RECONCEPTUALISING THE CAPITAL ADEQUACY REQUIREMENT OF SHORT-TERM INSURANCE COMPANIES WITHIN THE CALL OPTION FRAMEWORK

James Howard Christopher Britten

A dissertation submitted to the Faculty of Commerce, Law and Management, University of the Witwatersrand, in fulfilment of the requirements for the degree of Master of Commerce in Finance

Johannesburg, South Africa
October 2010

RECONCEPTUALISING THE CAPITAL ADEQUACY REQUIREMENT OF SHORT-TERM INSURANCE COMPANIES WITHIN THE CALL OPTION FRAMEWORK

Abstract

Conventional wisdom decrees that in order for insurers to provide cover, they require capital. One of the many methods of calculating capital requirements of short-term insurers is the insolvency put option framework. This technique was originally introduced by Merton (1977). The general argument is that bankruptcy occurs when shareholders exercise a valuable put option. Indeed, the corporation was introduced to protect shareholders from, mainly contractual, liabilities of persons who trade with the corporation. The corporation thus introduced the idea of limited liability of shareholders or as is often called the corporate veil. However, if a company defaults on its debt then equity holders have decided to allow an embedded *call option* to expire unexercised. As a result shareholders will behave as if they in fact hold a call option, which creates a different incentive than that suggested by the insolvency put idea. This study examines the role of capital and the influence of the insolvency put option within a short-term insurer. Specifically, it is argued that capital is not the cornerstone of a short-term insurer. Moreover, using Brownian motion and Itō calculus as well as continuous time financial models a more complete mathematical description of an insurance company is articulated by explicitly taking the embedded equity call option into account.

JEL Classification: G22, G28, G33

Keywords: Insolvency put option, call option, put-call parity, insurance, limited liability, regulation, capital adequacy.

Declaration

I, James Howard Christopher Britten, declare that this research report is my own, unaided work. It is submitted in fulfilment of the requirements for the degree of Master of Commerce in Finance at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

James Howard Christopher Britten

October 2010

Acknowledgements

There are many people to whom I owe a great deal of thanks for their guidance and support over the past few years. My parents, Tony and Diana have been a substantial source of support and encouragement and I am very grateful to both of them for encouraging me to further my education and supporting my choices in life. My supervisor, Hugh-David Hutcheson has been incredibly patient and supportive. Without his guidance I would be stuck at square one, in the doldrums. I will always be grateful for his considered critiques and the enlightening conversations I had with him not only during the course of this research but throughout my time at Wits. Prof Robert Vivian deserves much credit in helping me get over the finish line. He initially introduced me to this topic and was always extremely generous with his time. His kind encouragement and clarity of thought helped me flesh-out the finer points of my research. I thoroughly enjoyed our wide-ranging discussions. Prof Greg Lee was another source of morale-boosting support. His observations enabled me to gain perspective on my work and spurred me on to test my ideas on a wider audience. Prof Christo Auret was incredibly patient with me whenever I would report that my thesis was still a work in progress, not to mention the numerous revisions as to when I would finally complete it. His support has made working within the finance department at SEBS an enjoyable and rewarding experience. Thank you to my friends and colleagues (past and present), Ryan Blumenow, Nic Spearman and Claire Sebastiao who have put up with my lamentations during the course of my work. They were always forthcoming with helpful advice and our seemingly random conversations were always a welcome distraction.

Finally, a sincere thank-you to all my family and friends who showed an interest in my work and nagged me to get it done.

Table of Contents

List of Tables

List of Figures

1. Introduction

Presently, capital allocation is regarded as an essential component of risk management and the financial health of a short-term insurer in general. Naturally, regulatory authorities have embraced Basel style capital rules that regard capital as the foundation of a short-term insurer (Sherris, 2006; 2007). Traditionally in the United States (US) regulatory authorities focused solely on pricing (Joskow, 1973; Ippolito, 1979; Frech and Samprone, 1980), however, in the early 1990s the regulatory focus shifted to capital-based solvency measures, leading to the introduction of the risk-based capital (RBC) system in 1993 (Cummins, Harrington and Klein, 1995; Kwon, 2007; Grace and Klein, 2008). In Europe, Solvency I, a capital-based solvency measure, was introduced by regulators in January 2004. Solvency I has been developed further into Solvency II, where an even greater emphasis on risk-based capital is made (Eling, Schmeiser and Schmit, 2006). As a result, here in South Africa, regulators have followed the example of US and European regulators by proposing the capital-based Solvency Assessment and Management (SAM) System, which is expected to be implemented by 2014.

In spite of the regulatory focus on capital, there is questionable evidence that capital adequacy rules are appropriate for the short-term insurance industry. If one takes a broader view of the general regulatory process, there is considerable debate as to what are the motivations driving regulatory policy. Indeed, there is conflicting evidence on whether regulatory policy is developed with either the interests of consumers or that of the industry in question expressly in mind (Stigler, 1971; Posner, 1974; Peltzman, 1976; Ippoltio; 1979; Becker, 1983; Peltzman, Levine and Noll, 1989; Levine and Forrence, 1990; Grace and Phillips, 2008), and specifically, the effects of capital-based regulation in the US insurance industry appear to be ambiguous (Cummins, Harrington and Klein, 1995; Grace, Harrington and Klein, 1998; Pottier and Sommer, 2002; Klein, Phillips and Shiu, 2002). However, there is general agreement that regulation has distorted both pricing and quantity of insurance provided (Joskow, 1973; Ippolito, 1979; Grabowski, Viscusi and Evans, 1989; Harrington, 1992).

As a result, a general sense of ambiguity underscores insurance regulation, which has been complicated further by many competing interests. For instance, certain aspects of the development of US insurance regulatory policy, throughout its history, have been characterised by either stagnation or digression. What is more, long-standing anti-competitive legislative amendments have prevented the growth of more efficient and less costly insurance systems (Kimball, 1961; Joskow, 1973; Kwon, 2007; Grace and Klein, 2008). This ambiguity stretches into capital-based solvency measures such as RBC and Solvency I and II because despite the growth, and ensuing entrenchment, of risk-based capital requirements for the short-term insurance industry, there is no uniform risk measure to determine capital (Francis, Heckman and Mango, 2005). By the same token, Sherris (2007, p. 1) admits, "… there is limited theoretical guidance on which risk measure is consistent with value maximisation and no well developed economic theory underlying the risk measures." For example, several definitions of capital exist (Merton and Perold, 1993).

In fact, it can be argued that the limited theoretical guidance that Sherris (2007) describes has resulted in a regulatory framework that is not necessarily the most considered or appropriate approach in every instance. Moreover, another complicating factor with risk-based capital models is the rise of a risk measure used to determine the appropriate level of capital, the insolvency put option framework. The insolvency put option framework has become considerably influential over the last several years and is intertwined with capital allocation (Doherty and Garven, 1986; Babbel 1998; Myers and Read, 2001; Venter, 2004; Mildenhall, 2004; Babbel and Merrill, 2005; Sherris, 2006; Sherris and van der Hoek, 2006).

The origin of the insolvency put option framework can be traced back to the initial option pricing models of Black and Scholes (1973) and Merton (1973b). While option pricing theory is now well established, generally, the application of option theory to companies is relatively recent. After the initial option pricing theory breakthroughs made by Black and Scholes (1973) and Merton (1973b), a wide range of applications of option methodologies became possible. This led Merton (1974) to apply option pricing theory to analyse risky debt. Later, Merton (1977) explicitly applied put option valuation techniques to loan guarantees such as deposit insurance, which set the idea of the insolvency put option in motion. Since Merton (1977)'s seminal work, the insolvency put option has been applied to insurance contracts and used to explain why default occurs. In other words, a firm will default on its obligations when its shareholders exercise a valuable insolvency put option. Thus, the insolvency put option framework is used to determine the amount of capital needed to maintain the solvency of an insurer at levels prescribed by the regulator (Butsic, 1994; Babbel, 1998; Myers and Read, 2001; Babbel, Gold and Merrill, 2002; Babbel and Merrill, 2005). However, as will be pointed out later in this study, Merton (1977) never associated default with the insolvency put. Rather, he used put option valuation methods to describe a specific mechanism such as deposit insurance provided by a third party. That is, the payoff to equity holders remains unchanged irrespective of any third party guarantees.

The purpose of this thesis is to examine the role of capital and the influence of the insolvency put option framework within a short-term insurer. Specifically, it is argued that capital is not the cornerstone of a short-term insurer. In particular, capital-based solvency measures have resulted in an improper conceptualisation of the essential makeup of a short-term insurer. In fact, capital is relatively unimportant to the solvency of a short-term insurer when compared with premium income. Furthermore, the conceptual problem of the insolvency put option is explored and it will be argued that the notion of an insolvency put is spurious. That is, the idea that bankruptcy can be described via the mechanism of an insolvency put option is questioned. If a company defaults on its debt it is because shareholders have allowed an embedded *call option* to expire unexercised. Viewing insolvency from the perspective of a call option has important implications for solvency regulation because equity holder incentives will be distinctly different than that predicted by the insolvency put option framework. Furthermore, this study provides a more complete mathematical description of an insurance company by explicitly taking the embedded equity call option into account where Brownian motion and Itō calculus as well as continuous time financial models are used to reconceptualise capital adequacy within a call option framework.

This thesis is set out as follows. Chapter Two is devoted to the theoretical and empirical analyses of the influence of regulation. Here the two central theories proposed to explain the pattern of government intervention, the public interest theory and capture theory of regulation, are discussed. The regulatory experience of the US property-liability insurance industry as well as the history of insurance legislation and development of risk-based capital rules is examined. Furthermore, the efficacy of the risk-based capital formula as an indicator of solvency is surveyed.

Chapter Three addresses the initial mathematical descriptions of insurers, which led to the development of the insurance capital asset pricing model (CAPM) and the formalisation of return on equity calculations for insurance companies. The link between cost of capital estimates and capital-based models such as risk-adjusted return on capital (RAROC), economic value added (EVA) and value at risk (VaR) are also examined.

Chapter Four serves as a general overview of the evolution of the insolvency put option from its initial articulation by Merton (1974, 1977) to its most recent iterations set out by Butsic (1994), Myers and Read (2001), Sherris (2006), Sherris and van der Hoek (2006) and Yow and Sherris (2007) as well as its application to insurance pricing and capital adequacy. Important details of the link between the insolvency put model and risk-based capital are discussed. Moreover, the impact that the insolvency put is said to have on the market value of a firm is examined.

Chapter Five critically evaluates the insolvency put option framework and sets out the arguments that counter the notion that the insolvency put option stems from the introduction of the limited liability company. The general argument is that limited liability enables shareholders to put the losses of the company to their creditors, hence the association of bankruptcy with a put option. However, when one examines the development of the limited liability firm in English Company Law, it is clear that limited liability granted shareholders an implicit call option. Furthermore, Stoll (1969)'s seminal derivation of the put-call parity relationship is used to explain why even if shareholders do indeed own a so called insolvency put option they will only ever view their investment from the perspective of a call option. In addition, other conceptual flaws of the insolvency put option are examined, with specific reference to Modigliani and Miller (1958), Black and Scholes (1973), Merton (1973b) and Myers (1977).

Chapter Six details the counter arguments to capital-based solvency regulation and Myers and Read (2001)'s influential capital-based insolvency put model in particular. Here the technical aspects of Myers and Read (2001)'s framework are examined as well as the practical limitations of the model. In addition, by drawing on the observations of Adam Smith (1776/1976), and the work of Hill (1979) and Vivian (2007a, b), the usefulness of capital allocation itself is questioned.

Chapter Seven examines the implications of the capital-based insolvency put option framework for the short-term insurance industry. Drawing from the observations and reasoning of the preceding chapters, it is argued the interaction of the insolvency put option and the regulatory emphasis on Basel style capital adequacy can trigger bankruptcies in severe market downturns when access to capital and liquidity becomes scarce. Without available capital, an insurer will be unable to meet the capital requirements set out by the regulator, forcing the firm into technical insolvency. It is maintained that such a situation will be tantamount to eliminating time value of shareholders' call option on the assets of the insurer, which therefore makes insolvency more likely.

Chapter Eight articulates a mathematical description of a short-term insurer. This chapter draws on the observations of Ferrari (1968), Fairly (1979), Hill (1979) and Vivian (2007a, b), and expresses Adam Smith (1776/1976)'s description of a successful insurer mathematically. Here the importance of premium income is demonstrated. Furthermore, the call option perspective is incorporated and analysed using Brownian motion and Itō calculus as well as continuous time financial models.

The objective of this mathematical description is to articulate the effect that time value has on the embedded call option and ultimately, the behaviour of shareholders.

Last, Chapter Nine concludes by discussing the contributions and theoretical implications of this study as well as potential avenues of future research.

2. Regulation and Insurance

2.1. Theoretical and Empirical Analyses of Regulatory Influences

Economic theory suggests that the appropriateness of government intervention, whereby the operating activities of a firm (or an industry for that matter) are regulated, should be judged solely on situations where market distortions exist. Such distortions include, but are not limited to, monopoly pricing, destructive competition, information asymmetries and environmental externalities (Levine and Forrence, 1990; Laffont and Tirole, 1991). Within economic literature, two major theories have been proposed to explain the motivation behind government intervention. The first, the public interest theory underscores the role of a central authority in correcting these market distortions. In other words, under the public interest theory, government intervention and the implementation of a regulatory framework aims to present "…a governmental corrective device that improves market performance in instances in which competition 'fails'" (Ippolito, 1979, p. 55). Moreover, while it can be argued that these market failures underpin the notion that regulation itself will be in the public interest, the ultimate goal of regulation according to the public interest theory is to maximise social welfare. The second major theory to explain the pattern of government intervention is capture theory and its extension, the economic theory of regulation. The capture theory of regulation (which was formalised by Stigler (1971)) is a counter argument to the public interest theory in that it calls attention to the role of special interest groups in the creation of a regulatory framework. That is, the capture theory contends that special interest groups "capture" the regulator to ensure that the regulatory framework is imposed primarily for their benefit and not the general public.

As mentioned in the introduction, within the field of insurance regulation, capital-based rules have become the most important regulatory instrument to gauge a short-term insurer's solvency. Given that capital is regarded as pivotal to the successful operation of a short-term insurer, it is deemed to be best practice in terms of risk management to monitor and maintain an adequate capital cushion. It is therefore considered to be in the public interest to regulate a short-term insurer's capital levels by introducing capital-based solvency standards. However, various studies over the last several years have questioned the efficacy of capital-based solvency standards (Cummins, Harrington and Klein, 1995; Grace, Harrington and Klein, 1998; Pottier and Sommer, 2002). Despite these questions raised about the effectiveness of the risk-based capital framework, it is clear that some form of insurer solvency regulation will always exist. Indeed, irrespective of the growing influence of the capture theory of regulation, it is argued that even in the absence of any market distortions regulation is appropriate because insurance is vested in the public interest (Kimball, 1961; Adams and Tower, 1994). Thus, regardless of a Pareto optimal market, the general public relies on insurers to such an extent that the insurance industry requires government supervision (Eling, Schmeiser and Schmit, 2006; Grace and Klein, 2008).

The question, however, remains as to what form of regulation or regulatory emphasis is appropriate for the short term insurance industry. It follows that understanding the effect of regulation as well as any potential unintended consequences is vitally important when establishing a regulatory framework. Indeed, regulation itself can exacerbate existing market problems or introduce distortions that were absent in the first place (Frech and Samprone, 1980; Harrington, 1984; Grabowski, Viscusi and Evans, 1989; Nelson, 2000; Grace and Leverty, 2009).

2.1.1. Market Socialism

Before discussing the merits of the public interest theory of regulation, it is instructive to take a step back and briefly touch on regulation in a public owned economy; market socialism. Although this economic system does not have a straightforward definition, it is viewed generally as an extreme version of the public interest doctrine where the government takes on an omnipresent role. Although, Bonbright (1961) maintains that rate regulation in a socialist economy is simply a means to carry out socialist strategy, making it distinct from principle based utility regulation.

Lange and Taylor (1938)'s seminal "On the Economic Theory of Socialism" extols the benefits of a centrally planned economy. Interestingly, their motivation for a centrally planned economy was economic rather than political: "The economic problem is a problem of *choice* between alternatives. To solve the problem three data are needed: (1) a preference scale which guides the acts of choice; (2) knowledge of the 'terms on which alternatives are offered'; and (3) knowledge of the amount of resources available. Those three data being given, the problem of choice is soluble … The question remains whether the data under 2 are accessible to the administrators of a socialist economy" (Lange and Taylor, 1938, p. 60). Thus, they argue that a centrally planned and informed market regulator would be in a much better position to determine consumer preferences and costs that occur within the whole economy than any individual could. Furthermore, Lange and Taylor (1938) argue that knowledge of the available alternatives is attainable just as it would be in a free market economy. Lange and Taylor (1938, p. 98) comment; "Competition forces entrepreneurs to act much as they would have to act were they managers of production in a socialist system." At first glance, they appear to suggest that setting up a command economy would be redundant, as competitive forces will allocate resources in a similar manner to that of a socialist economy. However, Lange and Taylor (1938) stress that in reality, the actual allocation of resources of a competitive economy will be markedly different to that of a socialist economy. In fact Lange and Taylor (1938, p. 99) declare; "Only a socialist economy can distribute incomes so as to attain the maximum social welfare." As a result, such a system would be able to make more efficient allocations. The major implication of this framework is that the regulator would replace the market's role as a processor and disseminator of information and the government would manage every facet of the market.

The market socialist system has not gained much traction in the West as it is generally accepted that the framework of private ownership and enterprise is the most feasible way to disseminate information and provide meaningful incentives to react to that information. However, while many accept the principle of private ownership without question, most believe that government intervention can correct market deficiencies. Thus, it is argued that a benevolent far-sighted regulator can improve the market mechanism.

2.1.2. The Public Interest Theory of Regulation

As indicated, the primary purpose of regulation is to protect the general public and promote social welfare. Yet, the idea of the pubic interest does not have a concrete definition. For instance, Redford (1954, p. 1108) conceptualises the public interest as "…the best response to a situation in terms of all the interests and of the concepts of value which are generally accepted in our society." In other words, the concept of the public interest can be thought of as the embodiment of common beliefs or values.[1] These shared values or beliefs suggest an adaptable set of principles rather than a rigid set of rules. Thus, it is open to interpretation as to what actions could be considered to be in the public interest. Schubert (1957) maintains that common interests and ideals propel the public interest. Schubert (1957) takes Redford (1954)'s assertions a step further and argues that benevolent public minded regulators provide the "creative" force necessary to pursue the public interest.[2] Similarly, Friendly (1962) maintains that the legislation governing the conduct and scope of the regulator has a muted impact on its behaviour compared to the actual individuals in charge that set the agenda of the public interest. He suggests the best way to promote the public interest is via "…the appointment of commissioners of higher intellectual power and moral courage" (Friendly, 1962, p. 1294). Of course this can only be true if executive discretion rests with the regulator and not legislators. However, Friendly (1962) argues that regulatory commissions should have more of an arms-length relationship with that of the industry being regulated. That is, they should act more as "… an adjudicator, not a business manager" (Friendly, 1962, p. 1285).[3]

[1] Redford (1954) points out that a compromise of common values will be needed when it comes to conflicts between immediate and longer-term interests, and in particular, conflicts created by special interest groups. The latter observation is elaborated on in the section Regulatory Capture and the Economic Theory of Regulation.

[2] Redford (1954), in contrast to Schubert (1957), pays a great deal of attention to the role played by experts in pursuing the public interest. He maintains that the complexity of the issues faced by a government and society necessitate the assistance of experts, as they are the most likely to provide pragmatic solutions. Thus, in the context of Schubert (1957)'s argument, experts could be regarded as one of the creative forces that help pursue the public interest.

[3] Graham (2000) suggests that the discretion of a regulator should be limited to protect against capture, which will provide an incentive for regulated firms to improve their internal efficiency.

Friendly (1962)'s review of decisions taken by the regulatory commissions in the US uncovers concern for both anti-competitive monopoly pricing practises[4] and pricing systems that would encourage damaging competition. In fact, it appears that the central issue facing regulators at the time was how to balance the tension created by new entrants driving price and service advantages with that of established firms' (or an entire industry) inability to compete. Thus, it was believed that the public interest was best served by prescribing a pricing scheme that maintained a fair rate of return, which discouraged "…unfair and destructive practice" (Friendly, 1962, p. 1284).[5]

Redford (1954), Schubert (1957) and Friendly (1962) pay little attention to fundamental economic ideas and deal mainly with the legal and moral issues associated with the public interest. Yet, even the economic explorations of regulation in the public interest also pay a great deal of attention to normative issues. Bonbright (1961), for instance, maintains that it is the necessity of intervention itself that justifies regulation rather than the necessity of the item being produced. Consequently, it is necessary to decide what activity could broadly be interpreted as being in the public interest. Bonbright (1961)'s view is that the protection of consumers is paramount and any regulatory action that does so can, by definition, be regarded as acting in the public interest. Furthermore, government intervention will play a central role. Bonbright (1961, p. 29) insists that it is an important mechanism of sensible public policy; "[T]he identification of the public interest with the welfare of the people in the community or nation, [will necessitate] the state being regarded merely as an instrument for the attainment of this welfare".

[4] Demsetz (1968, p. 59), however, contends that threat of monopoly pricing does not justify regulation; "To the extent that utility regulation is based on the fear of monopoly price, *merely because one firm will serve each market*, it is not based on any deducible economic theorem."

[5] Friendly (1962) questions how such a policy can be considered to be in the public interest. After all, how is it that a lower rate for a service offered voluntarily by a firm can be deemed counterproductive? Furthermore, a "fair" rate of return is subjective at best. In fact, Friendly (1962) hints that special interest groups have a notable influence in the formation of regulatory policy, pushing the outcome away from the public interest. This point is discussed in detail in the section Regulatory Capture and the Economic Theory of Regulation.

However, simply arguing that any activity that protects consumers to be in the public interest pushes the theory of public interest into a normative corner. Indeed, Buchanan and Tullock (1962/2004) contend that the notion of public interest is ambiguous. In fact, it can be argued that this ambiguity had a notable influence on the evolution of insurance regulation. Kimball (1961, p. 476) argues that the development of insurance regulation in the US at the time was "…much influenced by the factors of inertia and drift." Kimball (1961) contends that only with tangible objectives of regulation can government intervention be truly made to serve the public interest. In the case of insurance, he argues that primary role of regulation be to "…facilitate the successful operation of the enterprise itself" (Kimball, 1961, p. 477). Central to the successful operation of an insurer will be its solvency, which Kimball (1961) maintains will be best managed via the regulation of insurance premiums. Indeed, the concern for solvency manifests itself via the stipulation that premiums not be "inadequate" (Kimball, 1961, p. 482).[6] Given the ambiguity of what can be regarded as acting in the public interest, Buchanan and Tullock (1962/2004) argue that regulation in the public interest should be concerned only with the efficiency (or lack thereof) of the market and not the fairness of the distribution of wealth between producers and consumers. Yet, regulatory controls introduced in the US, were not solely concerned with market efficiency. MacAvoy (1979) notes that regulatory controls were introduced for various reasons in different industries, indicative of the ambiguous conception of the public interest. He observes that legislators were concerned that producers failed to take into account the full social costs of their activities and enacted many regulatory controls to police environmental and health and safety standards.

Adams and Tower (1994) maintain that a lack of information and disclosure can justify regulatory intervention. Adams and Tower (1994) point out that consumers might not realise the importance and value of certain information thereby causing a breakdown of market signals necessary for the autonomous functioning of the market. They argue that all the regulator should attempt to achieve is to ensure that consumers

[6] MacAvoy (1979) and Ippolito (1979) point out that externalities arising from insolvency can be so costly and disruptive that the threat of insolvency alone justifies government intervention.

have enough information to make optimal decisions.[7] In other words, a regulator should not try to manage the activity of either producers or consumers (Adams and Tower, 1994; Booth, 1997). Moreover, such a regulatory approach is not designed to mitigate the chance of insolvency. Even if the regulator facilitates the free flow of information efficiently it is possible that that consumers will not interpret information correctly all the time. In fact, there may be situations where information is ignored. Hence, the risk of insolvency is always present and "[a]n insurance failure should, therefore, never be regarded as a prima facie case for further regulation" (Booth, 1997, p. 680).

Furthermore, Booth (1997) maintains that there are three central shortcomings of the public interest view. First, it only considers one aspect of market failure. That is, the inability of the market to reach the same state of a simplified model under perfect competition. The second problem is that it is assumed that the correction of a market imperfection will automatically improve economic welfare. Finally, the third issue with the public interest approach to regulation is that any issue could be regarded as a market failure requiring regulatory intervention. In other words, there will be a strong temptation to introduce laws to direct any activity that authorities deem to be in the public interest. Booth (1997, p. 683) stresses this point: "There is more incentive for the law to grow, finding more and more 'special cases', which require regulation to correct failures in the market. If the interests of regulators are different from those of consumers, it may be tempting for regulators to intervene to try to perfect what they regard as an imperfect market at every possible opportunity and perceived failure." Of course this scepticism is not new. John Stuart Mill succinctly articulates the idea that authorities will not act appropriately: "But the strongest of all the arguments against the interference of the public with purely personal conduct is that when it does interfere, the odds are that is interferes wrongly and in the wrong place" (Mill, 1859/1966, p. 109).

[7] Joskow (1973), for example, argues that consumers are generally unaware of the nuances of the insurance market and recommends that regulators focus on consumer protection and education.

However, Graham (2000) dismisses the argument that the normative, subjective nature of social issues disqualifies them from being incorporated into regulatory policy. While the emphasis on economic efficiency is paramount, social issues should also be of concern; "Doubtless not everyone will agree with the conclusions reached but it should, in theory, be possible to assess whether the arguments and evidence cited do support a particular policy decision. In any event, economic approaches do not seem to be in any better position. It is equally possible to have major disagreements between economic schools, such as the debate between Keynesians and monetarists in the 1970s and 1980s, which schools will have very different criteria for assessing the success of a policy" (Graham, 2000, pp. 25-26).

2.1.3. Regulatory Capture and the Theory of Economic Regulation

Arguably, one of the most important works to articulate the notion that regulators are not benevolent far sighted institutions is Stigler (1971)'s seminal work, "The Theory of Economic Regulation". He recognised that regulation is an economic good and thus, those involved in the regulatory process are themselves utility maximising agents. Stigler (1971) articulates a framework for the demand and supply of regulation where certain industries will acquire regulation so they can mould it to their own benefit. In other words, those with measurable political clout will acquire regulation to control the entry of new firms and consequently, competition as well.

Stigler (1971)'s main contention with the notion of a benign regulator was its inconsistency with actual experience. For example, if the introduction of oil import quotas in the US was designed to protect the public (the idea was to ensure a stable oil supply in times of conflict) a tariff would have been the best course of action. This would generate a significant stream of income for the government, which could be put to use in the achievement of the goals that brought about the imposition of regulation in the first place. Yet it is the oil producers that reap all the benefits at the expense of consumers. However, another question remains. If a well-organised and powerful special interest group like the oil industry can bend regulatory policy to its advantage, why does it seek the coercive powers of the state rather than a cash subsidy? The reason provided by Stigler (1971) is that if the state provided a generous cash subsidy it would attract new entrants and therefore threaten the incumbents ability to extract economic rents. Indeed, this question forms the basis of Stigler (1971)'s illustration of his approach to regulatory theory.

While the threat of new entrants piling into an industry, lured by an attractive cash subsidy, is a strong motivating factor for entrenched firms to seek out an alternative measure to stunt competition, the gains from cultivating the coercive powers of the state will be an even bigger incentive.[8] Stigler (1971, p. 4) emphasises this point: "The state has one basic resource which in pure principle is not shared with even the mightiest of its citizens: the power to coerce. The state can seize money by the only method which is permitted by the laws of a civilised society, by taxation. The state can ordain the physical movements of resources and the economic decisions of households and firms without their consent. These powers provide the possibilities for the utilisation of the state by an industry to increase its profitability."

One of the earlier examples of restrictive regulation in favour of a certain industry that highlights Stigler (1971)'s assertions was the railroad industry's aggressive action against the trucking industry in the US. As interstate and highway infrastructure began to improve in the late 1920s and early 1930s, the trucking industry's share of long-haul freight rose steadily, prompting the railroad industry to take action by lobbying for stricter regulation. In this instance, like that of import restrictions on oil, an industry was able to gain assistance and protection of the state, which resulted in a cost to society that was not offset by the benefit received by the favoured industry. Furthermore, with such an arrangement, Stigler (1971, p. 10) argues, "…one might expect a democratic society to reject such industry requests unless the industry controlled a majority of the votes."

[8] Stigler (1971) proposes four central policies sought by an industry. The first is a direct subsidy. Second, a firm may seek regulation of new entrants. Third, the suppression and control of substitutes may be sought. The fourth and most pervasive policy sought is price fixing.

In order to evaluate how some industries are able to acquire the benefit of the machinery of the state, it is necessary to examine the political process within a democracy and its interaction with market processes. Stigler (1971) points out that the most important distinction is that the political process is starkly different from that of the market. For instance, if the public is asked to choose between two goods[9] the public's decision must be followed by everyone regardless if an individual voted in favour of the chosen good. Thus, the political process requires everyone to make a decision at the same time. This necessity of simultaneity makes voting on issues a costly endeavour and to deal with this, representatives are employed by voters to make decisions on their behalf. As a result, unlike the market, "…the political decision does not predict voter desires and make preparations to fulfil them in advance of their realisation" (Stigler, 1971, p. 10).

The second salient feature of the political process is that it must involve society as a whole and not merely individuals directly involved or concerned with a particular development. By contrast the market does not require this. Using Stigler (1971)'s choice between air travel and rail as an example, a non-traveller never participates in the preference between the two but a freight company "votes" on the issue continually. As the political framework must include everyone, it cannot "…allow participation in proportion to interest and knowledge" (Stigler, 1971, p. 11).

To overcome this problem of non-participation of those unfamiliar or uninvolved in a particular issue a government will segment itself. Stigler (1971, p. 11) explains, "…I have somewhat more incentive to learn about local schools than about the whole state school system." This broad segmentation ultimately results in the segmentation of political representatives themselves. That is, they will organise themselves into political parties. This makes it easier to gain the support of interested individuals and those that undertake to implement the wishes of their constituency can expect to win an election and open the door to the trappings of office. Indeed, Stigler (1971) points

[9] Stigler (1971) illustrates this with the example of the choice between two modes of transport; rail and air travel.

out that if a representative could bank on re-election by voting against economic policy that harmed society, he or she would undoubtedly do so. However, this is where the special interest groups enter the frame. By voting against legislation that would have afforded large industries subsidies or other government support, those industries will move to support a more malleable representative. In short, "The industry which seeks political power must go to the appropriate seller, the political party" (Stigler, 1971, p. 12). As indicated, regulation is an economic good and the political process that supports it has costs which must be paid by the industry seeking state support, and those costs are votes and resources. Taken as a whole, Stigler (1971)'s framework illustrates that a government's intention to regulate an industry is mainly motivated by the desire to ensure political survival.

Posner (1974) explores the implications of Stigler (1971)'s economic view of regulation and while he acknowledges that the economic theory of regulation requires refinement and has little empirical support, it is far more compelling than the traditional public interest view of regulation. Yet, Posner (1974) does not dismiss the notion that regulatory authorities are set up with good intent. In fact he argues that regulation can be generally viewed as an honest attempt to pursue the public interest. However, the mandate of most regulators is beyond them. Hence, even if a market failure has been correctly identified, the regulator may not be in a position to bring about a more efficient functioning market. Posner (1974) observes that a fundamental lack of skills and knowledge, coupled with the sheer complexity of the industry being regulated can cause the regulator to become overwhelmed. In the instance of utility price regulation, regulators are required to determine the cost of the industry and keep price levels in line with those costs. Posner (1974) contends that the tools to perform such a task do not exist, which makes the objectives of the regulator unattainable. What is more, the costs of legislative supervision of regulators are prohibitive. Legislators interact with the private sector by means of negotiation, and bargaining with a large amount of individuals is a costly process, especially if the size of the group to negotiate with increases. To cope with this, legislators will delegate an increasing amount of duties to agencies (i.e. regulators) and consequently will have much less control over them. Thus, as Posner (1974) asserts, in the early stages of the creation of a regulator, legislators will have notable interest in tackling the issues that the regulator was created to deal with. However, in later years and as other problems arise that legislators must address, legislators cannot devote the same amount of time to monitoring a previously created regulator. Taking these factors into account, Posner (1974, p. 339) comments, “The agencies are asked to do the impossible and it is not surprising that they fail, and in attempting to succeed distort the efficient functioning of the regulated markets.”

Peltzman (1976) formalises Stigler (1971)’s theories yet, reaches a different conclusion. Central to Peltzman (1976)’s theory is the recognition that frictions associated with the political process will constrain the dominant group’s size as well as its potential benefits. In other words, a powerful group cannot completely capture a regulatory agency because the regulator will have significant political incentive to avoid “…exclusively [serving] a single economic interest” (Peltzman, 1976, p. 211).

The cornerstone of Peltzman (1976)’s theory is expressed by the equation

$$M = (n)(f) - (N - n)h. \qquad (2.1)$$

In words, the legislator will maximise the chance of re-election (M) by taking into account competing interest groups n (the beneficiary group) and the total number of voters, N. The legislator will also need to assess the relative probabilities of political support provided by the beneficiary group and opposition from the non-beneficiary group, denoted by f and h, respectively.

It is this that underpins Peltzman (1976)’s view that a regulator will not serve a single economic interest. Moreover, the other significant implication is that as the public’s involvement ($N - n$) in political activity increases, there will be a greater chance of their interests being pursued. However, if the probability of non-beneficiary opposition, *h*, declines then it is likely that the special interest group will mould legislation to their benefit (Peltzman, 1976; Becker, 1986).

Becker (1983) presents a model similar to Peltzman (1976) where various interest groups influence the political process. The general outcome of Becker (1983)'s framework is that a regulator will accommodate broader economic interests, which "…contrasts with the all-or-nothing outcomes implied by many other formal models of political behaviour, where the 'majority' clearly wins and the 'minority' clearly loses" (Becker, 1983, pp. 372 – 373).

Becker (1983) assumes that there are two competing groups, those that pay tax and those that receive subsidies, and each group seeks to maximise their income. Fundamental to his framework is the role that deadweight costs have on the competition between interest groups. For example, an increase in deadweight cost of a subsidy will result in lower revenue from taxes to contribute to the subsidised group. Thus, in such a case there will be less lobbying by the subsidised group to increase their income. Conversely, high deadweight costs associated with tax will stimulate greater pressure from the taxed group to reduce their tax burden. Ultimately, though, the ability of a group to gain political influence is determined by its relative efficiency (compared to rival groups). The efficiency of an interest group in turn is determined by its capacity to control free riders. Becker (1983) explains that free riding essentially increases the cost of lobbying due to the propensity of members to evade their obligations. Hence, keeping free riders in check will improve efficiency and be a notable advantage to the relevant lobby group. This observation suggests that the more successful interest groups will tend to be small and homogenous.[10] In particular, groups that lobby for subsidies will generally be small relative the number of tax payers; Becker (1983) points out that deadweight costs of tax fall as the number of tax payers rise due to a reduction in the tax per capita. Indeed, agricultural subsidies in the US and Europe are a prime example of Becker (1983)'s framework. General

[10] Becker (1983) does concede that in certain circumstances size may be an advantage; relatively bigger firms will be able to exploit economies of scale. However, the more successful groups will be smaller relative to adversely affected groups.

equilibrium will be reached when the deadweight costs of the competing groups even out. If this were not the case, both groups would pressure the regulator for a change (Becker, 1983; Peltzman, Levine and Noll, 1989).

Clearly, Peltzman (1976)'s and Becker (1983)'s extensions to Stigler (1971)'s model demonstrate that the public interest philosophy of regulation is not unattainable. In fact, Becker (1986) for instance pursues a novel line of enquiry on the issue of regulatory capture by examining the reciprocity of regulation of dentists. He finds support for Peltzman (1976)'s theory in that while the power of the industry acquiring regulation is important, he concludes, "…it is not always decisive. The public's interest can be and is maintained in most states" (Becker, 1986, p. 230). Although, empirical findings on this issue within the insurance industry is mixed. For example, Joskow (1973), Frech and Samprone (1980), find support for Stigler (1971)'s hypothesis, whereas Ippolito (1979)'s and Grabowski, Viscusi and Evans (1989)'s results are more consistent with Peltzman (1976)'s assertions. These studies amongst others are discussed in greater detail in section 2.2.

However, the key ingredient to Peltzman (1976)'s view is an informed and active public. Indeed, this is a central issue articulated by Stigler (1971); without a vocal public, a regulator does not have a strong incentive to act in the public interest. For instance, Booth (1997) argues that successful regulation of insurance is likely to be a trivial issue for voters when considering the re-election of a politician.

It is plain to see that Stigler (1971)'s assertions have been highly influential, however, his ideas are not universally venerated. For instance, Meier (1988) dismisses Stigler (1971)'s view of regulation. He is particularly critical of Stigler (1971)'s one-dimensional view of interest groups and charges that Stigler (1971) totally ignores the more nuanced incentives that drive them. In particular, Meier (1988) maintains that the insurance industry lacks the unity to influence the regulatory process. For example, the interests of a direct writer will be different to that of an agency insurer and the same applies to property-liability insurers and life insurers. Moreover, Stigler

(1971) makes no distinction between legislators and administrative regulators even though regulatory rules will differ from legislative policy. Meier (1988, p. 166) asserts, "…interests other than the industry have an impact on regulatory policy… Any good explanation of regulatory policy, therefore, must be based on more complex models than George Stigler's simple supply and demand model."

Peltzman, Levine and Noll (1989) attempt to gauge the efficacy of the economic theory of regulation and compare its development with that of the actual regulatory environment in various industries.[11] On reflection, they contend that the economic theory of regulation generally fits the regulatory experience of the 1970s and 1980s better than the public interest theory.

Although certain industries experienced a wave of deregulation in the 1970s and 1980s, Peltzman et al (1989) counter the notion that the regulatory experience contradicts the veracity of the economic theory of regulation. While the main assertion of the economic theory of regulation is that legislators and industries will be partial to regulation because of the benefits that can be extracted from a cooperative relationship, they demonstrate that the economic theory of regulation is useful not only in analysing the reasons behind additional regulation but it can also help explain why deregulation has occurred in certain industries[12]. Furthermore, the deregulation experienced over the period in question was not wholesale. Generally, the deregulation agenda was selective and regulation grew in other areas, which is consistent with the picture painted by the economic theory of regulation.[13]

The economic theory of regulation suggests that deregulation will become more probable if two significant economic changes occur. First, the gap between regulated

[11] Peltzman et al (1989) examine the regulatory experience of rail, trucking, airline, telecommunication, stock broking, banking and oil industries.

[12] Peltzman et al (1989) acknowledge that both Stigler (1971) and Peltzman (1976)'s explanations of regulation are incomplete. For example, neither adequately explains the continual regulation observed in a cluster of certain industries.

[13] Peltzman et al (1989) cite increased regulation of labour contracts and healthcare as examples of increasing regulation.

prices and deregulated must prices converge, making continued regulation redundant. Second, a declining level of wealth on hand for redistribution would not provide much political payoff if regulation were to be imposed (Becker, 1983; Peltzman et al, 1989). Peltzman et al (1989) contend that these two changes are related. For instance, should demand for a regulated industry's product decline then the regulated price should drift towards marginal cost. Resultantly, the potential wealth available to be extracted from regulation will decline as well. Furthermore, the decline in wealth is the central precursor to possible deregulation. Peltzman et al (1989) argue that this will be especially true if regulation itself led to higher prices; the evaporation of producer rents combined with consumer pressure for lower prices will make continued regulation unlikely. Upon examination of various industries in the US that were deregulated, the predictions of the economic theory of regulation fit the general experience well.[14]

2.1.4. Agency Theory and Regulatory Capture

Within the economic theory of regulation there have been some notable applications of agency theory to the public interest versus regulatory capture view. Levine and Forrence (1990) apply agency and information theory to examine the merits of the public interest and capture theory of regulation. They maintain that both the public interest and capture theory view governing institutions as a "black box" (p. 171). However, they argue that the capture theory describes a notable portion of regulatory activity and is an important foundation for future theoretical frameworks. Levine and Forrence (1990)'s point of departure stems from the view that in a democracy most government institutions are able to operate without much oversight by the public or by legislators that ultimately must answer to the public. In short, the issue with regulation is "…the inability of voters or their intermediaries to effectively limit and control regulation within the complex political system" (Levine and Forrence, 1990, p. 171). Like Posner (1974), they recognise the costliness and uncertainty of

[14] There are a couple of exceptions; the trucking and stock broking industries. In fact, Peltzman et al (1989, p. 40) comment, "Even though the [economic theory of regulation] can tell a coherent story about most of the examples of deregulation, it still cannot answer some important questions about them. Specifically, some of the examples raise questions about the design of institutions and their adaptability that have so far eluded the grasp of economists."

information gathered on a regulated industry as well as the uncertainty of the interaction among the regulator, the industry and the electorate. Thus, Levine and Forrence (1990) use agency and information theory to examine how the electorate's concerns are transformed by the outcome of the political process into policy.

