# Frictional costs in insurance

**Dissertation**
**submitted to the**
**Faculty of Business, Economics and Informatics**
**of the University of Zurich**

to obtain the degree of
Doktor der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Andrea Bergesio
from Italy

approved in February 2023 at the request of

Prof. Dr. Cosimo-Andrea Munari
Prof. Dr. Pablo Koch-Medina
Prof. Dr. Loriano Mancini
Prof. Dr. Hansjoerg Albrecher

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 15.02.2023

The Chairman of the Doctoral Board: Prof. Dr. Steven Ongena

## Acknowledgments

I am truly grateful to my supervisor Professor Cosimo-Andrea Munari. Professor Cosimo-Andrea Munari has always been available to provide his feedback and his guidance has greatly improved this thesis. My research benefited significantly from his expertise on insurance economics, as well as from our discussions, which helped me to deepen my understanding of many topics. I also enjoyed working together on our common project, which resulted in the second chapter of this thesis; this gave life also to many enjoyable moments during our multiple meetings.

I am also especially thankful to Professor Pablo Koch-Medina and Professor Loriano Mancini. Professor Pablo Koch-Medina has continuously offered his support and taught me how to keep on challenging standard modeling assumptions and sharpen my research focus. I am grateful for our numerous discussions, which gave me the opportunity to learn from his experience and knowledge of insurance markets, valuation and mathematics. Professor Loriano Mancini supported me since the beginning of our project. I am deeply grateful not only for his patience, but also for his insights that greatly improved my understanding of empirical methods and helped me to gain intuition on many aspects of my research. On a personal level, I benefited from their kindness, patience and helpfulness, which made me enjoy every meeting together.

I would also like to thank my co-authors, Dr. Mario Šikić, Dr. Paul Huber, Professor Cosimo-Andrea Munari, Professor Pablo Koch-Medina and Professor Loriano Mancini. I am grateful for their support, patience and constructive feedback. Our repeated interactions not only stimulated my critical thinking but also improved my writing skills and, thus, the readability of this thesis.

I would also like to thank Dr. Mario Šikić and Dr. Paul Huber. Sharing part of my research with them was a great opportunity for me to learn much about topics I was unfamiliar with. Not only are they brilliant professionals, but also pleasant colleagues with whom I enjoyed sharing friendly meetings. I would like to thank Dr. Mario Šikić for his patience in assisting me and for improving considerably my programming skills. I am much thankful to Dr. Paul Huber for his inspiring insights on the dynamics of the insurance industry and for the fresh perspective he brought in our common project.

I would also like to thank Santiago Moreno-Bromberg; I remember fondly of the delightful moments shared during our meetings.

I would also like to thank my cohort-mates, Marlon Azinović, Olga Briukhova, Marco Ceccarelli, Gazi Kabas, Jonathan Krakow, Mrinal Mishra, Michele Pelli, Stefano Ramelli, Alan Roncoroni, Li Yang and Isabelle Zheng. A special thank to Marlon Azinović for his friendship and never-ending kindness.

A special thank is dedicated to my partner, Francesca, who has constantly supported me and provided me with her caring love. I am also grateful for the affectionate support of her family.

I would also like to thank my friends Lubna Allam, Maria Arduca, Daniela Boiano, Iasson Chryssikos, Minji Sung and Riccardo Turin, who are a permanent part of my life.

Finally, I would like to thank my family for their unconditional support and encouragement. They have been a constant point of reference during my studies.

Andrea Bergesio, Zurich, September 2022

# Contents

# Part I

# Introduction

# Introduction

In its original form, the celebrated theorem by Modigliani and Miller (1958) stated that, in a complete market and in the absence of market imperfections, the capital structure of a firm has no impact on firm value. It is immediate to see that this implies that, in such a setting, also risk management fails to have an impact on firm value. As a result, capital and risk management are irrelevant for value-maximizing firms. However, markets are neither complete nor perfect. Asymmetries of information between managers and firm owners, corporate and personal taxes, transaction inefficiencies, financial distress and regulatory constraints all give rise to deadweight, or frictional, costs that have a negative impact on firm value. Many of these costs are directly or indirectly related to bearing risk so that, ultimately, managing risk becomes an effective means to oversee these costs and raise firm value. Hence, risk management was recognized to be relevant for value-maximizing firms not because risk reduction is per se desirable, but rather because it helps to reduce the deadweight costs associated with bearing risk, thereby increasing firm value.

