

Stochastic Modelling – Boon or Bane for Insurance Industry Capital Regulation?

Oliver Bäte, Philipp von Plato and Günther Thallinger

McKinsey & Company, Inc., Magnusstraße 11, 50672 Köln, Germany.

E-mail: Guenther-Thallinger@mckinsey.com

Regulation of the financial industry should pursue three key objectives: consumer protection, market stability, and competitive efficiency. This article discusses core elements of a capital regime that could be used to develop regulation that meets these objectives while fostering an industry-wide enhancement of risk management. The authors argue that a pre-commitment approach can have considerable advantages over regulation based on (stochastic) risk models, as the latter can have adverse effects, especially on market stability and competitive efficiency, while consumer protection would have to be supplemented by additional requirements (such as scenario tests) in any case. Academic studies on capital regulation based on stochastic models have focused more on banking and less on insurance, while work by insurance practitioners has concentrated on the implications for management. The authors, therefore, wish to contribute to a more fundamental discussion of the design of capital and risk management regulation of the insurance industry.

The Geneva Papers (2006) 31, 57–82. doi:10.1057/palgrave.gpp.2510071

Keywords: capital management; regulation; risk management; stochastic modelling; value at risk; pre-commitment approach

Introduction

In recent years, insurance regulators at both national and supranational levels have initiated numerous efforts to reframe risk management and capital regulation regimes. These efforts have contributed strongly to the enhancement of risk taking and have introduced important new elements into the discussion of insurers' risk management organizations and processes. This is a welcome development as the industry shows an occasional tendency to undervalue the cost of risk taking (both consciously, e.g., P&C pricing cycles, and unconsciously, e.g., optionalities in life products), which may gradually lead to a depletion of capital and, thus, of risk-bearing capacity.

From this perspective, it appears that the industry needs to be prodded to price and manage risk appropriately. This need may well be served by regulation. Making use of risk quantification methods that are more sophisticated than factor models is a good step for insurers to take towards better risk management. In this sense, stochastic modelling as an element of insurers' risk assessment efforts may be a boon for the industry. However, within this larger context, proposals to base the quantification of regulatory capital requirements directly on stochastic models may result in adverse effects on the industry: the results could be a false sense of security on the part of policyholders, potential amplification of systemic risk, and polarization of the market's competitive structure – in addition to a huge implementation effort by the

industry and by the supervisory authorities themselves. Used in a regulatory regime, stochastic modelling may be a bane.

This downside and how it might be prevented are the subject of the present paper. We discuss the adverse effects of embedding stochastic modelling in capital regulation for the insurance industry and suggest a fundamentally different design approach to achieve regulatory objectives without triggering such effects. Our remarks are structured in three parts.

First, to establish common ground for our assessment, we summarize the current views of scholars regarding the fundamental objectives that risk management regulation in general should pursue, namely market stability, market efficiency, and consumer protection.

Second, we show how regulation based on stochastic capital models is more likely to defeat these objectives than meet them. The use of stochastic capital modelling for the purpose of capital regulation poses a number of problems that could hamper the fulfilment of the objectives of regulation. It is not our intention to evaluate stochastic models across the board. Our focus here is solely on their specific use for measuring tail risk and deriving regulatory capital requirements. As an expression of this distinction, we use the term “stochastic capital models” instead of “stochastic risk models”. The former could be understood as an application of the latter.

Third, we describe the main elements of a capital regime that we believe can contribute to the further development of risk management regulation approaches. Our basic recommendation is to consider a pre-commitment approach (PCA). While further analysis is needed, we argue that a PCA has advantages over the (prescriptive) use of stochastic capital models in the context of regulation and may offer a means to achieve the objectives of regulation, without diminishing the potential advantages of risk assessment performed internally by companies.

In form and content, recent reforms of risk and capital management regulation proposed and implemented for insurance have been influenced by Basel II. Most prominently, a number of them comprise three core building blocks, modules, or pillars. The first building block addresses financial requirements, that is, the quantitative requirements for determining regulatory capital. It may cover more than this – for example, to reflect a comprehensive “total balance sheet approach”, this first building block under Solvency II as currently proposed contains measurement rules for technical provisions, investments, and shareholder equity. In order to reflect this definition, insurance regulators increasingly promote the use of stochastic capital models as a target approach to determine regulatory capital.¹

The second building block comprises elements subject to qualitative assessment by regulatory authorities, including insurers’ organizations and processes. Finally, the third building block aims at fostering market self-regulation by defining companies’ reporting and disclosure duties.

Throughout the document, we focus on the first building block of regulation – financial requirements – and address the other two building blocks to the extent that they are affected. We want to stress that we are in general agreement with the three

¹ For a comprehensive overview of recent regulatory solvency initiatives cf. CEA (2005).

building blocks. In particular, we consider the second building block to be at the crux of future insurance supervision since it focuses on the core of corporate risk management – the organization’s capability to actively manage risks and avoid company failure.

Objectives of insurance regulation

The macroeconomic importance of the insurance industry for bearing and transforming risk makes the smooth functioning of the insurance market a matter of public and political interest and, hence, an object of regulation. Three widely accepted overriding objectives for the regulation of financial institutions and, hence, insurance companies are (1) consumer protection, (2) market stability, and (3) competitive efficiency. As a large body of literature exists on the derivation and justification of these three objectives, we only briefly recapitulate them here.²

Consumer protection

Protecting consumers from excessively high prices or opportunistic behaviour by insurers is often considered the primary objective of insurance regulation and typically receives more attention in the literature. Consumers face the problem of asymmetric information and are, thus, vulnerable to moral hazard or adverse selection if they choose an incompetent or opportunistic insurer.³ To protect the interests of consumers, the regulator, therefore, aims at reducing objective information asymmetries (e.g., by increasing transparency) and/or at preventing the resulting risks (e.g., by requiring a minimum level of solvency capital, evaluating product design, and monitoring reserving).

Market stability

As a regulatory objective, market stability means protecting the insurance and financial system from shocks that could unleash systemic risk, that is, a sudden, unexpected event that adversely affects a large part of the financial system.⁴ The drivers of systemic risk can range from fairly obvious factors, for example, investment management or underwriting cycles that generate insufficient premiums, to somewhat more unexpected factors, for example, changes in the legal environment. Furthermore, due to the long tails of certain contracts, the effect of systemic risk may initially be hidden (e.g., asbestos).⁵

Competitive efficiency

In addition to the safeguarding objectives above, it is generally accepted that the regulator should uphold and strengthen efficient competitive structures in the financial

² Cf. Merton (1989); Benston (1998); Llewellyn (1999).

³ Cf. Grace and Klein (1999).

⁴ Cf. Hartmann and De Brandt (2000).

⁵ Systemic risk in insurance differs in this regard from the classic financial industry example of the “bank run”.

system⁶ – particularly in the areas of public trust and the quality of information.⁷ Public trust is an intangible public good that increases capital flows to the financial markets and improves the effectiveness of resource allocation. However, since these positive externalities are not fully reflected in the earnings of the individual financial institution, companies willing to break the rules of conduct can increase their earnings as “free riders” who exploit the trust generated by those companies that do abide by the rules.⁸ The quality of the information generated on the financial markets influences competitive efficiency because it coordinates decentralised decisions throughout the entire economy. Higher-quality information increases the efficiency of price signals and, thus, the ensuing resource allocations.⁹ This effect holds true not only for the issuers of financial contracts, but also for the financial institutions themselves and their products. To increase trust and access to high-quality information, regulators have taken steps to define and monitor codes of ethics, foster internal controlling mechanisms, and introduce transparency and disclosure rules aimed at reinforcing market discipline and self-regulation on the financial markets. Finally, the regulator should also seek to promote efficiency in a broader sense by applying efficiency criteria to its own activity, for example, by reviewing new regulation in terms of its costs and benefits.¹⁰

Stochastic capital models for insurance regulation

The potential effects of dependencies and trade-offs between the objectives of consumer protection, market stability, and competitive efficiency have yet to be studied in detail. Nevertheless, discussing stochastic capital models as a basis for risk management and capital regulation and their suitability for meeting the objectives outlined can serve as a form of “back testing”. The following discussion of some drawbacks of prescribing these models is intended to unearth latent risks, which come to light only when the proposed methods are considered in their entirety as an integrated system.

We will begin by analysing the effects of stochastic capital models on market stability and then competitive efficiency, as these two regulatory objectives have so far been commented on, if at all, mainly or only as boundary conditions, before turning finally to consumer protection issues, where we delve into some technical details of stochastic modelling.

⁶ Cf. Herring and Santomero (2000).

⁷ With public trust, financial institutions operate more fairly, more transparently, and more in the customer’s interest.

⁸ Cf. Freixas and Santomero (2003).

