Solvency Regulation of Insurers: A Regulatory Failure?

Peter Zweifel

Abstract: This paper puts forth a critique of European solvency regulation of the type imposed on insurers by Solvency I and II. Insurers’ underwriting and investment divisions seek to maximize the expected risk-adjusted rate of return on capital (RAROC) in period 0. For them, higher solvency serves to increase demand and hence premium income but ties costly capital. Sequential decision making by insurers is tracked over three periods. In period 1, exogenous changes in expected returns and in volatility occur, causing optimal adjustments of solvency in period 2. In period 3, the actual change in solvency triggers adjustments in underwriting and investment, resulting in new values of expected returns and volatility. These changes create an endogenous efficiency frontier in (μ,σ)-space for the insurer. Both Solvency I and II are shown to modify the slope of this frontier, inducing senior management to opt for a higher volatility in several situations. Therefore, both types of solvency regulation can run counter their stated objective, which may also be true of Solvency III. [JEL codes: G15, G21, G28, L51. Key words: regulation, insurers, solvency, Solvency I, Solvency II, Solvency III.]

INTRODUCTION

Risk-adjusted return on capital (RAROC) has increasingly become the benchmark for assessing the performance and governance of insurers’ underwriting and investment activities. While a higher solvency level has the benefit of increasing demand and hence premium income, it ties costly capital. At the same time, public regulation is concerned about solvency to

1Prof. em., Economics Dept., University of Zurich (Switzerland), Kreuth 371, A-9531 Bad Bleiberg (Austria), phone ++43 4244 20319; peter.zweifel@econ.uzh.ch
Acknowledgment: This paper has benefitted from criticisms and helpful suggestions from Roland Eisen (Munich), the participants of the Insurance Seminar of the University of Hannover, the Finance Seminar of the University of Zurich, the Management & Economics seminar of the University of Munich (especially Richard Peter), and management seminars at “Zurich” Insurance headquarters and SCOR (Paris) as well as by two anonymous referees. The usual disclaimer applies.
ensure the continuity of insurance operations. Solvency I was promulgated in 2002 in the guise of two ordinances: 2000/13/EC concerning property-liability insurance, and 2000/83/EC, concerning life insurance (European Commission, 2013). These ordinances updated regulation that had been in effect since the 1970s in major European insurance markets but extended it to all member countries of the European Union (EU). Insurers were to have free eligible (i.e., low-risk) assets relative to annual premiums or the 3-year average of claims outstanding, resulting in two solvability margins, the higher of which was to apply. In 2009, European Parliament and EU Ministers of Finance passed Solvency II regulation as part of the Single Market Initiative of the European Commission. In analogy to Basel II regulation applying to banks, Solvency II consists of three pillars, concerning solvency capital, risk management, and reporting requirements modelled after the International Financial Report Standards. The solvency margin is to consist of a minimal capital requirement and an additional solvency capital requirement, now relative to eligible own funds only. Solvency II was to be implemented by 30 June 2013 but is expected to come into effect no sooner than 1 January 2016, with Solvency I remaining in force in the meantime (Lloyd’s, 2013).

This paper deals with the conflict between optimal solvency as determined by insurers themselves and exogenously imposed solvency levels, as epitomized by Solvency I and the first pillar of Solvency II. It depicts an insurer in the process of its sequential decision making. In a first period, exogenous changes in expected returns and volatility on the capital market \( (d\hat{\mu}, d\hat{\sigma}) \) impinge on its underwriting and investment departments, which are assumed to form one division (“the Division” henceforth) for simplicity. A typical cause could be investments made in the previous period that turn out to have a lower rate of return or a higher volatility than expected. In the second period, the Division adjusts the insurers’ solvency level by \( dS^*/d\hat{\mu} \) and \( dS^*/d\hat{\sigma} \), respectively, in the aim of attaining again its efficient amount of solvency capital prior to rebalancing its underwriting and investment portfolios. In the third period, the Division proposes to senior management a rebalancing in response to the changed solvency level through endogenous adjustments in expected returns and volatility \( d\hat{\mu}/dS^* \) and \( d\sigma/dS^* \). This defines the slope \( d\hat{\mu}/d\sigma \) of an internal efficiency frontier on which senior management chooses the optimum, taking into account its degree of risk aversion.

This efficiency frontier is modified by solvency regulation such as Solvency I and II. It will be argued that Solvency I neglects that the relationships between premium income and solvency and between solvency capital and premium income depend on exogenous \( \mu \) and \( \sigma \). As to Solvency II, it addresses solvency directly but still fails to take into account the fact
The Division in charge of underwriting and investment maximizes RAROC by optimally selecting solvency $S$ (see text for a definition)

1. Shocks $d\hat{\mu}$, $d\hat{\sigma}$ occur
2. The Division adjusts solvency level by $dS^*$
3. Adjustment $dS^*$ moves the Division along the $(\mu, \sigma)$-frontier which is modified by Solvency I, II, and III regulation; senior management of insurer selects optimum on $(\mu, \sigma)$-frontier

$t$

Fig. 1. Timeline of the model.

that for an insurer that initially just met this standard, the amount of risk capital needed to improve solvency changes when $\hat{\mu}$ falls or $\hat{\sigma}$ increases. It will be shown that both Solvency I and II modify the slope of the efficiency frontier $d\hat{\mu}/d\hat{\sigma}$ as perceived by regulated insurers. While one might expect that these regulations serve to reduce the slope of the frontier (thus inducing senior management to opt for lower $\mu$ and lower $\sigma$), it turns out that the opposite can be the case. Indeed, through their neglect of parameters of importance to insurers themselves, both Solvency I and II may have the unexpected consequence of making at least some insurers opt for a higher value of $\sigma$ (i.e., higher volatility of the rate of return on their assets) than without it, causing regulation to miss its target. This risk is likely to increase again in the case of impending Solvency III regulation, without, however, attaining the original level of Solvency I.