The bulk of the literature on corporate risk management focuses on hedging financial market risk exposures and only to a lesser extent on reducing insurable risks by the purchase of insurance. Two of the three projects in this thesis explore how frictional costs influence the decision of non-financial, value-maximizing firms to buy insurance. At the same time, frictions affect not only the value of firms that purchase insurance coverage, but also insurance firms themselves. For an insurance firm, these frictions translate into a cost of bearing risk that ultimately needs to be viewed as a component of the value of its insurance liabilities.

The three research projects rely on the premise that firms maximize value, a concept that requires some justification. We develop this notion within an expected utility paradigm and make assumptions on the ownership structure of the firm. Specifically, all research projects focus on publicly-held firms with a diffuse ownership base.

In this setting, it is possible to argue that these firms will make decisions aimed at maximizing firm value, defined as the present value of future expected cash flows to their owners. Therefore, it is necessary to provide a working definition of "present value" and "expected cash flows". In our framework, expectations are determined with respect to a unique "economic" valuation measure (that is consistent with prices in financial markets and with a diffuse ownership base), whereas the present value is obtained by discounting cash flows at the risk-free rate. For firms without exposure to financial market risk, expectations under the economic valuation measure coincide with expectations under the "physical" probability measure.

Detailed introductions to the individual projects are included at the beginning of each chapter; here we provide only a bird's-eye view. The first chapter reviews the theoretical arguments underpinning the assumption that the objective function of a firm with a diffuse ownership is expected profit, or equivalently, firm value. Hence, such a firm will behave *as if* it were risk neutral. This chapter lays the theoretical foundations for the following chapters. The next two chapters study the demand for insurance coverage by non-financial, value-maximizing firms subject to frictional costs. The second chapter studies, in a static setting, when firms with limited liability and no access to capital markets purchase liability insurance. Our results extend the literature on optimal insurance contracts beyond a binary loss setting. The third chapter develops a dynamic firm model in which firms can buy insurance against property losses, while retaining exposure to profit shocks. This model allows us to analyze the impact of a rich set of frictional costs on the demand for property insurance by firms highlighting, among other things, how insurance can be seen as an alternative source of financing. The fourth and final chapter is concerned with the cost of bearing risk for insurers with a broad ownership. It shows how the "risk margin" in the valuation of multi-period liabilities under the Solvency II regime can be interpreted as a provision for frictional costs. Using data for U.S. insurers, it also provides an estimate of the size of this margin as implied by market expectations.

# Part II

# Preferences of widely-held corporations

# Chapter 1

# On the risk neutrality of widely-held corporations

Andrea Bergesio[1]

**Abstract** We review the basic arguments underlying the renowned Arrow-Lind Theorem, according to which risky projects financed by a sufficiently large number of investors and whose return is uncorrelated with the investors' wealth should be evaluated using the risk-free rate of return. When applied to corporations, the theorem implies that any firm with a diffuse shareholder base will behave *as if* it were risk-neutral. This result follows from observing that for well diversified investors the riskiness of a single, marginal investment in a firm decreases at a faster rate than the size of the shareholder base, driving the total return required by investors to the risk-free interest rate.

[1]A special thank to Pablo Koch-Medina and Cosimo-Andrea Munari for helpful discussions on this topic, from which this chapter has greatly benefited. Bergesio at University of Zurich and Swiss Finance Institute.

## 1.1 Introduction

In an expected utility framework, agents are assumed to take decisions that deliver the maximal expected utility under their budget constraint. While the objective function for individuals clearly depends on their own preferences, and thus utility function, when a firm is owned by more than one agent it is not always clear a priori which preferences should drive its decisions. The answer depends on many factors, including whether markets are complete or not, the number of investors, how their preferences differ and how each investor's wealth correlates with the return on firms' projects. In two instances the objective function of a firm is unambiguous. If markets are complete and arbitrage-free, then any cash flow (including firms') has a unique value and by the Fisher Separation Theorem firms should simply maximize the net present value of their expected profits regardless of investors' preferences (see e.g., MacMinn (2005)). Alternatively, if all investors share the same utility, the firm's preferences will coincide with those of any of its shareholders. Clearly this encompasses firms owned by a single agent as a special case.