⁹ The prices set on the capital and insurance markets are used by private households and companies to decide how to divide their income between consumption and savings, which investment projects to select, and how to finance them.

¹⁰ Such procedures have been widely established, for example, by the Financial Services Authority in the U.K. and by regulatory authorities in the U.S. In the European Union, the regulator has recently realized the need for improved cost-effectiveness assessments and issued a white paper calling upon the EU legislator to assess the potential economic impact and costs and benefits of new regulations. Nebel gives a good overview of the impact assessment practices applied by regulatory authorities, see Nebel (2004). Cf. FSA (1999); Grace and Klein (1999); OMB (2002); FAME (2003).

Market stability: harmonization of risk-sensitive capital regulation may increase systemic risk

Opinion is divided on the relevance of systemic risk in the insurance industry, but the empirical results of recent research indicate that its relevance is increasing.¹¹ Ensuring market stability primarily implies the prevention or reduction of conditions that could trigger systemic risk. In recent years, the combined pressure of insurers' increased capital market exposure, the formation of complex financial conglomerates, and increasing consolidation, which increase dependencies and the concentration of risks,¹² have created a situation that warrants even more supervisory attention to ensuring insurance market stability and forestalling systemic risk.

In its buffer function, regulatory capital protects the insurer against losses in collapsing markets, reduces the insurer's risk of ruin, and thus guards the market system as a whole against further destabilization. Opinions differ, however, regarding the form that regulatory capital requirements should take so that capital performs its buffer functions as efficiently as possible. Most insurance capital regulations currently underway aim at introducing risk-based capital requirements. The basic idea is that the more sensitive capital requirements are to an insurer's risk exposure, the more tightly this links capital in its buffer function with the potential for loss and the more rigorously moral hazard conduct can be discouraged, since excessive risk taking is only possible in exchange for a higher capital stake on the part of the company owners.

The development in the direction of a risk-sensitive capital regulation has a number of obvious advantages, one of which is the enhancement of risk taking at individual company level.¹³ We fear, however, that contrary to its intention this approach can increase risk at the systemic level. To explain how this danger arises, we distinguish three sequential components in the genesis and development of systemic risk: the trigger event or shock, the propagation mechanism, and the impact or effect.¹⁴

In any case, even if systemic risk is perceived as a minor danger by a regulator, it might be appropriate for the regulator itself to define contingency plans along the three components (in order to be able to respond quickly and effectively if a systemic failure appears to develop). Also, as Goodhart suggests, a plan should be in place for crisis management, including the restructuring of (parts of) the industry.¹⁵

Risk-sensitive capital regulation as propagation mechanism

The key characteristic of systemic risk is the propagation mechanism, which transmits the initial shock from one financial institution to the next and, potentially, to the financial system as a whole, and which ultimately determines the impact of the risk.

¹¹ For a more in-depth discussion of the relevance of systemic risk in the insurance industry, cf. Group of Thirty (1997), Wilmarth (2002); Herring and Schuermann (2003); Minderhoud (2003a, b); Harrington (2005).

¹² Cf. Morrison (2002) and Wilmarth (2002).

¹³ Cf. Jorion (2001); Cruz (2002); Szegö (2004).

¹⁴ Cf. Hartmann and De Brandt (2000).

¹⁵ Cf. Goodhart (1998).

The degree of contagion risk, that is, the extent to which the trigger event can spread system-wide, depends on the risk perception of the individual market participants. In general, insurers assume risk as an exogenous variable and the stochastic risk models they use are fundamentally based on this assumption.¹⁶

For many of the risks to which insurers are exposed, however, the perception of risk as an exogenous phenomenon is wrong. Most such risks are formed via market mechanisms, that is, through the interplay of the conduct of individual market participants. Risk in this form is by definition endogenous¹⁷ and can be seen, for instance, in the persistence of underwriting cycles. The asset-side risks are relentlessly exposed to capital market pricing mechanisms. The nature of the risk is, thus, not static as assumed by current risk measurement techniques, but the result of a dynamic system of interlinked individual decisions.¹⁸

Endogenous risk is a critical factor in a crisis when market participants all adopt the same risk preferences and take the same actions. Then, instead of cancelling one another out, the actions of the individual players derived from their largely uniform models reinforce one another. With risk-sensitive capital regulation, this “herd” or procyclic behaviour becomes critically relevant for the insurance industry. Through risk-sensitive capital regulation, the risk preferences and actions of the individual market participants are effectively homogenized by the regulator.

Thus, regulation-induced convergence of risk preferences itself might function as a propagation mechanism for systemic risk by transforming exogenous risks into endogenous ones. In technical terms, this effect violates a central assumption of all stochastic capital models, the stationarity of risk (independently and identically distributed (i.i.d.) assumption, i.e., the assumption that risk is independently and identically distributed). The modelling of market feedback effects calls for game theory approaches, but these are still only in their infancy in the form of a few initial studies.¹⁹ It is, therefore, doubtful that stochastic capital models will be able to reflect these dynamics under stress conditions any time soon. Risk-sensitive capital regulation, thus, might fail precisely at the point at which its protective buffer function is most needed.

Regime shifts in capital regulation as trigger of system shocks

In the context of regulatory design for market stability, two forms of instability can be distinguished depending on the type of trigger event: institutional and market. Institutional instability occurs when “the failure of one or a few financial institutions spreads and causes more widespread economic damage”. Market instability, on the other hand, is characterized by a sharp contraction of liquidity or by price volatility in the market as a whole and is defined “in terms of the wider impact that volatility in

¹⁶ Cf. Morris and Shin (1999).

¹⁷ Cf. Danielsson *et al.* (2001); Danielsson and Song Shin (2000).

¹⁸ Danielsson notes an analogy with gaming: “Current risk management practices rest on the roulette view of uncertainty... However, when the outcome depends on the actions of others..., risk modelling resembles poker more than roulette”. Cf. Danielsson and Song Shin (2000).

¹⁹ Cf. Morris and Shin (1999) and Persaud (2000).

prices and flows can have on the economy”.²⁰ The two types of instability are equally relevant for the regulator’s objective of ensuring market stability.

Risk-based capital requirements can trigger such shocks by influencing price formation and the liquidity situation in capital and insurance markets. A regime shift in the direction of risk-sensitive capital regulation represents a one-time preference shift of the market participants in that the regulatory assessment of their risk exposure changes. While such a shift might be desirable (as it corrects for the assessment of risk so far undetected or unquantified), it can be especially critical for the objective of market stability when the shift to the new regulatory regime takes place under tense market conditions. This one-time effect can be cushioned by granting financial institutions a relatively long transition period.

The effects of a shift in a regulatory regime can be illustrated by the experiences of some EU Member States that have already converted their capital regulation to a more risk-sensitive system. Denmark, for example, was one of the first European countries to introduce a system of risk-based solvency tests. Starting in 2000, pension funds and life insurers were required to prove their ability to meet underwriting claims under various stress scenarios. Because of the long duration on the liability side and comparatively short-term nature of the assets, the new regulation surfaced significant asset-liability mismatch risk. Many pension funds and life insurers were forced to extend asset durations through long-end buying. In late 1999, the demand for long-term bonds and hedging instruments surged dramatically. Given the limited liquidity of the Danish bond market, the movements rapidly extended to the European market. The pressure on the long end of the interest curve increased and, within a few months, the interest rate curve flattened and long-term bond yields dropped precipitously. The falling interest rate fuelled the asset-liability mismatch risk so that more insurers were forced to act.

Amplification of impact due to harmonized capital requirements

The effect of a systemic risk occurrence can be classified on a continuum from a single impact to multiple impacts. A single impact occurs when the trigger event affects only one or a few financial institutions or an individual market (systemic risk in a broad sense). A multiple-impact shock occurs when a number of financial institutions or markets are adversely affected at the same time (systemic risk in the strict sense). Multiple-impact systemic risk is obviously the more severe form and more important for the regulatory objective of ensuring a stable and functioning financial system.

In light of how systemic risk unfolds with risk-sensitive capital regulation, the regulator’s effort to bring about greater harmonization of capital requirements across borders and sectors, in itself laudable, can have dangerous side effects. The greater the extent of harmonization, the more institutions and markets may become victims of a systemic shock.

²⁰ Cf. Crockett (1997).

