TO WHAT EXTENT DOES A CONFLICT EXIST BETWEEN RISK MANAGEMENT THAT BENEFITS REGULATORS COMPARED TO RISK MANAGEMENT THAT BENEFITS SHAREHOLDERS?

CON KEATING
FINANCE DEVELOPMENT CENTRE
LONDON

JANUARY 2007

INTRODUCTION

The question posed raises the question of why regulation may be originally justified, given that unconstrained markets left to their own devices will usually produce the welfare optimal result. The case for regulation depends upon the uninhibited actions of the private sector resulting in market failure or more relevantly to this paper, sup-optimal results which are worse than the results of regulation. Three main strands of justification for regulation emerge from the academic literature:

1 Protection from monopolistic exploitation
2 Protection of the less well-informed (usually retail)
3 Assurance of systemic stability

There is an enormous literature which examines regulation and banking where all three justifications may be applicable but insurance, and particularly life companies, differ from banks in some significant ways. Insurance companies exist as institutions precisely to write contracts with consumers which capital markets alone cannot credibly write – for instance promising to pay individual-specific pensions at many dates far into the future. Futures markets exist because an institution, the clearing house, lends credibility to enforcement of the contract.

Long term contracts with a consumer which may be dependent upon the behaviour of the insurance contract after their agreement; information asymmetries and agency problems are the justification for prudential regulation. The situation with insurance companies differs in many important regards from the situation with banks.

1 Insurance companies are not susceptible to “runs”
2 The inter-institution linkages necessary for contagion are largely absent
3 There is no direct involvement in the payments system
4 Their assets are predominantly marketable

It is important to realise that the activity undertaken by an insurance company is fundamentally different from that of a bank; an insurance company transforms illiquid liabilities into liquid assets while a bank transforms liquid liabilities into illiquid assets. The case for systemic risk with insurers must lie with the existence of bancassurers and group reputational effects, or with the interaction with asset markets. In fact the predominant stream of arguments for an asset market channel to systemic risk lies in arguments that markets are imperfect. Predatory pricing of assets, given the knowledge that an insurer has suffered a substantial loss and requires cash, is perhaps the classic exemplar here. This paper is overwhelmingly concerned with prudential rather than systemic regulation. It also uses the term regulation to include supervision.

FORM OF REGULATION

The predominant form of regulation of insurance companies today is risk-based regulation of the institution, though other forms of regulation, such as product regulation, may be more effective and efficient in resolving the informational concerns which justify regulation in the first place. The overwhelming majority of academic discussion is concerned with the precise risk measures applied to the institution, but there is an emergent discussion of the effects of regulation upon the economy or asset markets.

In the simple perfect markets world of Markowitz's mean-variance asset selection the existence of regulatory constraints can be shown to affect portfolios selected, which should surprise no-one as this was the intended effect of the regulation. It can be demonstrated that under an assumption of Normality of returns the Mean-Value at Risk (VaR) efficient set is a proper sub-set[1] of the efficient frontier, though some care is advised with the choice of VaR quantile, α, of the risk measure $\alpha\sigma - \mu$. The measure is broadly consistent with the expected utility framework[2], and results generalize to the class of elliptic distributions. The optimal portfolios of the agent are, however, now dependent upon their level of risk aversion, and the effect of the constraint for agents

[1] This may be an empty subset.
[2] See Basak & Shapiro 1999

with low risk aversion can be an increase, no change, or decrease in the variance of the portfolio selected. A highly risk averse agent will increase the variance of the portfolio selected. If the VaR confidence level is sufficiently small, the subset may be empty, which means that equilibrium does not exist and requires market prices to adjust. In the presence of a risk-free asset Tobin's two fund separation theorem will, though with some constraint on the VaR quantile chosen, still hold. These results extend to the expected shortfall (ES) criterion and the selection results may be illustrated by the diagram below:

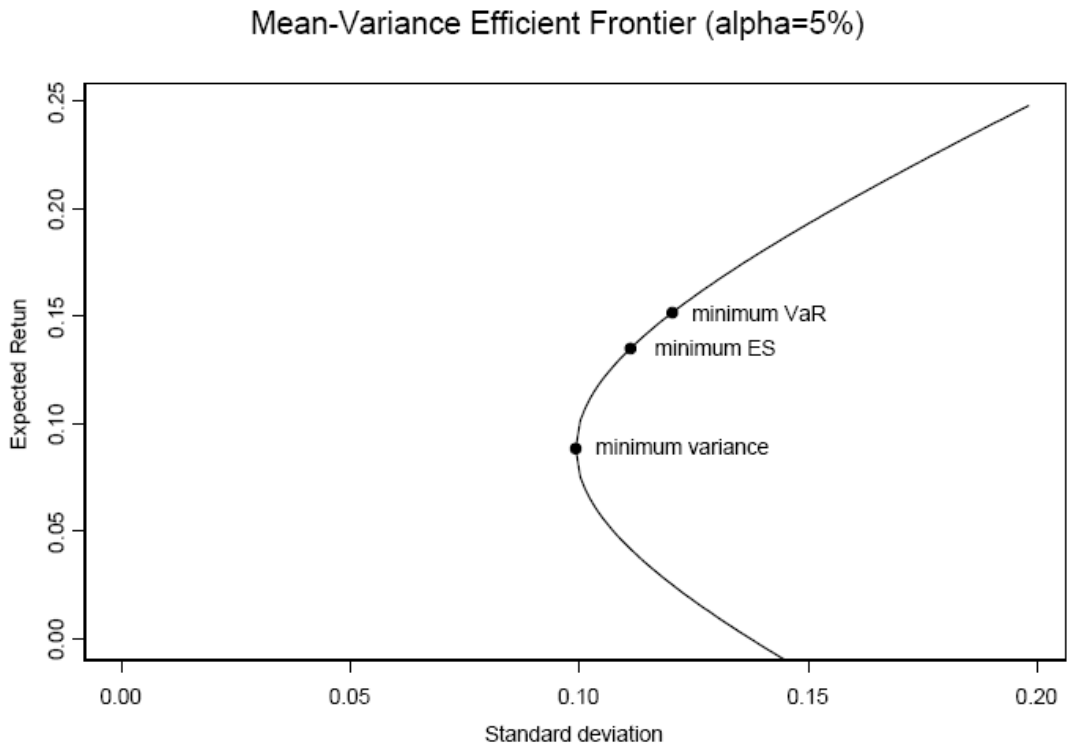

Figure 1: From E Giorgi " A Note on Portfolio Selection under Various Risk Measures" We see that under $VaR_{5\%}$ and $ES_{5\%}$ the set of efficient portfolios is reduced with respect to the variance.

The portfolio selection implications of a VaR constraint are illustrated below as Figure 2. Of course regulators currently utilize more quantitative tools than just the simple VaR constraint; prominent among these is the concept of stress-testing. The effect of a stress-test has been examined in the context of the stress-test as a trading limit in a paper[3] by

[3] Portfolio Management with Stress-Testing: Implications for Portfolio Selection and Asset Pricing

Alexander and Baptista, which incidentally won the PRMIA best risk management paper award for 2006.