When one departs from these two situations, however, the preferences of a firm are not clear. This is because in either case the objective function of the firm depends on the investors' preferences and thus on the firm ownership structure. Nonetheless, it is possible to show that for firms with a diffuse shareholder base, or equivalently with an infinite number of investors, one can ignore the preferences of its shareholders under some conditions.[2] This result, which was originally derived in Arrow and Lind (1970) for the valuation of public projects, is the focus of this chapter.

The next sections are concerned with showing how the preferences of firms with a diffuse shareholder base converge to those of agents that are indifferent towards risk, also known as risk-neutral agents. In particular, we review first the fundamental assumptions underlying the Arrow-Lind Theorem, discuss their importance and how the literature has challenged them. In addition, we briefly argue how firms ownership is related to their valuation.

To show that widely-held firms behave *as if* they were risk neutral, we focus on firms owned by investors with the same, constant wealth. This setting allows us to simplify the derivation of the Arrow-Lind Theorem while gaining insights on the

[2] To refer to firms with infinitely many investors, we use interchangeably the expressions "diffuse" and "broad" ownership, as well as "widely-held firms".

factors driving this result. We rely on the Arrow-Pratt approximation for strictly-risk averse agents and use it to provide intuition behind the risk neutrality of widely-held firms: this follows from the the fact that the minimum excess return required by investors decreases at a faster rate than the rate at which the number of investors increases. We also explain how the theorem applies to insurance companies and show the implications for the minimal profit margin required by widely-held insurers.

**Structure of the paper**

The organization of the remaining sections is as follows. Section 1.2 provides an overview of the Arrow-Lind Theorem and a discussion of the underlying assumptions, including a review of the related literature. Section 1.3 discusses the relation between firm ownership and valuation. Section 1.4 derives the Arrow-Lind Theorem for firms with heterogeneous investors, showing that as a firm ownership approaches a diffuse shareholder base, the firm will tend to behave *as if* it were risk neutral. We show also how to adjust this result to the case of an insurance firm. Section 1.5 concludes.

## 1.2 Background on the Arrow-Lind Theorem

According to the seminal paper by Arrow and Lind (1970), any public project financed by a sufficiently large number of taxpayers should be evaluated on the basis of its expected net benefit alone, provided the project's payoff is uncorrelated with the aggregate economic activity (Foldes and Rees (1977)). The theorem relies on a few assumptions that have been at the center of a longstanding contentious debate in the literature. Specifically, the Arrow-Lind Theorem holds if the following conditions are met: (i) the government finances through taxpayers and distributes the net returns from the public project based on changes in the tax levels, (ii) the net returns are statistically independent of any financier's wealth absent the project and (iii) the share of net returns for every person financing the project tends to zero as the overall number of financiers tends to infinity. Despite the theorem being proved for agents with the same utility, identically distributed income and subject to the same tax rate, the result is argued to hold also in a more general setting; for this to hold, however, it is necessary that the share of the public investment becomes sufficiently small when the number of investors increases (Arrow and Lind (1970), p. 371).

It is important to note that the Arrow-Lind Theorem has implications beyond public projects: the same conclusions are valid also for corporations with a broad shareholder base and projects' payoffs uncorrelated with investors' own portfolios, in which case the decisions of these firms will be observationally equivalent to those of risk-neutral agents (Arrow and Lind (1970), p. 375-376). This implies that similar to these agents, who are indifferent towards risk and do not require any compensation for taking on risk, firms with a diffuse shareholder base will evaluate risky projects based solely on the expected value of the projects' payoff.

The first assumption represents the main difference between public and private projects. While the Arrow-Lind Theorem can be applied also to projects financed by widely-held firms, private projects are financed by investors who may or may not be willing to take on the firm's risk. On the other hand, if public projects are financed by taxpayers, who then act as investors, there is no possibility for them to opt out and forego the public investment. This consideration has led some authors to conclude that public projects may be attractive precisely because of the lower financing costs that stem from access to cheaper taxpayer funds (for a discussion, see McKean and Moore (1972) and Grant and Quiggin (2003)).

Assumption (ii), according to which the returns from a public investment should be independent of investors' wealth, has been a major point of contention. In other terms, the Arrow-Lind Theorem is valid only if public projects have zero beta according to the Consumption Capital Asset Pricing Model (CCAPM), or equivalently if public investments yield returns that are independent of other components of the national income (Baumstark and Gollier (2014), Gollier (2021)). Clearly, this assumption might not hold by the very own nature of public projects, which can contribute to one nation's economic growth and thus generate correlation between the projects and economic output. For instance, public infrastructures and financing the reconstruction of areas hit by natural disasters are among the projects that may very well break assumption (ii). The same conclusion holds true for widely-held corporations: only if investors' wealth is uncorrelated with the firm's returns the Arrow-Lind Theorem applies. In effect, whether this requirement fails depends on the degree to which investors' portfolios and firms' returns contain financial market risk and thus systematic, or non diversifiable, risk.