Levine and Forrence (1990)'s framework is generally based on Kalt and Zupan (1984), as cited by Levine and Forrence (1990), where the principal-agent relationship between the voter and the politician ultimately leads to what they term as slack (i.e. frictions resulting from agency issues), which allows the politician to avoid his or her responsibility to the electorate. Importantly, slack enables special interest groups to persuade legislators to treat them favourably. However, Levine and Forrence (1990) point out an important distinction between this principal-agent view and the capture theory of regulation: "…unlike special-interest policies, they are not 'sold' to subsets of the polity in return for support. Rather, these acts or policies are other-regarding" (Levine and Forrence, 1990, p. 177). In other words, there will be circumstances where agency costs, coupled with other general information costs, prevent the general public from recognising an issue to be in the public interest and as such will not provide the regulator with their support. As a result, regulatory policy is not formed simply to cultivate support of a particular industry in order to gain political advantage.

The influence of slack in Levine and Forrence (1990)'s model is not straightforward. Consistent with Stigler (1971), Levine and Forrence (1990) maintain that slack can predispose politicians to maximise their own private utility. Moreover, they point out that cultivating industry support will not only help bureaucrats maintain office but it will also augment their chances of being employed in the private sector. However, Levine and Forrence (1990) argue that regulators have a choice and will not simply pursue the 'default' capture route. Rather, they have the choice of either investing further in slack or consuming it: "…when a regulator has slack, she can invest it in office holding or wealth by pursuing special interest policies, or she can consume it by pursuing other-regarding policies not favoured by her relevant polity. This slack is valuable either way, and it should not be surprising that regulation is often conducted so as to create or increase it" (Levine and Forrence, 1990, p. 180).

In conclusion, Levine and Forrence (1990) regard the propensity for regulatory capture as a function of the level of slack. That is, if publicity of the issue has not meaningfully reduced the level of slack then the possibility of capture will remain. However, the presence of slack will not necessarily result in regulatory capture as legislators may also pursue their own ideological agenda that is not held by the public.

Laffont and Tirole (1991) put forward a formalised principal-agent framework that describes the behaviour of interest groups. Their chief contribution is the incorporation of information asymmetries, which Stigler (1971), Peltzman (1976) and Becker (1983) ignore completely. It is these information asymmetries, argue Laffont and Tirole (1991), that enable a regulator to side with the industry or consumers. The private knowledge gained by the regulator enables officials to entrench their positions, whether it is getting elected to a higher office or a potential private sector career, because such specialist information will not be widely disseminated (Laffont and Tirole, 1991; Besley and Coate, 2003). Laffont and Martimort (1997) explore the implications of collusion between agents and its impact on the regulatory process and market behaviour. Consistent with Laffont and Tirole (1991), Laffont and Martimort (1997) stress the importance of the information structure between agents; private knowledge is a significant determinant of agents' ability to extract rents. Martimort (1999) argues that a dynamic life cycle process can model regulatory capture. Martimort (1999) contends that as time passes there will be greater opportunities for an interest group to collude with a regulator due to the accumulation of private information that is inherent in the relationship between the industry and the regulator. As a result, legislators will respond by implementing more stringent rules and limit regulators' discretion. This will ultimately lead to the regulated industry being tied up in more bureaucratic layers that Martimort (1999) describes as bureaucratisation.[15] This growth in red tape is "…an optimal dynamic response to the threat of capture" (Martimort, 1999, p. 931).

[15] Martimort (1999, p. 931) defines bureaucratisation as, "…the tendency of regulation to leave less and less discretion to regulators over time."

Grace and Phillips (2008) explore a recent avenue of the economic theory of regulation and the capture theory in general. Their line of inquiry is inspired by earlier research of Besley and Coate (2003) who find that elected regulators side with consumers rather than the industry they regulate. Besley and Coate (2003, p. 1177) explain; "When regulators are appointed, parties may be tempted to field candidates who would appoint pro-stakeholder regulators to further their interests in the public spending dimension … By contrast, if regulators are elected, their stance on regulation is the only salient issue so that the electoral incentive is to run a pro-consumer candidate." Grace and Phillips (2008)'s study is centred on the auto insurance industry where they find that the career prospects of legislators have a notable influence on how an industry is regulated. Those involved in the regulation of an industry with plans to move into the private sector will most likely be lenient and side with the industry in question.[16] Such a move will enhance their prospects of employment in the private sector, although simply applying a light touch to regulation on its own will not necessarily lead to an assured private sector career. Grace and Phillips (2008) point out that a regulator will need to demonstrate technical competence along with a favourable disposition.

On the other hand, it is argued that career politicians who do not rely on private sector employment in later years are most likely to side with consumers and favour strict regulation (Besley and Coate, 2003; Grace and Phillips, 2008). Indeed, Grace and Phillips (2008)'s results show that unit insurance prices are lower by almost 5% if the regulator has been identified as a consumer advocate. Pleasing the electorate will ensure a much more prosperous career path for someone seeking higher office compared to currying favour with an industry. Moreover, Grace and Phillips (2008) argue that this will be particularly true of elected rather than appointed officials. However, they find that an elected legislator will not simply be a pro-consumer regulator. In fact, their results call into question the widely held belief that elected bureaucrats will always act in the interests of consumers. Grace and Phillips (2008, p. 129) explain this contradiction by pointing out the institutional differences between

[16] This is not limited to regulators who intend to pursue a private sector career. Grace and Phillips (2008) find that when a commissioner sought a higher position within the regulator, the unit price of insurance rose by almost 6.5%.

insurance regulation and public utility regulation in the US: “State public utility commissions are composed of multiple commissioners and the size of the board typically ranges from three to seven commissioners. In contrast, the number of commissioners who exercise authority over insurance rate regulation is generally one.” It is easier for an industry to monitor and influence the behaviour of a single commissioner. If the complexity of the agency grows and number of commissioners increase, the marginal cost of influencing the regulatory process will rise as well. Hence, a regulator’s propensity to be strict or lenient also depends on the degree of monopoly power they hold over private information. Grace and Phillips (2008) posit that in the case of insurance regulation it is unlikely that other state officials (i.e. those not directly involved in the regulatory process) will learn the true profitability of the industry. This lowers the marginal cost of influencing the regulator, making capture and collusive behaviour more probable.

2.1.5. The Austrian Economic School of Thought and Public Interest Theory

Besides the “capture” view of regulation, arguably the Austrian economic school of thought provides the strongest counter argument to the ideas of public interest theory.[17] According to the Austrian philosophy, the public interest view is unattainable. Central to this conclusion is that knowledge of costs and consumer preferences are unlikely to be obtained by an omnipresent regulator. Furthermore, should inappropriate regulation be introduced, the public interest will be pushed aside by self-interested regulators being led by distorted incentives. Hence, Austrian economics maintains that the market is the most effective disseminator of information: “Variety can exist in a market which can lead to the satisfaction of more preferences than in a regulated market” (Booth, 1997, p. 693).

As mentioned earlier, some form of insurance regulation will always exist yet the preceding discussion illustrates how unlikely it is that regulation will always be the white knight of market ills. However, the central issue is the nature of government

[17] For a comprehensive discussion on the economic approach of the Austrian school see Taylor (1980).

activity and not its size. In fact, "[a] functioning market economy presupposes certain activities on the part of the state" (Hayek, 1960, p. 222). Despite this, Hayek (1960) argues that a centrally planned economy will result in a market that does not function adequately.

Central to the Austrian school of thought is the important relationship with economics and law. Specifically, the solution to the problem of too much or too little regulation is given by the degree of detail within regulatory laws. Hayek (1960) argues that a framework of general laws is the most prudent course of action because its influence will be more predictable. An important aspect of such a generalised legal framework is that it must not allow for significant administrative discretion. Again the argument here is that any bureaucrat cannot have the information in hand to produce a more efficient market by taking discretionary steps. Furthermore, a high degree of administrative discretion will result in ambiguous guidance as to how market participants should operate. It appears that the best regulatory approach is a principle-based framework without too many specifics which will likely become obsolete as the market evolves. Furthermore, although Hayek (1960, p. 224) maintains, "…a free system does not exclude on principle all those general regulations of economic activity…" such regulations of an economy will be exceptions rather than the rule. Rather, regulation for the most part will simply increase the cost of production and limit innovation. Hence, Hayek (1960)'s argument that the goal of regulation must indeed be meaningful given the costs it imposes.

The reliance on general rules and the principle of the rule of law can create an environment where market participants can learn from their mistakes. According to Hayek (1976) the ability to garner knowledge is central to the development of a sophisticated economy. Moreover, such knowledge cannot be accumulated by a central authority: "Even in the modern welfare societies the great majority and the most important of the daily needs of the great masses are met as a result of processes whose particulars government does not and cannot know" (Hayek, 1976, p. 2). Hayek (1976) stresses the importance of allowing a market to develop within a framework of appropriate rules. The advancement of an economy is a dynamic process whereby individuals are constantly adapting to changing circumstances, evaluating promising avenues of commerce and abandoning the unsuccessful ones (Hayek, 1976).

Hayek (1988) takes this idea further and applies an evolutionary approach to market development. He argues that civilisation has progressed only through individuals learning from their interactions with others via trade. In fact, Hayek (1988) maintains that government interference in Ancient Greece was the downfall of natural improvement and cultural evolution. Essential to the development of any institution is, "[t]he experimental process of adaptation to unforeseen change by the observation of abstract rules which, when successful could lead to the increase of numbers and the formation of regular patterns…" (Hayek, 1988, p. 46). In other words, markets are driven by individual knowledge rather than by prodding from a central authority.

Booth (1997) maintains the existence of deficiencies that prevent participants from learning from adverse events is the true test to determine if regulation is warranted. He argues that any introduction of law that undermines the mechanisms through which the market learns from its mistakes will be undesirable and detrimental. Specifically, introducing measures to prevent repetition of a past failure may produce such a result. Also, Booth (1997) is critical of any regulation that attempts to protect investors from risk as this can easily be construed as protecting investors from market mechanisms, which are in fact beneficial. In fact, Grace and Klein (2008) argue that the goal of insurance regulation should not be to prevent insolvencies but to mitigate the social costs resulting from insolvency. Of course the main motivation for shielding investors from risk is political. Booth (1997) points out that the necessary learning process may take such a long time that it is untenable politically, and the desire to respond promptly to financial scandals will make regulatory intervention essential.

2.2. The US Property-Liability Insurance Industry and the Effects of Regulation

Over the course of the 1960s and 1970s some states in the US moved to deregulate automobile insurers' rates. At the time of this regulatory shift within the property-liability insurance industry, research into this area had been scant. Joskow (1973) was the first to undertake an extensive examination of the property-liability industry and unearthed the distorting effects of state regulation on price levels and coverage.

In the US insurance market there are several organisational forms[18] but of particular interest are the two main methods of selling insurance. Joskow (1973) explains that the majority of short term insurance is marketed through the American Agency System where independent agents sell insurance policies to the public on the behalf of the insurers that they represent. In return, these agents receive a commission. The other method of selling insurance is direct writing. As the term implies, rather than using an independent agent, direct writers sell their own policies directly to the public. The direct writing scheme was a competitive response to the agency system, which was perceived to have prohibitive sales costs. Joskow (1973) regards direct writing as more efficient and seeks to uncover why it has not dominated the American Agency System. As will be discussed below, Joskow (1973, p. 384) argues that the major inhibitor of direct writers' growth is regulation; "It [the American Agency System] has been preserved as a combined result of price regulation by state commissions, price making in concert, and a quirk in the insurance law."

[18] There are four basic types of insurance companies. The most prominent are stock companies, which are owned by the shareholders who have provided equity capital for the firm. Conversely, mutual companies are owned by their policyholders and any retained income is distributed to them. A reciprocal exchange is an organisational form where members pool and share specific risks to which they are exposed. Last, Lloyd's Organisations consist of smaller groups of underwriters that share the risks of the policyholder (Joskow, 1973).

2.2.1. A Brief discussion of the Historical Development of the Insurance Regulatory Framework in the US

One important aspect of the property-liability insurance industry that is seldom discussed is the historical development of the regulatory framework that governs it. Joskow (1973) explains that property-liability insurance regulation stems from fire insurance. The most noteworthy trait of fire insurance regulatory policy was the imposition of monopoly-like price controls. Joskow (1973, pp. 391-392) elaborates; "The history of fire insurance and regulation is a direct consequence of the essentially non-competitive fire insurance market, dominated by cartels and essentially exempt from the federal antitrust laws, that existed through most of the first half of this [20th] century."

The move to create a non-competitive fire insurance industry in the US has an established history dating back to the colonial era of the early 1800s (Kwon, 2007). The idea was to insulate fire insurers from the destructive influence of competition (the concern was that fierce competition could lead to insolvency) but the initial regulatory setup had trouble succeeding. However, legislative amendments in the late 19th and early 20th century cemented the industry's ability to charge uniform rates (Joskow, 1973). Joskow (1973) also points out, at the same time that anti-competitive insurance regulation began to establish itself, there was fevered anti-cartel sentiment churning in most of the US. In fact, the insurance industry was subject to two significant legal challenges. The first, *Paul vs. Virginia*, in 1869 dealt with the issue of state versus federal regulation. The plaintiff, Mr Paul (an agent for a group of New York fire insurers), challenged the constitutionality of the state of Virginia's jurisdiction over out-of-state insurers and their agents, arguing that Virginia's insurance laws obstructed interstate commerce. The US Supreme Court ruled that insurance could not be regarded as commerce and as such Virginia's regulatory laws were not unconstitutional. The significance of this ruling was that not only did the authority to regulate insurers rest with individual states but more importantly, the insurance industry was exempt from federal antitrust laws (Harrington and Niehaus, 1999). The second legal challenge was brought about in 1944 by the US Department of Justice in *United States vs. South-Eastern Underwriters Association*. The US government charged that insurance rating bureaus[19] encouraged price fixing. The success of the challenge rested on whether the selling of insurance policies across states could be defined as interstate commerce. This time the US Supreme Court overturned *Paul vs. Virginia* and found that insurance is indeed commerce and therefore subject to federal antitrust laws (Harrington and Niehaus, 1999). As a result, legislators intervened (arguing that it was in the public interest to protect the industry from possible insolvencies brought about by the destructive forces of price competition) and a separate act, the McCarran-Ferguson Act, was created in 1945 re-establishing the insurance industry's indemnity from anti-competitive laws (Joskow, 1973; Harrington and Niehaus, 1999). Thus, it is this "legal quirk" that has fostered the growth of a costly, inefficient system.

The McCarran-Ferguson Act has had a significant influence on the shape of insurance regulation over the years. However, the more recent legal developments allow for increased competition between state regulated insurers with the regulatory focus on capital-based solvency ratios (Kwon, 2007; Grace and Klein, 2008). The notion of capital-based regulation is to reaffirm the central goal of regulation, the protection of policyholders' interest. Nonetheless, Kwon (2007, p. 13) points out, "…the history of insurance regulation evidences that, like other parties of interest, regulators have political and economic motives, especially a revenue generation motive – revenues from both premium and corporate income taxation."

2.2.2. The Growth of Direct Writers and Independent Pricing

In spite of the direct writing method being held back by a pervasive regulatory framework, Joskow (1973) cites the increase in concentration ratios[20] within the automobile insurance industry as a sign that insureds were moving towards the less costly direct writers throughout the 1960s and early 1970s. In fact, his empirical

[19] Insurance rating bureaus were created in response to a wave of insurer insolvencies that occurred in the late 19th century. The purpose of the rating bureaus was to ensure adequate pricing of insurance contracts in order to mitigate the risk of insolvency (Harrington and Niehaus, 1999).

[20] Joskow (1973) uses the percentage of total premiums written by agency and direct writing companies as a measure of firm concentration.

analysis confirms that direct writers have a considerable cost advantage over agency firms despite not having any significant economies of scale.

Another conundrum uncovered by Joskow (1973)'s analysis is that despite the property-liability insurance industry's inherently competitive market structure, in reality little competition exists. While, the greatest signs of price competition were displayed by the auto insurance sector (which was primarily a result of direct writers setting prices below fixed bureau rates[21]) it ultimately was very limited. Yet, Joskow (1973)'s data do show a general, albeit limited, shift towards independent pricing. However, he does not get carried away with this finding: "Even so, less than a third of the companies were filing deviating or independent rates, and both the number of companies and proportions of premiums written at off-bureau rates remained approximately constant for three years" (Joskow, 1973, p. 397). Of course a potential explanation for this lack of variation from bureau rates is that they are at or near market equilibrium levels. However, the evidence does not support this contention. Joskow (1973) examines the experience of insurers in California, where competitive pricing had been encouraged for some time, and New York, which abolished its prior pricing approval framework. The data for insurers operating in these states show that a significantly larger number of insurers charged premiums at off-bureau rates. In particular, once prior approval was removed in New York, off-bureau pricing jumped significantly. Consequently, Joskow (1973, p. 398) concludes; "The evidence supports the hypothesis that rate making in concert, combined with prior approval rate regulation, tended to discourage price competition."

As stated above, the broker insurance sales system is regarded as the costlier of the two sales methods. According to Joskow (1973), agents' organisations, along with rating bureaus, would prevent agents from selling policies of non-bureau companies. Consequently, the agents pressed for high commission rates in return for their cooperation. Overall, the costs generated by this scheme were higher than they need

[21] A bureau rate categorises risk and implements insurance rates, typically on the grounds of empirical evidence submitted to it (Frech and Samprone, 1980).

be. This explains the key strength of direct writers; by doing away with the broker middlemen they are able to operate at a lower cost and charge lower premiums. In aggregate, the reduction of insurance agents would lead to substantial cost savings. Of course this assertion is not conjecture. Joskow (1973)'s empirical analysis of expense ratios confirms that direct writers have significantly lower ratios than agency firms. In fact, taken as a whole (i.e. all insurance segments), direct writers expense ratios are lower by almost 11%. This result indicates that substantial cost savings would have been realised if direct writers were the chief vehicles of selling insurance in the 1960s and 1970s. Yet, the American Agency System remained dominant. There are several answers for this.

Generally, the personal interaction that brokers have with customers is deemed to be beneficial. An individual can receive much needed advice, particularly for less straightforward insurance issues. However, Joskow (1973) argues that agents do not provide a beneficial service to the public. When it comes to the more complicated insurance situations, Joskow (1973) charges that agents do not have the competency to deal with them. This is due either to a lack of training or experience. In fact, Joskow (1973, p. 403) asserts that "[t]he kinds of things which the small independent agents handle well are the standard recurring day-to-day insurance coverage applications." Consequently, the elimination of the agency system would not hurt the public. By Contrast, Frech and Samprone (1980) argue that agents provide consumers with a valuable service by selecting an insurer best suited to their needs. However, Frech and Samprone (1980) concede that the agency system is the costlier sales method.

Another possible reason why the growth of direct writers has been muted is that no insurer moved from the agency system to direct writing during Joskow (1973)'s period of study. The major reason for this static environment was a law preventing an insurer (using the agency system) from dealing with customers themselves; the brokers could claim property rights over the policyholders. If an insurer shifted to direct writing they would have to forfeit all their existing customers. A further hindrance on the growth of direct writing, which is almost certainly relevant today, is consumer information. Joskow (1973) points out that consumers are ignorant of price differences because comparisons are difficult and information is not readily available.

Not only has regulation stunted the expansion of direct writers but it also has restricted the amount of insurance provided to the public in general. Across the US, assigned risk plans[22] were introduced to ensure that everyone could, at the very least, obtain a basic level of coverage. However, for most members of the public it was not possible to get the coverage they wanted. Within the automobile insurance sector in particular, young drivers and individuals living in "ghetto" areas were appreciably cut off from insurance coverage in the voluntary market[23]. The cause of this restriction of coverage, as Joskow (1973) asserts, is that the regulatory framework did not allow for the creation of additional homogenous risk classes. Thus, a separate risk pool (which charged a different premium) for riskier insureds could not be established without justifying the new rate to the regulator.

2.2.3. An Examination of Stigler (1971)'s Hypotheses: Evidence from Regulated and Deregulated States

Given the insurance industry's active role in negotiating protection from antitrust laws in addition to the industry's natural predisposition to anticompetitive behaviour, Ippolito (1979) explores the motivation of the regulatory framework for US insurers. He seeks to determine if price regulation was used as "…a legal means by which the insurance industry could successfully evade antitrust laws against price fixing" (Ippolito, 1979, pp. 84 – 86). Moreover, it is the seminal work of Stigler (1971) that propels Ippolito (1979)'s line of enquiry.

[22] Assigned risk plans enable individuals, who would otherwise be denied coverage due to their risk category, to obtain automobile insurance from a group of insurers. Assignments to an individual insurer are proportional to total volume of auto insurance cover they provide and the rates charged for this plan are approved by regulators (Mowbray and Blanchard, 1955; Harrington and Niehaus, 1999).

[23] The voluntary market refers to the practice of providing cover to the desirable risks, while rejecting the less desirable ones. Individuals that are denied coverage in the voluntary market must be provided insurance through assigned risk plans (Harrington and Niehaus, 1999).

The gradual shift in legislation in the mid 1960s towards allowing greater competition within the property-liability insurance industry (and specifically the automobile insurance business) provided Ippolito (1979) with an opportunity to examine the difference between regulated states in the US and those that have been deregulated. It also provided an opportunity to investigate some of the peripheral effects associated with regulation. For instance, and in keeping with Joskow (1973)'s findings, Ippolito (1979) charges that regulatory interference disrupted insurers attempts to charge different rates to different risk groups. This according to Ippolito (1979) meant that riskier customers could only be insured through the substandard market[24] and "…[the end] result is that rates of the high-priced substandard market became more populated" (Ippolito, 1979, p. 60).

Ippolito (1979)'s study analyses annual premiums of various automobile insurance policies (including regulated and unregulated states) spanning from 1966 to 1972. As mentioned, the key proposition explored by Ippolito (1979) is that of Stigler (1971)'s economic theory of regulation. To assess Stigler (1971)'s assertions, both price dispersion and price level need to be examined. Recall that if Stigler (1971)'s theory (that regulation is sought by an industry in order to mould policy to its benefit) is true, then one should expect to observe little variation in premiums within regulated states as opposed to unregulated states (Ippolito, 1979). Similarly, a regulatory establishment that fosters anticompetitive behaviour should lead to higher insurance premiums vis-à-vis the more liberal states.

His investigation shows, contrary to Stigler (1971)'s theory, that regulation has a positive influence on price dispersion and significantly, there is no negative consequence on price distribution detected within a single category. Also, the overall impact of regulation on price levels is negative, which again contradicts Stigler (1971). However, Ippolito (1979) maintains that simply examining price levels is not

[24] In terms of the automobile insurance market, the substandard market includes drivers that are not accepted within an Insurance Rating Organisation (ISO) category. Thus, these individuals have to obtain insurance in the high-risk (substandard) market and pay a higher premium that is not necessarily commensurate with their actual risk (Mehr, 1986).

sufficient on its own to evaluate Stigler (1971)'s assertions. Even if regulation does not result in higher premiums, it is possible that insurers could benefit from lower loss ratios. That is, regulation could improve profitability by reducing claims costs or other associated expenses.[25] The total loss ratio is defined as claims-to-premiums (C/P) and consistent with the preliminary findings on price dispersion and price level, Ippolito (1979) finds that regulation has no influence on total loss ratios.[26]

By and large, the results demonstrate that regulation does not improve an insurer's pricing power. Yet, Ippolito (1979) explores other potential effects of regulation. To achieve this, he re-examines Joskow (1973)'s assertion that regulation prevents more efficient firms from charging lower, more competitive premiums. If this is indeed the case, the more efficient direct writers operating in regulated states should have lower loss ratios. In addition, such a situation will limit direct writers' market share (Ippolito, 1979).

Using a sample period of 1971 to 1973, Ippolito (1979) examines the loss ratios of insurance companies different organisational forms and finds that direct writers did indeed experience lower loss ratios over the sample period. Moreover, he finds that the lower loss ratios of direct writers are associated with a smaller market share. Indeed, this repressed market share (caused by a lack of different risk groups) inadvertently led to the overpopulated substandard market mentioned earlier. This results in a substantial benefit for smaller insurers as they are protected by the inhibited growth of more efficient insurers.

[25] Conversely, premiums could increase relative to a constant level of claims costs. Hence, examination of loss ratios is more informative (Ippolito, 1979).

[26] Ippolito (1979) also considers the effect of regulatory lag on the loss ratio. It is likely that regulators do not respond rapidly to shifting market conditions, thus it is possible that over the short run premiums could be eroded during a period of high, unexpected inflation. When regulators eventually allow a price increase, the increase should compensate for losses incurred in the periods of deflated premiums. In the context of Stigler (1971)'s theory, higher rates may be observed over the long run. However, even when the loss ratio data is extended over a longer period, no regulatory impact is found, confirming the initial results.

Given that Ippolito (1979) was unable to find any evidence that regulation impacts price levels in the long run, he concludes that his findings are more consistent with Peltzman (1976) rather than Stigler (1971). As indicated, unlike Stigler (1971), Peltzman (1976) contends that regulators are motivated chiefly by broad political support and are thus unlikely to side with insurance companies in a way that enables them to exercise significant market power.

Conversely, Frech and Samprone (1980) find evidence in support of Stigler (1971)'s excessive price hypothesis; they discovered higher unit prices for automobile insurance in regulated states. Moreover, they reveal that the impact on consumer welfare is negative. Frech and Samprone (1980) measure the impact on consumer welfare by comparing consumer surplus in states that impose price regulation and those that allow competitive pricing. An important consideration in their analysis is the value consumers place on the extra services received from non-price competition. Higher regulated rates will spur non-price competition making service-intensive sales methods (like the agency system) more attractive. However, as discussed previously, the agency system is markedly the costlier sales method which has a notable influence on welfare loss estimations. Also, Frech and Samprone (1980) find a higher concentration of independent brokers operating in regulated states. This leads them to hypothesise that if consumers valued the greater services provided by insurers, consumers would also pay for the extra services in competitive states, which is not what Frech and Samprone (1980) find. Indeed, their regression analysis shows that non-price competition has no influence on the quantity of insurance demanded leading them to conclude that consumers place no value on additional costly services provided. As a result, they demonstrate that price regulation in the property-liability industry caused notable welfare loss and a move towards freer competition would produce significant efficiency gains.

Harrington (1984), on the other hand, contends that regulation may not have resulted in excessive prices. He cites a survey by Miles (1980) who found that state insurance commissioners were most concerned with the affordability of automobile insurance. Moreover, Stelzer and Alpert (1982) as cited by Harrington (1984) suggest that property-liability regulation has more in common with natural gas regulation which is characterised by rate suppression. Pauly, Kunreuther and Kleindorfer (1986) as cited by Grabowski, Viscusi and Evans (1989) find over the period 1975 to 1980, regulation resulted in lower unit prices. They find the most notable price suppression with agency writers and consistent with Joskow (1973), Pauly et al (1986)'s evidence indicates that regulation restricted the market share of direct writers.

Another idea put forward in support of the notion that regulation does not cause higher rates is the regulatory lag hypothesis. According to Harrington (1984) regulatory lag will give the impression of inappropriate insurance prices. An insurer that requires prior rate approval will only react to changing costs with delay. For example, compared to a deregulated insurance market, if costs decline the insurer operating in the regulated market will only reduce its rates with measurable delay. Hence, in an environment characterised by falling costs, regulatory lag will give the effect of inflated premiums, which will ultimately adjust to a level consistent with the cost of the service. At the heart of the regulatory lag hypothesis is the public interest theory of regulation. That is, for insurance premiums to, in the long run, adjust to more competitive levels, it is assumed that regulators will act to protect consumers. However, Stigler (1971), Joskow (1973) and Posner (1974) amongst others have demonstrated that this is not always the case. Furthermore, Harrington (1984) cites government studies that acknowledge the problem with political interference leading to excessive insurance premiums. Nevertheless, Hanson, Dineen and Johnson (1974) as cited by Harrington (1984) find evidence in support of the regulatory lag hypothesis. They find nothing to suggest that automobile insurers engaged in collusive pricing and argue that regulation simply exacerbates the underwriting profit cycle. Witt and Miller (1981) as cited by Harrington (1984) also investigate the regulatory lag hypothesis. They examine auto insurance loss ratios in regulated and deregulated states over the period 1971 to 1979 for both agency and direct writers and find that direct writers make up a larger proportion of the market in deregulated states. However, they find that while the mean loss ratios are higher in regulated states the difference is not significant. Also, a notable feature of Witt and Miller (1981)'s study is the application of beta (extracted from time-series regressions on loss ratios spanning 1971 to 1979) as a measure of systematic underwriting risk. The weighted average beta in competitive states is lower than regulated states, indicating that the automobile insurance underwriting cycle is more pronounced in regulated states, which is consistent with Hanson et al (1974)'s findings.

These mixed results illustrate that the full effect of the move towards deregulation in the automobile insurance industry may have not been wholly reflected by the data. This is not surprising given that these studies examined the initial stages of the shift to deregulation. Hence, Grabowski et al (1989) revisit the issue of insurance rates and availability across regulated and deregulated states. In the context of these inconsistent findings, Grabowski et al (1989, p. 278) conclude; "There appears to be no 'generic' time-invariant effect of regulation on the automobile insurance industry."

Their study is one of significance because rather than only examining how regulation impacts price levels and coverage, they explore the post-deregulation experience of insurance firms as well. What is more, they find that over the period 1974 to 1981, regulation lowered insurance premiums.

Grabowski et al (1989) ascribe the mixed results of previous literature to the changing nature of regulation since the 1970s. As the 1970s was a period of rapid inflation, Grabowski et al (1989) argue that regulators were pressured to constrain premiums. They maintain the best response by policy makers would be to deregulate insurers, as competitive forces would bring premiums down. Alternatively, if regulators constrain insurance premiums, it is likely that prices would be restricted to a point below competitive levels, leading to excess demand, which in turn will lead to quantity rationing. As a result, Grabowski et al (1989) expect to find a large proportion of individuals in assigned risk plans and a large amount of uninsured motorists in regulated states. Moreover, Grabowski et al (1989) argue that should regulation result in below average rates one may observe a deterioration in quality and service levels.

Grabowski et al (1989) argue that in a competitive market equilibrium premium payments will equal expected claims costs as well as sales and various administrative costs. Thus, when the market clears the following general relationship will hold

$$P_{ij} = L_{ij} + e_{ij}P_{ij} + r_{ij}P_{ij} \tag{2.2}$$

Where, P_{ij} represents the premiums for state i for category j drivers, L_{ij} is the expected cost of claims to insurers in state i for category j drivers, e_{ij} represents miscellaneous expenses (such as sales, service and administrative costs) in state i for category j drivers and r_{ij} is the competitive return for risk bearing in state i for category j drivers.

Grabowski et al (1989) employ a fixed-effects panel data model where the major dependent variable is price per unit of benefit payments (i.e. insurance premium) to insurance holders. In addition, they make use of time dummy variables for each year to control for business cycle effects and any other shocks unique to a specific period. Their results show that the coefficient controlling for regulation is significantly negative. Thus, regulation results in lower unit prices.[27] Furthermore, they find that states with more stringent regulation have significantly lower unit prices. This effect is still present even when the variable for stringent regulation is included along with the general, less stringent regulation variable. However, Grabowski et al (1989) concede that the regulation coefficients are similar across all estimated equations, which could indicate that some omitted variable could be causing the differences observed in the states with stringent regulation.

Grabowski et al (1989) also examine whether regulation restricts insurance availability. To determine if this is the case, they use the percentage of a state's drivers in the involuntary market as their dependent variable. They find that

[27] The unit price (p_{ij}) = P_{ij}/L_{ij} = $[1/(1 - e_{ij} - r_{ij})]$, which is a slight variation of the premiums in state i for category j.

regulation has a significant and positive relationship with the size of the involuntary market, which indicates that regulation does indeed limit the availability of insurance. The effect is particularly strong for the states that impose stringent regulations. Hence, one can infer that particularly strict regulation will depress insurance premiums, which will lead to quantity rationing.

These results contradict the earlier studies that found support for Stigler (1971)'s hypothesis that regulation should result in higher insurance premiums. However, Grabowski et al (1989) explain this incongruity by reiterating the pressure that legislators and regulators faced to restrain insurance rates given the intense inflationary environment that characterised the 1970s. Furthermore, Grabowski et al (1989)'s results show that regulators responded by restraining prices and they argue that deregulation was an alternative strategy pursued by legislators as a means to lower premiums. Indeed, their results show that after the switch to competitive markets unit prices fell and this became more pronounced over time. The effect of deregulation was especially strong for direct writers as their low cost structure enabled them to take full advantage of setting their premiums at below bureau rates. Thus in this context, regulation actually provided a price floor for direct writers.

Harrington (1992) again examines the issue of rate suppression. He argues that the political pressure brought to bear on insurers to limit rate increases has caused insurance premiums to be pressured below levels that would occur under a more competitive environment.

According to Harrington (1992) rate suppression incentivises insurers to reduce the quality of their services and shrink their coverage. The reduction in coverage will result in a larger involuntary market, exposing insurers to a riskier group of insureds without being able to charge a premium in line with the exposure to risk as premiums in the involuntary market are set at bureau rates. In the extreme, rate suppression will result in withdrawal from the market altogether.

Restricting premiums below competitive levels also has a significant impact on investment. Harrington (1992) points out that insurers need to make substantial investments in distribution networks, human capital and other outlays that are not related to specific policies. The costs of such investments are usually recovered by charging higher rates over several years. Rate suppression prevents the firm from receiving an adequate return. Moreover, rate suppression will also hurt the value of existing investments. In short, rate suppression will discourage insurers from making future investments. Insurers could respond by reducing their cost structures as, generally, forcing premiums below levels that provide an adequate return is unlikely to result in bankruptcy provided premiums are enough to cover variable costs. However, Harrington (1992) argues that this is only likely to be a temporary solution as future premiums may be reduced further to reflect efficiency gains. Consequently, the effect of rate suppression is not immediate and insurers will be gradually squeezed and forced into bankruptcy.

Harrington and Danzon (1994) argue that regulatory price suppression was so severe that the resulting property-liability insurance underpricing was a significant precursor to insurance crises. What is more, its effect was compounded by the insurance industry's response. Harrington and Danzon (1994) argue that insurers underestimated their loss forecasts in order to drive growth and hide deteriorating financial conditions. The general narrowing of margins over the 1979 to 1984 period resulted in premiums that were too low to cover operating costs and anticipated claims. Ultimately though, this unsustainable underpricing resulted in a notable jump in premiums in the latter half of the 1980s.

Apart from regulatory impetus, Harrington and Danzon (1994) suggest two main reasons for the increase in underpricing. First, underpricing may have gained momentum due to weak incentives to guard against insolvency and the moral hazard issue arising from limited liability. Second, insurers with more optimistic information could have fallen into the trap of charging premiums below what would have occurred under full information conditional expectations.

Under the moral hazard view of underpricing, an insurer will pursue a strategy of rapid growth by gaining market share without giving much cognisance to default risk (given shareholder limited liability and third-party guarantees). However, the incentive to underprice policies will be strongest when franchise value deteriorates to such an extent that it induces a propensity to gamble by the insurer (Harrington and Danzon, 1994; Babbel and Merrill, 2005). The moral hazard hypothesis predicts that riskier insurers will charge lower premiums and will initially grow faster than their more soundly managed counterparts. Another indicator of moral hazard induced underpricing suggested by Harrington and Danzon (1994) is a substantial use of reinsurance. Furthermore, they purport that stock companies are more predisposed to moral hazard than mutual companies.

On the other hand, it is possible that inexperienced firms simply placed too much weight on their private information. That is, these firms can be stung by the winners curse even if conditional estimates are unbiased. Harrington and Danzon (1994, p. 521) explain that “[t]he reason is that the conditional expectation of loss given that the insurer sells the policy (i.e., has the lowest bid) exceeds the conditional forecast based on the insurer's private information and public information.” Hence, insurers with lower loss estimations should be associated with lower premiums and characterised by rapid growth.

Harrington and Danzon (1994)’s empirical analysis suggests that the moral hazard hypothesis explains a fair proportion of underpricing. They find that loss forecast revisions are positively and significantly related to the amount of reinsurance acquired. Moreover, they find that stock companies experienced faster growth with lower rates compared to mutual companies, indicating that moral hazard exerts a greater influence on stock companies. Consequently, Harrington and Danzon (1994) contend that this propensity for risk taking combined with rate suppression resulted in the extreme adverse environment experienced in 1985 and 1986.

Nelson (2000), Beaver, McNichols and Nelson (2003) and Gaver and Paterson (2004) document loss reserve manipulation in order counteract the effect of rate suppression. Nelson (2000) focuses on the relation between the settlement lag and loss reserve manipulation. She argues that troubled insurers will apply a higher discount rate to their loss reserves, thereby biasing the loss estimate downward. Moreover, a stringent regulatory environment will exacerbate the problem, as insurers will be incentivised to demonstrate that they can charge lower premiums to pursue growth. Beaver et al (2003) investigate the extent to which property-liability insurers use their loss reserves to smooth earnings. For instance, they find that less profitable insurers understate their loss reserves whereas the more profitable firms tend to overstate their loss reserves. Beaver et al (2003) also examine the effect of ownership structure (e.g. private, public and mutual companies) on the propensity to understate loss reserves. Interestingly, they find that private companies do not appear to understate their loss reserves whereas public and mutual property-liability insurers do (even though Beaver et al (2003) find most of the private insurer's profits fall into the left hand side of the distribution tail, the effect is still present even when financial condition is controlled for).[28] However, they also find that mutual companies overstate their loss reserves across the entire profit distribution, which is consistent with the notion of mutual companies understating their profits so as to limit dividend payments to policyholders. Gaver and Paterson (2004), using a similar methodology to Beaver et al (2003), find that insurers manipulate their loss reserves to avoid breaching certain solvency ratio bounds. Consistent with Beaver et al (2003), Gaver and Paterson (2004) report a greater incidence of profitable insurers overstating their loss reserves but find "[d]ownward reserve bias is most prevalent in the subset of financially struggling firms…" (Gaver and Paterson, 2004, p. 395). Taken as a whole, this suggests that rate suppression may incentivise greater earnings management. However, Gaver and Paterson (2004) admit that their analysis cannot lead to a firm conclusion that this behaviour by insurers results in inappropriate decisions by regulators.

[28] Beaver et al (2003) caution that private insurers could be using other undetected methods to manage their earnings.

Grace and Leverty (2009) argue that onerous rate regulation will cause insurers to react by overstating their loss reserves. Moreover, they suggest that rate suppression implicitly transfers wealth from the insurer to consumers by forcing premiums below competitive levels. Grace and Leverty (2009, p. 9) disagree with Nelson (2000) and discount the notion that insurers would react by under-reserving; "If regulation suppresses rates below the economic cost of writing business, then under-reserving would validate rates that are not sufficient to provide a reasonable return." Hence, the overwhelming incentive for an insurer will be to overstate their loss reserves and thereby understate their earnings. Consistent with Beaver et al (2003) and Gaver and Paterson (2004)'s findings, Grace and Leverty (2009) indeed uncover loss reserve overestimation in response to stringent rate regulation. In addition, Grace and Leverty (2009), like Beaver et al (2003), find that distressed firms will understate their loss reserves compared to their stronger counterparts. On the face of it, this may appear to be a minor accounting issue, however, the propensity to manipulate loss reserves due to the regulatory environment can, ultimately, have real impact. The moral hazard inherent in this behaviour can lead to increased insolvencies because of untimely, or inappropriate, regulatory intervention (Harrington and Danzon, 1994; Gaver and Paterson, 2004).