This obviously poses a regulatory trade-off: the harmonization of capital requirements is necessary to prevent regulatory arbitrage on the part of the protagonists across sectors and markets and to establish a level playing field. However, if one pursues this objective by means of risk-sensitive capital regulation (measured with stochastic capital models), harmonization may entail an increased danger of systemic risk. “While the international harmonization of prudential standards has the benefit of creating a level playing field there could well be a destabilizing effect if the same rule were to apply everywhere, inducing the same pattern of market conduct”.²¹

Competitive efficiency: have-nots may face discrimination

When regulatory requirements reach a certain level of complexity – in terms of processes, managerial capacity, or expertise – they may trigger a polarization between those that can afford to set up the (implicitly) required “machinery” and those that cannot. This may hold true for companies as well as national regulators. It is beyond the scope of this text to enumerate the operational and financial requirements associated with capital regulation based on stochastic modelling. Apparently, regulators have already taken this obstacle into consideration as can be seen from proposals for the provision of standard models (for those who do not have internal models).

Since the co-existence of proprietary and standard models is of special interest in the present context, we analyze the possible effects briefly using a simplified and abbreviated example from game theory. Let us assume that the market consists of only two insurers: a large, sophisticated insurer with deep pockets and a small, less sophisticated insurer with scarce resources. Both companies and their risk management systems are supervised by a resource-constrained regulator that offers them two alternative models for determining their capital requirements: one is a sophisticated, complex model, which is resource-intensive in implementation and maintenance but provides high-quality support of risk and capital management and, thus, helps a company gain competitive advantages and market share, for example, via reduced capital requirements. The other model is a less sophisticated, off-the-shelf model that requires only limited resources on the part of the insurer. In this constellation, the two insurers’ decision space and the ensuing competitive dynamics are foreseeable. The large insurer with plenty of staff and financial resources decides in favour of a sophisticated model in order to expand its market share. Since the small insurer lacks the resources to follow suit, its dominant strategy to avoid competitive disadvantage and to achieve additional earnings will be to engage in speculation and underwrite excessive risks.

In the example, the effect of requiring the choice of one of two disparate models leads to an inefficient market equilibrium, resulting in significant competitive distortions diametrically opposed to the original intention of promoting efficiency.

²¹ Nebel highlighted the potential adverse side effects of increased regulatory harmonization. Cf. Nebel (2004).

How relevant the game theory perspective is in reality depends on the prevailing market and supervisory structures and, in particular, on the regulator's resourcefulness in ensuring intense (qualitative) oversight. In the example, we assume there is only very limited supervisory oversight that might mitigate the gaming incentive mentioned above. In all other respects, a certain similarity of the basic constellation and assumptions in the two-company example to the approach suggested by insurance regulators – allowing insurers either to apply a predefined standard model or to opt for using internal models – is intentional. It is well known that such a “gaming environment” can be avoided with regulation that is principle-based (e.g., stipulating what a company should measure, not how it should measure it) and ensures the right incentives.²²

In any case, it might be valuable to assess the cost–benefit ratio of a new regulatory regime from the individual company and the market perspective. Such an assessment should also take the consumer perspective into account as, in the end, consumers will have to finance implementation of new regulation and any resulting changes to capital requirements.

Consumer protection: when used for regulation, stochastic capital models may create a false sense of security

In financial theory, risk and capital have always been an inseparable pair. The question of how to map their relationship in practice with enough accuracy to steer capital management was long a largely unresolved issue for the financial industry. In recent years, however, an apparent solution to the problem for both sides arrived on the scene in the form of economic capital (usually calculated with stochastic capital models). While we focus here on the use of these models for capital regulation, many of the following points need to be taken into account at the corporate level.

The introduction of stochastic capital models for regulatory purposes also sparked the idea of converging capital notions – especially economic capital (in its original definition, reflecting the shareholder view) and regulatory capital (needed to protect policyholders).

With a view to policyholder protection, this development raises two questions: first, how useful is the convergence of regulatory and economic capital? Second, how capable are internal risk models of reliably guiding an insurer to setting an amount of capital sufficient to protect the insured? The first question is conceptual, while the second requires a technical analysis of the relevant models.

Who is served by the convergence of regulatory and economic capital?

The usefulness of the convergence of regulatory and economic capital perspectives needs to be assessed in light of the stakeholders who advocate it (“consider the

²² Cf. Goodhart (1998).

source”) and the function that each type of capital is intended to serve (“follow the money trail”).

Regulatory capital is designed to protect consumers from disadvantages resulting from their limited information compared with insurance company shareholders or management. The regulator’s capital requirements are essentially a way of pricing risk high enough to create an incentive for shareholders and management to shun opportunistic behaviour. Whether these prices are sufficiently accurate to satisfy supervisory capital requirements and protect consumers depends entirely on the underlying risk measurement approach.

Economic capital serves to secure the status of the insurance company as a going concern and reflects the shareholders’ interests. In the event of insolvency, shareholders do not receive any money until all other claimants, primarily the insured, have been satisfied. Nevertheless, because shareholders do have claims to residual profits, and their profit rises disproportionately as risk increases, they are normally interested in taking on more risk than the insured are. From the shareholder perspective, it is not profit-maximizing to allocate capital to cover the risk of extreme losses at the periphery of the loss distribution (“tail losses”).

Thus, when setting regulatory capital requirements to protect consumers, the regulator must focus primarily on these peripheral areas of loss distribution. Bringing about the convergence of regulatory and economic capital (the latter reflecting the shareholder view) would be counterproductive for a regulator trying to protect consumers. In practice, of course, nearly all insurance companies hold additional capital (a capital buffer) in order to safeguard regulatory capital.

How reliably do available models quantify regulatory capital for consumer protection?

Model risk in general can be defined as “the risk that a financial institution incurs losses because its risk management models are misspecified or because some of the assumptions underlying these models are not met in practice”.²³ As argued previously, in the present context, the discussion of model risk primarily needs to concentrate on the suitability and accuracy of stochastic capital models for measuring tail risk at the periphery of the loss distribution. Obviously, the higher the model risk, the less consumer protection it provides.

We confine this discussion to the most widely used models, specifically to non-parametric models, in particular, historical simulation and the (conditional) parametric models, that is, variance–covariance approaches. Recent developments, such as the semi-parametric Extreme Value Theory, have so far not been used on a broad basis; they suffer in part from the same issues that limit the reliability of non-parametric and parametric models.

While the sources of model risk, that is, the deficiencies of these models, are often closely intertwined in practice, they can be separated and classified in three broad dimensions: input, model assumptions, and output.

²³ Cf. McNeil *et al.* (2005).

(1) Model risk with respect to model input

Most forms of stochastic risk modelling are based on historical input data, regardless of the form in which they estimate the distribution moments.²⁴ The type and availability of historical input data depend primarily on the calibration of the risk measure with respect to two parameters: the time horizon and the confidence level of the risk measurement. In insurance regulation, these are frequently stipulated as a time horizon of 1 year and a confidence level higher than 99 per cent. Such considerations may be conceptually justified, but against the background of the specific risk structure and data situation in the insurance business compared with securities trading, we have to wonder to what extent a calibration to this order of magnitude can be applied in practice.

At present, two main approaches are commonly used to calculate a 1-year Value at Risk (VaR) or Expected Shortfall (ES): a direct approach (without using a scaling law) and an indirect approach using a scaling law.²⁵ The direct approach takes the time horizon stipulated for the risk measurement and uses input data consistent with the time horizon as the basis for quantifying the percentile measure either directly (non-parametric method) or uses it indirectly to estimate the parameters of an assumed distribution function (parametric method). The interval for the historical observation points must correspond to the time horizon of the risk measurement.

The statistical accuracy of the risk estimate depends on the available historical observation points, that is, on the sample period. The sample period defines the observation period for which scaled historical input data – corresponding to the time horizon – are needed. In order to achieve the same statistical significance, the length of the sample period, and thus the data requirements, must increase in proportion with increases in the time horizon of the risk measurement.

In the 1-year horizon, this requirement is practically impossible for an insurer to satisfy, as the following example of the direct approach illustrates. Let us assume the starting point is a time horizon of one day in which the 1-day VaR (or ES) is to be calculated with a confidence level of 99 per cent with a sample period of one year (250 days or observation points). If we now increase the time horizon to 10 days, we also need 10 years of historical data (with a 10-day interval) to be able to calculate the 10-day VaR with the same statistical accuracy. As even this small case requires a very extensive historical record of 2,500 days, it is sometimes recommended to calculate the 10-day VaR using a 1-year sample of 250 days. However, the 10-day interval provides only 25 observations for the calculation of an event that, by definition, should occur only once in one hundred observations! Thus, proposing to measure risk for a 1-year horizon is effectively a *reductio ad absurdum* argument against the direct approach.²⁶

²⁴ A noteworthy exception is semi-parametric methods (e.g., Extreme Value Theory), but to estimate the tail index, they, too, refer to historical data and, thus, suffer from the same limitations.

²⁵ Danielsson is a leading analyst of the calculation of a 10-day VaR for banks under Basel II and has already demonstrated the problems of practical implementation for this much shorter time frame. Cf. Danielsson (2002).