This paper is structured as follows. Section 2 contains a review of the pertinent literature to conclude that solvency regulation not only of banks but also of insurers may serve to avoid negative externalities. In Section 3, a higher level of solvency is found to have two effects for an insurer’s investment and underwriting decisions aiming to maximize RAROC. On the one hand, it serves to increase premium income; on the other, it ties capital that would have other, more productive uses. This initial optimum is disturbed by exogenous shocks in returns $d\hat{\mu}$ and volatility $d\hat{\sigma}$, respectively (see period 1 of Figure 1). In period 2, the Division adjusts the insurer’s solvency level to these shocks. These adjustments are derived in Section 4. However, there can be only one adjustment $dS^*$, which moves the insurer along an endogenous efficiency frontier in the third period. The slope $d\hat{\mu}/d\hat{\sigma}$ of this frontier is derived in Section 5. Senior management is presented with this tradeoff and makes its choice, taking account of its degree of risk aversion. The regulations imposed by Solvency I, II, and III
are introduced as parameter restrictions in Section 6 to show how $d\hat{\mu}/d\hat{\sigma}$ is modified, causing senior management of regulated companies to opt for a higher value of $\sigma$ than absent this regulation in a number of situations. A summary and conclusions follow in Section 7.

**LITERATURE REVIEW**

The solvency regulation of insurers has traditionally been justified by the external costs of insolvency (Cummins, 1988). This view was challenged by the proponents of the Capital Asset Pricing Model, who emphasized that for well-diversified investors, the solvency of an insurer does not constitute a reasonable objective. They are concerned with expected profitability, adjusted for the degree to which the insurer’s profitability systematically varies with the capital market (the Beta of the Capital Asset Pricing Model). By way of contrast, for little-diversified investors (among them, policyholders of the insurer), the insurer’s overall risk is relevant, which importantly includes the risk of insolvency (for the case of banks, see Goldberg and Hudgins, 1996; Park and Peristiani, 1998; Jordan, 2000; Goldberg and Hudgins, 2002). Option Pricing Theory in turn shows that due to their limited liability, shareholders of the insurer in fact have a put option that is written by the other stakeholders (notably policyholders as creditors) of the insurer (Cummins and Phillips, 2001; Zweifel and Eisen, 2012, ch. 6.3; for the case of banks, see Merton, 1974; Jensen and Meckling, 1976; Merton, 1977).

When a solvency risk materializes, internal and external costs need to be distinguished. Internal costs are borne by the insurer’s shareholders, who see the value of their shares drop to zero unless the insurer is in business again. However, in view of the loss of reputation, this re-entry would meet with high barriers to entry (Epermanis and Harrington, 2006; for the case of banks see Smith and Stulz, 1985; Stulz, 1996). In addition, insolvency has external costs (i.e. costs not borne by the insolvent insurer). First, the insolvency causes consumers to go without insurance coverage (Bauer and Ryser, 2004). Policyholders stand to lose at least part of their assets in the event of loss. Some of the policyholders may be insurers themselves involved in the coinsurance of large risks; therefore, the insolvent insurer may drive other financial institutions into insolvency, causing substantial external costs (Furfine, 2003). Second, investors in the capital market at large often are affected as well. They have occasion to re-evaluate the estimated risk of insolvency of insurance companies in general. In response to a revised estimate, shareholders demand a higher rate of
return, resulting in a higher cost of financing for insurers, while policyholders curtail demand, resulting in a fall in premium income.

There is a substantial body of empirical research substantiating these claims (Cummins and Sommer, 1996; Flannery and Sorescu, 1996; Park and Peristiani, 1998; Covitz et al., 2004). It suggests that a solvency level that is deemed optimal by the individual insurer is too low from a societal perspective because insolvency causes external costs. However, it may be worthwhile to emphasize that this conclusion does not suffice to justify public regulation to ensure solvency. One would have to first examine whether the expected benefit of the intervention exceeds its expected cost. An important component of this cost is caused by unintended behavioral adjustments. The present contribution belongs to this tradition of research, which in the case of banking dates back at least to Koehn and Santomero (1980). Characterizing a bank by its utility function and assuming it to optimize a portfolio containing both assets and liabilities, they find that imposing a simple equity-to-assets ratio constraint is ineffective on average.

In Rochet (1992), banks choose their asset portfolio taking into account limited liability, which may cause them to become risk-lovers. The same might be true of insurers. This makes imposing minimum capital requirements necessary to prevent them from choosing very inefficient portfolios. However, the effectiveness of this type of regulation is not guaranteed. John, Saunders, and Senbet (2000) show that U.S. capital-based regulation introduced in 1991 may have failed to prevent bank managers from shifting risk to outside financiers unless features of their compensation plans are taken into account along with the opportunity set of asset investments. More recently, Repullo (2004) explicitly has dealt with Basel II (whose provisions are very similar to Solvency II) in the context of an imperfectly competitive banking market. He derives conditions for two Nash equilibria to obtain, one in which banks invest in riskless and another where they invest in risky assets. While capital requirements on risky assets do enlarge the parameter space of the “prudent” equilibrium, depositors bear the burden of regulation in the guise of lower interest rates. In the case of insurance, policyholders would be confronted with a higher premium-loss ratio, indicating an increased price of coverage. Lower returns for depositors are the reason why in Repullo (2004) capital requirements are in general effective in preventing excessive risk-taking by banks. Furthermore, he shows that Basel II permits a reduction in the overall amount of capital required by regulation compared to Basel I. However, pointing to bank-specific problems of governance, Mülbert (2009) argues that prudential regulation of the Basel I and Basel II type may even induce rather than prevent banking crises. A similar finding will be presented in this paper for insurance. An argument that goes beyond the framework presented
here is proffered by Macey (2013: ch. 11), who links the recent financial crises to the substitution of the unambiguous reputation built up by financial firms by the more ambiguous reputation of regulatory authorities.