The third assumption requires that an investor's share of the project's net return, either she be a taxpayer or a private investor, shrinks virtually to zero as the total number of investors grows to infinity. As Arrow and Lind (1970) pointed out, the actual number of investors for which the theorem may be valid is an open question and in reality a broad ownership base can be hard to achieve, even for public projects financed by millions of taxpayers. While many authors have argued against the actual applicability of the theorem based on this potential pitfall, among which McKean and Moore (1972), Brealey et al. (1997) and Cherbonnier and Gollier (2022), whether a failure of assumption (iii) poses a real threat to the validity of the theorem depends on the *size* of the public project. According to Gardner (1979), if the project being financed is truly marginal, i.e. it represents a negligible portion of an investor's wealth, then the Arrow-Lind Theorem should still be valid whether or not the number of investors approaches a diffuse ownership.

### 1.2.1 Overview of the related literature

Since the seminal paper by Arrow and Lind (1970), the literature has investigated the reasons for which the Arrow-Lind Theorem might fail, or whether the theorem can be extended further.

A review of the contentious points concerning the assumptions of the Arrow-Lind Theorem can be found in Klein (1997), which presents many perspectives brought about in the literature as to its potential pitfalls. McKean and Moore (1972) review the basic assumptions with a special focus on the role of taxpayers and the relevance of the number of taxpayers for the validity of the theorem. Brealey et al. (1997) and Grant and Quiggin (2003) discuss the implications of market imperfections for the Arrow-Lind Theorem and argue that asymmetric information determines ultimately the efficiency of public projects relative to private investments.[3] Baumstark and Gollier (2014) relate the assumption of uncorrelatedness between public projects and the national income to the CCAPM and discuss the efficiency of governments in spreading and pooling risk compared to private firms.

The literature has also attempted to extend the scope of the Arrow-Lind Theorem. In particular, prior research has studied the applicability of the Arrow-Lind Theorem to public investments with uncertain environmental side effects (Fisher (1973)), public financing in economies with a continuum of agents (Gardner (1979)), the cost of

[3] On a related note, Hirshleifer (1965) argues that the risk premium for public and private projects should coincide in the absence of market imperfections.

financing for public-private partnerships (Grout (2003), Spackman (2002), Moszoro (2014), Geddes and Goldman (2022)), financial protection strategies against natural disasters (Ghesquiere and Mahul (2010), Mechler and Hochrainer-Stigler (2014)) and non-marginal projects (Gollier (2011)). Empirically, Bednarek and Moszoro (2014) have tested the predictions of the Arrow-Lind Theorem concerning ownership concentration and risk premium, as well as its implications for company valuation. Their results provide evidence in support of the risk neutrality of firms with a broad shareholder base.

## 1.3 Firm ownership and valuation

When studying the objective of a firm with a given ownership base, the Arrow-Lind Theorem indicates under which circumstances the behavior of the firm can be taken to be observationally equivalent to that of a risk-neutral agent. Moreover, the valuation of a firm is tightly related to the underlying ownership structure and preferences of its owners.

From an economic perspective, valuations models have their roots in utility theory. Agents assess whether to undertake a project generating a cash flow by evaluating the net impact of that cash flow on their utility, whereby "net" refers to the need to account for the cost of undertaking the project. Hence, firms assess a project by evaluating how the project changes cash flows and whether that change has a positive net impact on the owners' utility. In a complete and arbitrage-free market, any given cash flow can be replicated by traded assets and the prices of these assets imply a unique value for that cash flow. This means that, in these markets, there is a natural way to value any cash flow so that assessing the value of a firm does not require knowing the preferences of its owners.[4]

When markets are incomplete, however, only those assets that can be replicated by traded assets carry a unique value. For cash flows that cannot be replicated, e.g. for cash flows containing insurance risks, only a range of possible values that are

[4]This is a direct consequence of the Fisher Separation Theorem, according to which under complete markets any firm should aim to maximize the net present value of expected profits independently of its shareholders' preferences (Fisher (1930), MacMinn (2005)). This follows from observing that when markets are complete an investor can achieve any preferred consumption path, constrained only by her budget. This allows to *separate* her consumption choices from firms' investment decisions.

consistent with observed market prices of the traded securities can be inferred. It follows that valuing a firm under incomplete markets requires knowledge about the firms ownership structure and the owners' preferences.