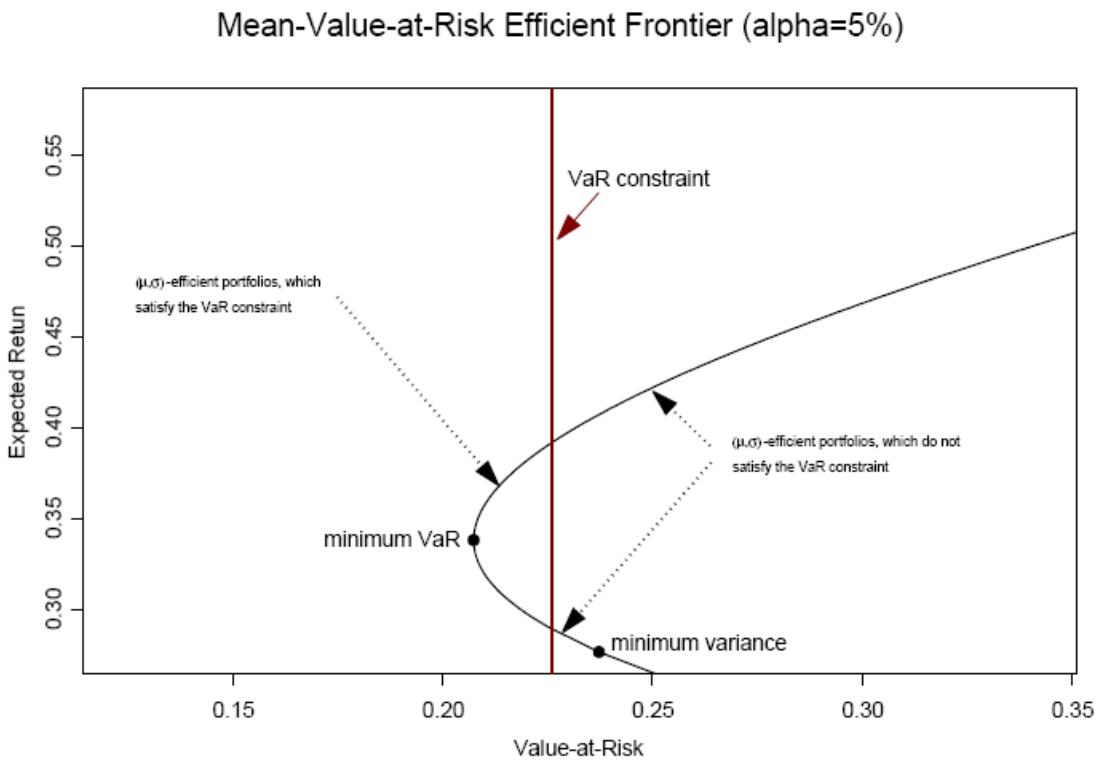

Figure 2: From E Giorgi " A Note on Portfolio Selection under Various Risk Measures". $(\mu, \text{VaR}_{5\%})$ boundary with the global minimum variance portfolio. Portfolios on the $(\mu, \text{VaR}_{5\%})$ boundary between the global minimum $\text{VaR}_{5\%}$ portfolio and the global minimum variance portfolio are mean-variance efficient. The VaR constraint (vertical line) could force mean-variance agents with high variance to reduce the variance but mean-variance investors with low variance to increase the variance to lie to the left of the VaR constraint.

When faced with K stress test constraints, and the same basic Markowitz set-up[4], a constrained agent's portfolio now exhibits K+2 fund separation regardless of whether there is a risk-free asset. Secondly the effect is identical to the effect of lowering the returns of risky securities and results in higher allocations to the risk-free or minimum variance security. Thirdly the market portfolio is now inefficient as only two of the four portfolios are mean-variance efficient. And perhaps most importantly an asset's return is

[4] It is worth noting that the presence of a binding stress-test constraint is not consistent with expected utility maximisation and that other utility functions may be more appropriate.

affected by both its systematic risk and by its idiosyncratic risk in those states for which the stress-test binds. In this circumstance security prices are no longer driven solely by systematic risk. (See also Box 1)

Box 1

This stress-test result extends in a straightforward manner to benchmark relative portfolio management with the single constraint being a tracking error. Before expanding on that it is worth touching on one additional aspect of benchmark relative management. The market standard tracking error calculation uses the standard deviation of the difference portfolio returns. This is an appropriate measure if, as was originally the case, the objective was index replication, but if added value is of interest it becomes necessary to use the correct more general measure which includes the expected return.

$$\text{Tracking Error} = \sqrt{\mu_{\text{diff}}^2 + \sigma_{\text{diff}}^2}$$, with obvious notation.

One consequence of this more general tracking error definition is that it places a bound on the expected investment mandate return, in that the tracking error admitted must be greater than the expected return other than in the deterministic case – and interestingly infeasible mandates with respect to this bound have been reported.