The present contribution adds to the literature in three ways. First, it takes into account the fact that insurers have a business model that differs from that of banks. Insurers derive profit from two activities—risk underwriting and capital investment. They receive a premium income prior to the occurrence of claims (which can be considered exogenous in the present context), creating scope for capital investment. Second, the paper clearly distinguishes between earlier Solvency I and later Solvency II regulation, showing that the more recent variant may have unintended consequences only for a subset of insurers rather than all of them. However, Solvency III is found to likely increase this subset again. In this respect, this work elaborates on and refines the contributions by Kim and Santomero (1988) as well as Rochet (1992). The third distinguishing feature of this paper is its emphasis on dynamics in the following way. Whereas earlier contributions analyzed optima or (in the case of Repullo, 2004) equilibria, here the insurer’s path of adjustment from one optimum to the next is analyzed. Adjustment to exogenous shocks will be shown to be conditioned by regulation of the Solvency I to III type. In return, welfare implications will not be spelled out; rather, the fact that insurers may be induced to act against the stated intentions of the regulator will be highlighted.

**OPTIMAL SOLVENCY IN A MODEL OF AN INSURER’S DIVISION IN CHARGE OF UNDERWRITING AND INVESTMENT**

An insurance company is seen to pursue two main activities—risk underwriting and capital investment. These two activities are amalgamated into a single division for simplicity. Admittedly, this construct is in tension with the fact that an insurance company pursues many more activities, ranging from marketing to sales, claims settlement, and on to risk management services. However, it has two benefits. First, it permits us to focus on the two activities—“risk underwriting” and “capital investment”—that are decisive for the company’s degree of solvency and hence solvency regulation. Second, distinguishing two divisions (say) would still result in two distinct efficiency frontiers in \((\mu,\sigma)\)-space, raising the question of how senior management is to choose between them. Thus, let this Division maximize the expected risk-adjusted rate of return on capital (RAROC) through its choice of solvency \(S\). The Division is assumed to act in a risk-neutral manner, which can be justified by noting that allowing risk
aversion to affect decisions of employees would result in inconsistencies, e.g., in risk underwriting. Employee A (who is strongly risk averse) might turn down a client seeking to obtain coverage while employee B (who has more risk appetite) of the same insurer would accept the client. Senior management needs to avoid such inconsistencies. Therefore, the expectation operator is dropped as long as the Division is being analyzed; risk aversion enters in period 3 when the insurer’s senior management selects its preferred position on the internal (m,s)-efficiency frontier generated by the Division (as argued below, there is no capital market line). A higher level of solvency S enables the insurer (and hence the Division) to obtain more funds through higher premium income (Sommer, 1996; for empirical evidence, see Epermanis and Harrington, 2006).

The use of a (μ, σ)-efficiency frontier can be criticized because these two parameters suffice to characterize a distribution only in the case of normality, whereas returns to underwriting and particularly investment are known to exhibit skewness and curtosis. However, being of fourth order, curtosis necessarily adds to variance, which is frequently true of skewness as well. Also, note that there is no need to define solvency in a formal way; however, it may be thought in terms of the likelihood of a shortfall (Leibowitz et al., 1992) or in terms of value-at-risk (VaR) or expected value-at-risk (EVaR) concepts (see Artzner et al., 1999, for a critique and Brandtner, 2013, for a comparison between so-called coherent risk measures and VaR). Whenever VaR or EVaR increases, solvency can be said to decrease. Also note that economics is replete with latent (i.e. not directly observed) variables ever since Keynes’s (1936) “state of confidence,” Friedman’s (1957) “permanent income,” and Barro’s (1977) “unanticipated money growth.” For the argument below, it is sufficient for the level of solvency to be a decision variable both for the insurer and for the regulatory authority.

In the case of insurance companies, Sommer (1996) as well as Cummins and Sommer (1996) have shown that a higher level of solvency serves to increase demand and hence premium income P. Assuming decreasing marginal returns as usual, one has

$$P = P(\cdot, S), \quad \frac{\partial}{\partial S} P(\cdot, S) > 0 \quad \text{and} \quad \frac{\partial^2}{\partial S^2} P(\cdot, S) < 0; \quad (1)$$

the arguments other than S are discussed in Section 4 below. Next, the amount of capital C > 0 increases with the solvency level S aimed at by the insurer,

$$C = C(\cdot, S) \quad \text{with} \quad \frac{\partial}{\partial S} C(\cdot, S) > 0, \quad \frac{\partial^2 C}{\partial S^2} > 0. \quad (2)$$
Note that equations (1) and (2) suffice to describe the effects a change in the solvency level has on the insurer.

For an insurer, RAROC depends on profits from two distinct activities—capital investment and risk underwriting. As to profits from investment activity, they have two components. The first is denoted by \( r_C C \) in eq. (3) below. Solvency capital \( C \) (which is equated to capital for simplicity) must predominantly be invested in gilt-edged securities (predominantly government bonds) at a rate of return \( r_C \). Note that this is not a risk-free interest rate. The financial crisis of 2007 has shown that such a rate does not really exist (see, e.g., Global Credit Research, 2010, in the case of Greek government bonds); accordingly, there is no capital market line complementing the efficiency frontier depicted in Figure 2 (see below). The second component is \( k\mu \cdot P(C,S) \), i.e., premium income carried over from the previous period (the time difference is neglected for simplicity), which is not matched by insurance claims yet. This makes funds available for investment according to the so-called funds-generating factor \( k \) (Cummins and Phillips, 2001). The higher \( k \), the longer the lag between premiums received and claims paid. These funds can be invested at the rate of return \( \mu \) prevailing on the capital market.