Firm ownership, however, can span a broad spectrum ranging from a single entrepreneur to a broad shareholder base in which each owner holds an infinitesimal fraction of the firm capital (see Figure 1.1).

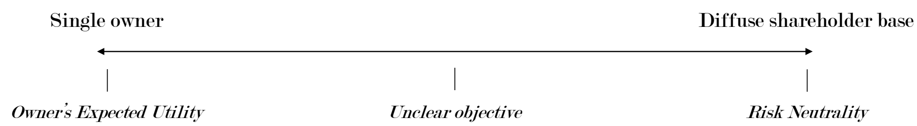



**Figure 1.1:** The figure shows the ownership of a firm, which can range from a single owner to a diffuse shareholder base, and the corresponding objective of the firm.

For a firm owned by a single risk-averse agent, the actions of the firm will fully reflect the individual preferences of the owner and thus her own risk aversion: the firm's preferences are identical with those of its owner (Figure 1.1). However, with more than one owner, it is unclear whose preferences should guide the actions of the firm. In such a situation, there is no obvious way to assign a single objective to the firm and any attempt in this direction leads to studying the allocation of voting rights (see e.g., Ekern and Wilson (1974), DeMarzo (1993), Kelsey and Milne (1996) and Harstad (2005)).

An instance in which specifying individual preferences proves unnecessary is the case of a firm with a diffuse ownership base, i.e., with an "infinitely large" number of owners. For such a firm, it is possible to show by means of a limiting argument that, despite the risk aversion of its individual owners, the firm will behave as if it were risk neutral. This result, which was originally derived in the context of the valuation of public projects, is an implication of the Arrow-Lind Theorem (Arrow and Lind (1970)). In the case of an insurance firm, there is another instance in which the valuation will not depend on the individual owners' preferences: it is not difficult to argue using the Law of Large Numbers that if an insurance firm has an "infinitely well-diversified" portfolio of policies, even a risk-averse investor will not require a risk premium for bearing insurance risk.

It would thus appear as though, in the two instances we have described, an insurance firm should view any compensation above the expected loss as pure profit and not as a risk premium. This seems to be at odds with actuarial valuation practices under which a risk loading is included above the expected loss.[5] While this approach is in line with what one would put forward when valuing a closely held insurance firm, it is not immediately compatible with a firm having a diffuse ownership base.

Our perspective, however, does not entail that insurance firms should ignore their exposure towards idiosyncratic risks. In effect, the existence of market imperfections imposes frictional costs for bearing risk. These "deadweight costs" come in the form of double-taxation, agency costs, regulatory costs and costs of financial distress.[6] For instance, taking more investment risk increases the regulatory capital held by insurers, which is costly to hold. These deadweight costs represent the channel that makes risk relevant for widely-held firms. Hence, our perspective does not imply that bearing insurance risks does not have a cost, but that the nature of these costs is different: they do not denote a risk premium but rather a compensation for incurred frictions.

The challenges posed by the two approaches are therefore different. Since risk premiums are preference dependent, if we interpret the cost of taking risk as a risk premium, it becomes necessary to specify the set of preferences with respect to which the risk premium is being determined. By contrast, given a diffuse ownership base, the "sole" requirement of our approach is to obtain accurate estimates of the firm's cash flows, including deadweight costs.

In the following we focus on firms with a diffuse ownership base and discuss not only how to establish their risk neutrality, but also the intuition behind this result and its application to the business of insurance companies.

## 1.4 The Arrow-Lind Theorem

Consider a one-period model with two dates, $t \in \{0, 1\}$. Uncertainty is resolved at time $t = 1$ and is described by a standard probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We derive a version of the Arrow-Lind Theorem (Arrow and Lind (1970)) showing that firms with a broad ownership base consisting of strictly risk-averse shareholders will act as if

[5] We refer the reader to the monographs Kaas et al. (2008) and Asmussen and Steffensen (2020) for an introduction to actuarial valuation approaches.

