the so-called "standard approach", the increase in the number of risk classes, the use of external ratings, or the explicit taking into account of credit derivatives and asset securitisation undoubtedly reduce the scope for arbitrage opportunities at a reasonable cost in terms of complication. On the other hand, under the more sophisticated "internal ratings based (IRB) approach", the degree of complication and hence the administrative costs has increased substantially without ensuring that the risk sensitivity of the requirements really improves. This is particularly true of the advanced IRB approach, where capital requirements are based on parameters estimated by the banks themselves (probability of default, exposure at default and loss given default in the banking book). The extended reliance on banks' own judgment regarding the riskiness of their assets is a two-edged sword.

On the one hand, the reliance on the banks' own judgment to solve the risk-measurement problem is a priori attractive. Banks have an incentive to collect detailed information on their risk profile and are in a better position to do so than outsiders. Therefore, it seems sensible for the supervisor to tap this source of information, in particular for those components of a bank's portfolio for which no public information is available. On the other hand, however, this approach is problematic for two reasons. First, in most cases it is an illusion to believe that the risk measurement undertaken is really precise. This is due to the fact that the banks themselves often lack the data that would enable them to measure the riskiness of their assets in a satisfactory manner. For instance, data on long-term loans are typically inadequate in terms of volume and frequency for estimating sufficiently precise risk measures. Hence, banks may in good faith underestimate or overestimate the riskiness of their assets. Second, it is difficult for supervisory authorities to validate the information provided by the banks. Based on a given set of data, the supervisory authority will usually be unable to make the crucial distinction between the profile of a risky bank and the profile of a safe but unfortunate bank. This uncertainty can be systematically exploited by banks. Hence, with the IRB approach, Basel II gives banks a new opportunity for regulatory arbitrage which is potentially more harmful than the one it was supposed to address. We will return to this problem of so-called validation in greater detail in the next section.

To sum up, in theory, capital adequacy rules which take the riskiness of banks' activities into account are better suited to aligning banks' capital ratios to socially desirable levels than a simple GRL. In practice, however, the problems associated with the measurement of banks' riskiness limit the usefulness and applicability of this approach.


16 See Prescott (2004) for a formal treatment of this point.
3.3 Full-model approach

A full-model approach to capital regulation takes the logic of RWCR to its extreme. Under this approach, banks calculate their total risk exposure and their required capital based on a model that takes into account all correlations across positions in their entire portfolio. In the process, each position is implicitly given an individual risk weight according to its marginal contribution to the riskiness of the banks' whole portfolio. For regulatory purposes, instead of trying to design a complicated set of rules, supervisors can simply rely on the banks' own calculations. In this case, the main task of the supervisory authorities is to ensure the quality of the models. Specifically, they need to ensure that the models accurately reflect the actual exposures of banks and their underlying risks.

An advantage of this approach is its cost effectiveness. Independent of any regulation, banks determine their risk position and calculate their desired "economic capital" anyway. Furthermore, this approach aligns regulatory capital with economic capital. In contrast to the more or less arbitrary risk weights set by a regulator, the full-model approach is flexible enough to fully reflect the true risk of banks' assets. Hence, portfolio distortions due to wrong risk weights can be avoided.

While a full-model approach seems very attractive at first sight - which might explain the growing number of supporters of this approach - it suffers from three major drawbacks. First, there are problems associated with missing or inadequate data. In comparison with RWCR à la Basel II, where only certain parameters have to be estimated by banks, in the context of full models the problems are further aggravated by a virtual lack of data on extreme events and on their correlations. However, in order to determine the default probabilities of banks - and hence the adequate capital requirements - precisely this type of data is essential. The missing data can only be replaced by ad-hoc assumptions whose appropriateness can generally neither be proved ex ante nor be falsified ex post. As a consequence, banks' models and their risk estimates can deviate substantially from the correct ones. Some banks may systematically underestimate their risk exposure without being aware of it.17

Second, the described data constraints pose a formidable challenge for validation by supervisory authorities. Not only are supervisors confronted with the problems that exist in connection with the validation of single parameters (see Section 3.2), they also have to assess the appropriateness of whole models. This represents a difficult task since there is no uniquely correct model and no generally accepted method to measure risk. In addition to testing the banks' models based on the available data, supervisors also have to verify the quality of the data provided by the banks themselves. Given the opaqueness of banks' assets, this is only possible to a limited extent. Overall, it is difficult or even impossible for supervisors to identify a bad model. And it is even more difficult for them to prove that a model or a risk estimate is wrong. Hence, banks may be tempted to exploit this uncertainty about the true parameters and models. They may choose their models and assumptions about underlying parameters in order to minimise their required capital.

Third, there is a potential conflict with the regulatory motive. Capital regulation is based on the idea that banks hold too little capital on a voluntary basis (see Section 2). Therefore, at least in its pure form, where regulatory capital is set equal to the banks' economic capital, a full-model approach is useless. To take into account the fact that unregulated banks tend to hold too little capital, the regulatory minimum has to be set higher than whatever the banks consider to be their desired economic capital. However, an appropriate mark-up over a bank's economic capital would depend on, for instance, the bank's risk profile and the quality of its risk management. As argued in the previous paragraph, however, such an assessment would be very demanding, both on supervisory authorities and on banks. Even if feasible, the approach would lose much of the cost effectiveness that constitutes its attractiveness.