2.2.4. Background to Risk-Based Capital (RBC) Standards and Subsequent Solvency Measures

As mentioned previously, the introduction of capital-based solvency rules is somewhat recent. Although, the idea of risk-based capital standards was not entirely new when the RBC framework was initially proposed. In fact, regulators had been mulling over new solvency rules for some time. Cummins, Harrington and Klein (1995) trace the motivation behind the RBC framework to a surge in property-liability insolvencies occurring in the 1980s. Moreover, in the initial years of the 1990s, the property-liability industry experienced a grim economic environment. Unprecedented catastrophe losses, caused by Hurricane Andrew in 1992 and the Northridge Earthquake in 1994, combined with depressed premiums caused a new wave of bankruptcies (Cummins and Xie, 2008). The general consensus was that the financial quality of the industry declined prompting regulators to implement new measures, which ultimately were inspired by the Basel framework on banking regulation (Grace and Klein, 2008).

Prior to the introduction of the RBC framework, solvency regulation depended on fixed minimum capital standards. Due to limitations of fixed minimum capital standards, in 1990 the NAIC began deliberations of its replacement with RBC standards. The advantage of the RBC framework, vis-à-vis fixed minimum capital standards, is that it is designed to vary with the type of exposure faced by insurers and it was formally adopted in 1993 for property-liability insurers (Klein, 1995).

Grace, Harrington and Klein (1998) list four key components of the RBC formula. The first risk factor is asset risk. This encompasses the risk of both default and market value declines. Second, credit risk, which captures uncollectible reinsurance and other receivables. Third, underwriting risk, which relates to pricing and reserve miscalculations, and fourth, off-balance sheet risk. Grace et al (1998, p. 217) explain; "The formulas apply factors to various amounts reported in (or related to) the annual statement to determine RBC charges for each type of risk. A covariance adjustment is made to the accumulated RBC charges to account for diversification between major risk categories." Should an insurer's total level of capital[29] decline below its measured level of RBC, various company and regulatory actions will be triggered.[30]

Klein (1995) observes that there was little opposition at the time of the implementation of the risk-based capital formula for life and health insurers yet there was wide concern voiced by the property-liability industry. The criticism centred on the RBC formula's ability to measure the unique risks faced by the property-liability

[29] The total level of capital is the total figure of RBC (after the adjustment for diversification effects). Hence, the overall capital figure is referred to Total Adjusted Capital (TAC) and is the yardstick for RBC comparisons (Klein, 1995).

[30] Under the RBC framework, the Company Action Level is triggered when capital falls below 200% of RBC and the Regulatory Action Level is triggered after declines below 150% of RBC. Any further declines below 100% will warrant regulatory control (Grace et al, 1998).

industry. For instance, some charged the RBC was inaccurate when it came to capturing the risk of long-tail property lines (relative to short tail property lines) and the charge for reinsurance was too high, which would discourage the use of reinsurance to mitigate risk (Klein, 1995).

In light of the conceptual short-comings of the NAIC RBC formula an ancillary measure, the NAIC's Financial Analysis Trading System (FAST) was developed in order to broaden the measure of an insurers overall financial strength rather than relying on a static capital measure. The FAST system is an expansion of the NAIC's Insurance Regulatory Information System (IRIS), which was a basic framework for vetting troubled insurers. Hence, the FAST system was created to identify financially weak insurers using a variety of financial ratios and variables that incorporate a greater range of risk factors than the RBC formula. Furthermore, the FAST ratios and ranking system were created using regulatory experience combined with statistical analysis with the goal of generating high rankings for insurers regarded as weak; a notable improvement over RBC (Grace et al, 1998).

Another notable development is the insurer rating agency A. M. Best's Capital Adequacy Relativity (BCAR) ratio. Unlike the RBC formula, the BCAR ratio's interest rate element captures the potential capital loss of a fixed income portfolio resulting from rising interest rates. In addition, the risk factors of the BCAR ratio vary markedly from that of the RBC formula. Most significantly, the BCAR ratio has a different methodological approach in that it considers qualitative factors such as financial flexibility and reinsurance quality (Pottier and Sommer, 2002). Also, as will be discussed in chapter four, probabilistic functions based on the expected policyholder deficit (EPD) approach, which is based on Merton (1977)'s deposit insurance insolvency put, have also been introduced (Butsic, 1994). However, Barth (2000) is rather dubious about the technique and argues that it will understate the cost of insolvency in certain situations.

Despite these additions to the RBC framework, the cornerstone of US insurer regulation remains rule-based with a focus on accounting data. Given the limitations of this solvency measure, which is discussed in the next section, there has been wide discussion in the US insurance industry of a shift away from rule-based regulation to a principle-based framework, such as Europe's Basel inspired Solvency II, where capital requirements rely more on economic rather than statutory capital (Grace and Klein, 2008; Vaughan, 2009).

2.2.5. The Efficacy of the Risk-Based Capital Formula

A previous examination of the RBC model by Grace, Harrington and Klein (1993) as cited in Cummins et al (1995) indicates that the RBC formula is not a robust indicator of solvency. In fact, their study shows that only a fraction of the insurers examined had RBC ratios that fell within the prescribed range warranting regulatory intervention. Consequently, Cummins et al (1995) attempt to discern if appropriate modifications to the RBC formula can improve its ability to promptly identify insolvent insurers. However, Cummins et al (1995) point out that simply applying a stricter solvency ratio will increase the number of financially sound firms being identified as needing regulatory action. In other words, Cummins et al (1995) seek to improve the RBC model by improving the Type I and Type II error trade-off[31] and therefore avoid interfering with the decisions of safe insurers.

Cummins et al (1995, p. 6) state; "They [insurers] seek to take maximum advantage of the law of large numbers by insuring the largest possible number of independent exposure units, using reinsurance to pool risk with other insurers and across national boundaries." However, they argue that imperfect liability estimation and long-tail policies necessitate the need for capital-based ratios because these unknowns threaten solvency. Moreover, a decline in premiums and unanticipated increases of liability

[31] A Type I error is the probability of incorrectly identifying a strong insurer as weak. Alternatively, a Type II error is the chance of identifying a weak insurer as financially sound. Implicit in the Type I/II error trade-off is power, which is the probability of correctly rejecting the null hypothesis when it is false (Gujarati, 2003). In other words, in the context of solvency regulation, power can be regarded as the probability of correctly classifying an insurer as weak (Grace et al, 1998).

claims will exacerbate the problem. Overall, Cummins et al (1995) maintain that risk-based capital regulation provides a compelling solution to these issues. However, they concede that the RBC framework does have some disadvantages. For instance, the chance of insolvency depends on many variables that are difficult to quantify, making the specification of the right amount of capital through a formula a challenging exercise. Furthermore, Cummins et al (1995, p. 11) point out that "[a]n inaccurate risk-based capital formula may distort investment, underwriting, and reinsurance decisions of well-managed insurers, leading to less effective diversification. This could actually reduce safety levels for financially sound insurers and lead to higher premium rates for any given level of safety."

Cummins et al (1995) examine insolvency experience of insurers and the actual levels of capital over the period 1989 to 1991. Specifically, using multiple logistic regression models, they compare RBC ratios each year for insurers that subsequently failed and those that survived in order to determine the within-sample Type I and Type II error rates. Hence, their objective is to isolate the relation between insolvency risk and RBC ratios and see if the Type I – Type II error rates cannot be improved with additional indicators such as firm size and organisational form.

Cummins et al (1995)'s results show that only a small fraction of surviving firms had RBC ratios less than or equal to the statutory requirement of 200%. The RBC ratios for the firms that failed vary from year-to-year in the sample period. For instance, 28% of insurers that failed had RBC ratios less than or equal to 200% in 1989 and 1990 but this rises to 49% in 1991. An intriguing finding of Cummins et al (1995) is that median RBC ratios are, overall, significantly higher for the failed insurer samples in each year. However, Cummins et al (1995) note two exceptions. First, the median RBC ratio is lower for failed insurers compared to those that survived in 1989. Second, the median level of loss reserve risk-based capital to surplus (RBC/S) for insolvent firms in 1989 is only marginally larger than surviving firms.

In the logistic regression analysis, Cummins et al (1995) find the estimated coefficients for RBC to be positive and significant. Also the addition of assets as an estimator is positive and significant, demonstrating the efficacy of RBC ratios and net assets as a predictor of insolvency. Moreover, they find Type I error rates to be significantly lower when the natural logarithm of total assets is introduced into the model. Although, the effect is much less pronounced for Type II error rates. Overall, Cummins et al (1995) conclude that the inclusion of the natural logarithm of assets will improve insolvency detection.

Grace, Harrington and Klein (1998) follow on from Cummins et al (1995) and try to identify further improvements to the RBC formula. Their study is motivated by the conceptual shortcomings of the RBC formula as well as the costs associated with incorrectly identifying a financially sound insurer as weak. Hence, the issue that Grace et al (1998) examine is whether the FAST solvency screening mechanism has more predictive power of insurer insolvency (and thus, lower Type I error rates) than the RBC ratio.

Grace et al (1998, p. 219) argue that an effective solvency framework will mitigate the costs associated with insolvency by "(1) helping to establish legal grounds for regulatory action against weak insurers, (2) encouraging regulators to take timely action even though political pressure may encourage forbearance, and (3) encouraging insurers to reduce risk for which private incentives for safety are suboptimal". However, as Grace et al (1998) mention, the critical element of any regulatory solvency framework is the ability to identify troubled firms early enough to take corrective action. Hence, the importance of a properly specified measure of insolvency.

Consequently, Grace et al (1998) test two main hypotheses. The first is that the RBC formula is at least as powerful as FAST in correctly identifying weak insurers. Their second hypothesis is that if RBC is less, or no more powerful than FAST, the combination of RBC and FAST should provide additional information, making the combined measures more powerful in correctly classifying weak insurers than FAST alone.

If the RBC ratio is an effective predictor of insolvency, on average financially weak insurers should have lower RBC ratios. However, Grace et al (1998) point out that this alone is not a sufficient condition for an efficient regulatory system. Grace et al (1998, p. 221) maintain that "…an efficient RBC system would equate the marginal benefits of increased accuracy in the formula with the marginal cost of information." Thus, Grace et al (1998) acknowledge that there will be practical limits to the level of accuracy of any measure. Despite this, they maintain that greater accuracy will soften the distortions of regulatory intervention and consequent adverse market reactions. Alternatively, as mentioned previously, the predictive ability of RBC and FAST could be augmented if the two measures are combined. Grace et al (1998) argue that a less efficient RBC ratio (compared to FAST) could still contribute additional information that is not contained in the FAST measure.

To test their hypotheses, Grace et al (1998) compare the predictive ability of an insurer's RBC to surplus ratio (RBC/S), its FAST score and its premium-to-surplus (P/S) ratio at a range of Type I error rates. Their sample period spans from 1989 to 1991 and the sample contains all stock, mutual, reciprocal, and Lloyd's property liability insurers included in the NAIC RBC database. They examine the new information hypothesis using a multiple logistic model that includes both the RBC/S ratio and the FAST score. Also, Grace et al (1998) are cognisant of Cummins et al (1995)'s empirical findings and include the individual components of RBC (as opposed to only including RBC/S) in other multiple logistic regression models.

Similar to Cummins et al (1995), Grace et al (1998) find that only a minority of insurers that failed had RBC/S ratios that would have necessitated regulatory intervention. Furthermore, in the failed company sample, the majority of failed firms had capitalisation rates greater than the 200% company action cut-off level. Another interesting result is that Grace et al (1998) find decidedly few insolvencies in the larger premium insurer subsample, which indicates the relative importance of premium income.

The results of the univariate model indicate that the RBC ratio performs less favourably than the FAST score. That is, the RBC ratio is less powerful and more susceptible to Type I errors and consequently, the hypothesis that the RBC formula is at least as powerful as the FAST score does not hold. Moreover, the FAST score has the highest power for the entire sample of failed insurers at each Type I error rate. The P/S ratio is found to be as informative as the FAST score for the same subsample at Type I error rates greater than 10%.

The most significant finding of Grace et al (1998) is that the combination of RBC/S and FAST in the multivariate model shows no improvement in power compared to the FAST score on its own. The weak performance of the RBC ratio is also evident in the multivariate models; the RBC/S coefficient is never significant and often negative. Moreover, this model included the subsample of failed insurers with direct premiums that exceeded $50 million, which is the same subsample where RBC/S performed well. However, Grace et al (1998) admit that the performance of RBC/S can be dependent on the sample year. Yet, taken as a whole Grace et al (1998, p. 232) conclude that such results "…undermines the already fragile evidence of superior performance of RBC/S in the univariate comparisons for this subsample."

Given the critical importance of the accuracy of a solvency measure, Pottier and Sommer (2002) re-examine the efficacy of the NAIC's RBC ratio and FAST score, and compare them to the proprietary models developed by the insurer rating agency A. M. Best; the Capital Adequacy Relativity (BCAR) ratios. The BCAR is of particular interest because unlike the NAIC's solvency measures, the BCAR incorporates qualitative as well as quantitative data (Pottier and Sommer, 2002). The RBC ratio, the FAST score and BCAR all use capital as their cornerstone of assessing insolvency risk, yet Pottier and Sommer (2002) acknowledge that capitalisation is only one indicator of financial strength. Thus, Pottier and Sommer (2002) attempt to discern if broader insolvency measures such as BCAR provide additional information.

Pottier and Sommer (2002)'s empirical analysis examines insurers that became insolvent during the period 1996 to 1998. Their predictor variables are based on solvency measures recorded in 1995. That is, they use data obtained in 1995 to predict insolvencies that occurred in the years 1996 to 1998. Consistent with Cummins et al (1995) and Grace et al (1998), Pottier and Sommer (2002) include firm size and organisational form as predictor variables in their logistic regressions.

Unlike Grace et al (1998), Pottier and Sommer (2002)'s results show that all the solvency risk measures are lower than average for the failed insurers. Moreover, Pottier and Sommer (2002) find that the insolvent insurers are generally smaller than the ones that survived and this difference is significant. However, Pottier and Sommer (2002) are not surprised by this result. They cite previous research that found a high rate of bankruptcy among smaller insurers and a possible explanation for this is that smaller insurers may be underpricing their contracts to gain market share and thereby expose themselves to greater variability in claims costs (Harrington and Danzon, 1994; Pottier and Sommer, 2002). One implication of this finding is that larger insurers could reach the same level of insolvency risk with a lower level of relative capital. Alternatively, the lower need for capital could be due to a lack of underpricing pursued by larger insurers.

Pottier and Sommer (2002) also use the capital-to-assets ratio in their analysis as a simple baseline proxy for insolvency risk to which the more complicated measures can be compared to. Surprisingly, their logistic regression results show that the simple non-risk adjusted capital-to-assets ratio has lower Type I error rates than the RBC ratio. This demonstrates that more sophisticated risk measures are not always the best predictors of insolvency. However, their results also demonstrate the superiority of the FAST score and BCAR relative to the RBC ratio. The FAST score and BCAR correctly identified 87% and 81% of insolvent insurers, respectively, whereas the

RBC ratio correctly singled out only 55% of insolvent insurers. Furthermore, consistent with Grace et al (1998) including the RBC ratio with BCAR or FAST in the regression model does not improve predictive ability. Thus, Pottier and Sommer (2002) arrive at the conclusion that RBC does not include any further information than that contained in BCAR or FAST. In fact, Pottier and Sommer (2002, p. 111) present a bleak pronouncement on the RBC ratio; "One obvious conclusion that can be drawn from the results thus far is that RBC is a poor predictor of insolvency."

It is unclear why the RBC ratio is a poor predictor of insolvency but Pottier and Sommer (2002) suggest that the greater dispersion of the RBC ratio along with the larger amount of outliers (compared to the FAST and BCAR data) could explain its poor showing. Also, there is a ray of light for the RBC ratio in that using the ranked RBC measures as independent variables in the regressions improves the Type I error rate thereby improving its use as an indicator of financial strength. Despite this, BCAR and FAST remain superior measures (Pottier and Sommer, 2002).

Generally, Pottier and Sommer (2002) maintain that the proprietary risk measures are better indicators of solvency than the public NAIC risk proxies. Interestingly, they note that capital should not be the only consideration in a solvency risk measure as "…substantial improvements in predictive ability are possible when broader measures of risk are used instead of risk-based capital measures" (Pottier and Sommer, 2002, p. 114).

Tougher regulation is unlikely to counter the pitfalls of the RBC ratio either. In fact, Klein, Phillips and Shiu (2002) find that stricter regulation actually increases the probability of insolvency. They maintain that regulation incentivises insurers to hold less capital. Furthermore, Klein et al (2002) argue that this reduction in capital encourages firms to increase their leverage.

Klein et al (2002) examine the responses of individual insurance firms to price regulation by analysing insurers' capital levels under different state regulatory systems. While the likes of Ippolito (1979), Grabowski et al (1989) and Harrington (1984, 1992) have shown that regulation in the US has resulted in significant suppression of insurance premiums, Klein et al (2002) argue that regulators may tolerate higher insurance premiums if it means they will avoid having to deal with an insolvent insurer under their supervision. In fact, some insurers may even augment their leverage so as to raise the likelihood of bankruptcy and thus, justify a higher premium (Spiegel and Spulber, 1994). In particularly onerous regulatory environments this tactic may be taken to such an extreme that the combination of less capital and increased leverage reduces the financial quality of the insurer. This according to Klein et al (2002) will encourage insurers with large initial amounts of capital to trim their businesses or exit markets completely.

The notion that regulated insurers may increase their debt levels to justify a higher premium is based on Spiegel and Spulber (1994)'s examination of price, capital structure and regulation. They assert that capital structure is a key factor in the interaction between the financing decisions undertaken by a regulated firm and the ultimate price imposed by regulators. In other words, Spiegel and Spulber (1994, p. 424) argue that "…the firm chooses its equity and debt strategically to affect the outcome of the regulatory process." Moreover, Spiegel and Spulber (1994)'s equilibrium model demonstrates that a firm will increase its leverage as a result of regulation and the regulator will react by increasing the regulated price to reduce the chance of presiding over a potential insolvency.

Therefore, Klein et al (2002) test two central hypotheses. The first is that insurers operating in states that impose price regulation will have higher leverage ratios compared to insurers doing business in unregulated environments. The second hypothesis purports that states with stricter regulation will have insurers with even greater leverage ratios than states that subject insurers to less strict regulation. Klein et al (2002) use cross-sectional data that is subject to varying degrees of price regulation. Their data set is drawn from all mandatory cost filings of property-liability insurers contained in the 1997 NAIC Data Tapes and their sample period spans from 1991 to 1997. Klein et al (2002) use two main proxies for regulatory stringency. The first is based on the size of the residual market in private passenger automobile and workers' compensation in a state.[32] The second regulatory stringency proxy is based on an index of regulatory intervention known as the External Climate Index (ECI).

Their results are consistent with the hypothesis that strictly regulated insurers have higher leverage ratios than those subject to less stringent regulation and unregulated insurers. Furthermore, Klein et al (2002)'s results show that regulated insurers are less profitable. Also, they find insurers that underwrite long-tail lines maintain higher surplus as a result of greater uncertainty inherent in long-tail lines. More importantly, Klein et al (2002, p. 96) argue that their study demonstrates that "…restrictive price regulation creates distortions and can thwart regulatory solvency goals by encouraging insurers to maintain higher levels of leverage and financial risk, all other things being equal."

[32] The purpose of the residual market is to prevent supply shortages resulting from insureds being denied coverage in the voluntary market. The residual market "…force[s] insurers that write a given type of coverage in a state to collectively supply coverage to most if not all applicants. An insurer must participate in the residual market if it wants to sell coverage in the voluntary market (Harrington and Niehaus, 1999, p. 157).

3. Capital Asset Pricing Model (CAPM) Based Property-Liability Insurance Frameworks

3.1. The Development of the Insurance Capital Asset Pricing Model (CAPM)

One of the initial formal articulations of the return on equity for property-liability insurers is provided by Ferrari (1968). He argues that despite the industry's traditionally low debt-to-equity ratio, property-liability insurers operate with a notable level of implicit leverage. The source of this inherent leverage is from the "…deferred nature of insurance liabilities" (Ferrari, 1968, p. 296). This form of insurance leverage is central to Ferrari (1968)'s understanding the relation between return on assets and return on equity. His formal argument is as follows. The total after-tax return to an insurer is a combination of its underwriting profit (*U*) and investment income (*I*). Thus, total return = $U + I$. Ferrari (1968) defines shareholders' equity as total assets (*A*) less reserves and other liabilities (*L*). In other words, shareholders' equity = $A - L$. From this it follows that the total return on equity is

$$T/S = (U + I)/S \qquad (3.1)$$

Where *T* is the total after-tax return to the insurer and *S* is shareholders equity.

When total assets and the investment income earned from reserves are included, total return on equity can be expressed as

$$T/S = [(A/A)(U + I)]/S = [A(U + I) + IL - IL]/AS \qquad (3.2)$$

This is then simplified to yield

$$T/S = [I/A(1 + L/S\ ) + (U/P)(P/S)] \qquad (3.3)$$

Where *P* is premium income.

Equation (3.3) shows that the return on equity for a property-liability insurer is determined by the insurance leverage factor (*1* + *L*/*S*), underwriting profit (*U*/*P*) and, what Ferrari (1968) describes as insurance exposure, (*P*/*S*). In fact, Ferrari (1968) argues that *P*/*S* ratio can be viewed as a proxy for solvency risk. Furthermore, he points out that the *P*/*S* and *U*/*P* ratios can be thought of in the same way as asset turnover and financial leverage influence return on equity calculations for non-insurance firms.

Ferrari (1968, p. 298) maintains that a minor manipulation of Equation (3.3) reveals an expression "…plainly analogous to the use of debt capital for financial leverage":

$$T/S = I/A + L/S\ (I/A + U/L) \qquad (3.4)$$

In fact, Ferrari (1968, p. 298) comments; "With this viewpoint, underwriting losses can be considered as the 'interest' that the insurer has paid for the use of [*L*] dollars of reserve capital." Overall, Ferrari (1968)'s framework illustrates that if the ratio of investment return on assets, *I*/*A*, exceeds the absolute ratio of underwriting profit, *U*/*L*, it will be in the company's interest to continue writing insurance, even in the event of underwriting losses.

Biger and Kahane (1978) propose a Capital Asset Pricing Model (CAPM) based model to analyse underwriting profit. Importantly, they extend their analysis from the single-line insurance case to the generalised situation of multiple-line insurers.

In a similar fashion to Ferrari (1968), Biger and Kahane (1978) present the single-line case where they define the return on equity of an insurer as

$$r_y = (1 + L) - Lr_u \qquad (3.5)$$

Where L is the premium to equity ratio, r_u is the underwriting profit or loss within a single period and r_p is the return on investment income generated within a single period.

Using their definition of an insurers return on equity, Biger and Kahane (1978) determine the systematic risk of equity (r_y). That is, they obtain the beta of the entire insurer. This in turn can be broken down into the systematic risk of the insurance portfolio (β_u) and that of the asset portfolio (β_p). Thus, the insurance beta is articulated using the familiar equation:

$$\beta_y = \frac{Cov(r_y, r_m)}{Var(r_m)} \qquad (3.6)$$

Substituting the return on equity equation (r_y) into β_y, yields

$$\begin{aligned}\beta_y & \frac{Cov\left[\left\{(1+L)r_p - Lr_u\right\}, r_m\right]}{Var\left(r_m\right)} \\ & = \frac{\left(1+L\right)Cov\left(r_p, r_m\right) - \left(L\right)Cov\left(r_u - r_m\right)}{Var\left(r_m\right)} \\ & = \left(1+L\right)\beta_p - L\beta_u\end{aligned} \qquad (3.7)$$

Substituting β_y into the original CAPM equation obtains

$$E(r_p) = r_f + \beta_p\left[E(r_m) - r_f\right] \qquad (3.8)$$

Biger and Kahane (1978) then articulate the expected underwriting return loss on an insurer

$$E(r_u) = r_f + \beta_u\left[E(r_m) - r_f\right] \qquad (3.9)$$

Equation (3.9) illustrates that, in equilibrium, expected underwriting loss has no relationship with the premium-to-equity ratio. Moreover, it shows that the expected underwriting loss will be equivalent to the expected return on the asset portfolio, provided they have the same betas. Another important feature of Biger and Kahane (1978)'s derivation is that for a zero beta underwriting portfolio, its expected loss will equal the negative of the risk-free rate.

The arguments above can be extended into a generalised multi-line insurance case. For instance the one period rate of return on equity can be generalised as

$$r_y = \sum_{i=m+1}^{n} (x_{ui})(r_{ui}) - \sum_{i=1}^{m} (x_{pi})(r_{pi}) \qquad (3.10)$$

Where r_{ui} is the loss rate on insurance line i, for $i = m + 1,\ldots, m$. r_{pi} is the return on investment for $i = m + 1,\ldots, n$. x_{ui} and x_{pi} are the premium-to-equity ratio and assets-to-equity ratio for $i = 1,\ldots, m$ and $i = m + 1,\ldots, n$, respectively.

Applying Equation (3.10) to the general CAPM risk-return framework and the following relationship results:

$$\sum_{i=1}^{m} x_i E(r_i) = r_f \sum_{i=1}^{m} x_i + \sum_{i=1}^{m} x_j \beta_i \left[E(r_m) - r_f\right] \qquad (3.11)$$

As in the single-line example, this multi-line derivation shows that the overall insurance portfolio has no relation to financial leverage (i.e. the premium-to-equity ratio). Also, should all insurance lines have a β_i equal to zero, then the expected multi-line underwriting return will equal the risk free rate. However, Biger and Kahane (1978) caution that this solution is not unique.[33] In addition, Biger and Kahane (1978, pp. 126 – 127) warn that the unobservable nature of underwriting activities can lead to estimation problems: "It follows that evaluation of the systematic risk of underwriting, which is not based on market returns but on reported

[33] See Biger and Kahane (1978, pp. 125 – 126). Furthermore, the multi-line extension results in several possible insurance pricing combinations for a regulator to monitor. Thus, Biger and Kahane (1978) maintain that a regulator will not be able to prescribe an insurance rate that will not disrupt market equilibrium.

profits, may result in biased estimates of the coefficients … In essence, the problem is similar to the well known problem of whether or not accounting betas are consistent substitutes for market betas.”

Fairly (1979) and Hill (1979) both examine the impact of corporate tax on insurance pricing. However, Fairly (1979) focuses on the influence of fractional ownership whereas Hill (1979)’s central concern is the estimation of a fair rate of return for a property-liability insurer.

Hill (1979) outlines his framework to determine the market premium as follows

$$p = \frac{PN - L - E}{N} \tag{3.12}$$

Where *p* is the premium for a single policy and *L* is the expected loss. *N* is the number of policies written and *E* refers to the expenses associated with writing insurance policies. Solving for *P*, the overall unit premium is obtained:

$$p = \frac{L + E}{1 - N} \tag{3.13}$$

This result is the basis for Hill (1979)’s derivation of a fair premium. However, he first expresses the individual premium as a rate of return, which requires the inclusion of shareholders equity (or capital), *K*:

$$r_i = \frac{r_j(PN + K) - L}{K} \tag{3.14}$$

To derive the equilibrium premium, Equation (3.14) is substituted into Equation (3.9) to arrive at

$$P = \frac{cy - \lambda Cov(L, r_m)/N}{r_f} \tag{3.15}$$

Where λ is a simplification of the systematic risk component $(r_m - r_f)/Var(r_m)$.

Hill (1979)'s derivation makes it plain to see that the fair premium is the sum of expected loss (*cy*) and a risk premium, and both are discounted by the risk-free rate. The risk premium in turn depends on the systematic risk of losses (i.e. the covariance of losses with that of the overall market). Moreover, it is also apparent that if losses are uncorrelated with the overall market returns, then Equation (3.15) reduces to the special case where $P = cy/R_f$. Another salient point of Hill (1979)'s fair premium formula is that the R_j (return on investments) term falls away because it is assumed that the insurer earns the best competitive rate from its investments. Another insight of the fair premium formula is its treatment of the inherent leverage an insurer obtains from the lag between receiving policyholder funds' and making actual claim payments. It shows that the expected rate of return of this implied loan is that of the risk-free rate. Hill (1979, p. 180) explains that "[t]he fair return to the loan is the riskless rate because, by assumption, the firm always holds enough capital to make the policies riskless against default." The only potential for risk is borne from the investment portfolio in that the insurer could choose to invest in riskier assets to boost returns. However, only the shareholders bear the extra risk, not the policyholders. Last, Hill (1979)'s model also illustrates the irrelevance of the premium-to-capital ratio. Hill (1979) maintains that in the case of insurance, capital is not "real" in that capital has no role in the creation of an insurance policy. In other words "…an insurance contract has no 'return to capital' dimension" (Hill, 1979, p. 180). Hence, there is no reason why the level of capital should have any influence on premiums.

However, this is only the case if taxes are ignored. Investment income of an insurance company faces a double layer of taxation, first at the corporate level and again at the shareholder level. As a result, insurers will be incentivised to hold the least amount of capital possible. Thus, in the presence of taxes, the premium-to-capital ratio will no longer be irrelevant and Hill (1979) reworks his fair premium model to incorporate this:[34]

$$P = \frac{cy - \lambda Cov(L, r_m)/N + (K/N)T(r_f - 1)}{tr_f + T} \tag{3.16}$$

Hill (1979)'s reworking of Equation (3.15) takes the capital per policy (*K*/*N*) explicitly into account, thus illustrating the importance of the premium-to-capital ratio. Furthermore, the presence of taxes also influence the discount rate; the denominator (tR_f+ T) is an after-tax discount rate.

Fairly (1979) considers the situation where an insurer operates as a mixture of a mutual and a stock company and the shareholders participate only in a predetermined fraction of investment returns. Fairly (1979) assumes an insurer will generate an investment return of r_A and if the firm is leveraged, which in this case is defined as the premium-to-capital ratio, *s*,[35] the firm can earn a multiple, *k*, of r_A. Furthermore,

[34] Several steps have been omitted. See Hill (1979, pp. 181 – 182).

[35] This definition differs somewhat to that of Biger and Kahane (1978) who define financial leverage as the premium-to-equity ratio.

Fairly (1979)'s model is expressed in terms of an insurer's underwriting profit margin. The underwriting profit margin is articulated as

$$P = -r_L L. \tag{3.17}$$

In order for a positive underwriting profit to be attained, the return on liabilities, r_L, must be negative. Moreover, Equation (3.17) also shows that in a competitive market, the expected underwriting profit margin is negative. It follows that if shareholders own a fraction, x, of the insurer they will have an expected return of xr_Aks on policyholders' funds. For N lines of insurance, the expected underwriting profit margin will be

$$P_N(x) = -xk_N r_f + \beta_{p,N}\left[r_m - r_f\right] \tag{3.18}$$

Fairly (1979) maintains that the trade-off of the different ownership structures (i.e. the difference between a mutual and a stock company) is that a shareholder owned insurer can operate with a lower underwriting margin, and resultantly offer a lower premium, due to the offsetting income earned on the investment portfolio. On the other hand, in the case of a mutual insurer, the policyholders will benefit from the asset portfolio but the firm will have a higher profit margin and a higher premium. However, they will be compensated with dividends based on the firm's investment performance.

Finally, if taxes are introduced in Fairly (1979)'s model, as in Hill (1979), the capital-to-premium ratio, s, becomes an important factor:

$$p_N = -kr_f + \beta_{p,N}\left(r_m - r_f\right) + \left(\frac{t}{(1-t)s}\right)r_f \qquad (3.19)$$

Here again, due to the double taxation faced by an insurer, Equation (3.19) demonstrates that the smallest possible holding of capital will be desirable.

Cummins (1991) expands and provides a more formal argument of Ferrari (1968)'s framework:

$$\begin{aligned} Y &= I + \pi_U \\ &= r_A A + r_U P \end{aligned} \qquad (3.20)$$

Where, *Y* is net income, *I* is Investment income, π_u is underwriting profit, *A* represents total assets, *P* represents premiums, r_A is return on assets of investment and r_U refers to underwriting return.

Furthermore, Equation (3.20) can be expressed in terms of return on equity:

$$\begin{aligned} r_E = \frac{Y}{E} &= r_A \frac{A}{E} + r_U \frac{P}{E} \\ &= r_A\left(\frac{L}{E} + 1\right) + r_U \frac{P}{E} \end{aligned} \qquad (3.21)$$

Where, *E* represents equity, *L* represents liabilities (*A* – *E*), *s* = *P/E,* which is the premiums-to-surplus ratio and *k* = *L/P*, which is the liabilities-to-premiums ratio (or funds generating factor).

As in Ferrari (1968)'s framework the return on equity for an insurer is made up of two components; one being investment income and the other underwriting income. Cummins (1991, p. 284) states; "The investment income part consists of the rate of return on assets multiplied by a leverage factor, (*ks* + 1), which is a function of the premiums-to-surplus ratio and the funds generating factor … The underwriting return component is the product of the underwriting profit ratio and the premiums-to-surplus ratio." Thus, Cummins (1991) articulates a formal expression for the leverage created by issuing insurance contracts.

Cummins (1991) demonstrates that Equation (3.21) can be manipulated as follows:

$$r_E = r_A + s(r_A k + r_U) \qquad (3.22)$$

Stated in words, an insurer will earn r_A on the investment of equity in addition to the net return on underwriting multiplied by the underwriting leverage ratio, *s*. If the insurer chooses not to offer any insurance policies (i.e. $s = 0$), it will essentially be an investment company that invests at the rate of r_A. Provided $r_A k > -r_U$, writing insurance at a negative underwriting profit will result in a higher return on equity[36]. In addition, Equation (3.22) illustrates that insurance lines with long tail payouts will support larger underwriting losses relative to lines with shorter payout tails. Longer tailed payouts enable an insurer to hold onto policyholders' funds for longer and will therefore result in a larger funds generating factor (*k*).

[36] Cummins (1991) points out that if $k = 1$ and $r_A = 0.1$ the insurer will make money by writing insurance provided $r_U > -0.1$ (or underwriting loss is less than 10% of premiums).

Doherty and Garven (1995) explore how the leverage generated from insurance policies will affect an insurer. Furthermore, they relax the assumption that interest rate changes will have no influence on insurers leverage, which the CAPM-based models of Biger and Kahane (1978), Fairly (1979), Hill (1979) and Cummins (1991) make no allowance for. The influence of leverage on an insurers underwriting return is central to Doherty and Garven (1995)'s analysis. Yet equally important is the relationship leverage has with the asset-liability mismatch of an insurer and the implications of the changing degrees of leverage on capital. Doherty and Garven (1995) explain the process as follows. A decline in the value of the insurer's assets, caused by an unanticipated spike in claims, will result in a lower level of surplus (or insurance equity), which in turn translates into higher leverage. An insurer facing an over-leveraged situation will be vulnerable to further adverse shocks in claims frequency. In addition, large declines in surplus will necessitate an injection of equity from external capital markets, which compared to internal financing is a costly source of capital (Myers and Majluf, 1984). Doherty and Garven (1995) argue that insurers will find it preferable to deviate from their target leverage ratio and gradually readjust using retained earnings to rebuild their surplus. However, the short run implications of a jump in leverage will be less coverage at a higher premium to compensate for the interim drop in surplus.

Implicit in the analysis above is the influence a change in interest rates has not only on liabilities but on equity as well. More to the point, the central issue, as mentioned earlier, is the asset-liability mismatch. Insurance equity depends on the durations of the assets and liabilities, as well as the leverage of the company. Hence, the appropriate measurement for the reaction of equity to shifting interest rates is duration. The duration of equity has a meaningful influence on the degree of the contraction in the supply of insurance, and consequently the level of premiums, for a given increase in interest rates. For example, insurers with high equity duration will experience a larger decline in the value of equity for a rise in interest rates. This will result in a larger increase in leverage, which will lead to a more pronounced decline in the amount of insurance supplied and higher premiums. On the other hand, Doherty and Garven (1995) point out that it is possible for insurers to mitigate this by matching the durations of their asset and liability portfolios.[37]

Changes in leverage, and consequently the riskiness, of an insurer also have important implications for potential conflict between shareholders and policyholders. While an increase in leverage could be unintentional, as a result of an adverse claims experience, Staking and Babbel (1995) argue that insurers will also intentionally alter their leverage. In fact, the ability to alter their risk structure once a policy has been issued results in a transfer of risk onto policyholders. In other words, an increase in risk levels will benefit shareholders at the expense of policyholders. Here again, the duration of the firm's assets and liabilities as well as the level of capital play a significant role. Stacking and Babbel (1995) maintain that one of the main methods of increasing leverage is by attaining greater exposure to interest rate risk, which brings in duration and the asset-liability mismatch. That is, the degree of the asset-liability mismatch and their respective durations will have a meaningful impact on leverage and risk. The solution to this problem is for shareholders to commit a meaningful amount of capital, or surplus, in order to bind them to maintain appropriate risk levels.[38]

On the other hand, an appropriate capital cushion will benefit shareholders as well. Cummins (2000) maintains that the single motivation for allocating capital is to maximise shareholder wealth. Moreover, Cummins (2000) points out that within the insurance industry the desire to show healthy accounting profits can cause a firm to lose sight of its main objective of boosting shareholder value. Hence, capital allocation is one way of ensuring that an insurer focuses and improves upon its

[37] Staking and Babbel (1995) argue that high levels of interest rate risk coupled with fluctuations in equity can be difficult to hedge.

[38] A meaningful amount of capital will also mitigate conflicts by serving as a safeguard against unexpected claims; "The greater the capital, the more certain policyholders are that they will receive compensation for insured losses" (Staking and Babbel, 1995, p. 692).

economic profitability. In fact, Cummins (2000) suggests that the literature on banking regulation and capital adequacy is applicable to insurers.[39]

Cummins (2000) revisits the application of the CAPM to the insurance industry and broadens his analysis to consider an insurance company that provides more than one line of insurance. For example, a firm that writes two lines of insurance would have the following net income:

$$I = r_A A + r_1 P_1 + r_2 P_2 \qquad (3.23)$$

Where i is net income, r_1, r_2 is the rate of return from lines 1 and 2, respectively, A represents assets and P_1, P_2 represent premiums from lines 1 and 2, respectively.

Return on equity is determined by equating the insurer's assets to its equity plus the liabilities created by the two business lines

$$r_E = r_A(E + L_1 + L_2)/E + r_1 P_1 /E + r_2 P_2/E \qquad (3.24)$$

Cummins (2000) then decomposes beta into

$$\beta_E = \beta_A(1 + k_1 + k_2) + \beta_1 s_1 + \beta_2 s_2 \qquad (3.25)$$

[39] Cummins (2000)'s view that Basel style banking regulation is applicable to insurance is motivated by the notion that policyholders can be viewed as quasi debt holders.

Where β_E, β_A, β_1, β_2 are the betas for assets and insurance risk of lines 1 and 2 of the firm, respectively. k_1, k_2 represent liability leverage ratios for lines 1 and 2 and s_1, s_2 are the premium leverage ratios for lines 1 and 2.

Overall, the decomposition of β_E illustrates that the required return on equity is the beta of the asset multiplied by one plus the leverage ratios for insurance lines 1 and 2. Following this, the formula aggregates the individual line's underwriting return betas, multiplied by a line-specific premium-to-surplus ratio. This, Cummins (2000, p. 13) argues, is a "…theoretical justification for the traditional rule of thumb leverage ratio that has been used for years in the insurance industry – the premium-to-surplus ratio."

The final derivation of the model indicates that for each line of insurance provided, an implicit interest payment must be made for the use of policyholder funds but the insurer will receive a rate of return determined by the systematic risk of the line. Moreover, Cummins (2000, p. 13) notes an important implication of this result: "It is not necessary to allocate capital by line using the CAPM, but rather to charge each line for at least the CAPM cost of capital, reflecting the lines beta coefficient and leverage ratio."

The developments of the preceding CAPM-based insurance frameworks can be consolidated and formalised into what is known as the insurance CAPM. The insurance CAPM presents a formal expression for the equilibrium underwriting return:

$$r_U = -kr_f + \beta_U (r_m - r_f) \tag{3.26}$$

Where, β_U is the underwriting beta, $\frac{Cov(r_U, r_m)}{\sigma^2_{r_m}}$

The central feature of this model is the treatment of insurance policies. That is, insurance policies are viewed as tantamount to debt and $-kr_f$ represents an interest credit for the use of policyholder funds. Specifically, "[t]he insurer borrows funds from policyholders, invests funds at r_A and pays claims (retires the debt) k periods later" (Cummins, 1991, p. 286). Moreover, the insurance CAPM will only reward a firm for bearing systematic risk, hence, increasing the number of policyholders will only increase the expected underwriting return if underwriting profits are positively correlated with the returns of the market.