When considering single tracking error constrained portfolio management three fund separation is observed with the third fund being the benchmark. Ordinarily this would not be an overwhelming problem since we can approximate the benchmark asset portfolio arbitrarily closely. However it presents a significant problem if we are to use a liability benchmark which is not well defined in terms of replicating assets[5]. To manage to this liability portfolio requires investment in that liability portfolio, which is not feasible. In the limit, in this frictionless world of elementary finance, if we wish to hedge liabilities perfectly we need a portfolio position which is exactly those liabilities.

Asset and liability management is uninteresting unless we view this from the standpoint of asset approximations to the liabilities or other sources of imperfection in the model.

[5] If the liabilities were exactly replicable in terms of capital market assets there would be no ex ante reason for the institution creating them to exist.

It is perhaps surprising that the elementary model of financial theory yields such a richness of results and this should serve as a caution to regulators seeking to determine intervention practice. Of course there is still a major dislocation between theoretical equilibrium asset pricing models and the real world. Markets are neither complete nor frictionless and as a result these standard methods may be inappropriate. These equilibrium models have a further difficulty from the perspective of this paper in that they fail to provide micro-economic foundations for the risk being regulated. Modeling asset prices in a first best world economy is modeling in an economy where regulation is unjustified. The exogenous nature of the risk in these models may be useful from the standpoint of simplicity and tractability but in the absence of a market failure it is not clear why this required regulation in the first place.

Danielsson and Zigrand[6] develop a constrained equilibrium model which does provides some micro-foundation (see below), where risk arises from an externality and warrants regulation.

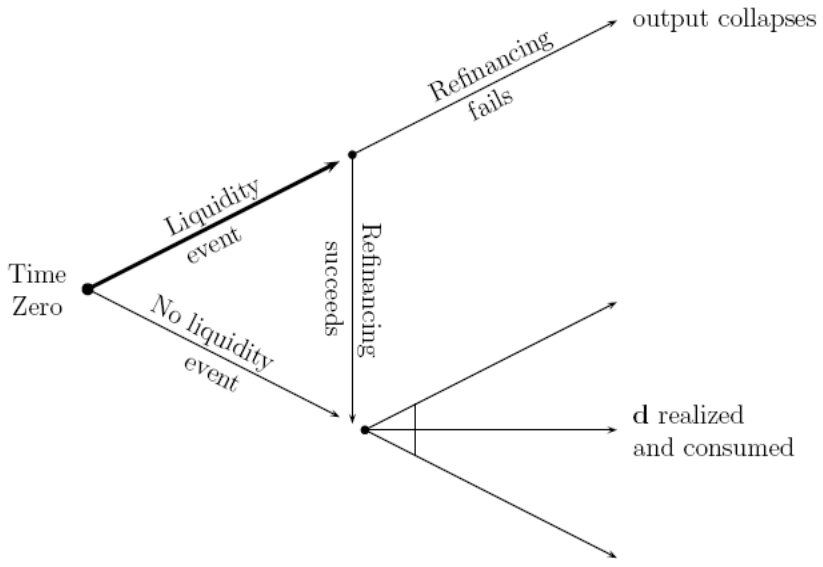

Figure 3: From Danielsson & Zigrand "Equilibrium Asset Pricing with Systemic Risk"

The model used further develops the feed-back effects of regulation on asset prices. The results of this study are that risk based regulation does mitigate risk reducing the

[6] Equilibrium Asset Pricing with Systemic Risk LSE working paper May 2006

probability of the systemic event and alleviating some of the free-riding externalities. But with risk pricing endogenous to the model and now subject to the Lucas critique, there are costs. Markets may again not clear if risk regulation is too all-encompassing. The statistical properties of risk change, and equity risk premia, illiquidity, volatility and co-variations all increase.