The insurer also derives profit from risk underwriting, which is simply given by the difference between premium income \( P(C,S) \) and losses paid \( L \). Assuming \( L \) to be exogenous and abstracting from operating costs and taxes, RAROC can therefore be expressed as follows,

\[
RORAC = \frac{r_C \cdot C(C,S) + k\mu \cdot P(C,S) + P(C,S) - L}{C(C,S)} = r_C + \frac{(1 + k\mu)P(C,S) - L}{C(C,S)} \tag{3}
\]

Maximization of RAROC (\( R \) for short) w.r.t. solvency \( S \) leads to the first-order condition (4) for optimal solvency. Here, \( e(P,S) = (\partial P/\partial S)(S/P) \) and \( e(P,C) = (\partial P/\partial C)(C/P) \) denote the elasticity of premium income and solvency capital w.r.t. the solvency level, respectively:

\[
\frac{dR}{dS} = \frac{(1 + k\mu)[\partial P/\partial S \cdot C - P \cdot \partial C/\partial S]}{C^2} =
(1 + k\mu) \left[e(P,S) \frac{P}{C} \cdot S - e(C,S) \cdot \frac{P}{C} \cdot S \right] = 0
\]

and hence

\[
\frac{dR}{dS} = e(P,S) - e(C,S) = 0. \tag{4}
\]

By eqs. (1) and (2), both elasticities are positive, justifying neglect of boundary solutions (\( S^* = 0 \) in particular).
Equation (4) can be interpreted as follows. Since \( e(P,S) > 0 \), the first term represents the marginal benefit of increased solvency in percentage terms. The Division needs to weigh this marginal return of solvency against its marginal cost, which is given by \( e(C,S) > 0 \), reflecting the capital needed for a higher solvency level.

However, the elasticities \( e(P,S) \) and \( e(P,C) \) depend not only on solvency \( S \) but also on the changing conditions on the capital market reflected by exogenous shocks \( d\bar{\mu} \) and \( d\bar{\sigma} \), respectively (see assumptions A6 and A7 of Table 1). This implies that the optimal adjustment to an exogenous change will not be given once and for all but importantly depends on parameters not yet specified, in particular the risk-return profile inherited from the past. Solvency regulation that fails to reflect this variability runs the risk of creating perverse incentives. Before substantiating this claim, however, it is worthwhile to note that regulation fixing a solvency level to be adhered to by all at all times also has its advantages. One is simplicity, although the Division may be hard put to operationalize “level of solvency” in all circumstances. Second, a fixed prescribed solvency level in fact permits separation of the insurer’s underwriting and investment policies, which again results in an important simplification of management tasks. On the downside, uniform regulation creates a similarity in the decision-making situation of regulated firms, which may result in a type of implicit collusion limiting competition. Firms now can easily predict the actions of their competitors, which are governed by regulation.

**ADJUSTMENT OF SOLVENCY TO EXOGENOUS SHOCKS**

In the first period, exogenous shocks impinging on rates of return \( d\bar{\mu} \) and volatility of returns \( d\bar{\sigma} \) occur (see Figure 1 again). As shown in the Appendix, the (sign of) optimal adjustment of the solvency level \( S^* \) to a shock in expected returns \( d\bar{\mu} > 0 \) is given by

\[
\frac{\partial^2 R}{\partial S \partial \bar{\mu}} = \frac{\partial^2 P}{\partial S \partial \bar{\mu}} \cdot S \left( \frac{S \cdot \partial P}{p^2} \right) - \frac{\partial^2 C}{\partial S \partial \bar{\mu}} \cdot S \left( \frac{S \cdot \partial C}{C^2} \right)
\]

\[
(+)(+) (+)(-)
\]

and hence

\[
\frac{dS^*}{d\bar{\mu}} \rightarrow 0 \text{ if } S \rightarrow 0;
\]
Table 1. Assumptions of the Model

A1: \[ \frac{\mu}{\sigma} = \frac{\mu^* + \mu}{\sigma + \sigma} \]
Returns and volatility \((\mu, \sigma)\) are additive in an exogenous \((\mu^*, \sigma^*)\) component determined on the capital market and an endogenous one.

A2: \[ \frac{\partial C}{\partial \mu} < 0 \]
The higher returns on the capital market, the less risk capital is needed to attain a given solvency level. A positive shock on returns makes positive net values of the company more likely, therefore reducing the need for risk capital.

A3: \[ \frac{\partial C}{\partial \sigma} > 0 \]
The higher volatility on the capital market, the more risk capital is needed to attain a given solvency level. Positive net values of the insurer are less likely, and this must be counteracted by more risk capital.

A4: \[ \frac{\partial P}{\partial \mu} < 0 \]
The (present value of) premium income depends negatively on the rate of return attainable on the capital market because policyholders now have more favorable investment alternatives.

A5: \[ \frac{\partial P}{\partial \sigma} > 0 \]
The (present value of) premium income depends positively on the volatility of returns on the capital market because the insurer now offers a comparatively safe investment alternative to risk-averse policyholders.

A6: \[ \frac{\partial^2 C}{\partial S \partial \mu} < 0 \]
A higher solvency level calls for more risk capital but to a lesser degree if higher market returns prevail, making positive net values of the company more likely.

A7: \[ \frac{\partial^2 C}{\partial S \partial \sigma} > 0 \]
A higher solvency level calls for more risk capital, especially when market volatility is high, making positive net values of the company less likely.

A8: \[ \frac{\partial^2 P}{\partial S \partial \mu} > 0 \]
While a higher rate of return on the capital market depresses premium income (see A4), this effect weakens if the insurer offers a high level of solvency.

A9: \[ \frac{\partial^2 P}{\partial S \partial \sigma} > 0 \]
Higher volatility on the capital market serves to increase premium income (see A5); this effect is reinforced if the solvency level is increased.

\[
\frac{dS^*}{d\mu} > 0 \text{ if } C \text{ large, since } C^2 \to \infty \text{ faster than } C \to \infty; \quad (5)
\]
\[
\frac{dS^*}{d\mu} < 0 \text{ otherwise.}
\]
These results are intuitive. In a situation where the solvency level is very low to begin with, its adjustment in response to increased returns in the capital market does not matter. However, with solvency capital C large, the opportunity cost of an increased solvency level is small, leading the Division to propose such an increase to senior management with the aim of boosting premium income. Since most insurers have excessive solvency capital (Nakada et al., 1999), \( dS^*/d\mu > 0 \) is considered the normal response.