To summarise, a full-model approach to capital regulation seems attractive because it minimizes distortions due to inappropriate risk weights and because it is compatible with banks' internal capital allocation systems. However, this approach suffers from major drawbacks, most notably those associated with the estimation and validation of banks' models. Until these issues are satisfactorily resolved, a full-model approach does not represent a feasible option for the regulation of bank capital.

17 As Rebonato (2003, p. S11) has put it:
"The percentiles often quoted in the economic capital context (99.75, or 99.90) are virtually impossible to estimate in a statistically robust manner, and of dubious relevance for the purpose of strategic decision-making."
4 The future of capital regulation: Three guiding principles

We have argued that, when left on their own, banks have a tendency to hold too little capital and to choose leverage ratios that are too high relative to the welfare-maximising optimum. This distortion is due to banks’ limited liability, the presence of financial safety nets, and externalities in the event of bank failures. Hence, capital adequacy requirements are desirable from a collective point of view.

We have stressed that risk sensitivity is a desirable property of capital requirements, i.e., the amount of required capital should be a function of banks’ asset risks. However, risk-measurement issues together with information asymmetries restrict the degree of precision in risk measurement that can be achieved and, as a consequence, the precision of risk weights. Furthermore, we have claimed that the risk-measurement problems cannot be solved by fully delegating risk measurement to banks, i.e., by relying on a full-model approach. Based on these observations, we identify three principles on which, in our view, future developments of capital adequacy rules should be based.

**Principle 1: Optimal capital adequacy requirements are not necessarily perfectly risk sensitive.**

Capital requirements should be risk sensitive. However, the costs as well as the benefits associated with a higher level of precision in risk measurement - and hence in risk sensitivity - should be taken into account when designing capital adequacy rules. On the one hand, improvements to risk-weighting schemes generate decreasing marginal benefits, in terms of both reduced opportunities for regulatory arbitrage and a lower incentive to take advantage of these opportunities. On the other hand, the marginal costs - including design, implementation, compliance, and monitoring costs - related to an improvement of the risk-weighting scheme are positive. Hence, optimal capital adequacy rules reflect the underlying risks only imperfectly - even if perfect risk measurement were possible. For practical purposes this implies that instead of aiming for perfect (but excessively costly) risk weights, it is more important to base future refinements in capital rules on careful cost-benefit analyses.

**Principle 2: Risk assessment should only be delegated to banks to the extent that banks’ assessments can be accurately verified by supervisors.**

From a cost-efficiency perspective, it would be desirable to base capital requirements on banks’ own risk assessments. However, while banks generally have an incentive to measure their own risks accurately, they also have an incentive to hold less capital than socially optimal. This creates a conflict of interest for the banks: If they truthfully report their risks to the supervisory authorities, they have to hold more capital than they would if they were free to choose. Banks can only be prevented from understating their actual risks if supervisors are able to verify the banks’ reports and impose penalties in case of misrepresentation. As a consequence, capital requirements for a given asset or asset class should be based on the banks’ own assessment if, and only if, (i) banks are able to assess the underlying risks adequately, and (ii) the supervisors are able to verify with sufficient confidence whether banks are reporting an appropriate risk measure for the assets concerned. The assumption that banks always behave in a socially efficient manner (against their own interests) and therefore truthfully report their non-verifiable risks, would contradict the motive for capital regulation.

**Principle 3: Risk-sensitive capital requirements should be complemented by a capital floor which is independent of banks’ risks.**

As should be clear from Principles 1 and 2, capital adequacy rules will always be imperfect in the sense that they will always leave banks with some room for regulatory arbitrage. While such imperfections may generally be relatively harmless, they can have serious consequences whenever banks operate at very low levels of capital. First, due to the moral-hazard effects induced by limited liability, the higher a bank’s leverage, the higher the discrepancy between the bank’s privately optimal level of risk and the socially optimal level of risk. As a consequence, the incentive to take advantage of regulatory arbitrage opportunities and to incur excessive risks will be strongest at low levels of capital. Accordingly, the supervisory authority is most likely to underestimate the true
riskiness of banks precisely when their capital bases are low. Second, the consequences of underestimating the riskiness of banks are particularly damaging when the capital base is low, i.e., when the buffer that protects a bank against the consequences of losses is small. For these reasons, it is essential that optimal risk-sensitive capital requirements be complemented by a capital floor which does not depend on the riskiness of banks' activities.18 By setting a floor to banks' absolute (unweighted) capital ratio, a limit can be set to the consequences arising out of the shortcomings of a risk-weighted capital requirement scheme.19

18 Such a combination of risk-weighted capital requirements and a gearing-ratio limit is in effect in the US. A similar approach is planned in the insurance sector, where the introduction of a sophisticated risk-based capital requirements scheme should be complemented by a simple capital floor. For a discussion of the future solvency framework (“Solvency II”) in the insurance sector, see http://europa.eu.int/comm/internal_market/insurance/

19 For a detailed analysis of the combination of risk-weighted capital requirements and a gearing-ratio limit, see Bichsel and Blum (2001).
5 References