[6] See Hancock et al. (2001), Zanjani (2002) and Babbel and Merrill (2005) for a discussion. Recent empirical literature on the frictional costs for insurers includes Ellul et al. (2015), Koijen and Yogo (2015) and Sen (2019).

they were risk neutral. This means that the hurdle rate for undertaking projects will be the risk-free rate, which, for convenience, we assume to be zero. Before deriving this result, we introduce formally the characteristics of the firm and of the investors.

We operate within the expected utility paradigm and consider a pool of infinitely many strictly risk-averse agents whose preferences are described by the sequence $(u_i)$ of strictly concave utility functions $u_i \colon \mathbb{R} \to \mathbb{R}$. We assume that there is one agent, say agent 0, who is more risk averse than all others. Denoting by $\pi_i(X)$ the risk premium that agent $i$ associates with a risky payoff $X$, i.e. the monetary amount that agent $i$ is willing to pay to get rid of the risk in $X$, this assumption translates into

$$0 \leq \pi_i(X) \leq \pi_0(X), \tag{1.1}$$

for every $i$ and $X : \Omega \to \mathbb{R}$. We fix $n$ and consider a situation in which the first $n$ agents have part of their wealth invested in a firm with initial capital $K > 0$. The firm is owned by these agents in equal shares. The remaining part of agent's $i$ wealth consists of a deterministic endowment $W > 0$, which is common across all agents.

The firm can invest $K$ in a risky project whose total return is $K + P + Z$, with $P$ representing the *profit margin*, i.e. the required return in excess of the risk-free rate (which is zero), and $Z$ an error term representing the project's risk, i.e. $Z : \Omega \to \mathbb{R}$ is a random variable with $\mathbb{E}[Z] = 0$. We assume the firm cannot go bankrupt independently of the size of the profit margin, namely

$$K + Z \geq 0, \quad a.s.$$

In particular, $Z$ is bounded below, i.e. $Z > -K$ a.s.. At date $t = 1$, terminal capital is returned to investors in equal shares. Terminal wealth for agent $i$ is then given by

$$W + \frac{K + P + Z}{n}. \tag{1.2}$$

The problem the firm faces is to determine the minimal profit margin $P(n)$ that is acceptable for every individual shareholder to take on the risk from the firm's project. Firstly, the minimal $P(n)$ needs to be such that

$$\mathbb{E}\left[u_i\left(W + \frac{K + P(n) + Z}{n}\right)\right] \geq u_i\left(W + \frac{K}{n}\right) \quad \text{for each } i, \tag{1.3}$$

i.e. the minimal profit margin must be such that, after the firm undertakes the project, each of the $n$ shareholders is at least as well-off as if the firm had not undertaken it. Since agent 0 is the most risk averse, it is clear that the minimal profit margin $P(n)$ needs to satisfy

$$\mathbb{E}\left[u_0\left(W + \frac{K + P(n) + Z}{n}\right)\right] = u_0\left(W + \frac{K}{n}\right). \tag{1.4}$$

It is then easy to see from (1.4) that

$$\lim_{n\to\infty} \frac{P(n)}{n} = 0. \tag{1.5}$$

Indeed, this simply reflects the fact that as the shareholder's stake in the firm $\frac{K}{n}$ decreases to zero, the additional compensation must also tend to zero in absolute terms. It is not clear a priori, however, that also the total excess return that the firm requires, $P(n)$, vanishes as the number of shareholders tends to infinity. This is, in fact, the conclusion reached by the Arrow-Lind Theorem. The intuition behind this result lies in the different rate at which the minimum excess return decreases compared to the rate at which the size of the investors base increases (Gravelle and Rees (2004), Gollier (2011)). To gain such an insight, we start by expressing the left-hand side of (1.4) as

$$\begin{aligned} &\mathbb{E}\left[u_0\left(W + \frac{K + Z}{n} + \frac{P(n)}{n}\right)\right] \\ &= u_0\left(W + \frac{K}{n} + \frac{P(n)}{n} - \pi_0\left(W + \frac{K + Z}{n} + \frac{P(n)}{n}\right)\right). \end{aligned} \tag{1.6}$$