This suggests that insurers should only be concerned with systematic underwriting risk, yet insurers must also be cognisant of extreme tail events that are not taken into account by the CAPM. Moreover, the funds generating factor, *k*, can only approximate the present value of payouts that occur at different periods of time, which can result in large errors in determining premiums. What is more, given the importance of interest rate risk in insurance pools, the CAPM's assumption of a constant risk-free rate oversimplifies a key risk factor (Cummins, 1991). Another issue with the CAPM is that it assumes any debt is riskless. In reality insurance debt carries an appreciable probability of default that will not be appropriately priced by the CAPM. Cummins (2000) notes that estimates of the underwriting beta can be difficult to obtain due to data limitations. Although recent advances in estimating the cost of capital have been made, which mitigates this problem somewhat (Lee and Cummins, 1998 as cited in Cummins, 2000; Cummins and Phillips, 2005). Also, the insurance CAPM will inherit the pitfalls of the original model (Roll, 1977). For instance, beta is not the sole driver of returns. Fama and French (1992), Lakonishok, Shleifer and Vishny (1994) and Fama and French (1996), amongst others, demonstrate that other economic factors such as size, book-to-market and the price-earnings ratio have significant explanatory power on equity returns. By relying only on beta, the CAPM will ignore other important determinants of the cost of capital (Cummins, 2000).[40]

3.2. General Risk-Based Capital Allocation Models: Risk-Adjusted Return on Capital (RAROC), Economic Value Added (EVA), Value at Risk (VaR)

Given the importance of risk-based capital systems in short-term insurance regulation, accurate estimations of the cost of capital are paramount. Indeed, the preceding discussion of the CAPM-based models of insurers illustrates this. Once the cost of capital has been determined it can be applied to capital allocation. Cummins (2000), Myers and Read (2001) and Sherris (2007) amongst others, argue that capital allocation is particularly salient to the insurance industry because capital is needed to absorb any larger than anticipated claims and mitigate the risk of insolvency.

Besides enabling an insurer to focus on maximising its market value, the pairing of capital allocation and value maximisation will allow an insurer to measure the performance of individual business lines and establish if each line is generating a return sufficient enough to compensate its cost of capital. Of course, in order to determine the cost of capital it is essential to ascertain the capital needed to provide different types of insurance, where the general rule of thumb is riskier lines require more capital than less risky lines (Cummins, 2000).

[40] Cummins and Phillips (2005) address these issues in their paper "Estimating the Cost of Equity Capital for Property-Liability Insurers". They demonstrate that different models can result in significantly different estimates of the cost of capital (and hence, a major impact of line-by-line capital allocation). Using the standard single-factor CAPM to generate beta estimates, they use the full information industry beta (FIB) methodology to decompose the cost of capital line-by-line. Furthermore, Cummins and Phillips (2005) expand the FIB methodology to incorporate the Fama and French (1992) three-factor model and find that the cost of capital for insurers is notably higher if the Fama and French (1992) model is used. The FIB methodology is somewhat new and is designed to overcome the limitations of the pure-play approach. Rather than ignoring cost of capital estimates for conglomerates, which is the case with the pure-play method, the FIB methodology makes use of a sample of conglomerate and specialist firms in order to isolate the effect of different business lines on a firm's cost of capital.

Cummins (2000) provides a general mathematical statement of the capital allocation problem. The amount of capital allocated to business *i,* is proportional to the company's total equity capital. That is, if x_i is the proportion of capital allocated to business i, the amount of capital allocated to business i is C_i. This is obtained by multiplying the total capital C by x_i. For an insurer with N lines, then $\sum_{i=1} x_i \leq 1$ and $\sum_{i=1} C_i \leq C$. The individual sum of all the insurer's lines will be less or equal to its total capital (Merton and Perold, 1993). This is an example of a basic method of allocating capital line-by-line.

Cummins (2000) illustrates that once capital has been allocated to a line of insurance, it is possible to calculate a firm's risk-adjusted return on capital (RAROC). Cummins (2000) defines RAROC as:

$$RAROC_i = Net\ Income_i / C_i \tag{3.27}$$

In other words, RAROC is net income earned from an individual line, divided by the capital allocated to that individual line. Following the banking literature, Cummins (2000) notes that net income is valued after taxes and interest expense. Despite interest expense being a banking term, Cummins (2000) argues that it can apply to insurers because it can be translated into an underwriting loss. That is, an underwriting loss is tantamount to interest expense, which needs to be taken into account when determining the rate of return of an individual business line.

Clearly, RAROC needs to be compared to an appropriate benchmark in order for a company to determine if its rate of return compensates for the risk taken. The simplest method suggested by Cummins (2000) is to compare the firm's risk-adjusted rate of return with its cost of capital, which can be determined using the insurance CAPM. Indeed, if RAROC matches or exceeds the insurer's cost of capital, then the commitment of further funds will lead to value maximisation. On the other hand, continuing to pour more resources into business lines that do not generate a sufficient return to compensate for risk will result in a lower market value. In this case it the best course of action may be to windup the business line (Cummins, 2000).

Another method of determining whether or not an individual line is enhancing market value is economic value-added (EVA). Cummins (2000, p. 6) observes that, "[e]conomic value-added measures the return on an investment in excess of its expected or required return." Thus, EVA can be expressed as:

$$EVA = Net\ Income - (WACC)(C_i) \qquad (3.28)$$

Where *WACC* is the weighted average cost of capital.

EVA expresses the cost of capital into a Rands and cents figure; hence, positive values of EVA indicate that an insurance line is adding value. Cummins (2000) also notes that EVA can be modified slightly to be expressed as a rate of return calculation. The modified formula is called economic value added on capital (EVAOC) and is defined as:

$$EVAOC_i = (Net\ Income_i/C_i) - r_i \qquad (3.29)$$

It follows that the calculation of EVA necessitates an estimation of the cost of capital. Besides the techniques of Biger and Kahane (1978), Fairly (1979), Hill (1979), Cummins (1991) and Doherty and Garven (1995) discussed previously, one long-standing approach is the pure play method. This method estimates the cost of capital by finding other firms that offer only one line of business where its cost of capital is applied to the firm in question (Hillier, Grinblatt and Titman, 2008). However, the pure play method is awkward for insurers as companies seldom write a single line of business (Cummins, 2000). A more feasible alternative is the use of full information betas (FIB). This technique uses data of firms that write multiple lines of business in order to run regressions so as to determine the cost of capital (Cummins and Phillips, 2005).

Another methodology that could provide a useful framework for capital allocation is value at risk (VaR). VaR is most widely used in the banking and investment banking sector. Moreover, it has become an essential tool for measuring the market risk exposure of a firm's trading book. Cummins (2000) maintains that VaR shares similar traits with other concepts used in the insurance industry. Thus, VaR is potentially a useful, easily understood application that can estimate the exposure of an insurance line to damaging losses.[41]

VaR can be applied to capital allocation by using exceedence probabilities. The exceedence probability is the chance that losses of a particular line will exceed any anticipated loss including any capital allocated to that line as well (Cummins, 2000). Specifically, the exceedence probability is defined as follows:

$$\Pr\left[Loss_i > E(Loss_i) + C_i\right] = \varepsilon \qquad (3.30)$$

[41] VaR can be interpreted in two different ways. It can be defined as the maximum amount that a company could lose over a predetermined time period for a specified probability. On the other hand, it can be interpreted as the minimum loss that will occur within a specified time period and probability. For example, a VaR of R100 million with a 5% probability can be interpreted as such; five days out of every hundred (i.e. 5% of the time) the firm can expect to lose more than R100 million. That is, the firm's minimum loss during the time period and given probability will be R100 million. The alternative interpretation is that R100 million is the maximum loss that will occur 95% of the time. In other words, a company can expect to lose no more than R100 million in 95 trading days out of a hundred.

Where $Loss_i$ = expected value of loss from line *i,* C_i = capital allocated to line *i.*

Cummins (2000) demonstrates that capital is allocated by equalising the exceedence probabilities across lines of insurance. For lines that have expected losses that differ in size, the exceedence loss is articulated in terms of ratios to expected losses. Overall, to attain the specified exceedence probability, the model shows that insurance lines with comparatively high risk must be given more capital relative to expected losses.

Despite the exceedence probability's usefulness, Cummins (2000) notes several issues with the methodology. First, for every insurance line, a low probability exceedence may result in a capital requirement that surpasses the total capital of the firm. In this case, the firm will have to raise more capital or accept a greater exceedence probability. Second, the exceedence probability ignores any diversification effects across business lines. Third, like VaR, the exceedence probability does not provide any information as to the size of loss should the exceedence level be breached.

4. The Insolvency Put Option

Within the field of capital allocation techniques, the doctrine of the insolvency put option has risen to prominence. This technique has considerable influence on line-by-line capital allocation and insurance premiums. It also determines the riskiness of the insurer itself (Myers and Read, 2001; Mildenhall, 2004; Venter, 2004; Sherris, 2006; Sherris and van der Hoek, 2006).

As mentioned in the introduction, the option pricing theory breakthroughs of Black and Scholes (1973) and Merton (1973b), opened the door to the nearly limitless applications of option methodologies to value any financial claim on a firm. Yet it was Merton (1974)'s use of option pricing techniques to value risky debt as well as his seminal paper "An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees" published in 1977 that set the idea of the insolvency put option in motion.

4.1. Merton (1974, 1977)'s Seminal Derivation of the Insolvency Put Option Model

Merton (1974) argues that the value of corporate debt is determined by three factors. First, and arguably the most important, is the rate of return generated by risk free securities such as sovereign bonds or high quality corporate bonds. The second factor relates to the particulars of the issue such as maturity, coupon rate, degree of seniority, etc. The third factor is the chance of default by the issuer.

Central to Merton (1974)'s framework is that the value of a firm can be expressed as a stochastic process by the following differential equation:

$$dV = (\alpha V - C)dt + \sigma V dz. \qquad (4.1)$$

Where α is the expected rate of return for a given time horizon, *C* represents payouts of either dividends or debt, σ is the standard deviation of returns (i.e. riskiness of the firm) for a given time horizon and *dz* describes a Gauss-Wiener process.

Merton (1974) considers the simple situation of a firm with only two types of assets; debt and equity. In the case of the debt claim the firm promises to repay the principal, *B* dollars, at a pre-specified maturity, *T*.[42] Should the firm default on its debt then the bondholders will assume control of the firm, in order to recover their loan, and shareholders will receive nothing. Stated alternatively, when the firm's debt matures, i.e. $\tau = 0$, shareholders will only repay it if $V(T) > B$, where their residual claim is worth $V(T) - B > 0$. If $V(T) \leq B$ shareholders will choose to default as they would have to pay in extra to make the debt claim whole. Thus, at maturity the value of debt can be expressed as

$$D(V, 0) = Min(V, B). \qquad (4.2)$$

In keeping with Modigliani and Miller (1958), Merton (1974) expresses the total value of a firm as the sum of its capital structure, $V = D(V, \tau) + E(V, \tau)$, where *D* and *E* is the value of debt and equity, respectively. Hence, $D = V - E$, and using Black and Scholes (1973) option pricing methodology, Merton (1974) argues that the value of debt can be expressed as

$$D(V,\tau) = Be^{-r\tau}\left[\Phi\left\{h_2\left(d,\sigma^2\tau\right)\right\} + \frac{1}{d}\Phi\left\{h_1\left(d,\sigma^2\tau\right)\right\}\right] \qquad (4.3)$$

[42] Merton (1974) assumes that the firm commits not to issue additional senior debt or pay any dividends to shareholders prior to maturity.

Where

$$d = \frac{Be^{-r\tau}}{V}$$

and

$$h_1\left(d,\sigma^2\tau\right) = -\left[\frac{\frac{1}{2}\sigma^2\tau - \log(d)}{\sigma\sqrt{\tau}}\right]$$

and

$$h_1\left(d,\sigma^2\tau\right) = -\left[\frac{\frac{1}{2}\sigma^2\tau + \log(d)}{\sigma\sqrt{\tau}}\right]$$

Finally, Merton (1974) expresses the value of debt in terms of its yield, *D*(*V*, *t*)/*B* and Equation (4.3) becomes,

$$R(\tau) - r = \frac{-1}{\tau}\log\left[\Phi\left\{h_2\left(d,\sigma^2\tau\right)\right\} + \frac{1}{d}\Phi\left\{h_2\left(d,\sigma^2\tau\right)\right\}\right] \qquad (4.4)$$

Merton (1974) argues that the term $R(\tau) - r$ is tantamount to a risk premium. Moreover, this risk premium is dependent on the firms riskiness, σ, and the debt burden of the firm, $Be^{-r\tau}/V$.

The above analysis of pure discount bonds can be extended to more complicated issues such as risky coupon bonds. Merton (1974) assumes that a firm will pay coupons continuously and that the debt issued by the firm is perpetual, i.e. $\tau = \infty$. However, unlike other interest bearing securities, the coupons on corporate bonds are not paid continuously but rather at specific discrete times of the year, either annually or semi-annually. To reconcile the assumption of continuous coupon payments, Merton (1974) defines $\overline{C}$ as the sum of previous discrete payments:

$$\overline{C} = \sum c_i \delta\left(\tau - \tau_i\right) \tag{4.5}$$

where δ (.) is the Dirac delta function[43] and τ_i is the time remaining until the i^{th} coupon payment.

[43] According to Kunkel and Mehrmann (2006) the Dirac delta is a generalised function representing an infinitely sharp peak bounding a unit area. Formally, the Dirac delta distribution is defined via

$$(\delta, \phi) = \phi(0) \text{ for all } \phi \in D.$$

$$\begin{aligned}\phi(0) &= -\left(\phi\left(\hat{t}\right) - \phi\left(0\right)\right)\\ &= -\int_0^{-\infty} \dot{\phi}\left(t\right) dt\\ &= -\int_R H\left(t\right)\dot{\phi}\left(t\right) dt\\ &= -\left\langle H, \dot{\phi}\right\rangle = \left\langle H, \dot{\phi}\right\rangle\end{aligned}$$

where

$$H(t) = \begin{Bmatrix} t \\ 0 \end{Bmatrix} \begin{matrix} \text{for } t < 0 \\ \text{for } t \geq 0 \end{matrix}$$

Taking all this into account, Merton (1974) formally articulates the value of a risky coupon bond as follows

$$D(V,\infty)=\frac{\overline{C}}{r}\left[1-\frac{\left(\dfrac{2\overline{C}}{\sigma^2 V}\right)^{2r/\sigma^2}}{\Gamma\left(2+\dfrac{2r}{\sigma^2}\right)}M\left(\frac{2r}{\sigma^2},2+\frac{2r}{\sigma^2},\frac{-2r}{\sigma^2 V}\right)\right] \qquad (4.6)$$

With the framework to describe corporate liabilities now mapped out, Merton (1977) directly applies put option valuation techniques to loan guarantees such as deposit insurance. Specifically, he maintains that not only can a bond be valued using option pricing theory but insurance contracts such as loan guarantees as well.

Merton (1977) outlines the logic of comparing a loan guarantee with that of a put option with the following example. A firm issues a loan with a face value of *B* dollars and as before, the value of the company's equity is *Max*(*0*, *V* – *B*) and its debt is *Min*(*V*, *B*). If the loan has been issued with an explicit guarantee "[i]n effect, the guarantor has ensured that the value of the firm's assets on the maturity date be at least *B* dollars" (Merton, 1977, p. 7). Should the firm default on its debt the guarantor will step in and honour the obligation to its insured bondholders. In the event of bankruptcy, the guarantee provides a cash inflow to the firm, which obviously will be an outflow for the guarantor. Hence, at maturity, the value of the guarantee can be expressed as

$$G(0)=Max(0,B-V). \qquad (4.7)$$

From here the Black and Scholes (1973) method can be used to value a loan guarantee with a specific maturity:

$$G(T) = Be^{-rT}\left[1-\Phi\left(x_1\right)\right] - V\left[1-\Phi\left(x_2\right)\right] \tag{4.8}$$

where

$$x_1 = \frac{\ln\left(B/V\right)+\left(r+\sigma^2/2\right)\left(T\right)}{\sigma\sqrt{T}}$$

$$x_2 = x_1 - \sigma\sqrt{T}$$

Attaching a guarantee to a bond can in theory make any issue riskless. Thus, if the market value of a risky bond is $Be^{-[R(T)T]}$, the addition of a guarantee will therefore result in $G(T) + Be^{-[R(T)T]} = Be^{-[r(T)]}$. This can be expressed as a yield and in keeping with Merton (1974) the cost of the guarantee will be

$$\frac{G(T)}{Be^{-rT}} = 1 - e^{-[R(T)-r]T} \tag{4.9}$$

Merton (1977) illustrates the application of Equation (4.8) specifically to the problem of determining the cost of deposit insurance. If bank deposits are guaranteed (as indeed they are in the US by the Federal Deposit Insurance Corporation (FDIC) and implicitly by the government itself) then they too can be viewed as riskless debt. Hence, $D = Be^{-rT}$. Furthermore, if the cost of the deposit insurance is expressed as per dollar of covered deposits, $G(T)/D$, then the guarantee can be manipulated slightly to generate

$$g(d,\tau) = \left[1 - \Phi(h_2)\right] - \frac{1}{d}\left[1 - \Phi(h_1)\right] \qquad (4.10)$$

where

$$h_1 = \frac{\ln(d) + \frac{\tau}{2}}{\sqrt{\tau}}$$

and

$$h_2 = h_1 - \sqrt{\tau}$$

Equation (4.10) demonstrates that a constant deposit-to-asset ratio, $d = D/V$, and variance, $\tau = \sigma^2 T$, results in a constant price for deposit insurance, just as one would expect in the case of a put option. Similarly, any rise in volatility and time horizon will cause higher insurance rates. However, unlike the original put option valuation model, interest rates will only have an indirect impact on the cost of deposit insurance via the influence they have on the deposit-to-asset ratio.

4.2. Doherty and Garven (1986)'s Seminal Application of Option Pricing Theory

Doherty and Garven (1986) provide another seminal application of option pricing theory to insurers. While their use option-pricing techniques to model financial claims of shareholders and policyholders was inspired by Merton (1974, 1977), growing dissatisfaction with the CAPM provided additional impetus. Doherty and Garven (1986) propose using option-pricing techniques to obtain the fair premium of an insurance contract as well as competitive rates of return on underwriting and shareholders' equity. The major advantage of this approach, they argue, is that it will take into account the chance of insolvency.

Doherty and Garven (1986) put forward a single-period model of an insurance firm where shareholders pay in equity of S_0 and policyholders pay premiums of P_0.[44] As a result, the initial cash flow to the insurer is

$$Y_0 = S_0 + P_0 \qquad (4.11)$$

It is assumed that claims payments and taxes are paid at the end of the period, thus the terminal cash flow for the period is defined as

$$\tilde{Y}_1 = S_0 + P_0 + (S_0 + kP_0)\tilde{r}_i \qquad (4.12)$$

Where $\tilde{r}_i$ is the rate of return generated from investments and k is the funds generating coefficient. Like the older CAPM-based models, Equation (4.12) takes into account the lag between when premium income is received and when claims payments are

[44] Doherty and Garven (1986) define premiums as net of production and marketing expenses.

actually paid. Doherty and Garven (1986, p. 1033) state; “The value $\tilde{Y}_1$ is allocated to various claimholders in a set of payoffs having the characteristics of call options.” They view shareholders as holding a long position in a call option on the insurer’s asset portfolio. However, equity holders also have a short position in a call option on the taxable income that will result from the insurer’s assets. From the perspective of policyholders, they have a long position in the insurer’s asset portfolio but have a short call option position on that portfolio. Essentially, the policyholders own the asset portfolio but have sold a call option on the portfolio to the insurer’s shareholders. The final claimant, the government, has a long call option position on the insurer’s taxable position. Overall, Doherty and Garven (1986) illustrate the payoffs to policyholders ($\tilde{H}_1$) and the government ($\tilde{T}_1$) as follows:

$$\tilde{H}_1 = Max\left(Min\left[\tilde{L}, \tilde{Y}_1\right], 0\right) \qquad (4.13)$$

and

$$\tilde{T}_1 = Max\left[\tau\left(\theta\left(\tilde{Y}_1 - Y_0\right) + P_0 - \tilde{L}\right), 0\right] \qquad (4.14)$$

Where $\tilde{L}$ denotes the claims cost that occurs at the end of the period and τ is the corporate tax rate.

It follows that Doherty and Garven (1986)’s framework of policyholder and government claims involve the valuation of call options. Thus, the proper expressions for the values of these claims are

$$H_0 = V\left(\tilde{Y}_1\right) - C\left(\tilde{Y}_1; \tilde{L}\right) \tag{4.15}$$

and

$$T_0 = \tau C\left[\theta\left(\tilde{Y}_1 - Y_0\right) + P_0; \tilde{L}\right] \tag{4.16}$$

Where V(.) is the valuation operator and *C*[*A*; *B*] is the current market value of a European call option written on an asset with a value of *A* at expiration and an exercise price of *B*. The market value of equity is the difference between the asset portfolio and the combination of policyholders' and governments' claims:

$$\begin{aligned} V_E &= V\left(\tilde{Y}_1\right) - \left[H_0 + T_0\right] \\ &= C\left[\tilde{Y}_1; \tilde{L}\right] - \tau C\left[\theta\left(\tilde{Y}_1 - Y_0\right) + P_0; \tilde{L}\right] \\ &= C_1 - \tau C_2 \end{aligned} \tag{4.17}$$

From Equation (4.17) it is possible to determine the fair premium. Doherty and Garven (1986) argue that premiums must be set in a manner that will generate a fair rate of return to shareholders. In order achieve this, the current market value of equity, V_E, must be equal to the initial equity investment, S_0. In their model, $\tilde{Y}_1$ and Y_0 are functions of P_0, which makes it possible to express the fair rate of return by a value of P_0^* that will satisfy

$$\begin{aligned} V_E &= C\left[\tilde{Y}_1\left(P_0^*\right)\tilde{L}\right] - \tau C\left[\theta\left(\tilde{Y}_1\left(P_0\right) - Y_0\left(P_0^*\right)\right) + P_0^*; \tilde{L}\right] \\ &= C_1^* - \tau C_2^* \\ &= S_0 \end{aligned} \tag{4.18}$$

Using this general formulation, Doherty and Garven (1986) develop more specific formulations based on a discrete-time, risk-neutral framework initially developed by Rubinstein (1974) as cited in Doherty and Garven (1986). The authors present two main cases. In the first model they assume that the returns on the insurer's asset portfolio, the value of the insurer's cost of claims and the wealth of a representative investor are all joint normally distributed. Also, they assume that the utility function of the representative investor displays constant absolute risk aversion. In the second model, the distributional assumption changes to joint lognormality and the representative investor is assumed to show constant relative risk aversion.

Under the first case, the value of equity is expressed as the discounted value of the certainty-equivalent terminal cash flow

$$\begin{aligned} V_E &= R_f^{-1}\int_{-\infty}^{\infty}\tilde{Y}_E\tilde{f}\left(\tilde{Y}_E\right)d\tilde{Y}_E \\ &= R_f^{-1}\tilde{E}\left(\tilde{Y}_E\right) \end{aligned} \qquad (4.19)$$

Where $\tilde{Y}_E$ is the uncertain cash flow to shareholders at the end of the period, $\tilde{f}(\tilde{Y}_E)$ is the risk-neutral normal density function and $\tilde{E}(\tilde{Y}_E)$ is the certainty-equivalent expectation of $\tilde{Y}_E$. The certainty-equivalent expectation of $\tilde{Y}_E$ is articulated as

$$\tilde{E}\left(\tilde{Y}_E\right) = S_0 + \left(1+\theta\tau\right)\tilde{E}\left(\tilde{r}_i\right)\left(S_0 + kP_0\right) + \left(1-\tau\right)\left(P_0 - \tilde{E}\left(\tilde{L}\right)\right) \qquad (4.20)$$

Where $\tilde{E}(\tilde{r}_i)$ is the certainty-equivalent expectation of the rate of return of insurer's investment portfolio and $\tilde{E}(\tilde{L})$ is the certainty-equivalent expectation of total claims costs.

Using Equations (4.19) and (4.20), i.e. substituting $\tilde{E}(\tilde{Y}_E)$ into V_E, Doherty and Garven (1986) derive expressions for the fair premium and rate of underwriting return

$$P_0 = \frac{E\left(\tilde{L}\right)}{1 - E\left(\tilde{r}_u\right)} \tag{4.21}$$

Where $E(\tilde{r}_u) = [P_0 - E(\tilde{L})]/P_0$.

Subsequently, the equity holder and tax claim call options are derived. The equity holder call option claim, *C₁*, is valued as follows:

$$\begin{aligned} C_1 &= C\left[\tilde{Y}_1; \tilde{L}\right] \\ &= R_f^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Max\left[\left(\tilde{Y}_1 - \tilde{L}\right), 0\right] \tilde{f}\left(\tilde{Y}_1, \tilde{L}\right) d\tilde{Y}_{1,}\, d\tilde{L} \end{aligned} \tag{4.22}$$

where $\tilde{f}(\tilde{Y}_1, \tilde{L})$ is the bivariate risk-neutral density function. This equation demonstrates that if the terminal value of equity exceeds total claims costs, then the call option will be valuable to shareholders. On the other hand, if the terminal value of equity is less than the total claims costs, "…shareholders will exercise their limited-liability option by declaring bankruptcy" (Doherty and Garven, 1986, p. 1037). Finally, the shareholder call option can be rewritten as[45]

[45] Several steps have been omitted. See Doherty and Garven (1986, p. 1037).

$$C_1 = R_f^{-1}\left(\hat{E}\left(\tilde{X}\right)N\left[\hat{E}\left(\tilde{X}\right)/\sigma_x\right] + \sigma_x n\left[\hat{E}\left(\tilde{X}\right)/\sigma_x\right]\right) \qquad (4.23)$$

where $N[\tilde{E}(\tilde{X})/\sigma_x]$ and $n[\tilde{E}(\tilde{X})/\sigma_x]$ are the standard normal distribution and standard normal density function, respectively.

The tax claim call option is based on the certainty-equivalent of the insurer's taxable income at the end of the period

$$\begin{aligned} C_2 &= C\left[\theta\left(\tilde{Y}_1 - Y_0\right) + P_0; \tilde{L}\right] \\ &= R_f^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Max\left[\theta\left(\tilde{Y}_1 - \tilde{L}\right) + P_0 - \tilde{L}, 0\right]\tilde{f}\left(\tilde{Y}_1, \tilde{L}\right)d\tilde{Y}_{1,}\, d\tilde{L} \qquad (4.24) \\ &= R_f^{-1}\left(\hat{E}\left(\tilde{W}\right)N\left[\hat{E}\left(\tilde{W}\right)/\sigma_w\right] + \sigma_w n\left[\hat{E}\left(\tilde{W}\right)/\sigma_w\right]\right) \end{aligned}$$

The conclusions drawn from Equation (4.24) are similar to the shareholder call option. Doherty and Garven (1986, p. 1038) state, "…if terminal value of taxable income is positive, the government will own a valuable claim. However, if taxable income assumes a negative value, shareholders will exercise their tax-exemption option. Thus, our model allows certain states of nature to arise in which shareholders' claims are less valuable due to the redundancy of tax shields related to the realisation of investment losses, underwriting losses, or both."

On the whole, the shareholder and tax claim call option, *C₁* and *C₂* can be incorporated into $V_E = C_1 - \tau C_2$ to generate an expression for the market value of equity:

$$V_E = R_f^{-1}\begin{pmatrix} \hat{E}\left(\tilde{X}\right)N\left[\hat{E}\left(\tilde{X}\right)/\sigma_x\right] - \tau\hat{E}\left(\tilde{W}\right)N\left[\hat{E}\left(\tilde{W}\right)/\sigma_w\right] \\ +\sigma_x n\left[\hat{E}\left(\tilde{X}\right)/\sigma_x\right] - \tau\sigma_w n\left[\hat{E}\left(\tilde{W}\right)/\sigma_w\right] \end{pmatrix} \quad (4.25)$$

As for the second case, Doherty and Garven (1986) argue that a risky security traded in a market with discrete-time lognormally distributed securities can be expressed as

$$\begin{aligned} V_0^j &= R_f^{-1}\hat{E}\left(\hat{Y}_1^j\right) \\ &= R_f^{-1}\hat{E}\left(\hat{Y}_1^j\right)e^{\left\{-\psi Cov\left[\ln\hat{R}_j,\ln\hat{R}_m\right]\right\}} \end{aligned} \quad (4.26)$$

Where $\hat{Y}_1^j$ is the cash flow paid to the holder of security *j* at the end of the period. The investors' relative risk aversion is denoted by *ψ*, which is equal to $\frac{E(\ln\hat{R}_m)-\ln R_f}{\mathrm{var}(\ln\hat{R}_m)}+\frac{1}{2}$

Furthermore, they demonstrate the relationship between the expected insurance firm's claims costs, $\hat{E}(\tilde{L})$, and the expected value of claims costs, $E(\tilde{L})$, can be shown as follows

$$\hat{E}\left(\tilde{L}\right) = \hat{E}\left(\tilde{L}\right)e^{\left\{-\psi Cov\left[\ln\hat{L},\ln\hat{R}_m\right]\right\}} \quad (4.27)$$

As in the previous case, Equation (4.27) is then used to derive the an expression for premium income and the competitive rate of return on underwriting and substituting

$\hat{E}(\tilde{Y}_E) = S_0 + (1-\theta\tau)\hat{E}(\tilde{r}_i)(S_0 + kP_0) + (1-\tau)(P_0 - \hat{E}(\tilde{L}))$ into $V_E = R_f^{-1}\hat{E}(\tilde{Y}_E)$, and equating *S₀* with *V_E*, the fair premium can be derived

$$P_0 = \frac{E(\tilde{L})}{1 - E(\tilde{r}_u)} \tag{4.28}$$

There is an important distinction between Equation (4.28) and Equation (4.21) in that $E(\tilde{r}_u)$ is joint lognormally distributed with a constant relative risk aversion parameter.

Next, Doherty and Garven (1986) develop expressions for the call options C_1 and τC_2. The value of the first call option is expressed as

$$C_1 = C\left[\tilde{Y}_1; \tilde{L}\right] \tag{4.29}$$

By, introducing a standardised normal term, $\tilde{z}$, C_1 can be expanded to become

$$\begin{aligned} C_1 = {} & V_0^U \int_{-d_2}^{\infty} \left(\sqrt{2\pi}\right)^{-1} e^{\left(-\sigma_u^2 + \tilde{z}\sigma_u - \frac{1}{2}\tilde{z}^2\right)d\tilde{z}} \\ & - R_f^{-1} P_0 \int_{-d_2}^{\infty} \left(\sqrt{2\pi}\right)^{-1} e^{\left(-\frac{1}{2}\tilde{z}\right)d\tilde{z}} \end{aligned} \tag{4.30}$$

This can be rewritten in terms of the cumulative standard normal distribution, which is equivalent to Black and Scholes (1973)'s option pricing formula:

$$C_1 = V_0^U N\left(d_1^U\right) - R_f^{-1} P_0\left(d_2^U\right) \tag{4.31}$$

Where,

N(*d*$_i$) is the cumulative standard normal density function for *d*$_i$.

V_0^U is the value of the claim *U* for that period.

$$d_1^U = \frac{\ln(V_0^U / P_0) + \ln R_f + \sigma_u^2 / 2}{\sigma_u}$$

$$d_2^U = d_1^U - \sigma_u$$

The second call option τC_2, the value of the tax claim is derived in a similar fashion to C_1. Using analysis the same as that illustrated above, the value of the tax claim is

$$C_2 = V_0^T N\left(d_1^T\right) - R_f^{-1} P_0\left(d_2^T\right) \tag{4.32}$$

Where,

V_0^T is the value of the tax claim, $\tilde{T}$, at the end of the period.

$$d_1^T = \frac{\ln(V_0^T / P_0) + \ln R_f + \sigma_t^2 / 2}{\sigma_t}$$

$$d_2^T = d_1^T - \sigma_t .$$

C_1 and τC_2 can now be combined to generate the market value of equity

$$\begin{aligned} V_E &= C_1 - \tau C_2 \\ &= V_0^U N\left(d_1^U\right) - \tau V_0^T N\left(d_1^T\right) - R_f^{-1} P_0 \left(\left(d_2^U\right) - \tau N\left(d_1^T\right)\right) \end{aligned} \qquad (4.33)$$

Doherty and Garven (1986) note several interesting points with regard to the option-based models. Generally, they find that their option framework generates higher rates of underwriting profit than the CAPM. More significantly, the differences in fair underwriting profit suggested by the option framework compared to the CAPM is that the option-pricing model takes the chance of bankruptcy, as well as redundant tax shields, into account. However, Doherty and Garven (1986) are cautious with regard to applying this framework to individual lines of insurance. They point out that the model is intended to determine the fair rate of return for a company offering a single line of business and problems occur when the analysis is extended to multiple insurance lines because insurers hold common equity and assets and incur common frictional costs over several business segments. This means that the company and not the insurance line is valued by the market. As a result, arbitrary allocations are made across insurance lines. In addition, Doherty and Garven (1986) express concern with the additivity property of option-pricing models in terms of capital allocation, as an option on a portfolio does not have the same value as a portfolio of options.

Garven (1992) reiterates and expands on Doherty and Garven (1986)'s option pricing model. His central interest is in examining the influence of limited liability and taxes on pricing and risk incentive of property-liability insurers. Furthermore, Garven (1992) compares and contrasts the option framework with that of the CAPM and argues that the CAPM is simply a special case of option-based models.

However, the most important distinction between Garven (1992) and Doherty and Garven (1986) is that Garven (1992) provides an initial description of surplus and its interaction with the insolvency put option. Thus, Garven (1992) is implicitly arguing that bankruptcy is brought about when shareholders (or any contractual counterparty benefiting from limited liability) exercise their limited liability put option. Surplus is a major determinant of the insolvency put option's value. Garven (1992, p. 40) elaborates; "Since the value of this [put] option is positively related to the probability of insolvency, a decrease (increase) in surplus increases (decreases) the put's value by increasing (decreasing) insolvency risk." In other words, if the risk of bankruptcy is minimal then by association, the option to default (exercise the insolvency put) will have little value. This point leads to Garven (1992)'s key distinction of the option framework to that of the CAPM; the CAPM makes no allowance for limited liability. Consequently, for low levels of insolvency risk, the prices generated by the CAPM-based methods will be the same as those produced by the option model, as the chance of default is trivial. However, as surplus deteriorates the possibility of insolvency will no longer be negligible and the value of the insolvency put will begin to rise.

This distinction between the CAPM-based and option-based models has an important implication for arbitrage arguments that ensure Modigliani and Miller (1958)'s first proposition that firm value will be independent of its capital structure. That is, the CAPM models assume that shareholders can replicate any structure undertaken by the insurer and as a result, they will be indifferent to the investment and underwriting policies of the firm. Conversely, in the case of option-based frameworks Garven (1992, p. 48) maintains, "…shareholders' portfolio decisions are no longer perfect substitutes for corporate decisions due to the effects of limited liability for shareholders and the government." This observation demonstrates that insurers are likely to alter their risk exposure based on the chances of taxation or bankruptcy and Garven (1992, p. 48) dubs this the "risk incentive hypothesis". He asserts that mutual insurance companies will seek less exposure than stock companies. The shareholders of stock company insurers faced with a rising probability of insolvency will be incentivised to take on greater risk because such a move will result in the transfer of wealth from policyholders. A mutual ownership structure does not suffer from this adverse incentive because as owners of the insurer, policyholders enjoy the benefits as well as bear the costs of the company's risk management policies. Thus, Garven (1992) maintains that mutual companies have a greater incentive to reduce risk.[46]

Despite Garven (1992)'s assertion that the CAPM is a special case of the option-based model, the choice between the CAPM-based and option-based models is not a trivial one. In fact, Garven (1992) notes several advantages that the option framework has over the CAPM. For instance, the option model clearly articulates the trade-off between return and insolvency risk and it quantifies the influence of underutilised tax shields. By not taking this into account, the CAPM-based models will understate actual insurer profits.

4.3. Capital Allocation and the Economic Balance Sheet - the Link Between the Insolvency Put Option and Risk Capital

More recently, influential research by Merton and Perold (1993) and Myers and Read (2001) has spurred on the idea of applying the insolvency put option to marginal capital allocation techniques. Merton and Perold (1993) combine the insolvency put option with the concept of risk capital, where they argue that the insolvency put is an asset of the firm (Francis, Heckman and Mango, 2005).

Merton and Perold (1993, p. 17) define risk capital as "…the smallest amount that can be invested to insure the value of the firm's net assets against a loss in value relative to the risk-fee investment of those net assets." Risk capital, they argue, must not be confused with regulatory capital or cash capital, which correspond to capital measurements in terms of a particular accounting standard and cash needed to carry out a transaction, respectively.

[46] It could be argued that a mutual ownership structure eliminates some agency costs and better aligns decision management and decision control (Jensen and Meckling, 1976; Fama and Jensen, 1983).

The key aim of Merton and Perold (1993)'s paper is to apply the combination of the insolvency put option and risk capital to general financing decisions of companies (insurers being one of them) such as capital budgeting and risk management. Allocating risk capital to individual business segments has a significant influence on the financing of a firm and its profitability. Indeed, Merton and Perold (1993, p. 17) argue that "…full allocation of risk capital across the individual businesses of the firm is generally not feasible, and attempts at such a full allocation can significantly distort the true profitability of individual businesses."

Merton and Perold (1993) illustrate their risk capital framework with hypothetical balance sheets of financial firms and use the tenet of replication to determine the valuation of risk capital under an assortment of configurations (Francis, Heckman and Mango, 2005). They begin by presenting the balance sheet of a hypothetical financial firm, Merchant Bank, which is a subsidiary of a large parent company that has virtually zero risk of insolvency. Merchant Bank has only one asset, a risky one-year bridge loan valued at $100 million promising a 20% interest rate. This asset is financed by the issuance of a default free one-year note of $100 million that pays 10% interest. Thus, the only risk in this case is the possible default of the bridge loan.

Furthermore, Merton and Perold (1993) consider three possible scenarios: The most likely Anticipated scenario in which the bridge loan is repaid in full with interest at the end of the year; an unlikely Disaster scenario where the borrower defaults and only half of the loan is repaid – that is, only $60 million will be recovered; finally a rare Catastrophe scenario in which nothing is recovered.

While Merchant Bank has two central methods of eliminating the chance of default on the note it issued, Merton and Perold (1993) point out that both methods will involve the purchase of insurance. First, Merchant Bank can do so indirectly by purchasing insurance on its assets. Second, it can use a more direct method through the purchase of insurance on its liabilities. Either way, Merton and Perold (1993) assume that

Merchant Bank can acquire insurance that will guarantee a return of $110 million on the bridge loan.

Table 4.1 to 4.9 illustrate Merchant Bank's accounting and risk capital balance sheets when a commercial insurer guarantees its assets:

Table 4.1: ***Accounting Balance Sheet A (all values in millions)***

Bridge loan	$100	Note (default free)	$100
Loan insurance (from insurance company)	5	Shareholders' equity	5

Table 4.2: ***Risk Capital Balance Sheet A (all values in millions)***

Bridge loan	$100	Note (default free)	$100
Loan insurance (from insurance company)	5	Risk capital	5

In this case, the noteholder is protected by Merchant Bank's note insurance obtained from a third party guarantor. The cost of this insurance is $5 million, which can be viewed as both the cash investment made by the parent and Merchant Bank's risk capital (Merton and Perold, 1993).

If the bridge loan does not default and is paid in full then Merchant Bank will be able to provide its parent with $10 million, which is the difference between income earned on the bridge loan ($20 million) and the interest payment to the noteholders ($10 million). On the other hand, if the bridge loan defaults, the asset insurance will cover any shortfall up to $110 million (the total value of the note) and Merchant Bank will have zero profits. Hence, the insurance company bears the risk of default of the bridge loan and Merchant Bank's parent bears the risk of losing its investment of $5 million.