²⁶ It is out of the scope of this article to discuss the time dimension of risk measurement in further detail. We emphasize, however, that especially in the context of long-term insurance business, risk should be considered throughout (economic/pricing) cycles; a rigid 1-year horizon could be inappropriate.

The data problem initially appears less troublesome for parametric models. They are less dependent on the data situation in the marginal distribution simply because they approximate this data using the relevant assumed distribution function. That said, because of the limited data available, a 1-year horizon also creates problems when estimating the variance, especially as soon as the covariances have to be estimated when extending to the multivariate distribution. Yet estimating correlations is decisive for risk and capital measurement, as they determine the extent of diversification effects that discount risk and, thus, capital requirements. Even apart from the issue of linear dependence, that is, correlation between the risk variables, there is the problem that the data requirements or the correlation coefficients to be estimated within the covariance matrix increase extremely fast when the number of risk factors to be observed increases. It often happens that no historical data at all are available for the covariance of multiple risk variables. In practice, such additions commonly take the form of manual “expert estimates”. Using VaR, Ju and Pearson show that risk measurement under these circumstances tends to significantly underestimate risk and that the actual VaR can be twice as large as the estimated VaR.²⁷

To overcome the problem of insufficient data for longer time horizons and higher confidence levels, practitioners using parametric models often take an indirect approach that makes use of the Square-Root-of-Time Rule.²⁸ However, this leads to significant estimation errors because the assumptions associated with this rule are false. Specifically, the scaling law requires that the risk variables observed over time are independently and identically distributed (the “i.i.d. assumption”).²⁹ In order to allow scaling over various time horizons and confidence levels, the risk variables must additionally have a normal distribution. These elementary assumptions are regularly violated in the insurance context, as discussed below. Furthermore, Danielsson and Zigrand show that “the Square-Root-of-Time Rule leads to a systematic underestimation of risk and can do so by a very substantial margin”.³⁰

Finally, a further problem is deciding which sample periods of historical data are best suited to predict the future risk profile. As a detailed discussion of this aspect would go beyond the scope of this article, we refer to Jorion: “VaR is highly sensitive to the specific period of time used to compute it, regardless of whether a historical simulation method or ‘delta-normal’ method is used.”³¹

²⁷ Cf. Ju and Pearson (1998). See also Krause (2002).

²⁸ Under assumptions specified in the parametric method, the scaling law enables high-frequency data to be translated into a lower frequency by means of multiplication with the Square-Root-of-Time horizon. The VaR or ES for a given time horizon t and a given risk probability α can thus be scaled up to a longer time interval T with $\text{VaR}_\alpha^T = \sqrt{T} \text{VaR}_\alpha^t$ or $\text{ES}_\alpha^T = \sqrt{T} \text{ES}_\alpha^t$.

²⁹ The abbreviation i.i.d. stands for independently, identically distributed. For a more detailed discussion of time scaling of risk (measures), see Diebold *et al.* (1997); Christoffersen *et al.* (1998); Danielsson and Zigrand (2004) and Los (2004).

³⁰ Danielsson and Zigrand (2004).

³¹ The “delta-normal” method is a parametric approach and another name for the variance-covariance approach. Cf. Jorion (1996).

(2) *Model risk with respect to model assumptions*

The assumptions underpinning stochastic risk models refer to the explicitly or implicitly assumed behaviour of risk variables in one- and multi-dimensional space. The role of assumptions is most pronounced in parametric methods as their functionality and practicability rests on the assumption of a certain distribution of the risk variables. While non-parametric methods do not make explicit distribution assumptions, thereby excluding a number of the problems discussed below, they do assume that the risk variables observed are constantly distributed in the time horizon of the risk measurement.

While attempts have been made to design methods not reliant on the assumptions discussed below, current practice is clearly dominated by approaches based on these assumptions, which do not adequately reflect reality and, thus, can hide enormous estimation risks.

Normal distribution actually doesn't apply Normal distribution is most commonly assumed for parametric methods of capital measurement. The practical explanation is that the normal distribution facilitates the aggregation of risks over time and across the total portfolio. With the normal distribution, one “merely” has to estimate variances and covariances to determine the risk of the entire portfolio. However, there are two troublesome issues here: (1) if the risk variables are not i.i.d., they cannot be approximated using a normal distribution and (2) since insurance risks – both asset and underwriting risks – are known for their fat tails, assuming a normal distribution can result in a significant underestimation of these events.³²

Concerning the i.i.d. assumption, it is known that for capital market risk the returns of all investment classes do not meet the definition of identically distributed random variables. Instead, their distribution profile is erratic. This is also the case for the large losses due to cumulative and trend risks that are of particular relevance in the insurance context (catastrophic risk, epidemics, longevity, etc.). That the loss distributions in these lines are unstable becomes clear when one considers the distribution of catastrophic losses.³³ The volatility of (inflation-adjusted) losses has clearly increased over the past 30 years.

As for the underestimation of tail risks, the magnitude of the underestimation of risk as a result of assuming a normal distribution can be illustrated by the daily gains and losses of the S&P 500 Index from 1929 to 2003 and their standard deviation. Assuming a normal distribution, a 5 sigma gain or loss should occur every 3×10^6 days, that is, once in 10,000 years. Yet the 30-year history of S&P contains five such sigma events.

These observations indicate that capital models based on a normal distribution cannot reliably map empirically observable extreme events and, thus, can lead to a severe underestimation of the risks in the tails of the distribution.³⁴ As capital

³² Cf. Danielsson and De Vries (2000). Engle and Manganelli discuss these issues for more advanced stochastic processes of the GARCH family and state that “the general finding is that these approaches (both normal GARCH and Risk Metrics) tend to underestimate the Value at Risk, because the normality assumption of standardised residuals seems not to be consistent with the behaviour of financial reforms”, Cf. Engle and Manganelli (2001).

³³ Cf. Swiss Re (2002); Mills (2005).

³⁴ Cf. Danielsson and De Vries (2000).

regulation in insurance is dedicated to protecting against tail risk, the models discussed can fail precisely at the point where their suitability for risk and capital measurement is most sorely needed.

Theoretically, it is possible to use parametric models for univariate risk measurement without assuming normal distribution, that is, by assuming more differentiated distributions (Lévy-distribution, Log-Laplace distribution, etc.). The disadvantage of using such distributions, however, lies in losing the relatively simple aggregation of risk over various positions and time. At the next level, it then becomes practically impossible to estimate a multivariate distribution, and the variance of an entire portfolio can no longer be derived linearly from the variance–covariance matrix.³⁵ Yet risk aggregation over multidimensional distributions is one of the core functions for which stochastic capital models are commonly used.

In light of the problems associated with parametric approaches, non-parametric methods may appear favourable as they do not explicitly require any distribution assumption. But non-parametric models are not free of problematic assumptions either, as they implicitly suppose that risk variables will behave in the future as they did in the past. As noted in Pritsker: “The principal disadvantage of the historical simulation method is that it computes the empirical cumulative density function ... by assigning an equal probability weight ... to each return. This is equivalent to assuming that the risk factors, and hence the historically simulated returns, are independently and identically distributed (i.i.d.) through time”.³⁶

Assuming Gaussian dependence leads to mis-estimates of diversification To aggregate risk, diversification effects between risk positions are mapped. Since diversification decreases capital requirements, this topic is under intense discussion.³⁷

When aggregating risks across multivariate distributions, the normality assumption automatically implies Gaussian dependence structures, i.e., linear correlations, between the risks. One problem with correlations is their variance over time, but this is likely the more minor problem for a regulator. The larger problem occurs when the volatilities of the risk variables and their correlations positively reinforce one another. Such a relationship has been demonstrated in various studies for capital markets and capital market instruments.³⁸

Hence, while the positive relationship between volatility and correlation leads to overestimation of the diversification effects under good market conditions, the supposed diversification effects shrink under stress – precisely when they are most urgently needed. The assumption of linear correlations thus harbours the danger of a massive underestimation of a company’s aggregate risk.

³⁵ “The disadvantages of non-normal innovations for the VaR exercise are several, for example, multivariate versions of such models are typically hard to estimate and recursive forecasts of multi-step ahead VaR levels are difficult to compute” Danielsson and De Vries (2000).

³⁶ Cf. Pritsker (2001).

³⁷ Cf. CRO-Forum (2005).

³⁸ See Karolyi and Stulz (1996); Longin and Solnik (1998); Ang and Bekaert (1999); Chow *et al.* (1999); Brooks and Persaud (2000).

The unsatisfactory results of the normal distribution assumption for modelling dependences suggest that one should turn to other, more sophisticated dependence structures with higher tail correlations such as, for example, copulas. In theory, these approaches indeed allow for a more differentiated modelling of the tail dependences between risk variables. In practice, however, their parameterization poses significant challenges as empirically, if at all, only very few observation points are available. Owing to the lack of data, most parameters need to be estimated based on individual judgement and therefore, the result of risk modelling becomes highly sensitive to the modeller's assumptions on dependence behaviour and can hardly be validated empirically.