Now consider a shock \( d\sigma > 0 \) (again, details are given in the Appendix), implying \( +(-)(-)(+) \) if \( C \) small since faster than \( C \to 0 \);

\[
dS^*/d\sigma > 0 \quad \text{if } C \text{ small since } C^2 \to 0 \quad \text{faster than } C \to 0 ;
\]

\[
dS^*/d\sigma < 0 \quad \text{if } P \text{ small since } P^2 \to 0 \quad \text{faster than } P \to 0 ;
\]

\[
dS^*/d\sigma < \quad \text{otherwise.}
\]

Again, these predictions are intuitive. If solvency is very low initially, adjusting it to an exogenous increase in the volatility of returns makes no difference. In a situation where solvency capital is scarce (\( C \text{ small} \)), it makes sense to increase the solvency level in order to boost premium income, which in turn contributes to RAROC. On the other hand, when premium income is quite low to begin with, increasing it through higher solvency has little leverage; therefore, preserving scarce capital takes precedence, suggesting a decrease in solvency.

**DETERMINATION OF THE ENDOGENOUS EFFICIENCY FRONTIER**

In the third period, the insurer inherits a net adjustment of solvency \( dS^* \) from the second period, \( dS^* \) being the result of responses to the shocks \( (d\mu, d\sigma) \) that occurred in the first period. The Division now proceeds to adjust \( \mu \) and \( \sigma \), the endogenous components of \( \mu \) and \( \sigma \), respectively. Optimal adjustments are described by eqs. (5) and (6), with \( dS^* \) assuming
the role of an exogenous shock. Therefore, comparative statics can now be performed in reverse to derive optimal endogenous adjustments \(d\hat{\mu}/d\bar{\sigma}\) and \(d\hat{\sigma}/dS\), respectively. The Division effects these changes by reshuffling assets and liabilities, creating an endogenous efficiency frontier with slope \(d\hat{\mu}/d\bar{\sigma}\). Senior management then chooses the optimum on this frontier (see below). The slope of this frontier can be obtained by dividing (6) by (5), yielding

\[
\left. \frac{d\hat{\mu}}{d\bar{\sigma}} \right|_{S^*} = \frac{\frac{\partial^2 P}{\partial S \partial \bar{\sigma}} \cdot \frac{S}{P} + \frac{\partial P}{\partial S} \left( \frac{-S \cdot \frac{\partial P}{\partial \bar{\sigma}}}{p^2} \right) - \frac{\partial^2 C}{\partial S \partial \bar{\mu}} \cdot \frac{S}{C} - \frac{\partial P}{\partial S} \left( \frac{-S \cdot \frac{\partial C}{\partial \bar{\mu}}}{C^2} \right) }{\frac{\partial^2 P}{\partial S \partial \bar{\mu}} \cdot \frac{S}{P} + \frac{\partial P}{\partial S} \left( \frac{-S \cdot \frac{\partial P}{\partial \bar{\mu}}}{p^2} \right) - \frac{\partial^2 C}{\partial S \partial \bar{\mu}} \cdot \frac{S}{C} - \frac{\partial P}{\partial S} \left( \frac{-S \cdot \frac{\partial C}{\partial \bar{\mu}}}{C^2} \right) } > 0. \tag{7}
\]

In principle, the sign of eq. (7) is indeterminate, even if normally its denominator is positive [see the comment below eq. (5)]. However, daily experience of investors in the capital market is that the slope of the efficiency frontier in \((\mu,\sigma)\)-space is positive. Therefore, a positive sign is assumed for eq. (7) in the following. Since the normal response makes its denominator positive, its numerator must be positive as well, an implication that will be of relevance in the section below. A crucial result to be noted already at this point is that the slope defined in eq. (7) depends not only on easily observable parameters such as \(C, P, S\) and first-order effects the regulator likely is aware of (such as \(\partial P/\partial \bar{\mu}, \partial P/\partial \bar{\sigma}\)) but also on terms such as \(\partial^2 P/\partial S \partial \bar{\mu}\) and \(\partial^2 P/\partial S \partial \bar{\sigma}\), which indicate that the relationship between premium income and solvency depends on conditions on the capital market (see assumptions A8 and A9 again).

Figure 2 shows three endogenous efficiency frontiers (minimum variance points are not shown to preserve space). Note that \(\mu\) and \(\bar{\mu}\) as well as \(\sigma\) and \(\bar{\sigma}\) are depicted on the same axis, reflecting the assumption that, e.g., a low first-period value of \(\bar{\sigma}\) tends to translate into a low third-period \(\sigma\). The first frontier (labeled \(S^*\)) holds prior to the influence of regulation. The two other frontiers (labeled I and II, respectively) are modified by Solvency I and Solvency II regulation in ways to be discussed in the section below.

**Conclusion 1:** Due to its responses to shocks in expected rate of return and volatility in the process of sequential adjustment, the underwriting/investment Division of the insurer induces an endogenous efficient frontier, whose slope is not stable. Rather, it varies under the influence of shocks
from the capital market that modify, notably, the relationship between solvency and premium income.

At least in the case of banks, there is some historical evidence supporting this conclusion. After adjusting for hidden reserves, Billings and Capie (2007) find that capital-asset ratios of the five major UK banks were particularly high during 1942–1946, when they had to finance the war effort. Their explanation is “that much of their lending to government was in the form of marketable securities, which have generated exposure to fluctuations in market prices” (p. 152, emphasis added). Banks apparently realized that their efficiency frontiers depended on changes in the relationship between solvency and risk capital due to changing market conditions during the Second World War. A similar awareness might well characterize the senior management of insurance companies, which is usually more prudent than that of banks.