The payoffs under the different scenarios envisaged by Merton and Perold (1993) are summarised as follows:

Table 4.3: ***Summary of Payoff Scenarios***

BRIDGE LOAN	*LOAN INSURANCE*	*BRIDGE LOAN & INSURANCE*	*NOTEHOLDERS*	*SHAREHOLDERS*
Anticipated Scenario (All values in Millions)				
$120	$110	$0	$110	$10
Disaster Scenario				
60	50	110	110	0
Catastrophe Scenario				
0	110	110	110	0

Next, Merton and Perold (1993) examine the situation where, rather than purchasing insurance, the parent of Merchant Bank guarantees the note. Compared to the previous case, the parent makes no cash investment in Merchant Bank. However, the parent makes an implicit $5 million contribution to risk capital by providing the guarantee. Thus, the guarantee is an additional asset of Merchant bank but this does not appear on the accounting balance sheet:

Table 4.4: ***Accounting Balance Sheet B; Parent Guarantee (all values in millions)***

Bridge loan	$100	Note (default free)	$100
		Shareholders' equity	0

If the guarantee is included, the risk capital balance sheet will be described by Table 4.5.

Table 4.5: ***Risk Capital Balance Sheet B; Parent Guarantee (all values in millions)***

Bridge loan	$100	Note (default free)	$100
Note guarantee (from parent)	G	Risk capital	G

To reiterate, the parent's cash capital is zero. If this situation is interpreted as the fair value of taking on default risk via an arms length transaction with a third party, then the parent's proceeds are the same as in the previous case (Francis, Heckman and Mango, 2005). Merton and Perold (1993) elaborate that should the bridge loan default, Merchant Bank will also default on its note. In the presence of the guarantee, the parent will cover the promised $110 million on the note in exchange for the value of the bridge loan. Overall, as guarantor of Merchant Bank's debt, the parent bears the risk of the asset and the possible loss of capital as shareholder of Merchant Bank. Comparing balance sheets A and B in Table 4.4 and 4.5, the economic equivalence of liability and asset insurance becomes clear. In both, the noteholders bears no risk (as they made a risk-free investment). The parent, as shareholder of Merchant Bank, will receive the same payoffs in all three scenarios: $10 million in the anticipated scenario and zero in the others.

Table 4.6: ***Summary of Payoff Scenarios with Parent Guarantee***

BRIDGE LOAN	*NOTE WITHOUT GUARANTEE*	*NOTE GUARANTEE*	*NOTEHOLDERS*	*SHAREHOLDERS*
Anticipated Scenario (All values in Millions)				
$120	$110	$0	$110	$10
Disaster Scenario				
60	60	50	110	0
Catastrophe Scenario				
0	0	110	110	0

In a subsequent case, Merton and Perold (1993) relax the assumption that noteholders are only interested in buying risk-free debt. Now the note has some chance of default. Without any guarantees, or any other credit enhancements for that matter, the risky note must sell at a discount, $D, to par. In the previous two cases, Merton and Perold (1993) have shown that the risk capital required to eliminate the risk of default was $5 million. Thus, the discount on the note will be equal to $5 million because this is the amount of cash equity required to meet the initial shortfall in funding (recall that the bank will need $100 million to buy the bridge loan). The accounting and risk balance sheets are as follows:

Table 4.7: ***Accounting Balance Sheet C; Risky Debt (all values in millions)***

Bridge loan	$100	Note (default free)	$100 – D
		Shareholders' equity	D

Table 4.8: ***Risk Capital Balance Sheet C; Risky Debt (all values in millions)***

Bridge loan	$100	Note (default free)	$100
Asset insurance (from noteholder)	5	Risk capital	5

The debt holder in this instance can interpret the purchase of the note as a long position in default free debt with a face value of $100 million and a short position in asset insurance valued at $5 million. In total, the risky note generates the general exposure of:

Risky note = Default-free Note – Asset Insurance.

Once again, the payoff to the parent will be a maximum of $10 million or zero otherwise:

Table 4.9: ***Summary of Payoff Scenarios with Risky Debt***

BRIDGE LOAN	*ASSET INSURANCE*	*DEFAULT FREE NOTE*	*RISKY NOTE (DEFAULT FREE NOTE – ASSET INSURANCE)*	*SHAREHOLDERS*
Anticipated Scenario (All values in Millions)				
$120	$0	$110	$110	$10
Disaster Scenario				
60	50	110	60	0
Catastrophe Scenario				
0	110	110	0	0

Merton and Perold (1993) provide a more general application of the concept of risk capital. That is, they argue that the risk capital needed to offset the possibility of default can be described as a put option. They demonstrate this by considering a firm with risky assets valued at $2.5 billion and customer liabilities outstanding that have been packaged as one-year guaranteed investment contracts (GICs) paying 10% on a face value of $1 billion. As the risk-free rate in this example is also 10% the default free value of the GICs is also $1 billion. Consequently, the net assets are worth $1.5 billion ($2.5 billion - $1 billion) and the risk capital required is $500 million. If the company's liabilities are segmented further into junior debt paying 10% on its par value of $1 billion, the total fixed liabilities at the end of the period will be $2.2 billion, which consist of $1.1 billion of GICs and $1.1 billion of the junior debt that is subordinate to the GICs.

Merton and Perold (1993) make an additional assumption that the firm decides to only partially insure its investment portfolio. That is, only the first $300 million decline in value, below $2.5 billion, is covered at a cost of $200 million. Clearly, the junior tranche of debt is significantly riskier than the senior GIC liability as only significantly large declines in the firm's portfolio value will put the GICs at risk. Merton and Perold (1993) assume that the GICs will trade at a discount of 1%, which implies a market value of $990 million. The riskier junior debt, on the other hand, is assumed to trade at a 10% discount, making its market value $900 million. Using Merton and Perold (1993)'s economic identity, the GIC holders can be seen to hold a long position in default free debt and a short position in asset insurance (which is sold to the firm) worth $10 million ($1 billion - $990 million). Similarly the junior debt holders have sold a much larger value of asset insurance to the firm of $100 million ($1 billion - $900 million).

In total, these implicit premiums for asset insurance that have been sold to the firm add up to $310 million (this includes $200 million paid to the third party guarantor and $100 million and $10 million from the junior debt and GICs, respectively). Yet as shown previously, the risk capital needed to fully insure the asset portfolio is $500 million. Hence, shareholders must provide the balance of $190 million. As indicated, Merton and Perold (1993) purport that this total insurance structure has the same payoffs as a put option. Furthermore, this put option has an exercise price equal to the current value of the portfolio, $2.5 billion, plus the risk-free rate of 10% bringing the total to $2.75 billion. Overall, the value of the put is $500 million; the risk capital of the firm.

Merton and Perold (1993) argue further that risk capital (i.e. the insolvency put option) can be used in capital budgeting and allocation decisions. When it is applied to capital allocation decisions the marginal benefit of risk capital must be weighed against its cost. If capital can be allocated to individual lines of business their profitability can be evaluated, which will provide guidance in expanding or exiting a business line. However, Merton and Perold (1993) warn against allocating risk capital based on stand-alone risks. They point out that individual business lines are not perfectly correlated with each other and as a result, there will be a diversification benefit. Consequently, the risk of the entire portfolio will be less than the sum of individual units and less capital will be required than that suggested by stand-alone risks. In short, allocating capital based on individual risks will lead to distortions of the true profitability of a business unit (Merton and Perold, 1993).

4.4. The Market Value of Equity in an Insurance Company and the Influence of the Insolvency Put Option

Babbel (1998) examines the relation between the market value of an insurer and the components that influence that value. He identifies four components of market value for an insurer, which can be demonstrated as follows:

Market value of equity = franchise value + market value of tangible assets – present value of liabilities + put option.

Franchise value is the present value of economic rents, which stems from intangible factors that enhance firm value. Moreover, it includes the value of the going concern of the business. Tangible assets are the assets of the firm that can be valued for sale, i.e. the company's liquidation value. Liabilities are the present value of projected obligations undertaken but not yet fulfilled. The last item, put option, is generally referred to as the insolvency put option, which stems from the notion that shareholders benefit from limited liability.

Babbel (1998) argues that the put option benefits shareholders by enabling them to gain from upside earnings while not bearing any costs associated with default. As a result, the insolvency put increases in value as the riskiness of the insurer increases. In some instances, Babbel (1998) notes that the insolvency put can be of significant value.

The central objective of Babbel (1998)'s study is to develop an abstract framework of insurance liability estimating. Moreover, he proposes that direct methods of liability estimation do not suffer from a significant number of drawbacks associated with indirect methods. For instance, Babbel (1998) argues that the indirect valuation approach cannot determine the amount of capital needed to totally satisfy any expected obligations that result from insurance policies. The reason for this as Babbel

(1998) argues, is that by subtracting the market value of equity from the market value of tangible assets, under the indirect method, any estimate of liabilities will be under estimated by the amount of franchise value and the insolvency put option.

On the other hand, by recognising that attention must be paid to the amount of current tangible assets required to cover any future liabilities, the direct method will result in a more considered approach. Babbel (1998) suggests that the present value of these anticipated obligations could be estimated directly by treasury-rate based lattices or simulations that factor in any interest rate sensitivities in the cash flows. The present value measure advocated by Babbel (1998) is augmented (relative to the traditional approach) by including stochastic interest rates and the related cash flows. Indeed, incorporating the stochastic nature of interest rates results in a more meaningful estimate of liabilities compared to methods that ignore this.

While including the stochastic nature of interest rates in liability estimation results in a sounder figure, Babbel (1998) argues that it is not enough to simply to set aside reserves that are just equal to anticipated liabilities. The prudent action to take would be to set reserves higher than liabilities. Babbel (1998) points out that, even when liabilities are duration and convexity matched, there is always the possibility for deviations from what is expected in the timing of claims. Furthermore, Babbel (1998) argues actual money is needed as a reserve cushion (he downplays the need to focus on accounting methods such as reserves, surplus, and risk-based capital) and specific amounts for actuarial risk is important to cover for any deviations from expected claims.

Overall, Babbel (1998) acknowledges that his proposed framework of insurance liability estimation is unlikely to be a precise measure of actual values of liabilities in the market. However, the analysis will provide a useful starting point for insures and regulators, given the simple nature of the calculations required, in addition to enabling a comparison among other insurers.

Babbel, Gold and Merrill (2002) illustrate the use of option pricing techniques to value fixed income securities such as mortgage-backed securities (MBS). In particular, they agree with Black and Scholes (1973)'s and Merton (1974)'s assertion that corporate securities can be viewed as options on the underlying assets of a company. Using Black and Scholes (1973)'s option pricing model, the value of a bond can be determined by subtracting the value of a call option from the firm's underlying assets.

They illustrate this by examining the payoffs to equity and debt holders of a hypothetical non-financial firm. The debt holders are only entitled to the value of the firm's assets up to the par value of their loan whereas the equity holders are entitled to any residual value of the firm's assets once the debt holders have been repaid. Thus, Babbel, Gold and Merrill (2002, p. 14) note, "…we can view equity as a call option on the assets with a strike price equal to the face value of the debt." Consequently, the value of a bond can be determined by subtracting the call option held by shareholders from the value of the company's assets. In other words, the value of the bond can be expressed as assets (*A*) less a call option (*C*). If the face value of the bond is viewed as the exercise price (*X*) that shareholders must pay to retain ownership of the firm, then the present value of the bond will be $A - C = Xe^{-r(T-t)}$.

From this, Babbel, Gold and Merrill (2002) make several salient observations. As can be seen by the expression, $A - C = Xe^{-r(T-t)}$, the bond will become risk-free as the assets of the firm increase in value. A rise in asset values will be beneficial to both debt and equity holders, however, debt holders will only benefit from rising asset values up to the face value of their debt. As for the shareholders, Babbel, Gold and Merrill (2002) point out that their holding will resemble a call option for large asset values as their upside potential is essentially limitless. Overall, a large enough asset value will result in a risk-free bond as the chance of default becomes negligible. Conversely, the bond will decrease in value if asset volatility increases. In such an instance, greater volatility implies a higher chance of the firm defaulting on its debt and thus, leaving debt holders with only a fraction of their claims.

However, Babbel, Gold and Merrill (2002) use the put-call parity relationship[47] to provide an alternative derivation of the arguments above. Substituting the company's total assets (*A*) for the share price (*S₀*), the formula can be rewritten as

$$A - C = Xe^{-r(T-t)} - P \qquad (4.34)$$

The left hand side of Equation (4.34) is the value of the bond described originally; however, the right hand side describes a risk-free bond less a put option. Hence, Babbel, Gold and Merrill (2002) assert that bondholders hold a risk-free bond and a short put on the assets of the firm. Moreover, Babbel, Gold and Merrill (2002) maintain that the opposite holds true for shareholders; they own the put option.

Babbel, Gold and Merrill (2002) argue that this restatement of a bond's value clearly shows the impact of changing interest rates and credit quality. Importantly, any changes in credit quality will be captured by the put option. This result is not trivial. Babbel, Gold and Merrill (2002) charge that other approaches do not take proper account of risk, whereas their approach does. Furthermore, when considering more complex examples they argue that the decomposition of debt into a risk-free security and a put option is particularly advantageous. They argue that this will improve transparency, which will benefit analysts, regulators, investors, and managers alike. Also, the consistent application valuation methodologies to determine risk-free interest rate dependent liabilities will facilitate comparison of liabilities (with varying structures) across different companies. Moreover, they argue that the put option can be calculated in a relatively straightforward manner using a method that is similar to valuing embedded options using the option adjusted spread.[48]

[47] The put-call parity relationship is discussed in detail in Chapter Five.

[48] For a discussion on option adjusted spreads see Kolb and Overdahl (2007).

Babbel and Merrill (2005) reiterate and expand on Babbel (1998)'s assertions. They demonstrate that as net tangible value of a firm falls, franchise value will decline in tandem but the value of the insolvency put option will increase due to greater risk. Like, Babbel, Gold and Merrill (2002), Babbel and Merrill (2005) argue that a bond issued by a risky company can be viewed as the issuer holding a long position in a put option and a short position in a risk-free bond. Moreover, the put option is an implicit asset on the firm's economic balance sheet and is thus owned by the firm's shareholders. As a result, the firm will have an incentive to increase the value of the put option by taking greater risks as net tangible value declines. Thus, Babbel and Merrill (2005) argue that if a firm's net tangible value declines to such a low level a situation will arise where the shareholders are more concerned with maximising the value of the put option.

Consequently, Babbel and Merrill (2005) contend insurers have an incentive to behave in a manner where they are more motivated by the insolvency put option. That is, they charge that an insurer will sell assets that have increased in value but will retain the assets that have unrealised losses so as to inflate the insurer's capital cushion. Indeed, Babbel and Merrill (2005) discuss various other methods that insurers use to inflate their capital cushion such as front-loaded earnings and surplus relief reinsurance.[49] This will enable the insurer to maintain enough of a capital cushion to satisfy regulators and remain in business. Furthermore, an insurer will have an incentive to pursue riskier strategies whenever net tangible asset value and franchise value decline. If franchise value drops significantly, the firm will take on greater risk to maximise shareholders' equity. It is the insolvency put option that boosts shareholders' equity in this instance because the put will increase as risk increases. Thus, insurers with minimal net tangible asset value and franchise value will have a high propensity to gamble (Babbel, 1998; Babbel and Merrill, 2005).

To mitigate the perverse incentives that result from declining franchise value, Babbel and Merrill (2005) suggest insurers focus on fundamental approaches to increasing

[49] See Babbel and Merrill (2005, pp. 8 – 10).

shareholder wealth such as seeking out positive net present value (NPV) projects and managing asset-liability mismatches to preserve value. While Babbel and Merrill (2005) note that positive NPV projects are difficult to come by in a competitive market, they contend that insurers have a relative advantage in their asset and liability side of their business.

The main advantage an insurer has is the issuance of policies. Babbel and Merrill (2005) maintain that insurance companies are able to charge more than the expected loss cost for insurance coverage, which will yield a positive NPV. Also, the ability to invest in illiquid assets such as property and private placement debt will help generate positive NPVs. However, as is the case with publicly traded securities, the competitive nature of these markets will result in zero NPVs and the company will require superior insight and analysis to reap any potential rewards. In addition, insurers can differentiate their products, which will enable them to earn above average profits. In this instance, general market segmentation and information asymmetries constrain competition to a degree.

4.5. The Insolvency Put Option and Capital Allocation: The Use of the Insolvency Put Option in Determining Default Risk and Capital

Butsic (1994) presents a simplified application of capital allocation to insurers using a formalisation of the insolvency put (which is practically identical to Merton (1977)'s deposit insurance pricing framework) known as the expected policyholder deficit (EPD). His development of the expected policyholder deficit is motivated by his dissatisfaction with the traditional probability of ruin approach.[50] Butsic (1994, p. 660) argues, "…the probability-of-ruin criterion is inadequate to express the policyholders' exposure to loss. It is not sufficient merely to consider the probability of ruin – its *severity* must also be appreciated."

[50] The probability of ruin approach quantifies the chance of a firm becoming insolvent. For a discussion on the various techniques involved see Embrechts and Veraverbeke (1982)

Butsic (1994) argues that the insolvency put option model has desirable characteristics as a solvency measure.[51] Indeed, Butsic (1994) maintains that the key feature of the insolvency put framework is its ability to quantify the severity of bankruptcy. The severity of insolvency can be determined by examining the claims that an insurer is obligated to pay its policyholders and the actual assets the insurer has available to meet those payments. In other words, the severity of insolvency is measured by $A - L$, as previously defined, and in the context of insurance, this is known as the policyholder deficit.

To arrive at a formal definition of the expected policyholder deficit, Butsic (1994) presents a simplified version of an insurance company whereby its balance sheet is composed of assets (A), a loss reserve (L) and capital (C), which is the difference between assets and the loss reserve. With this simplification, Butsic (1994) asserts that solvency risk can be measured in a consistent manner where the same level of capital is allocated across all insurance lines. Thus, for a discrete probability distribution, the expected policyholder deficit for losses is

$$D_L = \sum_{x>A} p(x)(x - A) \qquad (4.35)$$

where $p(x)$ is the probability density function for losses, $0 \leq x < \infty$.

Butsic (1994) extends this discrete version of the expected policyholder deficit to a more formal continuous probability distribution model:

[51] Butsic (1994) lists three main criteria that a desirable risk-based capital framework must satisfy. First, the solvency measure must be consistent for policyholders, insurers and capital providers alike. Second, the model should be objectively determined. Third, the model must be able to discern between different risk elements (e.g. shares and fixed income securities).

$$D_L = \int_A^\infty (x - A)p(x)dx \qquad (4.36)$$

The similarity of Butsic (1994) continuous probability distribution model to that of Merton (1977)'s deposit insurance pricing formula is plain to see. Like, Merton (1977)'s model, Butsic (1994) has up till now presented a risk-financing framework. In order to transform the expected policyholder deficit into a capital allocation model, an important modification is required. This change is achieved by expressing the expected policyholder deficit as a ratio of the expected value of insurance claims

$$d_L = \frac{D_L}{L} = k\phi\left(\frac{-c}{K}\right) - c\Phi\left(\frac{-c}{k}\right) \qquad (4.37)$$

Where k is the coefficient of variation of losses, c is the capital-to-liability ratio, $\Phi(.)$ and $\phi(.)$ are the cumulative standard normal distribution, and standard normal density function, respectively.

Capital can now be allocated in accordance with the EPD ratio. An insurer that follows this method will settle on the desired level of the EPD ratio, for example 1% or 0.1%. The chosen level of the EPD ratio can be interpreted as the chosen probability of insolvency. However, despite the simplicity of Butsic (1994)'s derivation of the expected policyholder deficit, the EPD ratio cannot be calculated directly and must be solved via an iterative process. That is, when the level of capital changes so too will the chance of ruin and the severity of insolvency.[52] Hence, capital is accumulated to a point where the common cost of insolvency reaches the predetermined EPD ratio (Butsic, 1994; Barth, 2000).

[52] The expected policyholder deficit can also be defined as the product of the probability of ruin (F) and the severity of insolvency (S): EPD = D_L = (F)(S).

Barth (2000), however, counters Butsic (1994)'s criticisms of the probability of ruin approach and argues that the technique does indeed take the severity of bankruptcy into account. Furthermore, he regards the indirect calculation of the EPD ratio a considerable weakness of the model. Unlike the ruin approach which always produces positive capital requirements, the EPD ratio, in low risk environments, suggests a negative capital requirement. In addition, Barth (2000) charges that the expected policyholder deficit method ignores indirect costs resulting from bankruptcy of large insurers and consequently understates the true cost of insolvency.[53]

Arguably one of the most influential articles that makes use of the insolvency put option framework is Myers and Read (2001)'s paper, "Capital Allocation for Insurance Companies". They argue that capital can be allocated across lines of insurance and that these allocations "…depend on the marginal contribution of each line to default value – that is, to the present value of the insurance company's option to default" (Myers and Read 2001, p. 545). Moreover, Myers and Read (2001, p. 545) argue that their line-by-line method of capital allocation adds up to the total capital requirement of an insurer, which is "…unique and not arbitrary."

Myers and Read (2001)'s analysis is built upon the foundation laid by Merton and Perold (1993). However, their conclusion that "unique and not arbitrary" capital allocations to individual insurance lines are possible, contrasts starkly to Merton and Perold (1993)'s view that full allocation techniques will result in the actual profitability of individual lines being distorted. Despite Merton and Perold (1993)'s general reservations with allocating capital to individual lines of insurance, Myers and Read (2001) note that it is the marginal capital allocations that result in a unique line-by-line allocation. Furthermore, they argue that full allocation of capital is unfeasible only for infra-marginal changes such as entering or exiting lines of business (which is the central focus of Merton and Perold (1993)'s analysis).

[53] See Barth (2000, pp. 404 – 407) for a detailed discussion.

Myers and Read (2001)'s full allocation solution is significant because it avoids the problem of apportioning the unallocated remaining capital to a business unit in some subjective manner. Venter (2003) explains that it is possible for an individual line to make enough money to cover its marginal costs (as well as its marginal cost of capital) but not enough to cover any allocated fixed costs. As a result, it is possible that the allotment of unallocated capital in this way can lead to an otherwise profitable insurance line to be wound up. Moreover, Venter (2003) points out that by providing a solution to the capital allocation problem, Myers and Read (2001) are actually allocating frictional costs. Frictional costs will arise even if capital is not put at risk. For example, frictional costs can include tax, agency costs (including the agency issues associated with free cash flow as pointed out by Jensen (1986)), liquidity issues, and diminished investment opportunities. Consequently, capital allocation is the method that represents the frictional cost allocation.

The option-pricing framework put forward by Myers and Read (2001) begins with a simple balance sheet of an insurer. However, Myers and Read (2001, p. 550) point out that such a situation is too simple as the possibility of insolvency is ignored and that future losses on outstanding policies "…are valued assuming claims will always be paid." Hence, the possibility of default gives rise to an implicit put option written on the insurer's assets. As a result, Myers and Read (2001) define the competitive premium as follows:

Premium = PV(Losses, assuming no default) + PV(Surplus costs)[54] – PV(Default option)

The last term, the PV(default option), reduces the value of the premium charged. The reasoning is that, the riskier the insurer (i.e. the greater the difference between the firm's liabilities and assets) policyholders should pay a lower premium to compensate them for insolvency risk. Naturally, when confronted with the potential for an adverse

[54] PV(Surplus costs) includes tax or other miscellaneous costs.

outcome people will expect a discount to offset the prospect of being handed a lemon (Akerlof, 1970).

The default option is not a simple plain vanilla put with a fixed exercise price but is rather an exchange option. Myers and Read (2001) argue that future losses cannot be known at the origination of a policy, therefore it will be impossible to identify a predetermined level at which the put will be exercised. Thus, the put option will be exercised (and hence, default will occur) at any point when liabilities, the obligations to policyholders, exceed the insurers assets. In this instance, the put option is expressed as:

Default value = PV(Default option) = PV(Option to exchange assets for liabilities)

Clearly an insurer's balance sheet must take default value into account. Thus, the default option is included as an asset. However, there is an offsetting liability to a guarantee fund. Based on Merton and Perold (1993)'s insights, such a liability can be regarded as a premium that would be paid by an insurer to secure payment of its losses. More formally the balance sheet of an insurer including default becomes:

Table 4.10: ***Market Value of Balance Sheet, Including Default Value***

Assets	PV(Losses)
Default Option	Guarantee Fund Liability
	Equity

Besides the addition of default value, another difference is that equity (E) is distinct from surplus (S). Surplus is equal to assets less liabilities ($S = A - L$) assuming no default. Myers and Read (2001, p. 552) state that "[s]urplus is an input and equity is an output," which means that whatever amount of assets is leftover after liabilities have been taken into account is the shareholders equity claim. It follows that the more surplus (or capital) there is, the greater the value of equity.

The actual option formulas are centred on a one-period model[55] and if the insurer's assets more-than cover its liabilities at the end of the period then shareholders will receive $A - L$. On the other hand, if liabilities exceed assets then the insurer will default and of course shareholders are left with nothing. Thus, shareholders' equity will be:

$$E_T = Max\{0, (A_T - L_T)\} \qquad (4.38)$$

Myers and Read (2001) go further and state that the payoff to shareholders also includes the default option, which is expressed as:

$$D_T = Max\{0, (L_T - A_T)\} \qquad (4.39)$$

Hence, shareholders' equity becomes:

$$E_T = A_T - L_T + D_T \qquad (4.40)$$

[55] Although, the risk associated with long tails of claims payments is not explicitly dealt with, PV(Losses) incorporates any outstanding losses incurred in previous periods. The losses arising from previously issued policies are treated as if they are newly created ones (Myers and Read, 2001).

The put option derives its value from the two underlying assets; a claim to the policy losses and a portfolio of investments. Thus, it is possible to determine the value of the option using a replicating portfolio by taking a combination of positions in the policy claim losses and the investment portfolio. As a result, according to Myers and Read (2001), the replicating portfolio can be expressed as:

$$D = (\partial D/\partial L)L + (\partial D/\partial A)A \qquad (4.41)$$

The size of the positions in the replicating portfolio is conditional on the present value of policy losses (L), the market value of assets (A) and their joint probability distribution.[56]

All this, however, assumes that it is possible to match the cash flows and the losses of the insurer in the market, which enables the calculation of the default option. Myers and Read (2001) do not believe that this assumption is unrealistic. In fact they argue that it is the normal assumption in corporate finance. Their argument is as follows; "… consider a corporation seeking to maximise market value and evaluating a proposed capital investment project by discounted cash flow. The objective will be endorsed by all shareholders if investing in the project does not change the scope of risk characteristics attainable by investors" (Myers and Read, 2001, p. 553). All that is needed for this situation to occur is a diverse securities market with ample choice. However, Madsen, Haastrup and Pedersen (2005) argue that it is impossible for the insurer to always readily replicate cash flows because the insurance market is incomplete. In other words, a lack of available derivative contracts result in an inability to eliminate discrepancies via arbitrage, which perpetuates a significant degree of information asymmetry.

[56] Myers and Read (2001) assume that policy losses and investment portfolio return are jointly lognormal and calculate the weights in the replicating portfolio to be $\partial D/\partial L$ = N(z) and $\partial D/\partial L$ = -N(z – σ), where N(z) is the normal cumulative probability density function. For a discussion on this subject see Margrabe (1978) as cited in Myers and Read (2001).

The ultimate aim of Myers and Read (2001)'s paper is to determine amount of capital required for individual lines of insurance. However, it is the value of the insolvency option that determines the level of surplus (or capital) required for these individual lines. Hence, the insolvency put option culminates into the capital needed for the insurer as a whole.

Consequently, increasing exposure to a particular insurance line has the potential to increase the value of the insolvency option for the entire firm. In order to maintain the initial value of the insolvency option, when expressed as a ratio of expected losses, more capital must be added to the line in question. In other words, Myers and Read (2001)'s method allocates the required extra capital to the additional exposure that generated it, which maintains *D*/*L* for the whole firm. Furthermore, these individual marginal values add up to the default value (value of the insolvency put) and capital requirement for the entire insurance company (Venter, 2003).

In short, total capital is the sum of the individual charges. Myers and Read (2001) illustrate this as follows:

An insurance company with *M* lines of insurance will have liabilities equal to:

$$L = \sum_{i=1}^{M} L_i \qquad (4.42)$$

Where $L \equiv PV(L_i)$

Thus, the total capital required for an insurer, which is the addition of the line-by-line allocations and is defined as surplus by Myers and Read (2001), is proportional to total liabilities:

$$\begin{aligned} S &= \sum_{i=1}^{M} L_i S_i \\ &= Ls \end{aligned} \tag{4.43}$$

Where $s_i \equiv \frac{\partial S}{\partial L_i}$

They then show that assets are the sum of liabilities and capital (surplus):

$$\begin{aligned} A &= \sum_{i=1}^{M} L_i (1 + s_i) \\ &= L(1 + s) \end{aligned} \tag{4.44}$$

Given that line-by-line allocations of capital add up to the capital requirement for the entire firm, it follows that the default value (or insolvency put value) of the entire firm is also the addition of the marginal contribution of individual lines of insurance:

$$D = \sum_{i=1}^{M} L_i d_i \tag{4.45}$$

Where $d_i \equiv \frac{\partial D}{\partial L_i}$

Hence, Myers and Read (2001, p. 554) reach the conclusion that "… line-by-line marginal default values 'add up'…"

The overall effect of a marginal change in losses, for a single line of insurance, on the default value is expressed as:

$$\begin{aligned}\sum_{i=1}^{M} L_i\left(\frac{\partial D}{\partial L_i}\right) = \left(\frac{\partial D}{\partial L}\right)\sum_{i=1}^{M} L_i\left(\frac{\partial D}{\partial L_i}\right) + \left(\frac{\partial D}{\partial A}\right)\sum_{i=1}^{M} L_i\left(\frac{\partial A}{\partial L_i}\right) \\ + \left(\frac{\partial D}{\partial \sigma}\right)\sum_{i=1}^{M} L_i\left(\frac{\partial \sigma}{\partial L_i}\right)\end{aligned} \tag{4.46}$$

From here the addition of the first two terms $\left(\frac{\partial D}{\partial L}\right)\sum_{i=1}^{M} L_i\left(\frac{\partial L}{\partial L_i}\right)$ and $\left(\frac{\partial D}{\partial A}\right)\sum_{i=1}^{M} L_i\left(\frac{\partial A}{\partial L_i}\right)$ equals the total default value for the entire firm; $D = (\partial D/\partial L)L + (\partial D/\partial A)A$. The last term $\left(\frac{\partial D}{\partial \sigma}\right)\sum_{i=1}^{M} L_i\left(\frac{\partial \sigma}{\partial L_i}\right)$ is zero.[57]

This culminates into the common term for marginal default values, where default, or the insolvency put, hinges on "…(1) the marginal surplus for the line of insurance, (2) the covariance of losses on the line with losses on other lines of insurance in the portfolio, and (3) the covariance of losses on the line with the returns on the insurance company's assets" (Myers and Read, 2001, p. 558). Thus, the put option, i.e. the option to default is expressed as follows

$$d_i = d + \left(\frac{\partial d}{\partial s}\right)(s_i - s) + \left(\frac{\partial d}{\partial \sigma}\right)\left(\frac{1}{\sigma}\left[(\sigma_{iL} - \sigma_L^2) - (\sigma_{iA} - \sigma_{LA})\right]\right) \tag{4.47}$$

[57] Several steps have been omitted. See Myers and Read (2001, pp. 555 – 557).

This equation demonstrates that the insolvency put option delta $\left(\frac{\partial d}{\partial s}\right)$ is negative. This suggests that larger marginal surplus values result in smaller marginal default (i.e. insolvency put) values. The positive vega $\left(\frac{\partial d}{\partial \sigma}\right)$ figure illustrates that higher covariances of losses with losses on other insurance lines will lead to a larger insolvency put option value. Last, a stronger covariance with asset returns will work in the insurer's favour by shrinking the value of the insolvency put.

If an insurer has a policy of keeping the same ratio of surplus to liabilities for every individual line then each line of business' marginal surplus will equal the marginal surplus for the company as a whole. That is in terms of Myers and Read (2001)'s framework, $s_i = \partial S/\partial L_i = s$. As a result, the marginal default value for an individual line will be

$$d_i = d + \left(\frac{\partial d}{\partial \sigma}\right)\left(\frac{1}{\sigma}\left[\left(\sigma_{iL} - \sigma_L^2\right) - \left(\sigma_{iA} - \sigma_{LA}\right)\right]\right) \tag{4.48}$$

Compared to the original marginal default equation, the insolvency put will depend only on the covariance of losses between insurance contracts and the covariance of losses with the returns of the asset portfolio. However, there is a problem with this because making the marginal surplus the same will result in marginal default values that vary by line. If an insurer defaults on a single line, it will default on all lines, so it makes no sense to have different marginal default values for every business line. It follows that capital should be allocated in such a manner where the marginal default value is the same for all lines of insurance; $d_i = \partial D/\partial L_i = d$. Hence, the surplus value to maintain the same marginal default value is

$$s_i = s - \left(\frac{\partial d}{\partial s}\right)^{-1}\left(\frac{\partial d}{\partial \sigma}\right)\left(\frac{1}{\sigma}\left[\left(\sigma_{iL} - \sigma_L^2\right) - \left(\sigma_{iA} - \sigma_{LA}\right)\right]\right) \qquad (4.49)$$

Once again delta and vega are negative and positive, respectively, meaning that higher covariance of losses between insurance lines necessitates greater capital, whereas higher covariance of losses with asset returns requires less capital.

Importantly, Myers and Read (2001) argue that their marginal capital allocation to individual insurance lines is significant, if not imperative, for pricing insurance contracts. The method has several implications for pricing. For a start, marginal default values for an insurer that provides many lines of cover will depend on the mix of insurance provided by the firm in addition to line-by-line risk characteristics. Thus, marginal default values will change as the mix of insurance changes. This means that the capital allocations derived from marginal default values will be subject to significant variation, especially if newer insurance lines grow to be a greater fraction of liabilities. However, Myers and Read (2001, p. 568) concede that for insurers with even a minimal amount of diversification, "…marginal surplus requirements for existing lines of business are reasonably robust to the introduction of a new line of business, and that marginal surplus requirements for existing lines can be approximately correct even as new lines are added or existing lines phased out."

Myers and Read (2001) suggest that increased diversification results in a trade-off between a smaller capital cushion needed and higher administrative and operating costs. Thus, it is possible to arrive at an efficient composition of business by striking a balance between required capital and operating and administrative costs. Greater diversification can lower the amount of capital required across individual insurance lines due to lower asset correlations. Myers and Read (2001) argue that gains from diversification will be large to begin with but adding additional lines of business will eventually result in higher costs. Hence, an insurer will continue to diversify until the marginal benefit of reduced capital equals the marginal cost of operating and administrative costs. While Myers and Read (2001) suggest regulators assume a base case of an efficient composition of business for calculating regulated premiums and capital requirements, they argue that no firm should be forced to adopt a particular composition of business. Indeed, some firms may specialise in a handful of closely related lines of insurance and effectively compete, despite the lack of diversification. The base case should be used simply as a guide and should an insurer deviate from it in a manner that increases default risk then the capital cushion must be increased to keep default values at a tolerable level. In reality, Myers and Read (2001) concede that the costs of diversification may go beyond greater administrative costs. In fact, conglomerates tend to trade at a discount to the sum of their individual parts (Smith, 1986a; Holmstrom and Kaplan, 2001; Laeven and Levine, 2007).

4.6. Further Development of the Insolvency Put Option and Capital Allocation Methods

Sherris (2006), in a similar vein to Myers and Read (2001), follows a single period model framework to determine fair pricing and capital allocation. Of critical importance is the insolvency put option. Moreover, Sherris (2006) argues the use of an arbitrage free model will result in allocated capital adding up to the total capital of the firm, which is consistent with the economic value of assets and liabilities on the balance sheet.

Using a discrete state complete markets model, Sherris (2006) contends that the ratio of default to liability values (*D*/*L)* for an individual insurance line does not vary with changing allocations of capital to these individual business segments. Rather, according to Sherris (2006, p. 73) “For any given insurance line the default value to liability ratio depends on the distribution of the liability value for the line of business and its correlation with the assets, and not the business surplus ratio.”

If liability estimations ignore the possibility of default on claims payments then the value of the option to default, held by shareholders as a result of limited liability, will be overlooked. Thus, in Sherris (2006)'s model, the insurance premium charged is dependent on the insolvency put option on claims payoffs under insurance contracts at the end of the period. Consequently, Sherris (2006) argues that a company's solvency ratio as well as its investment policy must be identified in advance in order for the model to determine a fair insurance premium. One of the central assumptions of the model is that the distribution of losses, as well the return distribution of assets, is known at the beginning of the period and a level of capital is maintained to meet this known solvency ratio. Furthermore, it is assumed that no reinsurance is purchased, or similarly, the cover provided from reinsurance is fixed and known in advance. That is, the results are only measured net of reinsurance. The assumption that the payoff distributions at the end of the period are known, enables the calculation of the fair rate of return for a liability when taking into account the risk of insolvency.

According to Sherris (2006) the end of period payoff for any asset will be determined by its market price, which will reflect any loss should the issuer default. The initial market value of assets is:

$$\begin{aligned} V_A &= \sum_{j=1}^{J} E^Q\left[\frac{A_j}{1+r}\right] \\ &= E^Q\left[\frac{A}{1+r}\right] \end{aligned} \tag{4.50}$$

Where $A_j = w_jA$ and w_j is the weight of an individual strategy within the insurance portfolio, Q is a risk-neutral equivalent probability measure equivalent to the real-world probability measure and r is the risk-free rate.

The initial value of the assets, V_A, is made up of policyholders' premiums and shareholders' capital. The above equation can be modified to include the return on asset A_j which incorporates credit risk as well. Overall, given the rate of return, R_A, the total value of the firm's assets will be:

$$\begin{aligned} A &= V_A\left(1 + r_A\right) \\ &= V_A\left(1 + \sum_{j=1}^{J} w_j r_{A_j}\right) \end{aligned} \tag{4.51}$$

In the event of insolvency, shareholders will exchange their obligation to pay insurance claims for the firm's assets. Thus, it is argued that default is triggered when shareholders exercise their insolvency put option (Babbel, 1998; Myers and Read, 2001; Sherris, 2006). Sherris (2006) regards insolvency risk and asset-liability mismatch risk, as one in the same; if assets matched liability payments exactly there would be no risk of bankruptcy. Furthermore, if an insurer defaults on its policies, all lines of insurance will experience default at the same time and it is assumed that losses caused by default are shared amongst policyholders on a pro rata basis. As a result, the distribution of claims, in addition to the covariance of claims of insurance lines with other business segments and assets, are the main determinant of default value (Sherris, 2006).

As for liabilities, Sherris (2006) argues that for any individual segment, *k*, assuming unlimited liability, the insurer can incur a random claim of L_k at the end of the period. Also, Sherris (2006) assumes that L_k is unaffected by capital structure, investment policy, reinsurance strategy or any other action taken by the company that may influence its ability to pay the liabilities of its insurance contracts. The total claim payments for an insurer at the end of the period is:

$$L = \sum_{k=1}^{K} L_k \tag{4.52}$$

If full payment is assumed, the value of the claim liability is expressed as:

$$\begin{aligned} V_A &= E^Q\left[\frac{L}{1+r}\right] \\ &= \sum_{k=1}^{K} E^Q\left[\frac{L_k}{1+r}\right] \end{aligned} \tag{4.53}$$

Despite the assumption that liabilities are assumed to be paid in full, claim payments are still risky due to the randomness of future payoffs. Hence, Sherris (2006) provides an expression for liabilities by incorporating historical probabilities:

$$\begin{aligned} V_L &= E^P[mL] \\ &= E^P[m]E^P[L] + Cov(m,L) \\ &= \frac{E^P[L]}{1+r} + Cov^P(m,L) \end{aligned} \tag{4.54}$$

Where m is a stochastic discount factor. Overall, liabilities are influenced by various economic risk factors, which are taken into account by the stochastic discount factor, m.

So far the derivation of liabilities does not include the impact of insolvency, only general economic risk factors captured by m. Thus, the liabilities of an insurer must reflect general underlying risk factors and its level of solvency. Similarly to Myers and Read (2001), under Sherris (2006)'s model, insurers with greater amounts of capital will charge higher premiums. In other words, a smaller insolvency option indicates a higher credit rating for the insurer. Thus, the premium charged will be equal to $V_L - D$. The insolvency put is expressed as:

$$\begin{aligned} D &= \frac{E^Q\left[Max(L-A,0)\right]}{1+r} \\ &= \frac{E^Q\left[L-A \mid L-A>0\right]\Pr^Q\left[L-A>0\right]}{1+r} \\ &= \frac{E^Q\left[L-A \mid \frac{A}{L}<1\right]\Pr^Q\left[\frac{A}{L}<1\right]}{1+r} \end{aligned} \tag{4.55}$$

Equation (4.55) includes both the probability of insolvency and the expected severity of insolvency based on risk neutral probabilities. Sherris (2006) notes that, given an extreme event, risk neutral probabilities tend to be greater than actual historical probabilities. As a result of this underestimation of extreme events by historical probabilities, Sherris (2006) contends that measures such as VaR are inadequate for evaluating insolvency risk. Risk neutral probabilities provide a more conservative approach, as they "…can be many times the actual probabilities for tail events" (Sherris, 2006, p. 77).