(3) Model risk with respect to model output

A number of further significant problems occur in the process of risk measurement itself and involve the technical features of the risk measure selected. In the context of regulation, the two most frequently discussed measures are VaR and ES.

VaR – not truly appropriate for measuring tail risks VaR is applicable to a large number of (elliptical) distributions and easy to use; it is also easy to communicate internally and externally. The technical criticisms of VaR centre on its non-coherence.³⁹ Colloquially, one might say coherence ensures that the greater the risk, the greater the risk measure. A central characteristic is subadditivity, which does not hold for VaR. The lack of subadditivity of VaR means that, for two risk positions A and B , it is possible that $\text{VaR}_{A+B} > \text{VaR}_A + \text{VaR}_B$. This violates the diversification principle. Dowd concludes that “if we accept the need for subadditivity, then we must immediately reject the VaR as a risk measure”.⁴⁰ The problem of VaR's missing subadditivity is irrelevant provided the focus is on elliptical distributions and – as their special case – normal distributions (in these cases, VaR is subadditive).⁴¹ The problem is, however, that most risks relevant to the insurance business are not elliptically distributed and have anything but normally distributed tails. Applying non-subadditive VaR to this kind of spiked distributions may not only misstate the quantum of risk and capital required but may, moreover, render their reasonable allocation across lines of business impossible.

A further problem for regulators is that VaR, as a point measure, does not take account of the distribution in the tails. Hence, it is possible for an insurer to legitimately manipulate VaR, for example, by using options to shift risks out of the quantile relevant for capital measurement and into the tail.⁴² This effectively increases the risk while reducing the capital requirements.

Expected shortfall – theoretically superior, but hard to implement ES is the conditional loss expectation in a defined time horizon, given that the loss exceeds a defined threshold value. In practice, the threshold is mostly defined using VaR for a given

³⁹ Cf. Artzner *et al.* (1999). See also Kondor *et al.* (2004).

⁴⁰ Cf. Dowd (2004).

⁴¹ Cf. Embrechts *et al.* (1999).

⁴² Among others, Ahn and Boudoukh (1999) and Krause (2002) show how VaR can be manipulated using options.

confidence level. ES provides a perspective on “How bad is bad?”⁴³ As the ES is by definition and *ceteris paribus* (i.e., for the same confidence level) greater than VaR, risk measurement and capital requirements determined with ES are more conservative than those determined with VaR.

Compared with VaR, ES offers the regulator an alternative that is at least theoretically superior. ES is a coherent risk measure.⁴⁴ As an interval measure, ES provides information about how high the average expected loss will be above VaR and, due to its coherence, is technically suitable for quantifying tails that are not normally distributed. However, these differences compared with VaR do not lead *per se* to efficient risk management and capital deployment from an economic perspective, as can be illustrated by the following example.⁴⁵ The example is an insurance portfolio P consisting of two individual risks X and Y that are independent of each other. The expected loss distribution for the two risks in millions of EUR is:

X	1	3	100	Y	1	5
Probability	0.90	0.09	0.01	Probability	0.20	0.80

The distribution of the total risk of $P=X+Y$ given their independence is then:

$X+Y$	2	4	6	8	101	105
Probability	0.18	0.018	0.72	0.072	0.002	0.008
Cumulative probability	0.18	0.198	0.918	0.990	0.992	1

For a default probability of $\alpha=1$ per cent we then get $VaR_x(P)=\text{EUR } 8$ million and $ES_x(P)=\text{EUR } 104.2$ million.⁴⁶ The capital requirement using ES is, thus, over 13 times higher than with VaR. This result is obviously attributable to the extreme values in the 1 per cent tail of the distribution. The example may overestimate the discreteness in the marginal distribution, but it makes clear that the choice of risk measure for risk variables with spiked distributions can have a dramatic impact on risk and capital measurement. Alongside this insight, the example also illustrates a second important point. If one asks to what extent the ES method shown here protects the company against insolvency, the following picture emerges: the secured recurrence period T is extended by ES in this example merely from 100 to 125 years.⁴⁷ That means that ES effectively increases the safeguard against insolvency by only 25 per cent for over 13 times the capital required when using VaR. While this is an extreme example, it shows

⁴³ Artzner *et al.* (1999).

⁴⁴ *Ibid.*

⁴⁵ Cf. Pfeifer (2004).

⁴⁶ $ES_x(P)=(101 \times 0.002 + 105 \times 0.008)/0.01=104.2$.

⁴⁷ The secured recurrence period is the reciprocal of the loss probability $T=1/\alpha$. For VaR at a confidence level of 99 per cent, it by definition equals 100 years. In the example, the ES of EUR 104.2 million secures 99.2 per cent of the worst cases. This results in an effective probability of ruin of 0.8 per cent and a secured recurrence period of 125 years.

that ES may tend to lead to economically doubtful and inefficient risk and capital figures.

Potentially, the most severe problem with using ES is that it is difficult to implement in practice: it results in significant estimation errors when applied to the data typically available to insurers. Yamai and Yoshiba analyse the estimation accuracy of ES in comparison with VaR and show that ES requires a much larger data sample than VaR does in order to achieve the same level of statistical accuracy in risk measurement.⁴⁸ Their study also shows very clearly that, for a constant sample size, the greater the distribution's deviation from normal distribution, i.e., the fatter the tails, the lower the estimation accuracy. Again, the problem becomes more severe when the required confidence level is increased.

We have examined the suitability of stochastic risk models to assess insurers' tail risks, which should be at the centre of regulatory capital regulation dedicated to consumer protection. As we have seen, model risk exists in multiple dimensions and, in sum, can lead to unreliable output in the form of severely underestimated risk and thus to a false sense of security.

Several regulators now require insurers to apply stress tests to check their solvency directly under extreme events. Where stress tests are required, they are generally required on top of risk modelling. This combination is very much in line with the current academic assessment of the reliability of stochastic modelling to assess tail risk, as Jorion states: "While VAR [stochastic modelling] focuses on the dispersion of revenues, stress testing instead examines the tails. Stress testing is an essential component of a risk management system because it can help to ensure the survival of an institution in times of turmoil".⁴⁹

In other words, to protect consumers, a supervisor needs to apply scenario stress testing in any case. The question then is whether a regulator needs to require or prescribe risk modelling at all or to what extent.

Risk management regulation that meets the key objectives

In this section, we present the essential elements of alternative regulation of capital in the insurance industry in order to stimulate discussion about the fundamental design of capital regulation. We believe it is obvious that the regulatory objectives of consumer protection, market stability, and competitive efficiency are not independent of one another. For example, after a certain point, the attempt to protect consumers by requiring insurers to hold ever more capital reserves will have a negative effect on competitive efficiency in the insurance market. It follows that one task of the regulator is to optimize the fulfilment of the objectives holistically as an inter-related group. This, in turn, implies prior discussion and prioritization of the objectives to create the basis for holistic agreement on, and coordination of, the forms of regulation to be applied.

⁴⁸ Cf. Yamai and Yoshiba (2002).

⁴⁹ Jorion (2001).

In the following, we operate on the assumption that ensuring company solvency and, thus, consumer protection has top priority, and that the attainment of market stability and competitive efficiency should also be supported or in any case not prevented. Consequently, our focus is on the first regulatory building block, as the design of financial requirements and, in particular, capital regulation is of central importance for this conception of the objectives.

For the design of insurance capital regulation, we propose adapting the PCA originally developed and analysed for banking in the mid- to late 1990s.⁵⁰ In its original form, PCA met with a mixed response from the various market participants. Because the approach depends on the effectiveness of “own commitments” by the companies under supervision, the initial debate centred on the question of how much faith one can have in the efficiency and self-regulation powers of the market. While market advocates welcomed the high degree of freedom offered by the PCA approach, critics were concerned about the potential for discretionary conduct with adverse effects for consumers and market stability.

Based on this discussion and attempting to tailor it for the insurance industry, we have developed the original PCA approach further. To express this differentiation, we use the term “principle-based PCA”. In the following, we briefly describe the key design elements of such an approach as applied to the insurance industry and comment on its ability to fulfil the main objectives of regulation.