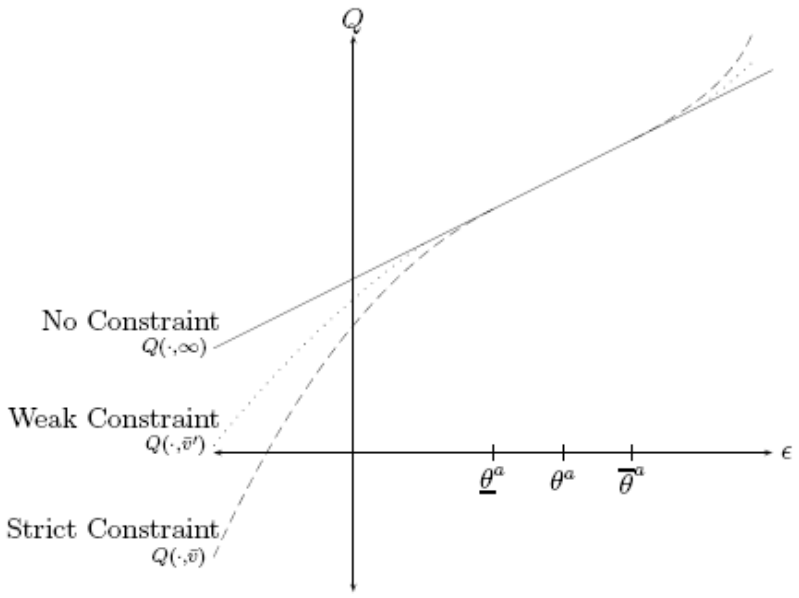

Figure 4: PRICING FUNCTION

The pricing function without constraints and with increasingly binding constraints, $\infty > \bar{v}' > \bar{v}$. The downside effects become more pronounced as the constraint becomes stricter.

Figure4: From Danielsson & Zigrand "Equilibrium Asset Pricing with Systemic Risk"
The x-axis thetas are market portfolios of risky assets.

Risk Measures

It is usual to place the origins of VaR in the techniques known as RiskMetrics advanced by J.P. Morgan in the early 1990s but it can in fact be traced back to 1963 and Baumol's expected gain – confidence limit criterion. Somewhat more contentiously it might be said to be a long held central tenet of actuarial Ruin Theory.

A number of studies[7] report that a VaR constraint may, in simple cases, be effective in lowering the riskiness of bank asset portfolios though questions remain over the incentives for bank management to report VaR truthfully and the possibility of substitution of high return (and risk) assets for low return assets or regulatory arbitrage more generally[8]. This latter possibility was one of the principal reasons cited for the move from Basel I to more complex regulation for banks. One obvious question must be whether these results may extend to insurance companies where the concern is failure. There have been a number of academic studies[9] of the predictive power of these forms of risk measure, as applied by insurance regulators, which have concluded that this is low. This may demonstrate a practical limit to statistical methods for regulation arising from the small sample of insurance failures.

It is also evident from the simplest ruin probability models that in addition to or substitution for capital requirements (C), an insurer or regulator might use other determinants of ruin probabilities as a VaR control mechanism. Increasing the number of insurance contracts written (N), accepting less volatile loss risks (σ), and raising the loading factor (λ) will all serve to lower ruin probabilities.

$$P_{ruin} \leq \frac{1}{\beta^2}, \qquad \text{where } \beta = \frac{C + \lambda N}{\sigma\sqrt{N}}.$$

The beta shown above is sometimes known as the coefficient of security.

Regulation of risk by control of the loading factor or magnitude of their volatility is product rather than institutional regulation. From the perspective of the regulator though this is problematic; such regulation makes directly evident its contribution to costs in the price of the product offered to the public. The agency problem emerging here, between regulator and consumer, is a first indication that regulatory systems may not be entirely benign in their design.

---

[7] Cuocco and Liu
[8] Kim & Santomero (1989 ?)
[9] Grace et al, Cummins et al

Control of the demand for a product, the aggregate number of insurance contracts, is not desirable and arguably not feasible for the regulator; this is best left to the supply and demand functions in a free market. Of course, whichever of the control variables is chosen demand is affected; we may expect fewer products of higher quality to be produced as result of effective regulation and sold at higher prices. The problem from a regulatory standpoint is that the use of the number of contracts written, the market share of a firm, provides incentives for firms to grow and merge to increase volume and diversify risk; it affects risk in distribution. Fewer firms ultimately will carry consequences for the competitiveness of the market for its products.