The last piece needed to determine the market value of equity, and along with it the full derivation of the economic balance sheet, is the solvency ratio. Sherris (2006) assumes that the total amount of assets is established in such a way that the solvency ratio is fixed and is a known proportion of liabilities excluding the insolvency put option. In other words, the solvency ratio, s, will be a fixed proportion of liabilities so that $V_A = (1 + s)V_L$. Clearly a solvent insurer will have a positive s, which is implicit in the price for cover provided by the insurer. Overall, the market value of equity is made up of surplus and the insolvency put:

$$V_E = V_A - V_L + D = sV_L + D > 0. \tag{4.56}$$

The market value of equity, $V_A - V_L + D$, also happens to be the capital subscribed to an individual line of insurance and the premium that will be charged is equal to $V_L - D$.

Sherris (2006), confirms and reiterates Myers and Read (2001)'s adding up property of capital allocation. He demonstrates this as follows. For *k* lines of insurance, the amount of capital required will be determined by the insolvency put for line *k*. That is, the default value for a business segment is denoted by D_k:

$$D_k = \frac{1}{1+r} E^Q \left[L_k Max\left[1 - \frac{A}{L}, 0\right]\right] \tag{4.57}$$

The adding up property of the insolvency option can be illustrated as follows:

$$\begin{aligned} \sum_{k=1}^{K} D_k &= \frac{1}{1+r} \sum_{k=1}^{K} E^Q \left[L_k Max\left[1 - \frac{A}{L}, 0\right]\right] \\ &= \frac{1}{1+r} E^Q \left[\sum_{k=1}^{K} L_k Max\left[1 - \frac{A}{L}, 0\right]\right] \\ &= \frac{1}{1+r} E^Q \left[LMax\left[1 - \frac{A}{L}, 0\right]\right] \\ &= \frac{1}{1+r} E^Q \left[Max\left[L - A, 0\right]\right] \\ &= D \end{aligned} \tag{4.58}$$

Although capital allocations based on the insolvency put option add up, Sherris (2006), in contrast to Myers and Read (2001), argues that there is no unique allocation of assets to individual business lines. Hence, various criteria can be considered when allocating assets to business segments and Sherris (2006) discusses two main alternatives. First, it is possible to assign capital to insurance lines in a manner that maintains the same solvency ratio as for the entire company. Alternatively, assets could be allocated so that the expected return on the assigned capital will be equal for all lines of insurance, as well as the whole firm.

Despite these varying methods of allocating assets, Sherris (2006) confirms Myers and Read (2001)'s approach by noting that irrespective of the way in which assets are allocated, the capital allocated to individual insurance lines will always sum to the insurers total capital. In other words, given the arbitrary manner in which the allocation of assets can be made, "…an internal allocation of capital has no economic impact" (Sherris, 2006, p. 83). However, this will only be the case provided that assets, liabilities and the insolvency option are valued using an arbitrage-free model. Thus, this is the main weakness of Sherris (2006)'s model, which he acknowledges and suggests a more realistic approach be followed in future to take into account market frictions such as taxes, bankruptcy and agency costs. Moreover, Sherris (2006)'s analysis only considers a single time period and hence, ignores the dynamic aspects of capital allocation.

Sherris and van der Hoek (2006) develop a closed form model using dependent lognormal distribution assumptions to determine the line-by-line allocation of the insolvency put option. Furthermore, their model is designed to take the actual payoff of individual lines in account in the event of bankruptcy. They argue that this approach can be generalised to other distributions for business segments, which makes Sherris and van der Hoek (2006)'s approach different to others such as Myers and Read (2001).

Sherris and van der Hoek (2006), like Sherris (2006), argue that there is no non-arbitrary way of allocating assets to insurance lines without the need for further measures such as a constant expected return from the capital allocated to a particular insurance line. Moreover, Sherris and van der Hoek (2006) differentiate their measure for the solvency ratio from Myers and Read (2001) who define the solvency ratio as a partial derivative in their paper. The authors denote the solvency ratio as:

$$\tilde{s}_i = \frac{S_i}{L_i} = \frac{A_i - L_i}{L_i} \tag{4.59}$$

Total surplus is allocated to individual lines so that

$$S = \sum_{i=1}^{M} S_i \tag{4.60}$$

With

$$S = \sum_{i=1}^{M} \tilde{s}_i L_i \tag{4.61}$$

So that if $x_i = L_i / L$ then

$$\tilde{s} = \sum_{i=1}^{M} x_i \tilde{s}_i \tag{4.62}$$

Where $\sum_{i=1}^{M} x_i = 1$. To differentiate their allocations of the insolvency option from Myers and Read (2001), Sherris and van der Hoek (2006) define

$$\tilde{d}_i = \frac{D_i}{L_i} \tag{4.63}$$

So that

$$D = \sum_{i=1}^{M} \tilde{d}_i L_i \tag{4.64}$$

This illustrates a significant difference in the approach taken by Sherris and van der Hoek (2006) to that of Myers and Read (2001). Sherris and van der Hoek (2006) explicitly establish the value of the insolvency option by individual business segments centred on the payoffs to each segment in insolvency. Myers and Read (2001) define d_i as a sensitivity $\partial D / \partial L_i$ whereas Sherris and van der Hoek (2006) base $\tilde{d}_i$ on the explicit payoffs for each line of insurance.

Overall, the allocation of the insolvency option and surplus will be such that (as in the Myers and Read (2001) case) the individual business segments add up to the total economic capital:

$$S + D = \sum_{i=1}^{M} \left(\tilde{s}_i + \tilde{d}_i\right) L_i \tag{4.65}$$

Sherris and van der Hoek (2006), like Sherris (2006), note that the internal allocation of capital to business segments has no direct economic consequence for the insolvency of the company. That is, the probability of insolvency is determined not by the internal allocation of capital, but by the total level of capital. The manner in which capital is allocated has no influence on insolvency itself.

Sherris and van der Hoek (2006) argue that their closed form derivations enable a more accommodating and broader approach to determining the insolvency put option value as well as its by-line allocation compared to Myers and Read (2001). Moreover, Sherris and van der Hoek (2006) assume that only the ratio of assets to liabilities, denoted $\Lambda(t)$, is log-normal, whereas Myers and Read (2001) assume that both assets and liabilities are log-normal in order to make use of Margrabe (1978)'s exchange option formula (as cited in Sherris and van der Hoek, 2006). They contend that the portfolios of individual assets and lines of business with lognormal properties can only be lognormal if the portfolio is rebalanced often enough to ensure constant weights. Indeed, they point out that under the Myers and Read (2001) framework, the size of the lines of business are assumed to be fixed at the beginning of the period and are thus not continuously rebalanced. Also, assets that closely match liabilities at the beginning of the period should result in a ratio of assets to liabilities at the end of the period that has a general lognormal distribution. However, Sherris and van der Hoek (2006) concede that these assumptions have yet to be tested empirically.

Thus, the more reasonable assumption of the ratio of assets to liabilities being approximately lognormal enables Sherris and van der Hoek (2006) to consider a greater range of processes for individual business segments. Importantly, the assumption that the ratio of assets to liabilities is lognormal enables the derivation of the insolvency option value for individual insurance lines in closed form, which will sum to the overall insolvency option value for the insurer.

Sherris and van der Hoek (2006) represent the value of liabilities for line $i = 1,\ldots, M$ at time t by $L_i(t)$ for $0 \leq t \leq T$ where T is the end of the period. They assume that the risk-neutral dynamics of $L_i(t)$ are

$$dL_i(t) = \mu_i L_i(t)dt + \sigma_i L_i(t)dB^i(t) \qquad (4.66)$$

Where $L_i(T)$ is the amount paid at time T. Each line has a lognormal distribution at time T. $B^i(t)$ are Brownian motions under the risk-neutral dynamics. Furthermore, they argue that if L_i can be mimicked by traded assets and if claim payments (other than those occurring at the end of the period) are absent then $\mu_i = r$, which is the risk-free rate. Total liabilities are defined as

$$L(t) = \sum_{i=1}^{M} L_i(t) \qquad (4.67)$$

Equation (4.66) ignores the possibility of default and thus assumes that all claims are paid in full. As a result, Sherris and van der Hoek (2006) derive a closed form for the default option for insurance line *i.* Consider the following

$$M(t) = E^Q\left[e^{-r(T-t)}[1 - \Lambda(T)]^+ \mid F_t\right] \qquad (4.68)$$

Where F_t is the filtration defined by the Brownian motions $B^{\Lambda}(t)$, $B^i(t)$. Q specifies that the expectation follows risk neutral dynamics.

Sherris and van der Hoek (2006) state that Equation (4.68) is the value of a European put option written on an underlying asset that pays a continuous dividend of r - μ_{Λ} with a current price of $\Lambda(t)$, exercise price equal to one and maturity equal to T. If it is assumed that $\Lambda(T)$ has a lognormal distribution, using the Black and Scholes (1973) model, Sherris and van der Hoek (2006) derive the closed form insolvency put option value[58]:

$$M(t) = e^{-r(T-t)}N\left(-d_{2t}\right) - \Lambda\left(t\right)e^{-\left(r-\mu_\Lambda\right)\left(T-t\right)}N\left(-d_{1t}\right) \qquad (4.69)$$

Where

$$d_{1t} = \frac{\ln \Lambda\left(t\right) + \left(\mu_\Lambda + 0.5\sigma_\Lambda^2\right)\left(T-t\right)}{\sigma_\Lambda\sqrt{\left(T-t\right)}}$$

and

$$d_{2t} = d_{1t} - \sigma_A\sqrt{(T-t)}$$

The insolvency option value for an individual business segment is given by

$$\begin{aligned} M(t) &= L_i(t)e^{\mu_i(T-t)}E^{Q_i}\left[e^{-r(T-t)}\left[1-\Lambda(T)\right]^+ \mid F_t\right] \\ &= L_i(t)e^{\mu_i(T-t)}M^i(t) \end{aligned} \qquad (4.70)$$

From Equation (4.70), the insolvency option for the entire insurer is given by

$$\begin{aligned} D(t) &= E^{Q_i}\left[e^{-r(T-t)}\left[L(T)-V(T)\right]^+ \mid F_t\right] \\ &= \sum_{i=1}^{M} D_i(t) = \sum_{i=1}^{M} L_i(t)e^{\mu_i(T-t)}M^i(t) \end{aligned} \qquad (4.71)$$

[58] Several steps have been omitted. See Sherris and van der Hoek (2006).

Sherris and van der Hoek (2006) contend that this approach results in fair insolvency option values, which in turn can be used to determine the fair premium for a multiline insurer. However, when compared to Myers and Read (2001), the allocation of capital to individual insurance lines differs significantly. As a result, the authors point out that different approaches of capital allocation can have substantially different results. For example, the Myers and Read (2001) method will add up to the total capital but Sherris and van der Hoek (2006) argue that the allocation will not result in fair premiums for individual insurance lines that reflect the allocation of the insolvency option value based on equal priority. Moreover, when evaluating the allocations that result from other risk measures such as the standard deviation, VaR and the TailVaR they find large variations in capital allocations across these measures. Also, the allocations generated by the standard deviation, VaR and TailVaR do not take the value of the insolvency put into account. Hence, they emphasise the importance of the method used and recommend a cautious approach when basing financial decisions on these methods.

Yow and Sherris (2007) develop a single period model, for a multi-line insurer, that takes frictional costs associated with capital, policyholder price elasticity and insurer financial quality (determined by the insolvency put option) into account. Their model illustrates the influence of these factors on profit margins and capital allocation.

Yow and Sherris (2007)'s model uses a value maximising approach based on enterprise or economic value added (EVA). However, their measure of EVA is different to the traditional EVA measure. It is defined as the difference between the initial value of equity and the amount of capital subscribed including frictional costs and insolvency. Their model is outlined as follows:

$$E_1 = (A_1 + L_1 + D_1)(1 - \tau_1)(\tau_1 - \tau_2)R_0 \qquad (4.72)$$

Where A_1 is the time 1 payoff from the assets of the insurer accumulated at random rate r_A, L_1 is the contractual time 1 liability payoff of the insurer. D_1 is the insolvency put option equal to $\max(L_1 - A_1, 0)$. R_0 is the initial cash capital subscribed at time 0. τ_1 is the corporate tax rate charged on insurer profits and τ_2 is the agency cost borne by shareholders at time 1.

The initial payoff to shareholders is expressed as

$$E_0 = (A_0 + L_0 + D_0)(1 - \tau_1) + e^{-r} (\tau_1 - \tau_2)R_0 \tag{4.73}$$

Where, $A_0 = R_0 + P_0 - c_0$, which is the net initial investable cash available from shareholders and policyholders. The term c_0 represents production costs of all issued policies. L_0 is the present value of insurance obligations.

The insurer will then assign capital to individual business lines to maximise EVA. Stated mathematically,

$$\begin{aligned} \underset{R_0, p_i, 0}{Max}\left\{EVA_0\right\} &= \underset{R_0, p_i, 0}{Max}\left\{E_0 - R_0\right\} \\ &= \underset{R_0, p_i, 0}{Max}\left\{\left(P_0 - c_0 - L_0\left(1 - d_0\right)\right)\left(1 - \tau_1\right) - \left(\left(1 - e^{-r}\right)\tau_1 + e^{-\gamma}\tau_2\right)R_0\right\} \end{aligned} \tag{4.74}$$

where d_0 is the reduction in the time 0 fair value of the liabilities resulting from insolvency risk as a proportion of the liability.

A critical input of Yow and Sherris (2007)'s model is the financial quality of the insurer, which is measured by the insolvency put option. Like the preceding models, the insolvency put captures the likelihood of default, hence it follows that it can be used as a measure of the financial quality of an insurer. Yow and Sherris (2007) argue that this has a direct effect on the insurance premium. They contend that policyholders will recognise undercapitalised insurers and will demand less insurance in response.

Similar to Myers and Read (2001) and Sherris and van der Hoek (2006), Yow and Sherris (2007) define the insolvency put option as tantamount to an exchange option. The insolvency put is outlined as

$$d_0 = \Phi(z) - \Lambda_0 \Phi(z - \sigma) \tag{4.75}$$

Where

$$\Lambda_0 = \frac{A_0}{L_0}$$

and

$$z = \frac{-\ln(\Lambda_0)}{\sigma} + \frac{1}{2}\sigma$$

Volatility is estimated using a similar method to Myers and Read (2001)

$$\sigma = \sqrt{\sigma_L^2 + \sigma_A^2 - 2\sigma_{LA}} \tag{4.76}$$

Where

$$\sigma_L^2 \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \sigma_i \sigma_j \rho_{ij}$$

and

$$\sigma_{LA} \sum_{i=1}^{N} x_i \sigma_i \sigma_A \rho_{iA}$$

From this, Yow and Sherris (2007) develop a demand function for individual insurance lines. The demand for insurance is assumed to be determined by price (the insurance premium), default risk (measured by the insolvency put) and frictional costs associated with the risk of bankruptcy:

$$q_{i,0} = q\left(p_{i,0}, d_0, f\right) \qquad (4.77)$$

Where $\partial q_{i,0}/\partial p_{i,0} < 0$, $\partial q_{i,0}/\partial d_0 < 0$ and $\partial q_{i,0}/\partial f < 0$. In other words, higher insurance premiums along with greater risk of default and associated frictional costs will result in lower demand. Thus the demand curve is downward sloping with respect to both price and default risk.

This indicates that in order to maximise insurer profitability, and by definition shareholder value, the firm must minimise the default put option to ensure that the company is sufficiently capitalised. Moreover, Yow and Sherris (2007) maintain that different methods of capital allocation yield different levels of profitability.

Specifically the method of marginal line-by-line allocations will result in a different profit margin to allocations based on the broader level of insurer risk. Another important factor Yow and Sherris (2007) draw attention to that influences profitability is insurance price elasticity. Varying degrees of price elasticity and allocation methods interact to produce an ambiguous impact on profit margins. For example, even if price elasticity is similar across most lines of insurance, the various commercial lines, compared to household and motor lines, will generate different levels of profitability. In general, low price elasticity lines are more profitable given that demand is less sensitive to rising prices.

Overall, Yow and Sherris (2007) maintain that capital allocation will only be consistent with maximising shareholder value if policyholder demand and preferences for financial quality are taken into account. Importantly, the key determinant of financial quality is the insolvency put option, which in turn is also an important component of the line-by-line capital allocation methods espoused by Myers and Read (2001), Sherris (2006) and Sherris and van der Hoek (2006).

5. A Critique of the Insolvency Put Option Framework

5.1. Bankruptcy, Limited Liability and the Call Option Perspective

As indicated, within the insolvency put option framework, bankruptcy is cited as an example where equity holders have exercised a valuable put option by defaulting on their debt. Indeed, this so called insolvency put option stems from the notion of limited liability. Importantly, the insolvency put option is said to exert significant influence on the shareholders of a firm. For instance Babbel and Merrill (2005) maintain that it is possible for a situation to arise where the firm is more concerned with maximising the value of the put option. When an insurer seeks to maximise the value of the put option, Babbel and Merrill (2005, p. 7) argue that shareholders can take risks without any downside: "Then if the insurer makes risky bets and wins, the equity holders reap the rewards. If the insurer loses its bets, others (i.e., the policyholders and the various state insurance insolvency guarantee programs) pick up the tab."

At first glance this makes sense. If the put option exists and if the risk of the company increases then so too will the value of the put option; higher risk results in higher option prices. Moreover, as the risk of the company increases and as its assets begin to fall short of its liabilities then the ability to default on those obligations becomes valuable.

However, if one examines the incentives faced by shareholders as well as what the introduction of limited liability achieved then it becomes clear that in situations where insolvency is imminent shareholders will not exercise a valuable put option. In fact, the complete opposite occurs. If a company defaults on its debt then equity holders have decided to allow an embedded *call option* to expire unexercised.

5.2. The Introduction of Limited Liability in Company Law

The key concept behind the notion of the call perspective is limited liability enjoyed by shareholders. Indeed, the legal concept of limited liability is well established and has been in effect for many years. However, in terms of legislative history, the limited liability company is a relatively recent development. In English Company Law, the limited liability company was only enacted in 1855.

In response to the economic disaster of the South Sea Bubble of 1720, which was fuelled by the abuse of the unincorporated joint stock company form, legislators of the era banned the legal form altogether. In its place a new act was brought about, creating the unincorporated deed of settlement company. The idea behind this legal entity was to combine trust and partnership; a company that was "…in general regulated by the law of partnership" (Cilliers and Benade, 1982, p. 24). It is this legal form that had a sizable influence on the evolution of company legislation.

By the early 19th century, wariness of the abuse of the unincorporated company form coupled with the influence of partnership law, prompted legislators to enact the first single corporate form. This was the first general Companies Act of 1844, which created the unlimited company. Cilliers and Benade (1982, p. 24) explain; "This kind of company, although duly incorporated in terms of the Act, lacked an important characteristic of legal personality in so far as its members were expressly – without any limitation – held liable for the debts of the company." Hence, any shareholder that invested in this corporate form was subject to a linear and symmetrical payoff, and could be held liable for an amount far exceeding the value of their share capital.

Approximately ten years later, in 1855, a major change in company law was enacted: the Limited Liability Act. For the first time in English Company Law, shareholders had separate legal personality from that of the firm. As a result, they were provided with a cap on their liability exposure to the amount of their invested share capital. Hence, a creditor could only seek funds from the capital of the company and not the shareholder to repay any loans. This meant it was up to the creditor to perform due diligence and ensure that their risk exposure was not unreasonable. In other words a new dispensation came about; caveat creditor (Cilliers and Benade, 1982).

As mentioned previously, the notion of a limited liability company was new in 1855, yet the concept of limited liability itself was well established in common law governing contractual agreements. Cilliers and Benade (1982, p. 25) explain the application of limited liability to a company: “The theory as applied in company law is that the word ‘limited’ in the name of the company gives notice that the liability of members is limited to the amount of their shares and that any person doing business with the company is deemed to have agreed to do so on the basis of limited liability.”

The introduction of the limited liability company resulted in an asymmetrical payoff profile for shareholders. From this point onwards, an equity investment in a limited liability company possessed an inherent call option quality. In fact, it is possible to compare Cilliers and Benade (1982)’s description of the Limited Liability Act to that of a call option. The notion that a member’s liability is limited to the amount of their invested share capital can be directly compared to the situation where a call option holder is liable only for the premium paid to acquire the option. That is, an investment in a limited liability company will result in a payoff of the greater of the firm’s assets less its liability payments or zero. The legal description of a shareholders exposure can be expressed mathematically, as $Max(A - L, 0)$, which is identical to the payoff of a call option. Of course the shareholder is exposed to substantial loss but the size of that loss is determined by the size of their shareholding in the company and more specifically, the price that the shareholder paid to acquire the shares.

Although the Limited Liability Act of 1855 has been amended several times over the course of time, it remains the basis of the English Companies Act of 1967 and the South African Companies Act of 1973. Moreover, it remains the fundamental basis as to why an investment in a company is tantamount to holding a call option.

5.3. The Insolvency Put Option Framework

5.3.1. Babbel (1998), Babbel and Merrill (2005)'s Components of Insurance Firm Value and the Present Value of Liabilities, Revisited

It is instructive to take a step back and re-examine the basic tenets of the insolvency put and its role within a short-term insurance company. As discussed previously, one of the more influential discussions of the insolvency put option and the general makeup of a short-term insurer is provided by Babbel (1998) and Babbel and Merrill (2005).

Recall Babbel (1998)'s general description of the market value of insurance equity:

Market value of equity = franchise value + market value of tangible assets – present value of liabilities + put option.

As indicated, bankruptcy is said to occur when equity holders exercise a valuable put option by defaulting on their debt. In keeping with this line of thought, Babbel (1998, p. 3) comments, "[a]s the firm increases in insolvency risk, market value increases, and as it decreases in risk, again there is an increase in firm value. This equity market value premium over net tangible value stems either from franchise value or from put option value, or from some combination of the two."

However, Francis, Heckman and Mango (2005) question Babbel (1998) and Babbel and Merrill (2005)'s assertion that the market value of equity could be made up of some combination of franchise value and put option value. They observe that it is uncommon for franchise value and the insolvency put option to have large positive values at the same time. Indeed, by the time bankruptcy becomes likely franchise value will be negligible. Yet this is beside the point. The concept of the insolvency put option is flawed.

5.3.2. Two Serious Flaws with the Notion of an Insolvency Put Option – Discussions of Modigliani and Miller (1958), Black and Scholes (1973), Merton (1973b) and Myers (1977)

There are two problems with Babbel (1998)'s and Babbel and Merrill (2005)'s reasoning. First, as pointed out above, equity holders are not placing any value on the put option. In fact, while the right to default will certainly be valued by investors, none of them would ever consider buying shares of a risky company because of a potentially valuable put option. Shareholders are the last in line when it comes to receiving claims on a bankrupt firm and generally, they never get any money back once the firm has been liquidated. If there was a put option to be exercised then one would expect to recover, at least, some of their funds and in reality this is not the case. So why do investors continue to buy shares in risky companies? They certainly do not buy risky shares to take part in a valuable put option. Investors will buy shares in risky companies because equity has an inherent call option quality. This idea is not new. Myers (1977)[59] maintains that any growth opportunity is tantamount to a call option and that the option's exercise price is the investment outlay required to own the asset. Thus, the call option will be exercised if the asset's future value exceeds the initial investment. Moreover, according to Myers (1977) the decision to exercise the growth opportunity call also depends upon the value of any loan promised to debt holders; if the value of the loan is greater than the present value of the asset then the call option will not be exercised.

All firms are valued as a going concern[60] and Myers (1977, p. 156) states; "The value of a going concern can be maintained only by positive action… There is continual effort devoted to advertising, sales, improving efficiency, incorporating new technology and recruiting and training employees. All of these activities require

[59] Although, Myers (1977) uses the inherent optionality of equity to illustrate debt holder and equity holder conflicts, his discussion on how growth opportunities are "exercised" lead to important conclusions about the incentives that influence shareholders. In particular, his theory demonstrates why shareholders try and delay bankruptcy as long as possible, whereas creditors would prefer to force a firm into liquidation while they can still recover their loans; shareholders are hoping to have a call option that expires in-the-money.

[60] One valuation method calculates the present value of equity based on its present value of growth opportunities: $V_0 = E / r + PVGO$. Where *E* = earnings per share and *r* = cost of capital.

discretionary outlays. They are options the firm may or may not exercise." This suggests that any investment decision relates to the exercise of a call option. Black and Scholes (1973) support this notion. They point out that the equity of an issuer of coupon bonds can be viewed as a compound option, where each coupon payment involves the active decision to exercise a call option. Black and Scholes (1973, p. 651) observe; "The common stock is an option on an option on … an option on the firm. After making the last interest payment, the stockholders have an option to buy the company from the bondholders for the face value of the bonds."

When insolvency risk increases, investors are betting that if the company survives and makes a profit, they will be handsomely rewarded. If, on the other hand, the firm fails then they forfeit their investment, which is exactly how a call option works. Anyone who purchases a call option contract will enjoy all the upside exposure to a rising share price while benefiting from limited downside risk, which is restricted to the value of the premium paid. This is precisely the situation investors will find themselves in when they consider buying shares of risky companies; they would not be interested in exercising an option that resulted in a loss of money as would be the case with the insolvency put option. In short, equity holders will be hoping that they will be able to exercise a valuable call option and if things do not work out then they can default on their liabilities and walk away.

Another problem with the insolvency put idea is that once the put becomes valuable, investors would want to exercise it immediately (Black and Scholes, 1973; Merton, 1973a). A put option is a deferred sale of an asset and should the put become deep-in-the-money then the exercising of the option is a mere formality (no rational investor would allow a put with positive intrinsic value to expire and if the put is deep-in-the-money then it is unlikely that its intrinsic value could fall to the extent that it becomes out-of-the-money). The problem is that if an investor holds a valuable put and has to wait to exercise it (as would be the case with a European[61] put option) he or she will

[61] European options can only be exercised on their expiration dates. Conversely, American options can be exercised anytime before and up to their expiration dates.

suffer an opportunity cost; the investor has essentially sold the asset but still has his or her capital tied up with it. Thus, the investor will lose out on interest that could have been earned had the put been exercised immediately and the funds deposited into an interest-bearing account. When it comes to bankruptcy, shareholders try to delay it as much as possible, which is at odds with the presence of a valuable put option (Myers, 1977; Brealey, Myers and Allen, 2006). Such behaviour is more in keeping with the idea of an embedded call option. Shareholders will delay bankruptcy not only because they will most likely receive nothing should it occur, but they are also hoping that the company could turn a corner and make a profit. They are implicitly hoping that the call option will end up in-the-money. A basic tenet of option valuation pointed out by Merton (1973b) is that a rational investor should never exercise a call option prior to expiration. This makes sense because a call option is a deferred purchase of an asset, which makes it optimal to delay exercising the call option. Moreover, if an investor wanted to dispose of a call option it would be better to sell the option rather than exercise it because early exercise will result in the loss of time value. Once a call is exercised prior to maturity an investor will receive $S_t - X$. However, if the call remains unexercised, the value of the call will be $S_t - Xe^{-r(T-t)}$. Hence, early exercise of a call will cause a loss of $X - Xe^{-r(T-t)}$, which is the time value of the option. This result demonstrates that early exercise is never optimal (Kolb and Overdahl, 2007). In addition, this explains why someone holding a put option, that is deep-in-the-money, is at a disadvantage if they have to wait to exercise it, whereas in the case of a call, an investor practically owns the asset without having paid for it. Thus, the investor benefits from not having his capital tied up with the asset and is able to earn interest.[62] It can also be argued that forcing a firm into bankruptcy (by exercising the insolvency put) causes time value as well as the compound option described by Black and Scholes (1973) to be lost.

The second problem with Babble (1998)'s statement is the notion that the market value of the firm will increase as risk increases. In their seminal paper, Modigliani

[62] There is another reason why it is not optimal to exercise a call option early. If the underlying asset is equity then it is expected to grow, at least, at the risk-free rate, while the fixed exercise price can be discounted at the risk-free rate. This will result in a higher intrinsic value for the call $\left(S_0e^{rT} - Xe^{-rT}\right)$.

and Miller (1958) demonstrate that in the absence of taxes capital structure does not matter. So no matter how one divides up a firm's cash flows or assets, they will still add up to the same total. Thus, to argue that increasing risk will increase the value of a firm is misplaced. In addition, Modigliani and Miller (1958)'s second proposition states that as risk increases, the return required by equity holders will increase in a linear fashion, which means that increasing risk will not increase shareholder wealth as Babbel (1998) and Babbel and Merrill (2005) suggest. It will only increase a firm's cost of equity.

Clearly, some equity holders like risk; after all they chose to invest in a security that does not have a certain payoff as in the case of a bond. One could argue that higher risk would be good for shareholders because it will allow them an opportunity to achieve higher returns that would otherwise be impossible (imagine a world where the only securities available were risk-free bonds – investment opportunities would certainly be truly limited). However, risk is not an end in itself. Simply increasing risk will not automatically translate into higher profitability for a firm and, likewise for shareholders, will not necessarily mean greater wealth. Increasing exposure to risk will only enable greater earnings, thus risk could be regarded as a tool to unlock greater profitability. However, higher risk will not guarantee higher profitability because risk cuts both ways. In fact, contrary to Babbel (1998)'s and Babbel and Merrill (2005)'s notion that higher firm risk will result in a higher market value, if a firm's franchise value declines to the point where the going concern of the company becomes uncertain, as a result of higher risk, then the market value of the firm will most certainly tumble. Investors and analysts will recognise the context of the greater firm risk and they will make their opinions known by downgrading the firm's share price and by making upward revisions of its risk premium.[63] In extreme cases, the only share value remaining will be speculative. Thus, in situations where risk is extremely high the only reason why someone would buy shares in such a firm is because he or she would be tempted by the possibility that things may improve and

[63] In terms of the intrinsic valuation model, higher risk without higher expected growth will result in a lower share price. $V_0 = \frac{D_1}{k-g}$, where k = cost of capital, g (growth) = return on equity x earnings retention.

thereby make a massive return on their investment. In short, he or she would be speculating that the implicit call option would end up in-the-money.

5.4. The Put-Call Parity Relationship

5.4.1. Stoll (1969)'s Seminal Derivation of the Relationship Between Put and Call Options – How the Combination of a Long Put Option and Equity is Equivalent to a Long Call Option

Despite all of these arguments, suppose that when a company defaults on its liabilities it does indeed exercise an insolvency put option. Certainly some might argue that the introduction of the Limited Liability Act of 1855 actually provided a specific mechanism to limit the losses of shareholders, enabling them to "put" the company's losses to its creditors. Nevertheless, given the legal description of what the Limited Liability Act of 1855 actually enabled, it is clear that investors would have viewed their shareholdings from the point of view of a call option. More significantly, the combination of equity and a put option will be tantamount to a call option because put and call option prices are related to each other via a theoretical model known as put-call parity.

In the seminal paper "The Relationship between Put and Call Option Prices," Stoll (1969) formulates a relationship between put and call option prices. Using a variety of combinations of option contracts and the underlying shares, Stoll (1969) demonstrates that options can be purchased directly or indirectly. For example, indirect exposure to calls can be created by taking a long position in the underlying share and buying a put. Similarly, an investor can gain indirect exposure to puts by shorting the underlying share and buying a call. The ability to gain indirect exposure to put and call options means that, overall, if their values move beyond what is considered to be reasonable (i.e. too high or too low relative to each other) investors will be able to exploit this imbalance and thus, keep put and call premiums closely related.

Some of the central assumptions of Stoll (1969)'s put-call parity formulation are that the options market is frictionless, competitive and risky. In addition, there are no restrictions on borrowing or short selling. Table 5.1 demonstrates the potential combinations that will enable arbitrage should mispricing occur. If an investor believes that call premiums are too high compared to put premiums then he or she can write a call, buy the underlying share and buy a put. Conversely, if it is believed that put premiums are too high then the appropriate strategy will be to write a put, short the underlying share and buy a call. These combinations will leave an investor without any net exposure to the put, call or the share price; hence the only exposure is to the relative mispricing.[64]

Table 5.1: ***Zero Net Exposure Arbitrage Combinations***

	Short call	***Long share***	***Long put***	***Net position***
$S_T \leq X$	*0*	S_T	$X - S_T$	X
$S_T > X$	$-(S_T - X)$	S_T	*0*	X

	Short put	***Short share***	***Long call***	***Net position***
$S_T \leq X$	$-(X - S_T)$	$-S_T$	*0*	$-X$
$S_T > X$	*0*	$-S_T$	$S_T - X$	$-X$

These combination trades are the basis of Stoll (1969)'s derivation of put-call parity. If the options are held to maturity then the following occurs. Writing a call option generates a cash inflow (C) whereas the long put results in an outflow (P). To fund the long share position, an amount equal to the share price (S_0) is borrowed over a time period equal to that of the maturity of the options at a cost of $S_0.i$. The long put, short call combination is a perfect hedge and as a result, the strategy will earn the risk-free rate. The cash flows of the strategy are described as:

[64] Both of the combinations described in Table 5.1 will mimic the payoff of a zero coupon bond. However, the short call, long share, long put combination will result in a *net long* position in a synthetic bond, whereas the short put, short share, long call combination will create a *net short* position in a synthetic bond. In order maintain exposure to the relative mispricing, the arbitrager will need to enter into a short position in the actual bond in the situation where he has created a long position in a synthetic bond. If a short position in a synthetic bond has been created, then a long position in the actual bond will be required. In both cases the arbitrager will only be exposed to the relative mispricing.

$$C - \frac{(S_0)i}{1+i} - P = M \tag{5.1}$$

Where *M* is the expected arbitrage profit.

Stoll (1969) points out that when market equilibrium is reached, *M* is equal to zero. Hence, Equation (5.1) can be rearranged to show that, in equilibrium, put and call premiums differ by the risk-free rate:

$$C - P = \frac{(S_0)i}{1+i} \tag{5.2}$$

Thus, Stoll (1969, p. 806) concludes; "If the interest cost is constant, any change in call prices is immediately and fully offset by an equal change in put prices, the call price always exceeding the put price by the interest cost."

Merton (1973a) modifies Stoll (1969)'s equilibrium equation to account for situations other than those where the share price (S_0) is equal to the exercise price (*X*):

$$C - P = S_0 - \frac{X}{(1+i)} \tag{5.3}$$

This is the standard form of the put-call parity that appears in Bodie, Kane and Marcus (2006), Reilly and Norton (2006) and, Kolb and Overdahl (2007), amongst others. Equation (5.3) can be rearranged and the notation changed slightly to accommodate situations where interest is compounded continuously:

$$S_0 + P = C + Xe^{-r(T-t)} \tag{5.4}$$

Where S_0 is the underlying asset,

P is the put premium,

C is the call premium, and

X is the face value of a bond discounted continuously at the risk-free rate.

Equation (5.4) shows that if one were to purchase a call option and a bond with the same maturity as the expiration date of the call (with a face value equal to the exercise price of the call) it would have the same payoff of a portfolio consisting of a long position in equity and a long put option (Stoll, 1969; Bodie, Kane and Marcus, 2006; Reilly and Norton, 2006; Kolb and Overdahl, 2007). Table 5.2 and 5.3 illustrates how the payoffs of a long position in a call and a bond are equivalent to the payoff of a protective put. If the call expires out-of-the-money, the investor will receive the face value of the bond, X. However, if the call expires in-the-money then the investor will exercise the call and use the proceeds of the bond to pay the exercise price, resulting in a position worth S_T.

Table 5.2: ***Payoff of a Long Position in a Call Option and a Long Position in a Bond***

	$S_T \leq X$	$S_T > X$
Value of call option	0	$S_T - X$
Value of bond	X	X
Total	X	S_T

The process is somewhat similar for the protective put. If the put is in-the-money at expiration, the security will be sold for *X*. On the other hand, if put expires worthless, the overall position will be worth S_T.

Table 5.3: ***Payoff of a Protective Put***

	$S_T \leq X$	$S_T > X$
Value of stock	S_T	S_T
Value of put	$X - S_T$	*0*
Total	*X*	S_T

Clearly, Stoll (1969)'s analysis provides the basis for Babbel, Gold and Merrill (2002)'s alternative articulation of the insolvency put, $A - C = Xe^{-r(T - t)} - P$, and according to Babbel, Gold and Merrill (2002) the bondholders of a firm hold a risk-free bond with a short put option, sold to the shareholders of the firm. However, bear in mind that Babbel, Gold and Merrill (2002, p. 14) describe equity "…as a call option on the assets with a strike price equal to the face value of the debt." If this is taken literally, the previous equation implies that bondholders are short both a call and a put option (the opposite of this argument is that shareholders own both a call and a put option) yet it is clear that this cannot be the case. It has to be one or the other; it cannot be both. Given the preceding arguments it is clear that shareholders will always behave as if they hold a tacit call option.

5.4.2. Merton (1974, 1977)'s Seminal Application of the Insolvency Put Model: Only Appropriate for Analysing Third Party Guarantees

As discussed previously, the idea that a debt obligation can be expressed as a risk-free bond less a put option was initially introduced by Merton (1974). While Merton (1974) provides a useful analysis for determining the factors that influence bond yields he does not suggest that default occurs when shareholders exercise a valuable put. Rather, Merton (1974) decomposes a risky corporate bond into a risk-free bond and a put option to determine the how the volatility of the underlying firm, the maturity of the bond and the overall leverage of the firm affect its value. The original association of default with the exercising of a put option was expressed by Merton (1977) to specifically describe deposit insurance.

Merton (1977) uses the put option framework to describe a specific mechanism such as deposit insurance because in such an instance there is a third party[65] guaranteeing the bank's deposits. In the case of deposit insurance, should a bank collapse, the guarantor will step in and make good on the bank's deposits. Hence, Merton (1977, p. 8) argues that, "…by guaranteeing the debt issue, the guarantor has issued a put option on the assets of the firm which gives management the right to sell those assets for B dollars on the maturity date of the debt." However, Merton (1977) acknowledges that in the event of default, a firm with a third party guarantee will default its assets to the guarantor. Indeed, irrespective of the guarantee, if the value of the firm's assets exceeds the value of its debt obligations then the bondholders will be paid X and shareholders will receive the residual, $A - X$. If the firm's assets are less than its obligations then it will default and the guarantor will pay the bondholders X, which translates into a loss for the guarantor of $X - A$. Shareholders, will receive nothing. Overall, the presence of a guarantee does not change the payoff to shareholders. Their payoff remains a $Max(0,\ A - X)$. Therefore, Merton (1977) illustrates that a firm has limited liability regardless of any third party guarantees of its debt or deposits in the case of a bank. Only the system of a third party guarantee, such as deposit insurance, can be described as an insolvency put and not the act of default by the firm itself because the shareholders still do not receive anything in the event of default nor are they liable to make any of the firm's obligations whole. Moreover, it can be argued that the debt holders or bank depositors and not the shareholders own the guarantor's short put.

Babbel (1998), Babbel, Gold and Merrill (2002) and Babbel and Merrill (2005) ignore the fact that Merton (1977) only applied the concept of the insolvency put to deposit insurance and did not use it to describe default itself. However, Babbel, Gold and

[65] The Federal Deposit Insurance Corporation (FDIC) for commercial banks is a separately funded entity that guarantees savings of depositors in the US.

Merrill (2002)'s reworking of the put-call parity formula to state, $A - C = Xe^{-r(T-t)} - P$, is conceptually correct. Despite this, shareholders will always behave as if they hold an implicit call option and it is the put-call parity relationship that explains why this is so.

5.4.3. Interpreting Collateralised Loans Using the Put-Call Parity Relationship

When the put-call parity relationship is applied to collateralised loans some interesting conclusions arise. Smith (1986b) and Bodie, Kane and Marcus (2006), like Babbel, Gold and Merrill (2002), illustrate how collateralised loans can be described using put or call options despite the fact that the payoffs of puts and calls are starkly different. The unifying thread is the put-call parity relationship.