The regulatory continuum

To compare the PCA with the design of stochastically based capital regulation, it is helpful to consider the possibilities for implementing regulation as occupying spaces along a continuum. At one end of the continuum is leaving the individual design of risk management to market forces. This approach is not only the simplest, but presumably also superior to flawed regulation. When the cost of regulatory intervention is likely to be greater than the benefit, the logical consequence is to decide in favour of “no regulation”. The other end of the continuum is what might be termed a “hard link” or “stringent” approach, that is, an exogenously defined relationship between risk exposure on the one hand and capital requirements on the other. By definition, such a hard link is very prescriptive since the regulator sets the parameters for all companies’ calculation of their capital requirements. For example, Solvency II in its currently proposed form is moving closer to this end of the continuum.⁵¹

Elements of a principle-based PCA

The basic idea of the PCA is that an insurer pre-commits that it will enforce internal controls, that its losses will not exceed certain thresholds, and that its staff will adhere to conduct-of-business rules.⁵² Building on suggestions by Kupiec and O’Brien, Taylor

⁵⁰ Cf. Kupiec and O’Brien (1997); Daripa and Varotto (1998).

⁵¹ For the development and current outline of Solvency II cf. EU Commission MARKT (2002, 2003, 2004, 2005) and EU Commission (2001, 2005).

⁵² Goodhart (1998).

proposes an enhanced PCA for banks⁵³ with only three essential variables: the bank's capital adequacy ratio, the potential loss, and the regulatory capital. We propose adapting this approach to insurance and making it "principle-based" by supplementing the three key components, both quantitative and qualitative, which are defined by the supervisor: (1) the capital adequacy ratio τ , which as a basic condition links insurers' pre-commitments on their loss thresholds L , on the one hand, with their capital thresholds C , on the other hand, such that $C/L \geq \tau$; (2) supplemental basic principles to guide decisions on the basic condition; and (3) stress test scenarios to back test insurers' commitments on L and C .

The interplay of these three components works as follows: the regulator sets in advance a capital adequacy ratio τ . A qualifying insurer proposes thresholds L and C , which must meet the basic condition $C/L \geq \tau$. If the regulator accepts the thresholds (we will return to this in a moment), the insurer commits to keeping its capital above the agreed level C and its losses below the agreed level L over an agreed period. In practice, a regulator may define and require L and C annually for each major type of risk (diversification will be discussed in the following section), and thus lay the basis for very detailed and transparent loss monitoring. Any insurance company with losses in excess of its L commitment or capital below its C commitment would be subject to regulatory and supervisory intervention. Several stages of intervention would be needed to reflect different degrees of severity of the breach – from "plan submitted to regulator" to "withdrawal of authorisation".⁵⁴ For the approach to work properly, it is vital that breaching a threshold results in prompt and rigorous corrective intervention. This will give an insurer a powerful incentive to keep its capital and manage its losses as committed.

The basic condition is at the heart of the regulating mechanism of the principle-based PCA. However, it is not free of risk. In particular, one might argue that the basic condition alone offers significant room for discretionary actions, particularly given the long-term nature of insurance business.

Thus, additional flanking support of the basic condition in the form of preventative supervisory instruments seems necessary in order to avert the danger of short-term PCA "malpractice". This function can be fulfilled efficiently if the regulator additionally stipulates some basic principles and defines stress tests to be applied by insurance companies. Both forms of support will strengthen the basic condition and are simple for the supervisor to apply.

Under "basic principles", we understand the fundamental conditions or criteria that an insurer must adhere to when setting L and C . The principles would include a catalogue of the risks that insurers must map when setting the threshold values. Enforcing adherence to these criteria is already the object of the second regulatory building block, the supervisory review processes. In effect, the basic principles would be a useful checklist for the requirements to be reviewed by the supervisor. Similarly, the basic principles would give insurers clear indications of the organizational and procedural requirements that their risk management functions must satisfy, for

⁵³ In Taylor (2002, 2003), the approach is called NGA (New General Approach).

⁵⁴ Cf. EU Commission MARKT (2002) and American Academy of Actuaries (2002).

example, what risk drivers need to be managed, for which events a contingency plan must be provided.

While the basic principles would ensure *ex ante* a fundamental/minimum degree of certainty and consistency for the individual specification of L and C , their *ex post* robustness can be reviewed by introducing the third element: stress test scenarios against which the sensitivity of the insurer's L and C thresholds would be measured (before acceptance by the supervisor). We have previously noted both the suitability of stress tests for pragmatically mapping the impact of tail risks and the increasing use of stress tests by regulators. The regulator may start with a short list of scenarios and supplement the list over time.⁵⁵

For the definition of stress tests, different escalation stages of regulation are conceivable, the intensity of which can be linked to an insurer's past performance with respect to the threshold values. For example, for insurers that have not violated the basic condition of $C/L \geq \tau$ for a longer period of time, the regulator might simply stipulate the set of scenario variables to be stress tested, allowing the insurer to decide on the selection and combination of the stress tests and the stress intensity – with the proviso that the choices are to be justified as part of the supervisory review process and possibly adjusted. For insurers that have breached thresholds, a higher level of supervisory attention would be necessary – for example, stipulating the specific test scenarios, including parameterization of the individual stress variables.

Quantification of the basic condition

In order to assess the basic condition $C/L \geq \tau$ for individual insurers (and risk types), the regulator starts with the quantification of the capital ratio τ . The regulator may choose to set different capital ratios for different types of insurance and types of risk. For every ratio, the regulator will set the time period (during which insurance companies need to meet the thresholds) and the confidence level, that is, what percentage of insolvencies the regulator “wants to expect”. Given these boundary conditions, the regulator is equipped to determine empirically a range for the adequacy ratio. Within this historically determined range, the adequacy ratio applicable for the actual period may then be determined considering the concentration in a certain market, the stage of the business cycle, or other specific market conditions, such as accounting practices. It should also reflect “newly” quantified risks or mitigation effects (such as diversification⁵⁶). Such effects could be identified in an industry-wide effort and then transparently incorporated in the requirements.

The specific figures for loss threshold L and capital threshold C commitments should come from the individual insurance company. L should express the maximum actual quarterly loss that company management can sustain. The company is, in principle, free to set these thresholds and, thus, may well choose to use an internal risk model to define L , while the regulator uses its capital ratio τ to determine whether L and C fulfil the basic requirements. A regulator may even choose to provide a standard

⁵⁵ For example, the scenarios defined under the Swiss Solvency Test. Cf. BPV (2004).

⁵⁶ For all these adaptations, a fact base accepted by the industry should be provided.

model that every insurance company may want to take and adapt to its needs on its own responsibility.

The decisive difference between the PCA described here and the approaches currently under discussion is two-fold:

- (1) The regulator remains at a low level of regulatory depth, that is, prescribes only the basic principles that an insurer must demonstrably use to set L and C , and does not prescribe technical requirements for models or specify model parameters such as risk factors, the risk measure or confidence level. By preserving the individual company's degrees of freedom in determining L and C , such a principle-based PCA succeeds to some extent in diversifying the model risk across the market participants and allowing individual companies' management skills and market competition to determine "quality". In addition, the regulator does not attempt to influence the individual risk preferences and considerations of the market participants or their handling of risk.
- (2) The regulator establishes a "leading indicator" of management capabilities, namely L . As L is set by management, management needs to be able to avoid breaching it. L is (or the multiple L s are) an indicator of the quality of a company's ability to steer its business.

Obviously, these are only initial thoughts about the quantification of (sets of) L , C , and τ . While further research and empirical analyses are needed, compared with proposals for capital regulation based on stochastic risk modelling, a PCA has clear advantages for fulfilling regulatory objectives, which we discuss in the following section.

Fulfilment of regulatory objectives under a PCA

A PCA applied to the insurance industry would strongly support the regulator in meeting the objectives of consumer protection, market stability, and competitive efficiency.

Fulfilment of the objective of consumer protection depends first on the quality of the requirements for capital. Since regulators can derive a clear minimum level of C based on historical data for each insurer, the requirements for regulatory capital and, thus, the level of protection achieved will certainly be no worse than in the past.

More importantly, a principle-based PCA can conceivably improve consumer protection – if the regulator provides for complementing C with a short-term counterpart L that is supplemented and validated through scenario testing procedures as mentioned. Today there is consensus that supervisory reliance solely on capital requirements does not sufficiently protect consumers against fraudulent behaviour and moral hazard as capital suffers from being a lagging indicator. Precisely this issue can be addressed under a PCA by introducing a short-term, leading indicator L that reduces information asymmetries and makes the credibility of management decisions visible on a high-frequency basis. Since L is a true measure of actual business results, the insurer's individual commitment to L will enable regulators to take much earlier action than when relying purely on the provision of capital.