It is interesting to note that regulators in the bulk annuity insurance sector have in recent times been attributing the properties of competitive free markets to the market for pension annuities. In reality bulk annuity insurers are highly regulated institutions and their product pricing is in large part determined by these regulations. The concept of valuation by market pricing, that is exchange among others, is contentious for traded assets; the concept of valuation by regulated market pricing for liabilities far more so.

The preliminary question addressed in this paper concerned the rationale for regulation; ruin theory fails to address that; the implicit assumption is that the probability of failure should be regulated and that capital requirements are the best way to achieve that. The question can be rephrased – are the imperfections of the insurance market sufficient to require regulation? The existence of a market failure is a necessary rather than sufficient condition; there is the further condition that the cost of regulation must be lower than its benefits.

The costs of course express themselves in many ways beyond the constrained equilibrium asset models covered earlier and for investigation of that we need a newer set of micro-economic techniques.

Firms, Contracts and Modern Economics

The neo-classical view of the firm is as a production function, a black box, with selfless managers maximizing profits by the choice of input and output levels. The average cost of a product initially falls as fixed overheads are distributed over greater output until some variable inputs which are not perfectly scalable rise increasingly and average costs begin to rise– the average price/output curve has the familiar U shape.

The theory is useful for the analysis of production choices, the aggregate behaviour of an industry and indeed, with relaxation of the perfect competition assumption, strategic interaction between firms. It offers no insight into incentives within a firm. In 1937 Ronald Coase also made the point that neo-classical theory is entirely consistent with there being just one huge firm.

Principal-Agent theory has emerged as the dominant stream of micro-economic analysis of incentives within firms. Under this approach one or more inputs of the agent, typically effort, is endogenous and unobservable, which means that any contract written upon it would be unenforceable. The problem becomes the design of contracts based upon observed effort and one of optimal incentives versus optimal risk sharing between principal and agent.

A form of land tenure known as share-cropping is illustrative; under this the tenant (agent) delivers some significant proportion (50%) of the crop to the landlord (principal) in exchange for the land use. The contract is clearly not first best. The tenant's effort in producing the crop cannot be directly observed and indeed the crop output is not perfectly correlated with the tenant's effort. This sharecropping contract may be decomposed into a rental agreement and an insurance contract under which the landlord refunds part of the rent if the crop proves poor. It has risk-sharing properties. Full insurance would be inadvisable since it would reduce all effort incentives, a moral hazard problem.

The principal-agent literature is huge, covering optimal and equilibrium incentive schemes in labour, capital and insurance markets. It encompasses repeated relationships, multiple agents, reputational effects, career concerns and much more; it is quite widely used in managerial compensation determination. It can not however address the concern expressed by Coase.

It is sometimes suggested that asymmetries of information are reduced within a firm enabling superior incentive contracts to be written, but that merely begs the question why should this be. An alternate variation of this suggestion, subject to the same concern, is that cost or profit sharing are possible within a firm but not across independent firms. This latter variant may bring with it competition concerns and be regulated.

Perhaps the principal failing of principal-agent theory is its incompleteness with respect to transaction or contracting costs; this theory attributes costs to observation. However it is necessary to write contracts to enable their subsequent legal enforcement and that involves negotiation and comprehensive foresight of future states of the world; this is costly and results in incomplete contracts. An incomplete contract may, of course, be renegotiated as times passes, but that incurs ex ante costs prior to renegotiation and ex post costs during the renegotiation process. The process may for example involve the courts. These ex-post costs are material in insurance because the inverted nature of insurance production, the long term of cover – costs (of production) are only known only with the claim arrival. Switching suppliers is not (usually) an option; though of course an unsatisfactory ex post negotiation will usually result in a switch of insurer. Transaction cost theories though are still not informative as to why the renegotiation process and "hold-up" problems may differ within a firm rather than between independent parties.