Consider a loan where, at maturity, the borrower is obligated to repay the principal, X. At maturity, the value of the pledged collateral will be worth A_T. This situation creates an embedded call option which is held by the borrower. In theory, the borrower will repay the loan only if the collateral (i.e. the firm's assets) is more valuable than the loan itself. If the collateral is worth less than the loan, the borrower will default on the loan and hand over the collateral in lieu of payment. Such an arrangement essentially allows the borrower to turn over the collateral to the lender but keep an option to repurchase it for X at maturity of the loan. Hence, the borrower will exercise the implicit call option if $A_T > X$, by repaying the loan. Indeed, under these circumstances Black and Scholes (1973) regard the shareholders as holders of a call option on the firm's assets. Black and Scholes (1973, p. 650) state; "In effect, the bond holders own the company's assets, but they have given options to the stockholders to buy the assets back."

There is another way of looking at this example. Suppose the borrower commits to repay the loan for X but has the right to sell the collateral to the lender for X to satisfy the loan (of course this will only happen if the collateral is worth less than the loan). Here the borrower has a put option, which will be exercised if $A_T < X$.

Viewed through the lens of the put-call parity theorem, these two examples are one in the same. In the call option example, the borrower turns over the asset but has a call option should the asset be worth more than the loan. In the put option case, the borrower is obligated to repay the loan but retains a put option.

From the lenders perspective this can be viewed as either a covered call or a short put. No matter what, the lender does not benefit from any asset values greater than *X*. The borrower will view this same situation as a long call regardless. Even if the borrower holds a put option that enables the firm to default on its obligation, ultimately, he or she will still hope that the firm's assets will be worth more than the loan. Consider the circumstances of a protective put. A protective put is one of the simplest forms of portfolio insurance, which allows an investor to create a minimum or a floor value for their portfolio. Hence, investors that use the protective put strategy anticipate that share prices will rise making their portfolio more valuable but have a put option that will only be relied upon should things go horribly wrong. If things do go badly and the put is triggered then the investor will have lost some wealth but will not have lost everything. As with all forms of insurance, people hope that they will never need it yet are still willing to pay for the potential cover. Thus, even if the insolvency put option does indeed exist, equity holders will exercise it only as a last resort upon insolvency. Here again, the hope is that the company's assets will be worth far more than its liabilities. As mentioned previously, shareholders only receive the residual value of a company. So, even if they hold a put option shareholders will behave as if they in fact hold a call option (this should not be surprising; the put-call parity states that a long stock plus a long put is equivalent to a call option), which creates a starkly different incentive than that suggested by the insolvency put idea.

Thus, the put-call parity relationship explains how shareholders that are said to hold an insolvency put option will actually behave as if they hold a call option. In addition, it can be argued that issuing fixed income securities is equivalent to having a short position in a bond. Thus, using the put-call parity formula, equity holders' payoff profile can be expressed as

$$C = A + P - Xe^{-r(T-t)} \qquad (5.5)$$

Therefore, this result demonstrates that shareholders will never view default from the perspective of a put option. Shareholders will always behave as if they hold an implicit call option regardless.

In fact, as a case in point, Myers and Read (2001) implicitly acknowledge this. They describe shareholders equity as $E_T = Max[0, (A_T - L_T)]$. Stated in words this says that the payoff to shareholders will be the excess of assets over liabilities or zero otherwise. This is precisely the payoff that a call option provides; if the call ends up in-the-money then an investor obtains the excess of the share price over the exercise price of the option contract. If the share price falls below the exercise price when the call option contract expires then the call is worthless. Hence, there is no need for an insolvency put option to introduce limited liability because Myers and Read (2001) have already indicated that shareholders enjoy limited liability by describing the payoff to shareholders as the maximum of $A_T - L_T$ or zero. In fact, Myers and Read (2001, p. 552) state "…shareholders have the option to payoff the insurance policies and thereby realise the residual value, if any, of the assets." They are clearly describing a call option.

Sherris (2007), also discusses the insolvency put option framework and considers a situation where shareholders have unlimited liability. Thus, shareholders would provide any shortfall of liability claim payments over premiums as they would be compelled to pay all claims. If shareholders were to purchase a guarantee from the outset that would cover such a shortfall, then the value of the guarantee will be that of the insolvency put option; the maximum of $L - A$ or zero otherwise. Consequently, this guarantee (or insolvency put option) will protect shareholders if policyholders' claims outstrip premium income. However, Sherris (2007) goes on to describe the payoff to equity holders as a *maximum of $A - L$ or zero otherwise.* Hence, Sherris (2007) is describing the payoff of a call option.

Ultimately, by defaulting, shareholders are allowing an out-of-the-money call option to expire. Again, Myers and Read (2001, p. 555) acknowledge this; "…default is a deep out-of-the-money option…" There appears to be a contradiction in that, Myers and Read (2001) and Sherris (2007), in particular, regard default as the exercise of a valuable insolvency put option but describe the payoffs to equity holders from the perspective of a call option. Similarly, it makes no sense to view default itself as a deep *out-of-the-money* option and illustrate the process from the standpoint of a put option. How can shareholders be exercising a valuable put option if default is a deep out of the money option, as the insolvency put option framework states? Any rational shareholder would only exercise such an option if it were *in-the-money*. Therefore it follows that shareholders will always view their expected return to equity from the point of view of a call option regardless if they enjoy limited liability from the outset or if it is created artificially via an insolvency put option.

5.4.4. The Insolvency Put Option and the Importance of Time Value

One of the key determinants of option pricing is time to maturity. Without knowing an option's maturity, the relative values of the underlying asset and the exercise price say little about the option's value. As discussed previously, an option's maturity has an important influence on the relative values of put and call options. Indeed, the simple assumption made by Black and Scholes (1973) and Merton (1973b) that a deep out-of-the-money option will always be left to expire worthless is only valid if maturity is held constant. In addition, maturity is also a key factor in Stoll (1969)'s put-call parity framework. He argues that an option holder will not rationally exercise their option prior to maturity.

Stoll (1969) illustrates this by demonstrating the instance where an investor converts puts into calls. The conversion of puts into calls is created by writing a call, buying the underlying asset and buying a put. Should the written call be exercised against the converter when there is still significant time to maturity, the converter will be left in the situation where the long put will have realisable value which is not included in the arbitrage operation $C – (S.r)e^{-r(T-t)} – P = M$. The value of the put stems from the non-zero probability that the share price will fall below the exercise price before the option expires. Similarly, the call will also have value that is not captured by the previous equation because it is just as likely that the share price may rise further, making the call even more valuable by the time it matures.[66] It follows that this analysis also holds for converting calls into puts, $P + (S.r)e^{-r(T-t)} – C = N$.

If investors anticipate the possibility of early exercise, Stoll (1969, p. 808) points out that the cash flows $–(S.r)e^{-r(T-t)}, –P, (S.r)e^{-r(T-t)}$ and $–C$ will be overstated "…by a factor which depends on the probability that the option written by the converter will be exercised before maturity," and as a result, "…profits are no longer certain; and the position, (1)[67] or (2)[68], is not, strictly speaking, a perfect hedge (even though the only 'risk' is that the converter does better than contracted for)." Yet this will only be true if investors ignore the benefit of time value. Hence, Stoll (1969) contends that any investor that exercises an option prior to maturity is essentially discarding time value. Thus, as mentioned previously, an option holder will sell his position rather than exercise it so as to capture time value, which can be expressed as $X – Xe^{-r(T-t)}$. As a result, the initial short call, long stock and long put combination will not be disrupted by early exercise.

Stoll (1969) demonstrates that if calls sold only for their intrinsic value (i.e. if they only sold for $S_t – X$, which ignores time value) whereas put option premiums priced in

[66] Stoll (1969) assumes that percentage changes in share prices are symmetrical. That is, a share has the same chance of rising above an option contract's exercise price as it does of falling below it.
[67] Position (1) refers to converting puts into calls: $C – [S.r(e^{-r(T-t)})] – P = M$.
[68] Position (2) refers to the conversion of calls into puts: $P + [S.r(e^{-r(T-t)})] – C = N$.

time value then there will be an incentive to convert calls into puts. That is, write put options, short the underlying shares and buy call options. The cash flows are described as follows:

$$-C_t + P_t + \left(S_t r\right)e^{-r(T-t)} + \Delta S_t e^{-r(T-t)} = N_t \qquad (5.6)$$

The term $\Delta S_t e^{-r(T-t)}$ is the expected profit from this arbitrage transaction. Thus, $\Delta S_t e^{-r(T-t)}$ is in actual fact time value. In aggregate, the forces of arbitrage will cause the price of calls to rise (relative to puts) and N_t will equal zero. Solving for C_t and dropping out $e^{-r(T-t)}$ yields:

$$C_t = \Delta S_t + S_t r + P_t \qquad (5.7)$$

Hence, Stoll (1969) establishes that the early exercise of a call would be irrational because exercising the call prior to maturity will only yield ΔS_t, whereas selling the call and keeping it "alive" will bring in $\Delta S_t + S_t r + P_t$. It is this proof that leads Stoll (1969, p. 810) to conclude "…the hedge position is maintained to maturity; and if it is not, with the knowledge that, in an rational market, the hedge will never be liquidated, so that the option purchased by the converter yields him additional profit."

Black and Scholes (1973) also demonstrate that early exercise is not optimal. Generally speaking, in terms of the Black and Scholes (1973) option pricing model, the values of puts and calls increase with maturity. Holding the value of the underlying asset constant, when expiration is exceptionally far into the future the value of a call option will tend towards the value of the underlying asset. Black and Scholes (1973) provide an intuitive explanation as to why this is so; the present value of a bond that has a face value equal to the exercise price of the call will be so low that the exercise price can almost be ignored. That is for exceptionally long maturities, $Xe^{-r(T-t)}$ will tend towards zero, hence the value of a call will be S_t which is the value of the underlying asset. Thus, given enough time value a deep out-of-the-money call can potentially end up in-the-money.

However, the influence of maturity on put options can be contradictory. As mentioned above, a put option is generally more valuable the longer its maturity because this gives the holder more time, and hence opportunity, for the underlying asset to move in the desired direction and therefore result in a valuable put. Yet, as is the case with European put options, longer maturities will make the put less valuable because the holder cannot exercise the option and realise its value immediately. The longer the maturity, the smaller $Xe^{-r(T-t)}$. Hence, The same property that makes delaying exercising a call option until maturity optimal is what ultimately destroys the value of a put option. Taken to the extreme, the value of a put option will tend to zero given a long enough time to maturity.

Given the incentive to exercise a put option early, Merton (1973a) asserts that Stoll (1969)'s general derivation is unique only for European options, which of course can only be exercised upon maturity. While Merton (1973a) agrees that early exercise of dividend-protected[69] call options is irrational, he points out in the case of American put options, early exercise will in most instances be optimal. According to Merton (1973a), Stoll (1969) is implicitly suggesting that American style options (where early exercise is possible) will have identical values to European options. However, the flexibility provided by American options will make them more valuable vis-à-vis European options, which makes Stoll (1969)'s assertion incorrect.

[69] Call option holders are not entitled to any dividends (or any other cash flows arising from the underlying asset), as they do not own the underlying asset, only the contract entitling them to buy the underlying asset for the exercise price. Consequently, a long call position will decline in value whenever a dividend is paid because the contract holder will not benefit from the payment. Furthermore, the call will be less valuable due to the drop in share price when it trades ex-dividend. Dividend-protected call option contracts make upward adjustments in the number of shares controlled by the contract to compensate for this loss of value.

Moreover, Stoll (1969)'s derivation assumes that the parity value of a put (i.e. the value of a European put) will be greater than the value of an American put only if $C > rXe^{-r(T-t)}$. However, Merton (1973a) demonstrates that this will not always be the case because given a small enough share price (S_t), $C < rXe^{-r(T-t)}$ and the put option's exercise value will exceed its parity value. Hence, an American put will exceed the value of a European put. Also, an American put option must be a non-decreasing function of time to maturity. The reason for this is that an investor can potentially do anything with puts that expire far into the future given the flexibility of early exercise. Thus, American puts with long maturities should be worth more than shorter-term puts. Merton (1973a) argues that if American and European put options have equal value then American put options must tend to zero as maturity tends to infinity. To suggest that an American put option's value will decline as maturity increases (which is what Stoll (1969) implicitly assumes) contradicts the fact that American options are non-decreasing functions of time. Hence, Merton (1973a, p. 183) states; "If the value of an American put option always equals the value of its European counterpart, then the value of the American put option must tend to zero as its time to maturity tends to infinity. But, the value of an American put option is a non-decreasing function of its time until expiration, from which it follows that all American (and hence, European) put options must have zero value which is clearly nonsense." Stoll (1973), in response to Merton (1973a), concedes this point but argues that Merton (1973a)'s example overstates the deviation from the put-call parity band because the share price decline necessary for $C < rXe^{-r(T-t)}$ is not likely to occur in reality. Overall, Stoll (1973) maintains that the derivation of the put-call parity relationship for American options will not differ greatly from the original European option case.

6. Counter Argument to the Capital-Based Allocation Framework

6.1. A Critique of Myers and Read (2001)

Despite the growing influence of the capital-based insolvency put option frameworks, several criticisms have emerged. In particular, much of the criticism is centred on Myers and Read (2001)'s influential model. For instance, Venter (2003) questions the appropriateness of Myers and Read (2001)'s assumption that aggregate losses are log-normally distributed. He asserts that it will not be a sensible assumption for every company examined. While it is feasible to use other distributions for Myers and Read (2001)'s method, Venter (2003) argues that the choice of possible alternatives introduces additional ambiguity as most of those suggested in the literature are picked without much justification. What is more, Venter (2003) suggests that the Myers and Read (2001) framework will not provide clear guidance about the profitability of different insurance lines and should not be used as a starting point for a return-on-capital calculation. The reason is that other sources of profits are included that are not proportional to the allocated capital and there is no theoretical justification to suggest that Myers and Read (2001)'s method would result in more appropriate pricing in this regard. However, in a paper published in 2004 Venter changes his mind and embraces the Myers and Read (2001) approach. He argues that it is not only appropriate for pricing insurance contracts but should also be the basis for allocating frictional capital costs.

On the other hand, Mildenhall (2004) shows that the adding up property of the Myers and Read (2001) framework will only occur if the loss distributions of the insurance line are homogenous. In reality, Mildenhall (2004) demonstrates that loss distributions are not homogenous, which means that it has little practical application in the insurance industry.

Gründl and Schmeiser (2005) question the need to allocate capital in order to price insurance contracts. They argue that the allocation of marginal surplus (i.e. capital) to individual lines of insurance are not unique and do not add up to the total capital requirement of an insurer as Myers and Read (2001) claim. In fact, they contend that Myers and Read (2001)'s method is unnecessary for insurance pricing and can result in improper premiums. Furthermore, they assert that net present value frameworks result in more considered and pragmatic capital budgeting decisions compared to capital allocation overall.

Gründl and Schmeiser (2005)'s main criticisms are centred on the usefulness of capital allocation methods. In addition, they argue that Myers and Read (2001)'s focus on the distributional and technical properties that capital allocation methods should engender ignore the "…context of the company's economic goals" (Gründl and Schmeiser, 2005, p. 1).

In keeping with Myers and Read (2001), Gründl and Schmeiser (2005) follow the one-period option-pricing outline where the insurance premium is dependent on the default value. Gründl and Schmeiser (2005) expand the asset term (A) to include equity, E, and the competitive insurance premium, P, which once invested earns a stochastic rate of return, r:

$$D^i = PV\left(Max\left\{L^i_{t+1} - \left(E^i_t + P^i\right)(1+r), 0\right\}\right) \qquad (6.1)$$

The competitive premium based on the initial insurance portfolio is dependent on the insolvency put value:

$$P^i = PV\left(L^i_{t+1} - D^i\right) \qquad (6.2)$$

Where,

$$D^i = PV\left(Max\left\{L_{t+1}^i - \left(E_t^i + P^i\right)(1+r), 0\right\}\right)$$

The premium of any new line of insurance will be dependent on the marginal default value, $d_i = \partial D/\partial L_i$. Holding the initial marginal default value constant the new premium will be proportional to the new line of insurance:

$$P^{i+1,n} = PV\left(L_{t+1}^{i+1,n}\right)\left(1 - d_i\right) \qquad (6.3)$$

Given the marginal default value:

$$d_i = \frac{D^i}{PV\left(L_{t+1}^i\right)} \qquad (6.4)$$

The implication of Equation (6.4) if D^i is zero (i.e. assets exceed liabilities) is that d^i will also be zero:

$$d_i = \frac{0}{PV\left(L_{t+1}^i\right)} = 0$$

Thus, the default option has no value and therefore the full, default free, premium is charged:

$$P^{i+1,n} = PV\left(L_{t+1}^{i+1,n}\right) \equiv P^{i} \qquad (6.5)$$

Conversely, if D^i is positive (assets do not cover liabilities) the default option has intrinsic value and as a result can potentially be exercised. This can be illustrated as follows:

$$d^{i} = \frac{PV\left(Max\left\{L_{t+1}^{i} - \left(E_{t}^{i} + P^{i}\right)(1+r), 0\right\}\right)}{PV\left(L_{t+1}^{i}\right)} \qquad (6.6)$$

Once more*,* $P^{i+1,n}$ is established solely on the marginal default value, d_i. If the riskiness of the insurer changes (due to changes in magnitude of the default option) then so too will the premium. Thus, $P^{i+1,n}$ is a market-clearing premium only if the marginal default value remains the same. For the marginal default value to remain the same, an insurer must uphold sufficient risk management measures[70] (Gründl and Schmeiser, 2005). Gründl and Schmeiser (2005) point out that if an insurer charged $P^{i+1,n}$ on a new line of insurance and carried out no further risk management the firm would generate a net present value of $\mathrm{PV}_{i+1,n}$. However, in a competitive market, $\mathrm{PV}_{i+1,n}$ is the price of an additional measure of risk management (which maintains the level of d_i), hence, $\mathrm{PV}_{i+1,n} = \mathrm{PV}_{risk\ man,n}$. The only risk management measure that Myers and Read (2001) consider is capital and $\mathrm{PV}_{risk\ man,n}$ is the competitive price for any change in capital. Gründl and Schmeiser (2005) go further and maintain that the size of the change in capital, given its price, can be determined by setting the overall capital of an insurer equal to future payments to its shareholders:

[70] For example, equity capital, reinsurance and hedging strategies.

$$E_t^i + E_t^{i+1,n} = PV\left(Max\left\{ \begin{array}{c} \left(E_t^i + E_t^{i+1,n} + P^i + P^{i+1,n}\right)(1+r) \\ -\left(L_{t+1}^{i+1,n}\right), 0 \end{array} \right\} \right) \quad (6.7)$$

The competitive premium, $P^{i+1,n}$, is now an input that determines the amount of extra capital required. This leads Gründl and Schmeiser (2005, p. 7) to conclude "…there is no need to allocate capital back to lines of business (or to single contracts) when making pricing decisions or determining the change in equity capital needed." Furthermore, Gründl and Schmeiser (2005, p. 8) assert that "…in the context of a perfect capital market, capital allocation to lines of business is neither needed for pricing insurance contracts nor for determining the change in the insurance companies equity capital after it writes a new contract." Hence, Gründl and Schmeiser (2005) reject Myers and Read (2001)'s notion that line-by-line capital allocations add up to the total capital requirement of an insurer and any adding-up property of such a capital allocation rule will not be "…unique and not arbitrary" (Myers and Read, 2001, p. 545).

Furthermore, Gründl and Schmeiser (2005) examine Myers and Read (2001)'s process of allocating frictional costs (e.g. issues arising from double taxation or agency issues) to individual insurance lines. They argue that allocating equity and frictional costs in the method suggested by Myers and Read (2001) will lead to inconsistent contract pricing. The reason behind this, as Gründl and Schmeiser (2005) argue, is that the allocation of frictional costs relies on the initial insurance portfolio. However, when a new contract is sold, the portfolio changes and so too will the allocation of frictional costs. This inconsistency could lead to inappropriate business decisions, influencing a firm to exit an insurance line because the market insurance premium is lower than the estimated price.

Gründl and Schmeiser (2005) contend that the use of the Myers and Read (2001) approach will necessitate that authorities assume a benchmark insurance company with certain risk management techniques in order to settle on the appropriate capital allocation. Gründl and Schmeiser (2005) acknowledge that Myers and Read (2001) address the issue of an efficient risk management mix but note that they provide no realistic answer to the problem. Moreover, even if equity-driven common costs are correctly allocated, the problem of allocating common costs that are not equity-driven remains. Overall, there is the potential for price regulation to result in too low a premium, resulting in an absence of insurance coverage (Harrington, 1984; Grabowski, Viscusi and Evans, 1989; Harrington, 1992; Klein, Phillips and Shiu, 2002; Gründl and Schmeiser, 2005).

Taken as a whole, the general premise of the capital-based, solvency models is that the more capital an insurer has, the smaller the default option. As illustrated by Gründl and Schmeiser (2005), a smaller default value results in a larger premium. This means that safer firms (i.e. insurers with more capital) charge a higher premium. Conversely, policyholders will expect to pay a lower premium to an undercapitalised, less secure insurer to reflect the possibility that their claims may not be covered. However, capital should have no influence on the premium charged by an insurer, nor is it necessary to allocate capital to individual lines of insurance as pointed out by Gründl and Schmeiser (2005).[71]

[71] Gründl and Schmeiser (2005) cite three typical steps of capital allocation techniques that can lead to problems. First, capital is allocated to the entire firm, which in the Myers and Read (2001) case will be based on the insolvency put option. Second, capital is then assigned to different insurance lines using any one of the methods discussed in Cummins (2000). Earnings from other insurance segments are then compared with the cost of the assigned capital. Third, comparing the cost of capital to earnings enables decisions to be made with respect to expanding or contracting insurance lines. Gründl and Schmeiser (2005) point out that the first step can be subject to significant ambiguity given the large array of capital allocation methodologies found in the literature. The second point can lead to what Gründl and Schmeiser (2005) describe as a common cost problem. Should insolvency occur, it will be the result of liabilities exceeding the assets of the entire company not just a particular business segment. The issue is that information limitations do not allow for a non-arbitrary method for allocating common costs for the aim to measure performance and determine pricing. According to Gründl and Schmeiser (2005, p. 16) "…the generally accepted response is to develop a set of desired properties for the allocation process itself and proceed with a method that best satisfies these properties." However, they point out that whatever method is chosen, distortions will result. Furthermore, the variety of possible performance measures means it is possible to create virtually any profit ranking of an insurance line. As for the third step, should the firm decide to windup what has been identified has the most unprofitable line, there is the risk that a natural hedge could be eliminated. This could necessitate additional, costly risk management measures that ultimately hurt overall profitability.

6.2. Point of Departure: Adam Smith (1776/1976)'s Statement on Insurance and the Irrelevance of Capital

The notion that capital is of peripheral importance to insurance pricing is not new. As discussed previously, Hill (1979) argues that capital should not have any sway in the creation of insurance contracts. Furthermore, recall Hill (1979, p. 180)'s observation that "…an insurance contract has no 'return to capital' dimension". Yet Hill (1979)'s arguments were not new at the time either. In fact, the relative unimportance of capital in short-term insurance pricing had long been established in financial economics. However, over the years this reasoning has simply been forgotten. Moreover, the influence of Basel-style capital-based regulation that has dominated thinking on the approach to regulating the short-term insurance industry for the past decade or so has, almost without question, set the prevailing wisdom that capital be central to the creation of insurance policies. Vivian (2007a) notes that it was Adam Smith (1776/1976) who initially pointed out the irrelevance of capital when he described the basic components for a successful insurance company: "In order to make insurance, either from fire or sea-risk, a trade at all, the common premium must be sufficient to compensate the common losses, expense of management, and afford such a profit as might have been drawn from an equal capital employed in any common trade. The person who pays no more than this, evidently pays no more than the real value of risk, or the lowest price at which he can reasonably expect to insure it" (Smith, 1776/1976, p. 121).

The idea that insurers require capital stems from the inappropriate application of the principles of banking to that of the short-term insurance industry. Vivian (2007b, p. 28) argues; "Banks are concerned with the preservation of depositors capital, the nations savings; short-term insurers are not." Furthermore, Vivian (2007a) highlights the fact that insurance claims are paid out of premium income and not capital. Consequently, an insurer "…is then operating exclusively in the service industry" (Vivian, 2007a, p. 2). Thus, if an insurer is deemed to be undercapitalised, and therefore needs to increase its equity cushion, the only way it can increase the amount of capital in the long run is to charge a higher premium. Moreover, an insurer will not mitigate the gap between potential losses that can arise from existing policies and current assets by charging a premium that does not compensate it for the common risk (Vivian, 2007a).[72]

Another significant reason why capital is not the cornerstone of an insurer is that a critical element of insurance is the basis of pooling. Vivian (2007a) maintains that the pooling phenomenon lends itself to the creating of multiple risk pools. He illustrates this by observing that with a single risk pool all policyholders would be charged a common premium. This means that an individual with lower risk will be charged the same rate as a much riskier policyholder. Hence, the lower risk policyholder will be paying a premium that is not consummate with the risk and "…in a free market, a new insurer will offer the appropriately rated policy to the low risk insured" (Vivian, 2007a, p. 4). It follows that policyholders will migrate to the insurance pool that reflects their risk, forming many risk pools. As a result, there will be various pools of insurance within any insurer, each with their own market clearing risk-adjusted premium. Hence, Vivian (2007a) points out that competitive premiums will be determined by the riskiness of the activity that is insured and not by the amount of capital allocated to any line of insurance.

In keeping with Vivian (2007a)'s argument, Sherris (2007, p. 8) states; "In a perfectly competitive market insurers will sell at market clearing risk adjusted premiums. If an insurer charges above the competitive market premium then it will not sell any business since other insurers will offer the competitive market premium." Thus, if a policyholder were to obtain insurance for his or her car, house, etc, he or she would be charged a common rate. That common rate would be based on the market clearing risk-adjusted premium described by Sherris (2007). Indeed, Sherris (2007) points out that the financial models (Myers and Read, 2001; Sherris, 2006; Sherris and van der Hoek, 2006; Yow and Sherris, 2007) developed to determine the competitive premium assume that insurers are price takers, which means that premiums reflect

[72] If it is possible for an insurer to charge a premium that generates an income that always surpasses expenses, such an insurer will always be profitable. This in itself will eliminate the need for capital. However, Vivian (2007a) concedes that in reality this will not always be the case as in certain years abnormal claims will occur, necessitating a margin of safety.

systematic risk factors. However, these models rely on capital in order to be implemented and any calculated premium is based on the allocation of capital to each line of business. Yet, Vivian (2007a, b)'s and Sherris (2007)'s observation that risk is the main driver of the premium charged contradicts the method of using capital to generate competitive market premiums. Consequently, it should be of no surprise that Cummins et al (1995), Grace et al (1998), Klein et al (2002), Pottier and Sommer (2002) find that RBC is not a robust indicator of a property-liability insurer's solvency. Capital is largely unimportant to the financial health of a short-term insurer (Vivian, 2007b).

If premiums calculated using a capital allocation framework are higher than the intrinsic risk of the line of insurance then a new insurer will charge a lower premium based on the actual risk that the policy engenders. This is consistent with the pooling phenomenon described by Vivian (2007a) and Sherris (2007)'s description of market clearing risk-adjusted premiums. The point is, however, that an insurer will not sell any insurance if it charges a premium that is higher than the market clearing rate of the insurance pool it underwrites.[73]

[73] While policyholders may be willing to pay a premium that is greater than the competitive market rate in certain circumstances, the overwhelming incentive will be for lower risk policyholders to migrate to an insurer charging a lower premium based on the actual risk (Sherris, 2007; Vivian, 2007a).

7. The Implications of the Capital-Based Insolvency Put Option Framework

As argued previously, the issue with the capital-based insolvency put option framework lies with the fact that the option to default, and limited liability in general, is viewed as a put option. Yet, if one arrives at the conclusion that shareholders hold an implicit call option (whether or not it arises directly out of limited liability or if it is fashioned synthetically by the combination of a put option and equity as illustrated by the put-call parity theorem), why then is there a problem with describing default as the exercise of a put option? The issue is that the incentives behind a put and a call option are starkly different. Furthermore, with the previous arguments of the irrelevance of capital in mind, the regulatory focus on capital adequacy will compound the problems associated with the insolvency put model.

If shareholders behave as if they hold an implicit call option on an insurer’s assets, their reactions to risk and potential insolvency will differ greatly than if they believed that default occurred when a valuable put option was exercised rather than leaving an out-of-the-money call option to expire worthless. In other words, when it comes to bankruptcy investors will not congratulate themselves on cashing in a valuable put option but will rather rue the fact that their investment is now worthless and that any potential for future growth has now evaporated. That is not to say that the value of limited liability is ignored by shareholders, however, an investor’s ultimate goal is to buy shares in companies that will appreciate in value, which indicates that shares are valued for their inherent optionality; specifically a call option (Myers 1977). Consequently, the interpretation of the capital-based insolvency put option framework changes.

To reiterate, the value of the default put option is $D = L - A$, which reflects the amount of capital needed to plug the shortfall between the assets of an insurer and its liabilities. If this is viewed as a call option then the excess of liabilities over assets can rather be interpreted as the amount saved by choosing not to exercise the call. Hence, shareholders do not have a profitable call option to exercise and they will defer any exercise of the call until it becomes profitable (i.e. they will wait until $A > L$). Therefore, if shareholders anticipate that the firm will be profitable in the future they will keep the firm operating as a going concern. Shareholders will not allow the call option to expire (i.e. the shareholders will not default on their obligations) because they have an incentive to keep the company running. Thus, if there is a situation where assets fall short of liabilities, the shortfall is not as great a threat as suggested by the insolvency put option framework. Provided there is a chance that assets will exceed liabilities in the future then shareholders will not default. A key point being alluded to here is time to expiration. The gap between assets and liabilities (or S_t and X as discussed previously) in itself does not dictate if a call option is valuable or not. Several other factors such as volatility, interest rates and time to expiration are hugely important (Black and Scholes, 1973; Merton, 1973a). Thus, the simple observation that assets are less than liabilities says little about the value of an option. The context of the situation is important just as in the case of higher firm risk. The relative values of an insurers assets and liabilities are more meaningful if one knows the values of the other variables; notably time to expiration. If assets are less than liabilities at expiration then the call will expire worthless and the insurer will default on its obligations. Thus, time to expiration can be interpreted as a company's going concern. As long as it is conceivable that the insurer will be profitable (without crippling short run expenses) then an out-of-the-money call option will remain in place in the anticipation of it becoming valuable at a future date.

Hence, shareholders of an insurer will react differently to that predicted by the insolvency put framework. As mentioned earlier, the insolvency put argument is at odds with how shareholders behave in practice when dealing with risky firms on the brink of bankruptcy. Consequently, the insolvency put argument sounds an alarm bell, pointing out the possibility of default exists because shareholders will rush to exercise a valuable put. If this is viewed as a call option, however, then the insurer is not faced with a valuable put option and imminent exercise but an out-of-the-money call. Thus, shareholders will try to keep the business alive in the hope that the call ends up in-the-money one day. Of course this will only happen if it is not prohibitively costly to do so. Such short-term costs could be tantamount to eliminating time value and once that happens, shareholders will choose to default and turnover the firm's assets to its creditors. As a result, an out-of-the-money call only becomes a risk at expiration. The chance of insolvency is not as likely as that suggested by the insolvency put model because it focuses on non-expiration situations, which overstates the risk of default. In short, the risk of insolvency is overstated, which in turn overstates the amount of capital needed for a safe insurer.

As indicated, it is the interaction of the insolvency put and the regulatory emphasis on Basel style capital adequacy that leads to problems. In fact, in severe economic downturns the capital-based insolvency put framework can actually act as a catalyst, increasing the likelihood of bankruptcy; precisely the opposite of what it is meant to do. For instance, Borio (2009) discusses the use of capital adequacy and minimum liquidity standards as buffers against market shocks and notes that such standards can actually lead to greater procyclicality during a crisis, which will exacerbate the lack of funding liquidity.[74] Moreover, Adrian and Shin (2009) outline how marking to market and active management of balance sheets by financial intermediaries causes procyclicality of leverage, which ultimately affects aggregate price levels. Such behaviour can result in severe declines in market and funding liquidity during a crisis. As asset prices decline, leverage will increase and firms will react by reducing the size of their balance sheet to reduce their leverage. In the extreme, this can result in a dismal cycle of balance sheet weakness, leading to downward leverage adjustments and further asset price declines.

Consider, for example, the credit crisis of 2007/2008, where a substantial decline in funding liquidity and depressed asset values was experienced. A short-term insurer faced with such an environment will likely experience a decline in its assets relative to its liabilities, which will put its capital cushion under pressure. If the decline in capitalisation is large enough, the minimum statutory capital requirement level will be breached and the capital based solvency measures will prescribe that the insurer

[74] Borio (2009, p. 1) defines funding liquidity as "…the ability to raise cash (or cash equivalents) either via the sale of an asset or by borrowing."

increase its capital. However, as illustrated by Adrian and Shin (2009) and Borio (2009) capital will be hard to come by during market crises. In a market where cash is king, those with ample capital will be reluctant to assist any firm that is subject to regulatory intervention. Although, the central issue is the lack of available capital and the regulatory attention will be of a secondary concern. When capital becomes scarce, funding will dry up regardless of any regulatory intervention. If an insurer cannot raise its capital level to that prescribed by the regulator then the firm will be declared technically insolvent and will be liquidated. However, being declared technically insolvent by the regulator does not necessarily mean the insurer is insolvent in the economic sense. For example, recall that under the RBC framework an insurer with a capital ratio below 200% will trigger regulatory action and the firm will be required to increase its capital cushion. If the insurer is unable to increase its capital then it is possible for the insurer to be declared technically insolvent. Yet any capitalisation level at 100% or above, means that the insurer in question could be liquidated and meet all its liabilities. If an insurer is forced to increase its capital at a time when funding liquidity has disappeared (as in the situations described by Adrian and Shin (2009) and Borio (2009)) it is possible that the insurer will be liquidated, as it will be deemed under capitalised and “insolvent”. Hence, capital-based regulatory frameworks can create an incentive to liquidate otherwise healthy insurers when faced with an inability to raise its capital cushion.

Figure 7.1 illustrates this process. Should a severe economic downturn occur, causing asset values to decline to point where an insurer’s capital cushion comes under pressure, the insolvency put option model (along with capital adequacy standards that dictate when recapitalisation will be required) will prescribe increases in the firm’s capital level. However, as firms move to reduce their leverage in reaction to asset price declines, funding liquidity will contract. When funding liquidity contracts, capital will become scarce, which will prevent the insurer from rebuilding its capital cushion. As a result, the insurer will be in breach of statutory capital adequacy requirements and declared technically insolvent. This will be tantamount to eliminating the time value of the inherent call option and by implication the going concern of the firm. Consequently, with no time value remaining the out-of-the-money call option will be left to expire worthless, shareholders will default on their liabilities and the insurer will be liquidated.

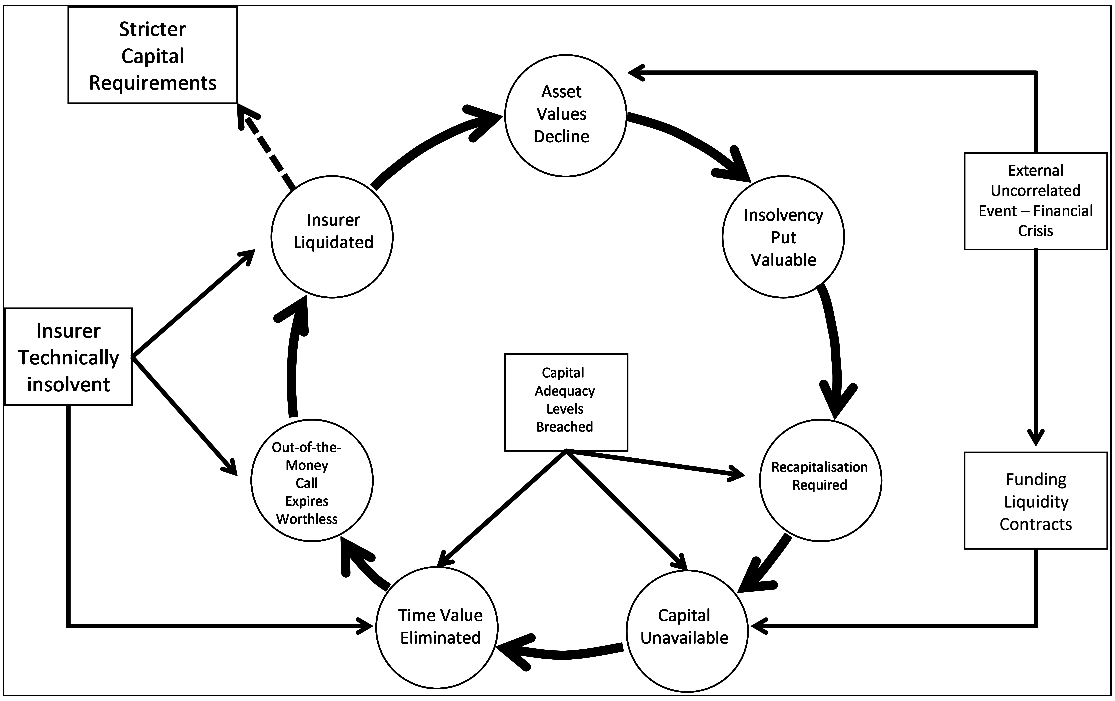
Stricter Capital Requirements
Asset Values Decline
External Uncorrelated Event – Financial Crisis
Insurer Liquidated
Insolvency Put Valuable
Insurer Technically insolvent
Capital Adequacy Levels Breached
Out-of-the-Money Call Expires Worthless
Recapitalisation Required
Funding Liquidity Contracts
Time Value Eliminated
Capital Unavailable

Figure 7.1: ***Market Procyclicality and the Role of Capital Requirements***

Furthermore, it can be argued that even if the insurer's assets actually exceed its liabilities the regulatory stipulation to raise capital when funding liquidity is severely diminished could incentivise shareholders to windup the insurer and payout a liquidating dividend; something far more attractive when faced with the alternative of raising extra capital in a depressed market. Unnecessary liquidations could have further negative knock-on effects leading to yet further asset price declines. In the long-run, capital-based regulation will result in the unnecessary winding up of undercapitalised insurers and those firms that remain will likely reap monopoly profits, which ultimately will hurt consumers. Moreover, the failure of the insurer will be ascribed to its lack of adequate capital, possibly prompting regulators to enforce even stricter capital standards in response (Vivian, 2007b). This combined with the insolvency put option model will exacerbate stresses faced by an insurer during a financial crisis rather than mitigate them.

8. A Mathematical Description of a Short-Term Insurer

8.1. The Fundamental Elements of Premium Income and the Irrelevance of Capital

To evaluate what the appropriate make up of a short-term insurer is, it is instructive to expresses Adam Smith (1776/1976)'s statement outlining the requirements for a successful insurer mathematically. Thus, gross premiums (*GP*) can be broken down into:

$$GP = C + E_u + RoI \qquad (8.1)$$

Where, *C* represents the cost of claims, E_u denotes underwriting expenses[75] and *RoI* is the return on investment to shareholders.