**EFFECTS OF SOLVENCY REGULATION ON THE EFFICIENCY FRONTIER**

*Solvency I*

*Solvency I* stipulates capital requirements as a function of risk-weighted assets and separately for off-balance sheet positions in the same...
way as for banks (Basel Committee on Banking Supervision, 1988). Its focus is on the relationship between solvency and capital. By defining four asset classes with fixed weights, Solvency I imposes a fixed relationship between solvency capital $C$ and solvency $S$ (see the locus $B$ of Figure 3 below). It therefore does not allow insurers to react to changes in market conditions affecting the risk characteristics of assets. In terms of the model, this neglect amounts to the restrictions

$$
\frac{\partial^2 C}{\partial S \partial \mu} = 0, \quad \frac{\partial^2 C}{\partial S \partial \sigma} = 0.
$$

(8)

Therefore, both the relationship between solvency and risk capital and that between solvency and premium income are seen as being independent of conditions prevailing in the capital market, in contradistinction with assumptions A6 to A9 (see Table 1 again). Inserting these restrictions in eq. (7), one immediately sees that the numerator increases. Concerning the denominator, the deletion of $\partial^2 C / (\partial S \partial \mu)$ causes a decrease in its value. The restrictions (8) thus result in a steeper slope of the endogenous efficiency frontier (subscript $I$ denoting Solvency I),

$$
\left. \frac{d \hat{\mu}}{d \sigma} \right|_{I} > \left. \frac{d \hat{\mu}}{d \sigma} \right|_{S^*}.
$$

(9)

In Figure 2, the Solvency I frontier therefore runs steeper than the original $S^*$ frontier, approaching but never crossing it for high values of $\mu$ because regulation cannot increase the insurer’s feasible set.

One might argue that the insurer can choose to act in accordance with parameters it knows to be of importance, contrary to the regulator’s decision rule. This would amount to neglecting the restrictions stated in (8). However, as emphasized by Power (2004: ch. 7), managers are responsibility-averse, leading them to use regulatory decision rules as a convenient justification of their own actions. For example, let there be a second-period upward adjustment in solvency indicating that the insurer should move away from the origin on the efficiency frontier. With the flat endogenous efficiency frontier $S^*$ of Figure 2 in view, the Division would propose to accept a substantial increase in volatility, whereas based on the steeper efficiency frontier induced by Solvency I, the suggested increase is smaller. If the insurer’s senior management were to move along $S^*$, it could be criticized by the regulator for taking on an excessive amount of risk. This threat causes a responsibility-averse management to adopt the restrictions (8), accepting the steeper Solvency I efficiency frontier as the relevant one.

For predicting optimal solutions, one needs two assumptions regarding the preferences of insurer’s senior management. The first is risk aven-
sion on the part of the insurer’s senior management. In principle, fully diversified owners would like management to accept a great deal of risk because this allows them to benefit from an increase in the value of their shares (which have properties of a call option; see e.g., Zweifel and Eisen, 2012, ch. 4.3). Managers however, being less than perfectly diversified, with their human capital tied to the employing firm, have reason to be risk-averse. Given imperfect governance by (dispersed) ownership, management can follow its preferences at least to some extent (Shrieves and Dahl, 1992). Second, homotheticity of risk preferences is imposed in order to obtain sharper predictions. Under these assumptions, Solvency I regulation induces the insurer to be less conservative regardless of degree of risk aversion (types A and B in Figure 2; see the movements from $\sigma^*_{S^*}$ to $\sigma^*_i$ and from $\sigma^*_{S^*}$, to $\sigma^*_{I^*}$, respectively).

**Conclusion 2:** Regulation of the Solvency I type is predicted to induce insurers to take a more risky position than they would on their own, thus having an effect running counter to stated regulatory objectives.

One could argue that according to Figure 1, this increase in volatility goes along with an increase in expected returns; therefore, solvency as defined by VaR or EVaR need not diminish. Using the lognormal distribution of returns as an approximation, an increase of $\sigma$ by $x$ percent would have to be associated with an increase in $\mu$ by $x$ percent for VaR and EVaR to remain constant, implying unitary elasticity (in analogy to Cummins and Nye, 1981). In Figure 1, the corresponding locus is a straight line through the origin with a 45 degree slope. However, much lower slopes (typically below 0.5) have been found in empirical research (see, e.g., Woehrmann et al., 2004). Therefore, the predicted increase in $\sigma$ indeed causes a reduction in solvency.

Note that these predictions hold even if the regulation-induced downward shift of the efficiency frontier is minimal (see Figure 2 again). The crucial point is that Solvency I signals to insurers that interaction parameters their underwriting departments would take into account can be neglected, causing their perceived efficiency frontier to indicate that more return can be achieved on expectation for accepting a given increment in volatility. As an example, consider a company heavily engaged in the financing of mortgages (a low-risk asset according to Solvency I regulation). When expected rates of return in the capital market increase ($d\mu > 0$), it can free risk capital (slightly decreasing its effective solvency level and premium income), invest it in a more risky asset, and derive (substantially) more investment income without violating the Solvency I norm.
Solvency II

Solvency II allows a choice of approach for the calculation of capital requirements, viz. the Standardized Approach and the Internal Ratings–Based Approach (Basel Committee on Insurance Supervision, 2004). Whilst the first approach is based on Solvency I, the second lets insurers choose their probability of default, their percentage loss at default, and the maturity of their credits. Large institutions with average and below-average credit risks mostly prefer the Internal Ratings–Based approach to save on capital despite its higher cost of implementation. In terms of the model, Solvency II permits these insurers to take all elements of eq. (7) into account, which amounts to a lifting of these restrictions as long as the constraint regarding the solvency level is not binding.

To show the effect of Solvency II, assume that the insurer has opted for the more flexible Internal Ratings–Based approach. Taken together, the rules promulgated by the Basel Committee on Insurance Supervision (2004, especially para. 40 to 44) establish a relationship between a targeted solvency level and required risk capital. This relationship (which the insurer cannot modify once it has selected its internal model) is depicted by the locus labeled $F$ in Figure 3. It has increasing slope to reflect decreasing marginal returns to risk capital as stated in eq. (2). Locus $F$ runs below locus $B$ of Solvency I, reflecting that Solvency II permits insurers to save on risk capital. The other difference is that Solvency II imposes a minimum degree of solvency, which is denoted by $\bar{S}_{II}$. 