Nevertheless, one might argue that a PCA in its pure form, unlike stochastic-model-based capital regulation, is still not entirely free of gaming risk. Speculative behaviour might not always be detected by the regulator given that large insurance risks typically materialize infrequently. For example, let us consider an insurer who excessively writes high-impact, low-frequency risks, but sets L artificially low in order to exploit discounts on C . It could well be that L would not reveal the insurer's manipulation strategy for a long time, and one might argue that L as a high-frequency measure does not sufficiently serve the regulator's need to supervise low-frequency risks. From our point of view, the lower level of L implied in such manipulation strategies would effectively mitigate the risk of gaming in most cases, as trigger points for regulatory intervention are generally tighter and, consequently, a closer review of the entire risk portfolio becomes more likely. In addition, diligently defined scenario testing may reveal a significant number of such strategies in advance. In certain cases, profit/risk testing of products (i.e., the pricing strategy in certain lines of business) might be needed.

The second objective, market stability, will, at minimum, not be worsened under the PCA. Regulation of risk management would be incentive-based, rather than prescriptive. As trivial as the approach may seem, it does not "create" drivers or sources of systemic risk, which means that the problems discussed in the section on propagation mechanism cannot occur. The PCA lays down only general principles for adequate capitalization. That means it regulates only the output, that is, the level and relationship of L and C , but not the input and processes required to measure and manage them. In particular, there are no prescriptive or even standardized rules on how risk should uniformly be modelled and measured. Thus, as capital regulation of this kind does not homogenize individual market participants' risk preferences, it does not run the risk of fostering the propagation of external shocks. Contagion driven by risk-sensitive capital requirements is not possible. By the same token, the introduction of a PCA regime does not entail the risk of triggering systemic shocks because its implementation does not imply a market-wide shift of risk preferences at one point in time. Finally, as systemic risk is not fostered, its impact cannot be amplified. Thus, EU-wide harmonization based on a PCA's general principles is not critical from a systemic risk point of view. Moreover, we believe that keeping to the level of principles would ease the political processes needed to achieve supranational and cross-sectoral agreement on future capital regulation.

Adopting a regulated PCA is also likely to improve fulfilment of the third objective – competitive efficiency. With the capital adequacy ratio τ , the PCA sets an important but basic standard for the output, that is, C and L . It does not regulate the means for determining and handling C and L , that is, management strategies and instruments. These are self-regulated by the competitive forces of the market. The use of τ would require insurers to make a judgement about how well they can manage risk and, therefore, how much capital they need. Thus, an insurer's choice of L describes to a certain extent its ability to manage risk. In this respect, the basic condition will work well with the design of supervisory review processes as currently proposed under Solvency II.⁵⁷ Furthermore, depending on the reporting and disclosure rules, a

⁵⁷ See EU Commission MARKT (2002, 2003, 2004).

threshold breach could be an instructive event, not only for the insurance company itself but also for its competitors.⁵⁸ In this sense, overall risk taking by the industry would also be enhanced since those insurers with a tendency to undervalue risk would be under the regulator's scrutiny as they would display too high a volatility in results (i.e., too-high losses and potentially violation of L).

In conclusion, a few words on a cost–benefit analysis of the proposed PCA. Part of the appeal of the PCA is that it would be much simpler to implement than the modelling process currently under discussion – both organizationally and technically.

It might also be argued that the limited operational requirements of a PCA will permit greater concentration on the second regulatory building block, that is, the supervisory review processes. Many companies have been working intensely to prepare for the first building block, the new financial requirements, but in some cases have yet to tackle their risk management organizations and processes. Assuming that a company first has to implement a risk management organization and processes before it can pursue risk management actively, the effect of shifting attention from the first to the second building block achieved with the PCA would represent refocusing towards the main building block of solvency regulation – hence returning to the notion of the early suggestions for the Solvency II reform.⁵⁹

In order to make the proposed PCA operational and exploit the associated advantages, however, further research and analysis are required, particularly in the field of principal–agency problems. The research should focus on both *ex ante* and *ex post* measures. First, as mentioned, the elements of the supervisory review building block should be specified, that is, risk management standards derived from the basic principles for the basic condition. Second, in order to put L effectively to work, study is also needed to identify the optimal design of trigger points for regulatory intervention and their interplay with corresponding sanctions and external reporting requirements.

Conclusion

The prescription of stochastic models for the quantification of regulatory capital requirements in the insurance industry appears more likely to be a bane than a boon for the objectives of consumer protection, market stability, and competitive efficiency. The use of stochastic models bears significant risks, which include the underestimation of the probability of tail losses and consequently insolvency, the propagation of systemic risk, and the reduction of competitive efficiency.

In view of these drawbacks, an alternative approach is needed. We advocate more study of the design elements of a principle-based PCA. A PCA may (1) serve as a nucleus for a pragmatic solution and (2) foster industry-wide improvement in risk management by promoting better internal risk assessment and loss management. At minimum, a principle-based PCA would avoid increasing systemic risk, quite simply because it does not prescribe a uniform risk measurement approach for the industry as a whole. Over time, a PCA also has the potential to foster competitive efficiency by

⁵⁸ Depending on the design of the third pillar of Solvency II, that is, reporting and disclosure rules.

⁵⁹ Cf. Sharma (2002).

making management performance transparent as insurers gain a “track record” of adherence to (or violations of) the loss and capital thresholds of the basic condition $C/L \geq \tau$. Finally, a PCA, combined with principles that do not prescribe or prohibit internal use of available risk models and supplemented by stress testing would improve consumer protection, driven by market forces.

While further work is needed, the PCA appears to avoid the problems associated with stochastic modelling as a basis for quantifying regulatory and economic capital and provides leeway for companies to optimize their risk and capital management as they see fit.

Against the background of the known limitations of stochastic capital models and the potential advantages of a PCA, a regulator should consider whether the benefits of using stochastic models for regulatory purposes justifies the risk for consumers.

Acknowledgements

We thank the anonymous referees for their valuable suggestions and remarks.

References

- Ahn, D.-H., Boudukh, J., Richardson, M. and Whitelaw, R.F. (1999) ‘Optimal risk management using options’, *Journal of Finance* **54**(1): 359–375.
- American Academy of Actuaries (2002) *Comparison of the NAIC Life, P&C and Health RBC Formulas*, Joint RBC Task Force Report, February 2002, available at: www.actuary.org/pdf/finreport/jrbc-12feb02.pdf.
- Ang, A. and Bekaert, G. (1999) *International asset allocation with time-varying correlations*, NBER working paper, 7056.
- Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999) ‘Coherent measures of risk’, *Mathematical Finance* **9**: 203–228.
- Benston, G. (1998) *Regulating financial markets: a critique and some proposals*, Hobart paper 135, published by the Institute of Economic Affairs, London.
- BPV, Bundesamt für Privatversicherungswesen (2004) *White paper on the Swiss solvency test*, Zürich/Bern.
- Brooks, C. and Persaud, G. (2000) *Value at risk and market crashes*, discussion papers in Finance 2000–01, ISMA Centre, University of Reading.
- CEA, Comité Européen des Assurances (2005) *Solvency Assessment Models Compared*, Report jointly produced with Mercer Oliver Wyman, Paris.
- Chow, G., Jacquier, E., Kritzman, M. and Lowry, K. (1999) ‘Optimal Portfolios in Good Times and Bad’, *Financial Analysts Journal* **55**(May/June): 65–73.
- Christoffersen, P.F., Diebold, F.X. and Schuermann, T. (1998) *Horizon problems and extreme events in financial risk management*, working paper, No. 98–16, The Wharton School, University of Pennsylvania.
- Crockett, A. (1997) *Why is Financial Stability a Goal of Public Policy?* Symposium on Financial Stability in a Global Economy, Federal Reserve Bank of Kansas City, Kansas.
- CRO-Forum, The Chief Risk Officer Forum (2005) *A Framework for Incorporating Diversification in the Solvency Assessment of Insurers*, 10 June 2005.
- Cruz, M. (2002) *Modelling, Measuring and Hedging Operational Risk*, New York: John Wiley.
- Danielsson, J. (2002) ‘The emperor has no clothes: Limits to risk modelling’, *Journal of Banking & Finance* **26**: 1273–1296.
- Danielsson, J. and De Vries, C. (2000) *Value-at-Risk and Extreme Returns*, Working Paper, London School of Economics, Erasmus University, Rotterdam.
- Danielsson, J., Embrechts, P. and Goodhart, C. (2001) *An Academic Response to Basel II*, Special Paper Series, No. 130, Economic and Social Research Council, LSE Financial Markets Group.
- Danielsson, J. and Song Shin, H. (2000) *Endogenous Risk*, working paper, London School of Economics.
- Danielsson, J. and Zigrand, J.-P. (2004) *On Time-Scaling of Risk and the Square-Root-of-Time Rule*, Working Paper, London School of Economics.
- Daripa, A. and Varotto, S. (1998) *Value at Risk and Precommitment: Approaches to Market Risk Regulation*, Federal Reserve Bank of New York Economic Policy Review, October 1998, New York.