Insurance deductibles and minimum thresholds[10] may be viewed in this framework. Transactions costs here are the costs of administration of claims processing. The trade-off between risk transfer and transaction costs optimal[11]. There are two further relevant

[10] See for example Synergy Insurance
[11] Ken Arrow (??)

aspects of an insurance contract, moral hazard and adverse selection. If an insured has full coverage, there may be no incentive to minimize risk; this is the source of moral hazard. Some forms may be relatively subtle; for example, the insured firm may no longer monitor closely (since this may be costly) the behaviour of (sub) agents, its workers. The residual risk exposure of the deductible is the optimal contractual mechanism; it minimizes the risk of the policyholder.

It is also worth noting that these principal-agent problems can also be addressed by newer contractual forms such as parametric rather than indemnity triggers; cover based upon an independent earthquake damage index is an example of a parametric trigger. The basis risk occupies the role of the deductible for the insured. Here the loss mitigation selected by the insured is outside of the insured's control or influence and consequently moral hazard is absent. This differs from the use of additional triggers where these are informative of the agent's effort.

In the case of parametric triggers adverse selection is also eliminated. The role of the deductible in the case of adverse selection, a form of hidden information about the insured, is to permit self selection among policyholders – this is a form of signalling. The worst risks will select the lowest deductibles, allowing the insurer to price these differentially. Note that to be credible a signal must also be a costly action to the signaler.

Implicit within the neo-classical model of equilibrium is the concept of perfect information. This model is most interesting for what does not matter within it; these include institutions (markets see through them), history and the distribution of wealth. An absence of surprises together with imperfect information as to price, quality and effort had no effect in a model where initial endowments, the production technology and preferences determined all future.

Imperfect information has some remarkable consequences. It limits, for example, the domain of the law of diminishing returns, the law of supply and demand, the law of the single price and the efficient markets hypothesis. Under quite general conditions it is

possible to show that it never pays to buy just a little information. Markets may no longer achieve equilibrium.

Information asymmetries may be inherent or may be created; the management of a firm has incentives to create asymmetries to enhance their bargaining power. The regulator earlier has an incentive to design regulation such that the cost is opaque to the consumer and the consumer will demand ever more of this “costless” good, entrenching the position of the regulator. The analysis of informational asymmetries concentrates upon the incentives and the mechanisms for information acquisition. This should be confused with the differences between non-contractible information and non-contractible action; even if it is possible to specify the circumstances where an action, such as restructuring, is advisable it is not usually possible to contract upon the detail of this action. The trade-off becomes one between imperfection of information and imperfection of markets. Information externalities can arise; private gains can exceed social gains. Markets here do not provide correct incentives for information disclosure and there is in result a role for government. And perhaps most importantly expenditures on information acquisition can be excessive. The Arrow-Debreu efficient market is one of perfect information in which beliefs cannot be endogenous; these cannot change because of the actions of any individual which includes the action of investing in information discovery. It is a most unrealistic view of the real world.

A market failure may admit the possibility of an institutional solution but that should not be taken to mean that, in equilibrium, it fulfills that function. Nor should it be taken to mean that these institutions can fulfill the role as well as government. In fact non-market institutions can actually make matters worse. Given that markets are only efficient under limited circumstances, there is certainly no reason to believe that an equilibrium involving both non-market institutions and markets should be efficient.

The separation of ownership and control within a firm creates problems; shareholders (principals) clearly cannot fully monitor the workers and managers (agents) within a firm that they own and in addition may not even know what these managers should be doing.

Managers may collect “rents” at the expense of shareholders and majority shareholders may disadvantage minority shareholders.

The lack of explanation of the boundaries of a firm requires yet further techniques. These centre on property rights and ownership. In this approach an owner has control of the non-human assets of the firm, control rights, rather than the perhaps more usual rights to residual income. It becomes possible to distinguish between leased and owned assets and relationship-specific investments figure prominently in this literature. We own our house because giving anyone else those control rights reduces our interest in the value of the house. Offering lower grade employees equity incentives will add little to their productivity since they have in their employment little by way of control rights. Firms with complementary assets will tend to merge but, where markets exist with many buyers and sellers, firms whose are assets are independent should not integrate. It can throw light onto the question of industry mutual-owned entities, a structure still wide-spread in insurance. It is also interesting to think about the use of credit rating agencies in the context of control rights, but not central to this paper.