![Fig. 3. Implications of Solvency II regulation.](image)
Now let a shock $d\tilde{\sigma} > 0$ occur (volatility of returns has increased). In keeping with assumption A7, this corresponds to a steepening of the locus, resulting in $F'$ from the insurer’s point of view. It indicates that a given capital $\bar{C}$ would now only suffice to guarantee a solvency level $S < \bar{S}_{II}$. Therefore, in order to keep to the Solvency II norm, a company that just satisfied it initially would have to come up with the full additional amount of capital $\bar{C}' - \bar{C}$. Absent Solvency II, the company might opt for a point such as $Q$ that entails a somewhat lower solvency level (compared to $\bar{S}_{II}$) in return for a substantial saving of costly risk capital.

An insurer with excess solvency, symbolized by the combination $(S^+, C^+)$, would not have to react to the shock $d\tilde{\sigma} > 0$. The same conditional responses are predicted for a shock $d\tilde{\mu} < 0$, i.e., a drop in the expected return on investments.

Conversely, consider a shock $d\tilde{\sigma} < 0$, i.e. capital markets have become less volatile. For the insurer, this causes the locus $F$ of Figure 2 to become flatter, such as $F''$. Now $\bar{C}'' < \bar{C}$ suffices to reach the prescribed solvency level, and the “marginal” insurer that was at $\bar{S}_{II}$ initially can reduce capital by as much as $\bar{C} - \bar{C}''$. This of course holds true of $d\tilde{\mu} > 0$ as well.

In sum, one has the following set of conditional predictions (in absolute value) for Solvency II, focusing on the critical changes $d\tilde{\mu} < 0$ and $d\tilde{\sigma} > 0$,

$$0_{\ell} < \left| \frac{\partial^2 C}{\partial S \partial \tilde{\mu}} \right|_{II} \left| \frac{\partial^2 C}{\partial S \partial \tilde{\mu}} \right|_{S^*} \quad \text{if } d\tilde{\mu} < 0 \text{ and } S = \bar{S}_{II};$$

$$0_{\ell} < \left| \frac{\partial^2 C}{\partial S \partial \tilde{\sigma}} \right|_{II} \left| \frac{\partial^2 C}{\partial S \partial \tilde{\sigma}} \right|_{S^*} \quad \text{if } d\tilde{\sigma} > 0 \text{ and } S = \bar{S}_{II}. \quad (10)$$

Here, $0_{\ell}$ symbolizes the zero restrictions imposed by Solvency I [see eq. (8) again]. Applied to eq. (7) and in view of assumptions A6 and A7, restrictions (10) cause the numerator to increase and the denominator to decrease. One therefore obtains,

$$\left| \frac{d\tilde{\mu}}{d\tilde{\sigma}} \right|_{II} > \left| \frac{d\tilde{\mu}}{d\tilde{\sigma}} \right|_{S^*} \quad \text{if } d\tilde{\sigma} > 0 \text{ or } d\tilde{\mu} < 0 \text{ and } S = \bar{S}_{II}. \quad (11)$$

Solvency II being less stringent (at least by intent), the frontier runs higher than that of Solvency I but still lower than absent regulation (see Figure 2 again). To make up for reduced expected returns, even a strongly risk-averse senior management (preferences of type $A$) is predicted to opt for a more risky allocation ($\sigma^*_{II} > \sigma^*_{S^*}$) provided the insurer just satisfied the Solvency II norm initially. This condition presumably holds as a rule for those companies with a less risk-averse management (preferences of type
B), again resulting in an investment policy that entails a higher volatility of returns than without regulation. In comparison with Solvency I, these counterproductive effects are less pronounced, since Solvency II causes a smaller downward shift of the efficiency frontier (see Figure 2 again).

In sum, Solvency I and Solvency II are predicted to have similar effects in one respect. Both may induce insurers to opt for a more rather than less risky exposure than if they were optimizing free of the respective restraints. However, the two regulations differ in another respect. Solvency I causes a "deformation" of the \((\mu, \sigma)\)-frontier for all insurers. By way of contrast, the "deformation effect" of Solvency II is limited to the subset of insurers who just satisfied the norm initially.

**Conclusion 3:** At least for insurers just compliant initially with the solvency norm, Solvency II may still cause insurers to pursue a riskier underwriting and investment policy than absent regulation, but less risky than under Solvency I.

**Solvency III**

The details of implementation of Solvency III regulation are not known yet at the time of writing. However, its objective clearly is to prescribe a higher level of solvency, to be attained by more solvency capital of which a greater part is to be equity (Bank for International Settlements, 2011; Basel Committee on Insurance Supervision, 2011). In terms of Figure 3, the mandated solvency level shifts toward \(S^*\) or even beyond. The consequence of this shift is that the set of insurers that does not need to react to a shock \(d\bar{\sigma} > 0\) shrinks while the set of insurers that have to come up with the full additional amount of capital to meet Solvency III requirements increases (see the discussion in the preceding section). Therefore, the steepening of the efficiency frontier predicted in (10) applies to a larger subset of insurers, at least during a (lengthy) time of transition that sees insurers struggling to increase their equity and reserves. Moreover, the fact that an increased share of solvency capital must be equity means an increase in regulatory stringency, causing the endogenous efficiency frontier pertaining to Solvency II to be shifted back down towards that of Solvency I (see Figure 2 again). Conclusion 3 therefore is predicted to hold more generally, implying that more insurers may in fact be induced to pursue a riskier investment policy than in the absence of regulation.

**SUMMARY AND CONCLUSION**

The basic hypothesis of this paper states that insurers’ underwriting and investment departments (combined into one “Division”) seek to attain
Solvency regulation of insurers: A regulatory failure?