- Diebold, F., Hickman, A., Inoue, A. and Schuermann, T. (1997) *Converting 1-day volatility to h-day volatility: scaling by h is worse than you think*, discussion paper series, No. 97–34, The Wharton School, University of Pennsylvania.
- Dowd, K. (2004) ‘VaR and Subadditivity’, *Financial Engineering News* 40(November/December): 7.
- Embrechts, P., McNeil, A. and Straumann, D. (1999) *Correlation and dependence in risk management: properties and pitfalls*, working papers, ETH Zürich.
- Engle, R.F. and Manganelli, S. (2001) *Value at risk models in finance*, working paper series, No. 75, European Central Bank, Frankfurt am Main.
- EU Commission (2001) *European governance – a white paper*, COM (2001) 428 final, July 25, 2001, Brussels.
- EU Commission (2005) *Amended framework for consultation on Solvency II*, July 2005, available at: www.europa.eu.int.
- EU Commission MARKT (2002) *Consideration on the design of a future prudential supervisory system*, paper for the Solvency Subcommittee, 2002, 2535/02-EN, available at: www.europa.eu.int.
- EU Commission MARKT (2003) *Solvency II – reflections on the general outline of a framework directive and mandates for further technical work*, 2539/03-EN, Notice to the IC Solvency Subcommittee, 2003, available at: www.europa.eu.int.
- EU Commission MARKT (2004) *Solvency II – organisation of work, discussion on pillar i work areas and suggestions of further work on Pillar II for CEIOPS*, 2543/04-EN, note to the IC Solvency Subcommittee, 2004, available at: www.europa.eu.int.
- EU Commission MARKT (2005) *Solvency II Roadmap*, note to the European Insurance and Occupational Pensions Committee (EIOPC): 2505/05-rev.2-EN, July 2005, available at: www.europa.eu.int.
- FAME, Foundation for Agency Management Excellence (2003) *Costs & Benefits of Future Regulatory Options for the U.S. Insurance Industry, An Analytical Framework*, Georgetown Economic Services.
- Freixas, X. and Santomero, A.M. (2003) *An overall perspective on banking regulation*, working paper, Universitat Pompeu Fabra, The Wharton School, University of Pennsylvania.
- FSA, Financial Services Authority (1999) *Cost-benefit analysis in financial regulation*, occasional paper, September 1999, London.
- Goodhart, C. (1998) *Financial Regulation: Why, How and Where Now?*, London and New York: Routledge.
- Grace, M.F. and Klein, R.W. (1999) *Efficiency Implications of Alternative Regulatory Structures for Insurance*, paper presented at the American Enterprise Institute, Center for Risk Management and Insurance Research, Georgia State University.
- Group of Thirty (1997) *Global Institutions, National Supervision, and Systemic Risk*, A Study Group Report.
- Harrington, S. (2005) ‘Capital adequacy in insurance and reinsurance’, in H. Scott (ed) *Capital Adequacy Beyond Basel*, Oxford: Oxford University Press.
- Hartmann, P. and De Brandt, O. (2000) *Systemic risk: A survey*, working paper series, No. 35, European Central Bank, Frankfurt am Main.
- Herring, R.J. and Santomero, A.M. (2000) *What is optimal financial regulation*, working paper 00-34, Wharton Financial Institutions Center, University of Pennsylvania.
- Herring, R.J. and Schuermann, T. (2003) *capital regulation for position risk in banks, securities firms and insurance companies*, working paper, Wharton Financial Institutions Center, University of Pennsylvania/Federal Reserve Bank of New York.
- Jorion, P. (1996) ‘Risk²: Measuring the risk in value at risk’, *Financial Analyst Journal* 52(6): 47–56.
- Jorion, P. (2001) *Value at Risk, The New Benchmark for Managing Financial Risk*, 2nd edn., New York: McGraw-Hill.
- Ju, X. and Pearson, N.D. (1998) *Using value-at-risk to control risk taking: how wrong can you be?*, OFOR Paper, No. 98–08, Department of Finance, University of Illinois.
- Karolyi, G.A. and Stulz, R.M. (1996) ‘Why do markets move together? An investigation of US-Japan stock return comovements’, *Journal of Finance* 51: 951–986.
- Kondor, J., Szepessy, A. and Ujvárosi, T. (2004) ‘Concave risk measures in international capital regulation’, in Szegö (ed) *Risk Measures for the 21st Century*, New York: John Wiley.
- Krause, A. (2002) *Exploring the limitations of value at risk – how good is it in practice?*, working paper, March 2002, University of Bath.
- Kupiec, P.H. and O’Brien, J. (1997) ‘The pre-commitment approach: using incentives to set market risk capital requirements’, *FEDS Paper 97–14*(March), Federal Reserve Board.

- Llewellyn, D.T. (1999) *The economic rationale for financial regulation*, occasional paper, No. 1, Financial Services Authority, London.
- Longin, F. and Solnik, B. (1998) *Correlation structure of international equity markets during extremely volatile periods*, working paper, CR 646/1998, HEC, Jouy-en-Josas.
- Los, C.A. (2004) *Why VAR fails: Long memory and extreme events in financial markets*, working paper, Department of Finance, Kent State University.
- McNeil, A., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management – Concepts, Techniques and Tools*, Chapter 2, Princeton: Princeton University Press.
- Merton, R.C. (1989) ‘On the application of the continuous-time theory of finance to financial intermediation and insurance’, *The Geneva Papers on Risk and Insurance* 14(52): 225–261.
- Mills, E. (2005) ‘Insurance in a climate of change’, *Science* 309(5737/August): 1040–1043.
- Minderhoud (2003a) *Extreme stock return co-movements of financial institutions: Contagion or independence?*, working paper, Monetary and Economic Policy Department, De Nederlandsche Bank, Amsterdam.
- Minderhoud, K. (2003b) *Systemic risk in the Dutch financial sector*, MEB series working papers, No. 2003-17, Monetary and Economic Policy Department, De Nederlandsche Bank, Amsterdam.
- Morris, S. and Shin, H.S. (1999) ‘Risk Management with interdependent choice’, *Oxford Review of Economic Policy* 15(3): 52–62.
- Morrison, A.D. (2002) *The economics of capital regulation in financial conglomerates*, working paper 08/2002, Saïd Business School, University of Oxford.
- Nebel, R. (2004) ‘Regulations as a source of systemic risk: the need for economic impact analysis’, *The Geneva Papers on Risk and Insurance* 29(2): 273–283.
- OMB, Office of Management and Budget (2002) *Stimulating Smarter Regulation: 2002 Report to Congress on the Costs and Benefits of Regulations*.
- Persaud, A. (2000) ‘Sending the herd off the cliff edge: The disturbing interaction between herding and market-sensitive risk management practices’, *Erisk* 12: 1–6.
- Pfeifer, D. (2004) *VaR oder Expected Shortfall – welche Risikomaße sind für Solvency II geeignet?*, Conference Proceedings, Handelsblatt Konferenz “Solvency II”, 18./19. November, Köln.
- Pritsker, M. (2001) *The hidden dangers of historical simulation*, working paper, The Federal Reserve Board and University of California, Washington, Berkeley.
- Sharma, P. (2002) *Prudential Supervision of Insurance Undertakings: Report of the London Working Group on Solvency II*, Conference of the Insurance Supervisory Services of the Member States of the European Union, Paris.
- Swiss Re (2002) ‘Natur- und man-made-katastrophen 2001, man-made-schäden einer neuen dimension’, *Sigma* 1(2002); Swiss Re Economic Research & Consulting, Zurich.
- Szegö, G. (2004) *Risk Measures for the 21st Century*, New York: John Wiley & Sons.
- Taylor, C. (2002) *A new general approach to capital adequacy: A simple and comprehensive alternative to Basel 2*, published by the Centre for the Study of Financial Innovation (CSFI), December 2002, London, New York.
- Taylor, C. (2003) *A new general approach to capital adequacy: A simple and comprehensive alternative to Basel 2, supplementary note, clarifications and illustrations*, submitted to Basel Committee on the Third Consultative Paper, July 31 2003.
- Wilmarth, A.E. (2002) *Controlling systemic risk in an era of financial consolidation*, working paper as of July 31, 2002, George Washington University, Washington DC.
- Yamai, Y. and Yoshihara, T. (2002) *Comparative analyses of expected shortfall and value-at-risk – their estimation error, decomposition, and optimization*, working paper, Institute for Monetary and Economic Studies, Bank of Japan, Tokyo.

About the Authors

Oliver S. Bäte is a director of McKinsey & Company, Inc., and the leader of McKinsey’s European Insurance Practice.

Dr Günther Thallinger is a principal and based in McKinsey’s Cologne office.

Philipp von Plato is a consultant and based in McKinsey’s Munich office.