As discussed earlier, Vivian (2007a) points out that an insurer will make money not from the investment of capital but for services provided. In fact, Vivian (2007a) contends that in the case of property insurance should a policyholder claim for a loss, the insurer can be regarded as tantamount to a contractor. Thus, with no capital investment, "…there can be no return on capital" (Vivian, 2007a, p. 2). The profit earned from providing a service is an underwriting profit (P_u). Hence, Vivian (2007a) modifies Equation (8.1) to account for this:

$$GP = C + E_u + P_u \qquad (8.2)$$

Rearranging Equation (8.2) in terms of P_u reflects profit before tax:

[75] Vivian (2007a) separates underwriting expenses further into three major categories: commissions paid to brokers, management costs associated with running the insurer and claims handling costs.

$$\begin{aligned} P_u &= GP \; - \; (C + E_u) \\ PbT &\equiv GP \; - \; (C + E_u) \end{aligned} \tag{8.3}$$

As claims are paid out of premium income, it is essential that premiums be banked as soon as possible once the risk is accepted. Furthermore, as mentioned previously, the pooling phenomenon is central to the operation of an insurer. Vivian (2007a, p. 2) explains; “The income [premium] can be regarded as funds deposited, in advance, in a risk pool and once claims arise these are settled out of that pool.” Consequently, it is necessary to allocate premiums throughout the insurance period to the annual risk pool, which means that premiums need to be portioned to the period they belong. That is, an annual premium cannot be booked in one lump sum. This gives rise to an unearned premium which must be accounted for. In terms of Equation (8.3), an adjustment for unearned premiums (*UP*) must be made:

$$PbT \; = (GP - \Delta UP) - (C + E_u) \tag{8.4}$$

Once the funds are received there is a lag before any claims that require payouts are made. This is the well-documented long-tail aspect of short-term insurance (Cummins and Weiss, 1991; Klein, 1995; Staking and Babbel, 1995; Colquitt, Sommer and Godwin, 1999). Indeed, this is regarded as a significant source of risk for an insurer. However, this lag enables insurers to invest their premium income. Hence, Vivian (2007a) modifies Equation (8.4) accordingly:

$$PbT = (GP - \Delta UP) - (C + E_u) + (I_{inc} - E_I) \tag{8.5}$$

Where, I_{inc} is investment income and E_I represents investment costs. Thus, $(I_{inc} – E_I)$ is the investment profit (P_I).

This illustrates that even in the absence of shareholders capital, an insurer will earn investment income. Thus, Vivian (2007a) articulates the two major sources of income for an insurer; investment profit, P_I, and underwriting profit, P_u.

Another element that mitigates the need for capital is reinsurance. Vivian (2007a) points out that rather than being paid out as an expense, reinsurance premiums are paid by ceding a portion of income to the reinsurer. Furthermore, any recoveries from reinsurers are offset against claims. Consequently, the insurers income and expenses can be expressed as:

$$PbT = (GP - \Delta UP - P_R) - (C + E_u - R) + P_I \qquad (8.6)$$

Where, P_R denotes the premium paid to reinsurers and R is the reduction in claims from reinsurance recoveries.

The next significant issue that needs to be dealt with is policyholders’ claims. Although the process with claims is somewhat more complicated, a basic procedure is followed to establish the size of a claim. Vivian (2007a, p. 5) states; “As soon as a claim is reported to an insurer, the income statement is debited with a claims provision based on the best estimate of the claim and the creditors’ ledger credited with the same amount.” Hence, it follows that C in the formula is in reality a provision for claims and not the actual payment of claims. Consequently, the accuracy (or inaccuracy) of the claims’ provision has a significant impact on the risk faced by an insurer. An inaccurate provision may require an upward adjustment and the objective of the claims’ provision is to provide as close an estimate of the final value as possible (Vivian, 2007a). However, despite the possibility of inaccurate provisions, the real risk faced by an insurer in not the full amount of the claim but the deviations from the actual final value of the claim payment.[76] An insurer will have a good idea as to what its claims will be and, as a result, make appropriate provisions.

Finally, an important source of risk is the possibility that a significant claim (for which the insurer is liable at the time of the event) has occurred without the insurer being aware of it at the time. Hence, an insurance company will need to compensate for this by making a provision for claims that are incurred but not reported (IBNR). Similar to the situation with claims provisions, the risk faced by the insurer is the change in IBNR and not the total amount. As a result, Equation (8.6) is modified accordingly:

$$PbT = [(GP - \Delta UP - P_R) - (C + E_u - R + \Delta IBNR)] + P_I \qquad (8.7)$$

Thus, an insurer's profit before tax, broadly speaking, comprises of underwriting profit (P_u) plus investment profit (P_I):

$$PbT = P_u + P_I \qquad (8.8)$$

[76] For example, if a claim for a stolen car is made and, based on the model and year of the vehicle, R200 000 is provisioned but the actual replacement value is in fact R250 000 then the risk is only R50 000 and not the entire R250 000. This suggests that the claims risk faced by an insurer is significantly lower than the capital allocation models (vis-à-vis Myers and Read, 2001; Sherris, 2006; Sherris and van der Hoek, 2006; Yow and Sherris, 2007) imply.

8.2. Reconceptualising the Insolvency Put Option Model as a Call Option Framework

Vivian (2007a, b)'s mathematical articulation of a short-term insurer can be modified yet further to include the implicit call option held by the shareholders of the firm. Specifically, the call option can be used to explore the influence of time value as well as capture the insurer's growth opportunities and by implication, the potential for bankruptcy. The main objective of the proposed reconceptualised option model of an insurer is to explicitly articulate the influence of time value and the incentives it creates if the shareholders of the firm are assumed to hold an implicit call option on the assets of the company.

In the case of option valuation, the factor of central interest is the random process that is driven by Brownian motion. That is, the main objective is to determine how changes in the value of the underlying share price (or asset) influences the value of the option. One of the tools developed to identify these processes is Itō's formula or lemma (Wiersema, 2008).

The description of a short-term insurer's solvency from the perspective of a call option framework requires that its equity value be a function of time and a Brownian motion:

$$f\left[t,B(t)\right] \qquad (8.9)$$

To derive the complete expression, Taylor's formula[77] for a function of two variables is applied to the random variable as if it were a deterministic variable:

$$\begin{aligned} df = \frac{df}{dt}dt + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(dt)^2 + \frac{df}{dB}dB + \frac{1}{2}\frac{\partial^2 f}{\partial B^2}(dB)^2 \\ + \frac{1}{2}\frac{\partial^2 f}{\partial t \partial B}dtdB + \frac{1}{2}\frac{\partial^2 f}{\partial B \partial t}dBdt \end{aligned} \qquad (8.10)$$

If $\frac{\partial^2 f}{\partial t \partial B} = \frac{\partial^2 f}{\partial B \partial t}$,

$$df = \frac{\partial f}{\partial t}dt + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(dt)^2 + \frac{\partial f}{\partial B}dB + \frac{1}{2}\frac{\partial^2 f}{\partial B^2}(dB)^2 + \frac{\partial^2 f}{\partial t \partial B}dtdB \qquad (8.11)$$

[77] The change in any smooth ordinary non-random variable can be approximated using a Taylor expansion. That is, $f(x_0 + h) - f(x_0)$ can be approximated as

$$\Delta f(x_0) = h\frac{df(x_0)}{dx} + \frac{1}{2}h^2\frac{d^2 f(x_0)}{dx^2}$$

There is also an approximation for a smooth function g of two variables x and y. Let

$$\Delta g(x_0, y_0) = g(x_0 + h, y_0 + k) - g(x_0, y_0)$$

then

$$\begin{aligned} \Delta g(x_0, y_0) = \left[\frac{\partial g(x_0,y_0)}{\partial x}\right]h + \left[\frac{\partial g(x_0,y_0)}{\partial y}\right]k \\ + \frac{1}{2}\left[\frac{\partial^2 g(x_0,y_0)}{\partial x^2}\right]h^2 + \frac{1}{2}\left[\frac{\partial^2 g(x_0,y_0)}{\partial y^2}\right]k^2 + \left[\frac{\partial^2 g(x_0,y_0)}{\partial x \partial y}\right]hk \end{aligned}$$

(Klein, 2002; Wiersema, 2008).

Substituting $(dt)^2 = 0$, $(dB)^2 = dt$, $dtdB = 0$;

$$\begin{aligned} df &= \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial B}dB + \frac{1}{2}\frac{\partial^2 f}{\partial B^2} \\ &= \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial B^2}\right)dt + \frac{\partial f}{\partial B}dB \end{aligned} \tag{8.12}$$

The function $f\left[t,\, B(t)\right]$ is commonly expressed as:

$$f\left[t, B(t)\right] = e^{[\mu t + \sigma B(t)]} \tag{8.13}$$

Rearranging Equation (8.11),

$$df = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial B}dB + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(dt)^2 + \frac{\partial^2 f}{\partial t \partial B}dtdB + \frac{1}{2}\frac{\partial^2 f}{\partial B^2}(dB)^2 \tag{8.11}$$

Again, substituting $(dt)^2 = 0$, $(dB)^2 = dt$, $dtdB = 0$:

$$\frac{\partial f}{\partial t} = \mu e^{[\mu t + \sigma B]}$$

$$\frac{\partial f}{\partial B} = \sigma e^{[\mu t + \sigma B]}$$

$$\frac{\partial^2 f}{(\partial B)^2} = \sigma^2 e^{[\mu t + \sigma B]}$$

Therefore;

$$\begin{aligned} df &= \left\{ \mu e^{[\mu t+\sigma B]} + \frac{1}{2}\sigma^2 e^{[\mu t+\sigma B]} \right\} dt + \left\{ e^{[\mu t+\sigma B]} \right\} dB \\ &= f \left\{ \left(\mu + \frac{1}{2}\sigma^2 \right) dt + \sigma B \right\} \end{aligned} \qquad (8.14)$$

Dividing by f (where $f \neq 0$), transforms Equation (8.14) into a rate of return:

$$\frac{df}{f} = \left(\mu + \frac{1}{2}\sigma^2 \right) dt + \sigma dB \qquad (8.15)$$

With the process of how changes in the value of an underlying asset will affect the value of an option set out by Itō’s formula, it is possible to examine the properties of a continuous time financial model (Bass, 2003; Wiersema, 2008). Any investment, be it for an individual or a company, begins with an initial level of wealth, X_{t0}. If an investor purchases Δ_0 share of assets initially and later buys more of the asset, say Δ_1 at time t, then the progressive change in wealth can be expressed as:

$$X_{t_0} + \Delta_0 \left(S_{t_1} - S_{t_0} \right) + \Delta_1 \left(S_{t_2} - S_{t_1} \right) + \ldots + \Delta_i \left(S_{t_{i+1}} - S_{t_i} \right) \qquad (8.16)$$

Where S_t denotes the price of the asset.

Equation (8.16) can be simplified to reveal

$$X_{t_0} + \int_{t_0}^{t} \Delta(s) dS_s \qquad (8.17)$$

Where $t > t_{i+1}$ and $\Delta(s) = \Delta_i$ for $t_i < s < t_{i+1}$.

Overall, the change in wealth is driven by a stochastic process (Bass, 2003). Thus, the value of a share price, or any other asset, assuming a non-zero interest rate, *r*, can be written as

$$P_t = S_t e^{-rt} \qquad (8.18)$$

If $P_0 = S_0$ and if Δ_i shares are held over the period t_i to t_{i+1}, then the present value of profits will be $\Delta_i \left(P_{t_{i+1}} - P_{t_i}\right)$. It follows then that the wealth expression becomes:

$$X_{t_0} + \int_{t_0}^{t} \Delta(s) dP_s \qquad (8.19)$$

And by Itō's formula:

$$\begin{aligned} dP_t &= e^{-rt} dS_t - re^{-rt} S_t dt \\ &= e^{-rt} \sigma S_t dW_t + e^{-rt} \mu S_t dt - re^{-rt} S_t dt \\ &= \sigma P_t dW_t + (\mu - r) P_t dt \end{aligned} \qquad (8.20)$$

Therefore the solution to the stochastic differential equation is

$$P_t = P_0 e^{\sigma W_t + \left(\mu - r - \sigma^2 / 2\right)} \qquad (8.21)$$

If P_t can be described by a geometric Brownian motion with a risk-neutral probability $\overline{\Pr}$, defined as

$$\frac{d\overline{\Pr}}{d\Pr} = M_t = e^{\left(aW_t - a^2 t / 2\right)} \qquad (8.22)$$

Where M_t is a martingale. Thus,

$$dP_t = \sigma P_t d\tilde{W}_t + \sigma P_t a dt + \left(\mu - r\right) P_t dt \qquad (8.23)$$

Where, $\tilde{W}_t = W_t - at$.

Furthermore, if $a = -(\mu - r)/\sigma$ then:

$$dP_t = \sigma P_t d\tilde{W}_t \qquad (8.24)$$

If $\tilde{W}_t$ can be described as a Brownian motion under $\overline{\Pr}$, then by definition P_t must be a martingale since it is a stochastic integral of a Brownian motion (Bass, 2003). Hence, Equation (8.24) can be rewritten as

$$d\tilde{W}_t = \sigma^{-1}P_t^{-1}dP_t \qquad (8.25)$$

Using the results obtained above, the wealth expression can be generalised. Hence, for any variable, V, there exists a constant and an adapted process K_s, which is articulated as

$$V = c + \int_0^t K_s dP_s \qquad (8.26)$$

This generalised wealth expression can be used to describe the value of equity of a short-term insurer. The major point of this result is that time value is implicit in the K_s and dP_s terms.

With the discussion of the consequences of capital-based insolvency put framework as set out in the preceding chapter in mind, Equation (8.26) shows that an insurer that is declared technically insolvent will not generate any additional business because P_s will decline to zero. However, a declining P_s on its own does not necessarily mean that time value of the embedded call option has been eliminated. It is entirely possible for insurance premiums to reach depressed levels without critically affecting the going concern of the business. Yet, as Harrington (1984, 1992) and Harrington and Danzon (1994) show, should depressed levels of premiums continue without future increases, in the face of onerous operating and fixed costs the prospects of an insurer will decline. In other words, time value will ebb away. Thus, Equation (8.26) demonstrates how regulatory intervention enforcing capital requirements during a financial crisis can lead to default; time value will be eliminated, leading to a zero value for P_s. All that will remain is c, which in the context of insolvency can be regarded as salvage value. Moreover, a high enough salvage value, c, in the face of technical insolvency will prompt the liquidation of an insurer rather than it being a mitigating factor, incentivising shareholders to keep it in business. Thus, the influence of limited liability can have a confounding effect. It can create the incentive to maintain a firm as a going concern in the face of significant losses. Yet it is also possible for an otherwise financially healthy firm to be liquidated even when its assets exceed its liabilities. Again the call option perspective can offer an explanation to this contradiction. Only positive time value combined with the potential for future profits will a firm be kept operating in the face of short-term losses. However, when time value is eliminated, a healthy firm will be liquidated because shareholders hold a valuable call option at expiration. Thus, inappropriate intervention can precipitate liquidation whether or not the implicit shareholder call option is in-the-money or out-the-money because they will ultimately be faced with zero time value, causing the option to expire. In short, Equation (8.26) shows that in the situation of positive time value, there will be a strong incentive for the shareholders to keep their firm operating. Moreover, as shareholders benefit from limited liability their payoff will never be negative. That is, their equity payoff will be $Max(V - L, 0)$.

Combining the preceding result with Vivian (2007a, b)'s mathematical description of an insurer, a complete articulation of an insurer from the call option perspective can be articulated as follows

$$PbT = (P_u + P_I)V_{0,\,t} \qquad (8.27)$$

Equation (8.27) shows that, for a short-term insurer, underwriting profit and investment income are two distinct earnings generators with the embedded equity call option implicit in both the underwriting side of the insurer and its investment portfolio. Thus, the embedded equity call option will have a real influence on the fundamental components of an insurer.

9. Discussion and Recommendations

9.1. Contributions and Theoretical Implications

Using Adam Smith (1776/1976)'s seminal description of the tenets of a successful insurance company as a point of departure and drawing on the work of Ferrari (1968), Fairly (1979), Hill (1979) and Vivian (2007a, b), it has been demonstrated that Basel style capital adequacy rules are inappropriate because capital is generally unimportant to the financial health of a short-term insurer. Instead, generating premiums that adequately compensate the insurer for the risk it is exposed to is the most important factor in maintaining a sustainable insurance firm. Thus, the regulatory focus on capital is misplaced (Vivian, 2007b). A case in point are the findings of Cummins et al (1995), Grace et al (1998), Klein et al (2002) and Pottier and Sommer (2002) that the risk-based capital measures are poor indicators of solvency. In fact, Cummins et al (1995) observe that an inaccurate risk-based capital formula will lead to destabilising distortions. Klein et al (2002) note that capital-based regulation incentivises insurers to increase their leverage. The objective of this behaviour is to justify charging a higher premium. Thus, this reaction indicates that short-term insurers are not particularly concerned with capital. Indeed, recall Cummins et al (1995, p. 6)'s general description of the workings of an insurer: "They [insurers] seek to take maximum advantage of the law of large numbers by insuring the largest possible number of independent exposure units, using reinsurance to pool risk with other insurers and across national boundaries." No mention of capital is made.

Moreover, it has been demonstrated that the notion of the insolvency put is spurious. The introduction of limited liability in company law implicitly and conceptually granted a call option to the shareholders of the firm on the value of the firm's assets. Thus, when a company defaults on its debts it is because shareholders have allowed an embedded call option to expire unexercised. Furthermore, by deconstructing the basic tenets of the insolvency put it has been shown that the behaviour of equity holders is inconsistent with the notion of an insolvency put. Indeed, the propensity for shareholders to delay bankruptcy for as long as possible is more consistent with the idea of an implicit call option, given the fact that the overwhelming incentive is to exercise a put option sooner rather than later. Delaying the exercising of a put option ultimately destroys value.

Yet the most compelling argument against the insolvency put stems from the analysis of Stoll (1969)'s put-call parity theorem along with its application to collateralised loans. That is, even if one supposes that limited liability granted shareholders an insolvency put, the combination of a put option and equity will always be tantamount to a call option. Despite the theoretical relationship between put and call options outlined by the put-call parity theorem, it has been illustrated that the incentives created by put and call options are starkly different. Importantly, this has significant implications for the capital-based insolvency put framework. First, a call option is only a risk at expiration. If positive time to expiration exists the call option will remain in place and default will not occur. Consequently, time value can be regarded as a firm's going concern. By contrast, the insolvency put framework emphasises the non-expiration periods and as a result overstates the risk of default. Reconceptualising the insolvency put option as a call option by using a relatively simple continuous time financial model has explicitly captured the effect of time value and, implicitly, the going concern of the firm. Second, the interaction of the insolvency put and capital adequacy rules can exacerbate market procyclicality during crises, leading to unnecessary insolvencies. Adrian and Shin (2009) and Borio (2009)'s illustration of how capital becomes scarce during liquidity crises contextualises how it is possible for capital rules to have destabilising effects. Forcing an insurer to recapitalise at a time when capital dries up will eliminate the time value of the implicit call option held by shareholders thereby forcing the firm into bankruptcy. Recognising that premium income is crucial to the long-term sustainability of a short-term insurer and not capital will help mitigate the disruptions caused by the procyclicality of the financial markets. This will ultimately benefit the industry and consumers alike by preventing unnecessary liquidations.

9.2. Recommendations for Future Research

This study is an initial step to providing a more complete mathematical description of a short-term insurer and the role of capital. Consequently, there are several possible avenues for future research. A potential line of enquiry is to examine the effect of the underwriting cycle as well as potential interactions with underwriting profits, investment income and the implicit shareholder call option on the behaviour of policyholders and shareholders. Also, an instructive addition to this research would be a simulation to calibrate the proposed model as well as identify how the interaction of the call option, underwriting profit and investment portfolio evolve over time. Indeed, while the effect of time to expiration is taken directly into account, it may be useful to explore the more nuanced effects of a multi-period time horizon on the framework of this study.

9.3. Concluding Remarks

It is clear that the pursuit of regulation in the public interest is at best extremely difficult, if not unattainable, given the complexities that surround the formation of regulatory policy. Indeed, the economic theory of regulation illustrates how it is possible for the public interest agenda to be subverted as the motivation to regulate an industry could be driven purely to ensure political survival (Stigler, 1971). Furthermore, when one considers the information asymmetries and agency issues that can arise within regulatory frameworks as outlined by Levine and Forrence (1990), Laffont and Tirole (1991) and Grace and Phillips (2008), the motivations behind regulation are complicated further due to the considerations of a regulator's career prospects within both the public and private sector. This crucible of competing interests between the industry, consumers, the regulator and legislators, in addition to a measure of political ideology, help explain why the early empirical tests of the public interest theory carried out by Ippolito (1979), Peltzman (1976) and Becker (1983) are mixed. There is no guarantee that regulatory policy will be carried out in the public interest. However, it is also clear that regulatory policy will not necessarily favour either the industry or consumers outright. Thus, it should be of no surprise that Kimball (1961) and Joskow (1973) find that the early years of insurance regulation in the US were characterised by inertia and drift and generally stifled the growth of the more efficient direct writers. What is more, there is clear evidence that overall, insurance regulation in the US has distorted price levels and coverage by preventing the creation of differentiated risk pools (Ippolito, 1979; Frech and Samprone, 1980; Harrington, 1984; Grabowski et al, 1989; Harrington and Danzon, 1994; Grace and Leverty, 2009).

The effects of rate suppression illustrate the central importance of premium income. The confluence of rising inflation and growing political pressures in the 1970s prompted regulators to limit rate increases to the point where premiums were actively constrained (Harrington, 1984). In the context of agency theory and the economic theory of regulation, consumers were able to pressure and incentivise politicians to champion their cause as significant political benefits could be extracted. However, while suppressed insurance premiums were undoubtedly welcomed by policyholders, Harrington (1984, 1992) and Harrington and Danzon (1994) argue convincingly that the rate suppression experienced over this period precipitated insurance failures in the 1980s. Ultimately, insurers were deprived of their central means of recovering the costs of providing cover. It was against this backdrop of rising insurer insolvencies that led to the introduction of capital-based regulation for short-term insurers.

Overall, despite the observations that regulation is not necessarily carried out in the public interest and that capital-based solvency rules are inappropriate for the short-term insurance industry, it is not the contention of this study that the short-term insurance industry be deregulated. Rather, in keeping with Hayek (1960)'s views, the assertion of this study is that it is the nature of regulation that is important. Thus, by examining the role of regulation and the essential components of a successful insurer and the conceptual issues brought about by the introduction of limited liability, this thesis presents an initial step towards a more considered regulatory approach.

References

Adams, M., & Tower, G. (1994). Theories of Regulation: Some Reflections on the Statutory Supervision of Insurance Companies in Anglo-American Countries. *Geneva Papers on Risk and Insurance: Issues and Practice*, *19* (2), 156-177.

Adrian, T., & Shin, H. (2009). *Liquidity and Leverage.* Staff Reports, No. 328, Federal Reserve Bank of New York.

Akerlof, G. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, *84* (3), 488-500.

Babbel, D. (1998). Components of Insurance Firm Value and the Present Value of Liabilities. *Wharton School, Financial Institutions Center*, *98* (18), 1-9.

Babbel, D., & Merrill, C. (2005). Real and Illusory Value Creation by Insurance Companies. *Journal of Risk and Insurance*, *72* (1), 1-21.

Babbel, D., Gold, J., & Merrill, C. (2002). Fair Value of Liabilities: The Financial Economics Perspective. *North American Actuarial Journal, 6* (1), 12-27.

Barth, M. (2000). A Comparison of Risk-Based Capital Standards under the Expected Policyholder Deficit and the Probability of Ruin Approaches. *Journal of Risk and Insurance*, *67* (3), 397-413.

Bass, R. (2003 Spring). *The Basics of Financial Mathematics.* Retrieved 2009 9-March from Lecture Notes: http://www.math.uconn.edu/~bass

Beaver, W., McNichols, M., & Nelson, K. (2003). Management of the Loss Reserve Accrual and the Distribution of Earnings in the Property-Casualty Insurance Industry. *Journal of Accounting and Economics*, *35*, 347-376.

Becker, G. (1983). A Theory of Competition Among Pressure Groups for Political Influence. *Quarterly Journal of Economics*, *98* (3), 371-400.

Becker, G. (1986). The Public Interest Hypothesis Revisited: A New Test of Peltzman's Theory of Regulation. *Public Choice, 49*, 223-234.

Besley, T., & Coate, S. (2003). Elected versus Appointed Regulators: Theory and Evidence. *Journal of the European Economic Association*, *1* (5), 1176-1206.

Biger, N., & Kahane, Y. (1978). Risk Considerations in Insurance Ratemaking. *Journal of Risk and Insurance*, *45* (1), 121-132.

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, *81* (3), 637-654.

Bodie, Z., Kane, A., & Marcus, A. (2006). *Investments* (6th ed.). McGraw Hill.

Bonbright, J. (1961). *Principles of Public Utility Rates.* New York: Columbia University Press.

Booth, P. (1997). The Political Economy of Regulation. *British Actuarial Journal*, *3* (3), 675-707.

Borio, C. (2009). Ten Propositions About Liquidity Crises. *Bank for International Settlements Papers* (293), 1-27.

Brealey, R., Myers, S., & Allen, F. (2006). *Corporate Finance* (8th ed.). McGraw Hill.

Buchanan, J., & Tullock, G. (1962/2004). *The Calculus of Consent: Logical Foundations of Constitutional Democracy* (Vol. 2). (C. Rowley, Ed.) Indianapolis: Liberty Fund, Inc.

Butsic, R. (1994). Solvency Measurement for Property-Liability Risk-Based Capital Applications. *Journal of Risk and Insurance*, *61* (4), 656-690.

Cilliers, H. S., & Benade, M. L. (1982). *Company Law* (4th ed.). Durban Pretoria, South Africa: Butterworths.

Colquitt, L., Sommer, D., & Godwin, N. (1999). Determinants of Cash Holdings by Property-Liability Insurers. *Journal of Risk and Insurance*, *66* (3), 401-415.

Cummins, J. D. (2000). Allocation of Capital in the Insurance Industry. *Risk Management and Insurance Review*, *3*, 7-28.

Cummins, J. D. (1991). Statistical and Financial Models of Insurance Pricing and the Insurance Firm. *Journal of Risk and Insurance*, *58* (2), 261-302.

Cummins, J. D., & Phillips, R. (2005). Estimating the Cost of Equity Capital for Property-Liability Insurers. *Journal of Risk and Insurance*, *72* (3), 441-478.

Cummins, J. D., & Weiss, M. (1991). The Structure, Conduct, and Regulation of the Property-Liability Insurance Industry. *Financial Condition and Regulation of Insurance Companies*, pp. 117-154.

Cummins, J. D., & Xie, X. (2008). Mergers and Acquisitions in the US Property-Liability Insurance Industry: Productivity and Efficiency Effects. *Journal of Banking and Finance*, *32*, 30-55.

Cummins, J. D., Harrington, S., & Klein, R. (1995). Insolvency Experience, Risk-Based Capital and Prompt Corrective Action in Property-Liability Insurance. *Journal of Banking and Finance*, *19* (3), 511-527.

Demsetz, H. (1968). Why Regulate Utilities? *Journal of Law and Economics*, *11* (1), 55-65.

Doherty, N., & Garven, J. (1995). Insurance Cycles: Interest Rates and the Capacity Constraint Model. *Journal of Business*, *68* (3), 383-404.

Doherty, N., & Garven, J. (1986). Price Regulation in Property-Liability Insurance: A Contingent-Claims Approach. *Journal of Finance*, *41* (5), 1031-1050.

Eling, M., Schmeiser, H., & Schmit, J. (2006). The Solvency II Process: Overview and Critical Analysis. *University of St Gallen: Working Papers on Risk Management and Insurance*, *20*, 1-24.

Embrechts, P., & Veraverbeke, N. (1982). Estimates for the Probability of Ruin with Special Emphasis on the Possibility of Large Claims. *Insurance: Mathematics and Economics*, *1*, 55-72.

Fairly, W. (1979). Investment Income and Profit Margins in Property-Liability Insurance: Theory and Empirical Results. *Bell Journal of Economics*, *10* (1), 192-210.

Fama, E., & French, K. (1996). Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance*, *51* (1), 55-84.

Fama, E., & French, K. (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance*, *47* (2), 427-465.

Fama, E., & Jensen, M. (1983). Separation of Ownership and Control. *Journal of Law and Economics, 26*, 301-325.

Ferrari, J. (1968). The Relationship of Underwriting, Investment, Leverage and Exposure to Total Return on Owners' Equity. *Proceedings of the Casualty Actuarial Society*, *55*, 295-302.

Francis, L., Heckman, P., & Mango, D. (2005). The Insolvency Put: Whose Asset? *Working Paper*, 1-37.

Frech, H., & Samprone, J. J. (1980). The Welfare Loss of Excess Nonprice Competition: The Case of Property-Liability Insurance Regulation. *Journal of Law and Economics*, *23* (2), 429-440.

Friendly, H. (1962). The Federal Administrative Agencies: The Need for Better Definition of Standards. *Harvard Law Review*, *75* (7), 1263-1318.

Garven, J. (1992). An Exposition of the Implications of Limited Liability and Asymmetric Taxes for Property-Liability Insurance. *Journal of Risk and Insurance*, *59* (1), 34-56.

Gaver, J., & Paterson, J. (2004). Do Insurers Manipulate Loss Reserves to Mask Solvency Problems? *Journal of Accounting and Economics, 37*, 393-416.

Grabowski, H., Viscusi, W. K., & Evans, W. (1989). Price and Availability Tradeoffs of Automobile Insurance Regulation. *Journal of Risk and Insurance*, *56* (2), 275-299.

Grace, M., & Klein, R. (2008). Insurance Regulation: The Need for Policy Reform. *The Future of Insurance Regulation Conference*, (pp. 1-53). Washington, D. C.

Grace, M., & Leverty, J. T. (2009). Political Cost Incentives for Managing the Property-Liability Insurer Loss Reserve. *Journal of Accounting Research*, *Forthcoming*.

Grace, M., & Phillips, R. (2008). Regulator Performance, Regulatory Environment and Outcomes: An Examination of Insurance Regulator Career Incentives on State Insurance Markets. *Journal of Banking and Finance*, *32*, 116-133.

Grace, M., Harrington, S., & Klein, R. (1998). Risk-Based Capital and Solvency Screening in Property-Liability Insurance: Hypotheses and Empirical Tests. *Journal of Risk and Insurance*, *65* (2), 213-243.

Graham, C. (2000). *Regulating Public Utilities: A Constitutional Approach.* Oxford: Hart Publishing.

Gründl, H., & Schmeiser, H. (2005). Capital Allocation for Insurance Companies - What Good is it? *Working Papers on Risk Management and Insurance*, *3*, 1-23.

Gujarati, D. (2003). *Basic Econometrics* (4th ed.). McGraw-Hill.

Harrington, S. (1992). Presidential Address: Rate Suppression. *Journal of Risk and Insurance*, *59* (2), 185-202.

Harrington, S. (1984). The Impact of Rate Regulation on Prices and Underwriting Results in the Property-Liability Insurance Industry: A Survey. *Journal of Risk and Insurance*, *51* (4), 577-623.

Harrington, S., & Danzon, P. (1994). Price Cutting in Liability Insurance Markets. *Journal of Business*, *67* (4), 511-538.

Harrington, S., & Niehaus, G. (1999). *Risk Management and Insurance.* Boston; London: Irwin McGraw-Hill.

Hayek, F. (1976). *Law, Legislation and Liberty* (Vol. 2). London and Henley: Routledge & Kegan Paul.

Hayek, F. (1960). *The Constitution of Liberty.* London: Routledge & Kegan Paul.

Hayek, F. (1988). *The Fatal Conceit: The Errors of Socialism.* (I. W. Bartley, Ed.) London: Routledge.

Hill, R. (1979). Profit Regulation in Property-Liability Insurance. *Bell Journal of Economics*, *10* (1), 172-191.

Hillier, D., Grinblatt, M., & Titman, S. (2008). *Financial Markets and Corporate Strategy* (European Edition ed.). London: McGraw-Hill Education.

Holmstrom, B., & Kaplan, S. (2001). Corporate Governance and Merger Activity in the United States: Making Sense of the 1980s and 1990s. *Journal of Economic Perspectives*, *15* (2), 121-144.

Ippolito, R. (1979). The Effects of Price Regulation in the Automobile Insurance Industry. *Journal of Law and Economics*, *22* (1), 55-89.

Jensen, M. (1986). Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *American Economic Review*, *76* (2), 323-329.

Jensen, M., & Meckling, W. (1976). Theory of the Firm: Managerial Behaviour, Agency Costs and Ownership Structure. *Journal of Financial Economics*, *3* (4), 305-360.

Joskow, P. (1973). Cartels, Competition and Regulation in the Property-Liability Insurance Industry. *Bell Journal of Economics and Management Science*, *4* (2), 375-427.

Kimball, S. (1961). The Purpose of Insurance Regulation: A Preliminary Inquiry in the Theory of Insurance Law. *Minnesota Law Review*, *45*, 471-524.

Klein, M. (2002). *Mathematical Methods for Economics* (2nd Edition ed.). Boston: Addison-Wesley.

Klein, R. (1995). Insurance Regulation in Transition. *Journal of Risk and Insurance, 62* (3), 363-404.

Klein, R., Phillips, R., & Shiu, W. (2002). The Capital Structure of Firms Subject to Price Regulation: Evidence from the Insurance Industry. *Journal of Financial Services Research*, *21* (1), 79-100.

Kolb, R., & Overdahl, J. (2007). *Futures, Options, and Swaps* (5th ed.). Blackwell Publishing.

Kunkel, P., & Mehrmann, V. (2006). *Differential-Algebraic Equations: Analysis and Numerical Solution.* Zurich: European Mathematical Society Publishing House.

Kwon, W. J. (2007). *Uniformity and Efficiency in Insurance Regulation: Consolidation and Outsourcing of Regulatory Activities at the State Level.* Networks Financial Institute, Indiana State University.

Laeven, L., & Levine, R. (2007). Is There a Diversification Discount in Financial Conglomerates? *Journal of Financial Economics*, *85* (2), 331-367.

Laffont, J. -J., & Martimort, D. (1997). Collusion Under Asymmetric Information. *Econometrica*, *65* (4), 875-911.

Laffont, J. -J., & Tirole, J. (1991). The Politics of Government Decision-Making: A Theory of Regulatory Capture. *Quarterly Journal of Economics*, *106* (4), 1089-1127.

Lakonishok, J., Shleifer, A., & Vishny, R. (1994). Contrarian Investment, Extrapolation, and Risk. *Journal of Finance*, *49* (5), 1541-1578.

Lange, O., & Taylor, F. (1938). *On the Economic Theory of Socialism.* (B. Lippincott, Ed.) Minneapolis, Minnesota, United States of America: The University of Minnesota Press.

Levine, M., & Forrence, J. (1990). Regulatory Capture, Public Interest, and the Public Agenda: Toward a Synthesis. *Journal of Law, Economics, & Organization*, *6*, 167-198.

MacAvoy, P. (1979). *The Regulated Industries and the Economy.* New York: W. W. Norton & Company, Inc.

Madsen, C., Haastrup, S., & Pedersen, H. (2005). A Further Examination of Insurance Pricing and Underwriting Cycles. *AFIR Conference*, (pp. 1-31). Zurich.

Martimort, D. (1999). The Life Cycle of Regulatory Agencies: Dynamic Capture and Transaction Costs. *Review of Economic Studies*, *66* (4), 929-947.

Mehr, R. (1986). *Fundamentals of Insurance* (2nd ed.). Homewood, Illinois: Irwin.

Meier, K. (1988). *The Political Economy of Regulation: The Case of Insurance.* New York: State University of New York Press.

Merton, R. (1977). An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees. *Journal of Banking and Finance*, *1*, 3-11.

Merton, R. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance*, *29* (2), 449-470.

Merton, R. (1973a). The Relationship Between Put and Call Option Prices: Comment. *Journal of Finance*, *28* (1), 183-184.

Merton, R. (1973b). Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science*, *4* (1), 141-183.

Merton, R., & Perold, A. (1993). Theory of Risk Capital in Financial Firms. *Journal of Applied Corporate Finance*, *6* (3), 16-32.

Mildenhall, S. (2004). A Note on the Myers and Read Capital Allocation Formula. *North American Actuarial Journal*, *8* (2), 32-44.

Mill, J. S. (1859/1966). *On Liberty.* (C. Shields, Ed.) New York: Forum Books, Pyramid Publications.

Modigliani, F., & Miller, M. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *American Economic Review*, *48* (3), 261-297.

Mowbray, A., & Blanchard, R. (1955). *Insurance* (4th ed.). New York: McGraw-Hill Book Company, Inc.

Myers, S. (1977). Determinants of Corporate Borrowing. *Journal of Financial Economics*, *5*, 147-175.

Myers, S., & Majluf, N. (1984). Corporate Financing and Investment Decisions when Firms Have Information that Investors Do Not Have. *NBER Working Paper Series*, *W1396*.

Myers, S., & Read, J. (2001). Capital Allocation for Insurance Companies. *Journal of Risk and Insurance*, *68* (4), 545-580.

Nelson, K. (2000). Rate Regulation, Competition, and Loss Reserve Discounting by Property-Casualty Insurers. *Accounting Review*, *75* (1), 115-138.

Peltzman, S. (1976). Toward a More General Theory of Regulation. *Journal of Law and Economics*, *19* (2), 211-240.

Peltzman, S., Levine, M., & Noll, R. (1989). The Economic Theory of Regulation after a Decade of Deregulation. *Brookings Papers on Economic Activity. Microeconomics*, *1989*, 1-59.

Posner, R. (1974). Theories of Economic Regulation. *Bell Journal of Economics and Management Science*, *5* (2), 335-358.

Pottier, S., & Sommer, D. (2002). The Effectiveness of Public and Private Sector Summary Risk Measures in Predicting Insurer Insolvencies. *Journal of Financial Services Research*, *21* (1), 101-116.

Redford, E. (1954). The Protection of the Public Interest with Special Reference to Administrative Regulation. *American Political Science Review, Vol. 48, No. 4 (Dec., 1954)*, *48* (4), 1103-1113.

Reilly, F., & Norton, E. (2006). *Investments* (7th ed.). Thompson South-Western.

Roll, R. (1977). A Critique of the Asset Pricing Theory's Tests. *Journal of Financial Economics*, *4*, 129-176.

Schubert, G. J. (1957). "The Public Interest" In Administrative Decision-Making: Theorem, Theosophy, or Theory? *American Political Science Review*, *51* (2), 346-368.

Sherris, M. (2007). Risk Based Capital and Capital Allocation in Insurance. *Institute of Actuaries of Australia*, (pp. 1-16). Christchurch, New Zealand.

Sherris, M. (2006). Solvency, Capital Allocation, and Fair Rate of Return in Insurance. *Journal of Risk and Insurance*, *73* (1), 71-96.

Sherris, M., & van der Hoek, J. (2006). Capital Allocation in Insurance: Economic Capital and the Allocation of the Default Option Value. *North American Actuarial Journal*, *10* (2), 39-61.

Smith, A. (1776/1976). *The Wealth of Nations.* (E. Cannan, & G. Stigler, Eds.) Chicago: University of Chicago Press.

Smith, C. (1986a). Investment Banking and the Capital Acquisition Process. *Journal of Financial Economics*, *15* (1), 3-29.

Smith, C. (1986b). On the Convergence of Insurance and Finance Research. *Journal of Risk and Insurance*, *53* (4), 693-717.

Spiegel, Y., & Spulber, D. (1994). The Capital Structure of a Regulated Firm. *RAND Journal of Economics*, *25* (3), 424-440.

Staking, K., & Babbel, D. (1995). The Relation between Capital Structure, Interest Rate Sensitivity, and Market Value in the Property-Liability Insurance Industry. *Journal of Risk and Insurance*, *62* (4), 690-718.

Stigler, G. (1971). The Theory of Economic Regulation. *Bell Journal of Economics and Management Science*, *2* (1), 3-21.

Stoll, H. (1969). The Relationship between Put and Call Option Prices. *Journal of Finance*, *24* (5), 801-824.

Stoll, H. (1973). The Relationship between Put and Call Option Prices: Reply. *Journal of Finance*, *28* (1), 185-187.

Taylor, T. (1980). *The Fundamentals of Austrian Economics* (Vol. 2). Brighton, England: The Adam Smith Institute in association with the Carl Menger Society.

Vaughan, T. (2009). The Implications of Solvency II for U.S. Insurance Regulation. *Social Science Research Network Working Paper no. 3505392009*, 1-26.

Venter, G. (2004). Capital Allocation Survey with Commentary. *North American Actuarial Journal*, *8* (2), 96-107.

Venter, G. (2003 Fall). Discussion of "Capital Allocation for Insurance Companies" by Stewart C. Myers and James R. Read Jr. *Casualty Actuarial Society Forum*, pp. 479-478.

Vivian, R. (2007a). *A Mathematical Description of a Short-Term Insurance Company.* University of the Witwatersrand, School of Economic and Business Science, Johannesburg.

Vivian, R. (2007b March). Financial Condition Reporting - Conceptually Flawed and Probably Disastrous. *Cover*, pp. 27-30.

Wiersema, U. (2008). *Brownian Motion Calculus.* Chichester: John Wiley & Sons, Ltd.

Yow, S., & Sherris, M. (2007 March). Enterprise Risk Management, Insurer Pricing, and Capital Allocation. *University of New South Wales Actuarial Studies Working Paper*.