153

a solvency level that balances the advantage of lower refinancing cost against the disadvantage of tying capital that would yield higher returns in other uses. However, this solvency level is too low from a societal point of view because it neglects the fact that insolvency causes substantial external costs. This analysis proceeds to assume that the Divisions maximize the risk-adjusted rate of return on capital \( (RAROC) \). For them, the marginal benefit of a higher solvency level is the lowered cost of refinancing, while its marginal cost consists of the return forgone on the extra capital allocated. Divisions establish the slope of their efficiency frontier in \((\mu,\sigma)\)-space in the course of three periods. In period 1, two shocks occur, viz., an exogenous change in expected returns \( (d\mu) \) and in their volatility \( (d\sigma) \). These shocks induce lagged adjustments \( (dS^*/d\mu, \; dS^*/d\sigma) \) in period 2. Net adjustment \( dS^* \) then triggers a reallocation of assets and liabilities and hence endogenous changes \( d\mu \) and \( d\sigma \) during period 3. This implies a perceived endogenous frontier in \((\mu,\sigma)\)-space prior to solvency regulation, with slope \( d\mu/d\sigma \). This slope is not a constant but depends importantly on the fact that the relationship between risk capital and solvency is modified by exogenous changes in expected returns and volatility occurring in the capital market (Conclusion 1). The regulations imposed by Solvency I are now shown to neglect these influences, causing a modification of the risk-return frontier as perceived by the regulated insurer. This modification is predicted to induce senior management to take a more risky position than it would absent regulation, resulting in a decrease of solvency as measured by (expected) value at risk (Conclusion 2). The implications of Solvency II are slightly more complex. Still, companies initially just attaining the prescribed solvency level are again predicted to react to regulation by taking a more risky position than they would have otherwise (Conclusion 3). As to Solvency III, its likely effect will be to increase the subset of insurers responding in the same way, running counter the very objective of the regulators, who want insurers to take on less rather than more risk.

All of these predicted adjustments may be argued to amount to regulatory failures. However, it would be inappropriate to conclude that Solvency I, II, or III or even solvency regulation in general should be revoked. First, the model analyzed in this paper might be too simplistic; insurers possibly pursue other objectives than just maximizing RAROC. Second, Solvency II already constitutes an improvement over Solvency I in that its self-defeating effect is limited to the (usually small) subset of insurers that initially had been just compliant with the prescribed solvency level. As to Solvency III, the short-run increase in this problematic subset effect has to
be weighed against the long-run increase in solvency levels generally achieved. Finally, assuming that solvency regulation does entail more benefit (in terms of external cost avoided) than cost (in terms of biasing insurers’ tradeoffs between $\mu$ and $\sigma$), one would have to find an alternative whose benefit-cost ratio beats that of the Solvency I to III type. According to this paper, one way to achieve this would be for the regulation to more closely mimic the decision-making situation of senior insurance managers. For the time being, it seems worthwhile to call attention to likely shortcomings of current and planned future solvency regulation of insurance companies.

REFERENCES


First, consider a shock $d\ddot{\mu}$ disturbing the first-order condition (4). With $R$ shorthand for RAROC, the comparative static equation reads,
\[
\frac{\partial^2 R}{\partial S^2} dS^* + \frac{\partial^2 R}{\partial S \partial \ddot{\mu}} d\ddot{\mu} = 0. \tag{A.1}
\]
Since $\frac{\partial^2 R}{\partial S^2} < 0$ in the neighborhood of a maximum, $\text{sgn} \left[ \frac{\partial^2 R}{\partial S \partial \ddot{\mu}} \right]$ determines $\text{sgn}[dS^*/d\ddot{\mu}]$. Differentiating eq. (4) w.r.t. $\ddot{\mu}$, one has
\[
\frac{\partial^2 R}{\partial S \partial \ddot{\mu}} = \frac{\partial \left( e(P,S) - e(C,S) \right)}{\partial \ddot{\mu}} = \frac{\partial}{\partial \ddot{\mu}} \left[ \frac{\partial P}{\partial S} \cdot \frac{S}{P} \right] - \frac{\partial C}{\partial \ddot{\mu}} \left( \frac{S}{C} \right).
\]
\[
= \frac{\partial^2 P}{\partial S \partial \ddot{\mu}} \cdot \frac{S}{P} + \frac{\partial P}{\partial S} \left( -\frac{S \cdot \partial P/\partial \ddot{\mu}}{p^2} \right) - \frac{\partial^2 C}{\partial S \partial \ddot{\mu}} \cdot \frac{S}{C} - \frac{\partial P}{\partial S} \left( -\frac{S \cdot \partial C/\partial \ddot{\mu}}{C^2} \right). \tag{A.2}
\]
The signs are based on assumptions A9, A4, and A2 as well as eqs. (1) and (2). This is expression (5) of the text.

Now consider $d\ddot{\sigma} > 0$. In full analogy to (A.1), one obtains from eq. (4),
\[
\frac{\partial^2 R}{\partial S \partial \ddot{\sigma}} = \frac{\partial \left( e(P,S) - e(C,S) \right)}{\partial \ddot{\sigma}} = \frac{\partial}{\partial \ddot{\sigma}} \left[ \frac{\partial P}{\partial S} \cdot \frac{S}{P} \right] - \frac{\partial C}{\partial \ddot{\sigma}} \left( \frac{S}{C} \right).
\]
\[
= \frac{\partial^2 P}{\partial S \partial \ddot{\sigma}} \cdot \frac{S}{P} + \frac{\partial P}{\partial S} \left( -\frac{S \cdot \partial P/\partial \ddot{\sigma}}{p^2} \right) - \frac{\partial^2 C}{\partial S \partial \ddot{\sigma}} \cdot \frac{S}{C} - \frac{\partial P}{\partial S} \left( -\frac{S \cdot \partial C/\partial \ddot{\sigma}}{C^2} \right). \tag{A.3}
\]
The signs are based on assumptions A9, A5, and A3 as well as eqs. (1) and (2) of the text. This is expression (6) of the text.