Insurance regulation, Capital and Information

The long term nature of insurance is a major source of informational asymmetry between insiders (management and significant shareholders) and outsiders (policyholders). One argument frequently advanced for the presence of significant financial institutions as shareholders in new insurance ventures is that these will have the skill to monitor management action; this offers comfort to the regulator and perhaps policyholders. It is possible to argue that an information asymmetry, managerial moral hazard, was a contributing factor in the Equitable debacle; this was a mutual with no significant expert financial shareholders. But these arguments are weak in the presence of myopia, few shareholders or monitors are interested in the very long term.

Capital, of course, is a signaling mechanism. A company’s shareholders, who are knowledgeable, may signal the quality of its risk managers by the amount of capital at

risk. But with diffuse shareholders this loses impact. Such companies though may signal their quality by maintaining capital in excess of the statutory requirements, in the expectation that policyholders will "screen" this information. But it should be noted that this is not an efficient mechanism for the resolution of informational issues.

Non-contractible information admits a role for corporate governance with residual control rights important. In good times it is optimal to have these residual control rights rest with shareholders but in bad they should rest with policyholders. There is though a problem with this simple remedy; policyholders are many and have a co-ordination of collective action problem. They cannot effectively exercise control rights. The role of the prudential regulator is as the representative of current policyholders in the governance structure of the insurer.

This view of the world leads to a set of regulations[12] which are very different from what we observe. The regulator should be a tough claim-holder, intervening early upon the basis of multiple triggers using simple public accounting ratios; this latter requirement arises from the need for ease and timeliness of verification. In fact the only circumstances where the regulator should adopt a soft stance are the rare occasions where the collective actions of insurers may have systemic consequence.

The world which we face is one in which the title question of this paper becomes relevant. The regulator has incentives to leave opaque the costs of his actions since this enhances his tenure; a principal-agent (representative) problem. The form of regulation, principle rather than prescription, will add to the difficulty of external verification of the regulator's effort. Explicit rules, where they exist, are based upon complex private information rather than public accounts; this adds to the difficulty of verification.

The management of an insurance company has incentives to collude with the regulator. The opacity of regulatory triggers and the severity with which the firm may be treated

[12] For a fuller set of regulatory policy recommendations see Plantin and Rochet "When insurers go bust" 2007

under a principles-based rather than prescriptive regulatory system add to the manager's opportunity to capture rents at the expense of shareholders. Regulation of the institution rather than regulation of the product also supports management in this rent-seeking behaviour.

The favoured argument supporting principles based regulation in general is that this admits innovation. The argument supporting institutional rather than product regulation is also that this admits innovation. But this misses a significant point of difference between insurers and other firms, the inverted nature of the product cycle. The work in progress of an insurer is its book of business in force and is substantially the entire insurer. The work in progress of another form of firm is typically small and while innovation may make this obsolete, this does not threaten the existence of the firm. The loss is born by the firm's capital in both cases but in the case of insurers this may threaten its continuing existence and the policyholder's claims.

Innovation as a justification for a particular form of regulatory regime should also find favour with management. Innovation is a risky business which brings with it exceptional returns to first movers. The quest for innovation allows the manager to speculate at the possible expense of shareholder and policyholder and to participate in the exceptional rewards when the speculation is successful. Shareholders of course have incentives to speculate when times are bad arising from the limited liability of their interest.

Management may desire regulation to constitute a material barrier to entry since this reduces competition and admits product pricing under market power. Provided the costs of regulation are difficult for the consumer to observe the regulator has incentives to construct a large and complex set or regulations. At the mundane level, this increases the value of his human capital. Left to their own devices regulators and management can be expected to develop large and complex regulatory systems which are costly to both the policyholder and shareholders.

The model which follows develops these ideas in a limited formal framework which considers only the extent to which regulators benefit at the expense of shareholders.