Since $u_0$ is invertible, Equations (1.4) and (1.6) together imply that

$$W + \frac{K}{n} = W + \frac{K}{n} + \frac{P(n)}{n} - \pi_0\left(W + \frac{K + Z}{n} + \frac{P(n)}{n}\right)$$

and, in turn, that

$$\frac{P(n)}{n} = \pi_0\left(W + \frac{K + Z}{n} + \frac{P(n)}{n}\right).$$

Further intuition on $\frac{P(n)}{n}$ can be gleaned from the Arrow-Pratt approximation for the agent's risk premium (Mas-Colell et al. (1995))

$$\frac{P(n)}{n} = \pi_0\left(W + \frac{K+Z}{n} + \frac{P(n)}{n}\right), \tag{1.7}$$

$$\approx -\frac{1}{2}\cdot\frac{u_0''\left(W + \frac{K+Z}{n} + \frac{P(n)}{n}\right)}{u_0'\left(W + \frac{K+Z}{n} + \frac{P(n)}{n}\right)}\cdot\frac{\mathrm{Var}(Z)}{n^2}. \tag{1.8}$$

The Arrow-Pratt approximation thus shows that, for sufficiently "small risks", the minimal profit margin is proportional to the riskiness, as measured by the variance, of the investor's share of the project payoff. We can now study the behavior of the *total* profit margin $P(n)$ required by investors as the firm ownership approaches a diffuse shareholder base

$$\lim_{n\to\infty} n\cdot\frac{P(n)}{n} = \lim_{n\to\infty} -\frac{n}{2}\cdot\frac{u_0''\left(W + \frac{K+Z}{n} + \frac{P(n)}{n}\right)}{u_0'\left(W + \frac{K+Z}{n} + \frac{P(n)}{n}\right)}\cdot\frac{\mathrm{Var}(Z)}{n^2} = 0, \tag{1.9}$$

where the second equality follows from (1.5). We record this result as the Arrow-Lind Theorem for firms with a diffuse shareholder base.

**Theorem 1.4.1** (Arrow-Lind theorem and Risk neutrality of firms with a diffuse shareholder base)**.** *Consider a firm owned by $n$ investors, each with utility $u_i\colon \mathbb{R}\to\mathbb{R}$ and deterministic wealth $W > 0$. Let $P(n)$ denote the minimal profit margin for which each individual shareholder accepts to take on the risk from a firm's risky project with zero expected return. Then*

$$\lim_{n\to\infty} P(n) = 0. \tag{1.10}$$

Theorem 1.4.1 shows that a firm with a diffuse shareholder base will behave *as if* it were risk neutral. The limit in (1.9) provides the intuition behind this result: the riskiness of $\frac{Z}{n}$, as captured by the individual risk premium, decreases at a faster rate, specifically $n^2$, than the rate $n$ at which the number of investors increases. In turn, this forces the total minimum excess return $P(n)$ to vanish as the ownership approaches a diffuse shareholder base.

It is important to point out that the result was derived for investors with the same, constant wealth. The non-randomness and equality of the agents' initial endowment can be relaxed; however, the derivation in this chapter allows us to develop a clearer interpretation of the theorem. When investors have random initial wealth, the result will hinge on the stochastic dependence between initial wealth and the project in which the firm invests (see e.g., Magill (1984)). In this case, the result continues to hold as far as the return on the firm's risky project is uncorrelated with its shareholders' wealth (Gravelle and Rees (2004)).

**Remark 1.4.2** (Insurance companies)**.** *Theorem 1.4.1 can be applied, in particular, to an insurance firm with a diffuse ownership base. Assume the insurer has capital* $K$ *and sells coverage against a bounded random loss* $L \geq 0$*, charging a premium* $\pi = \mathbb{E}[L] + P$*. This framework can be cast in the setting of the previous section: indeed, this is equivalent to investing in a risky project with profit margin* $P$ *and risk* $Z := \mathbb{E}[L] - L$ *(Gravelle and Rees (2004)). Hence, the Arrow-Lind Theorem implies that the minimal profit margin for insurers with a diffuse ownership base is* $P = 0$*.*

## 1.5 Conclusion

We have reviewed the assumptions and implications of the Arrow-Lind Theorem for firms with a diffuse shareholder base and heterogeneous investors. While firms owned by a single agent will inherit her preferences, corporations with a broad ownership will tend to behave as if they were risk-neutral even though each investor is strictly risk averse. This result follows from observing that the minimum compensation required by each investor to undertake a new risky project decreases at a faster rate than the rate at which the number of investors rises; in turn, this drives the *total* required return by investors to the risk-free interest rate. We have shown this result for investors with heterogeneous preferences but constant initial wealth. When wealth is random, an additional assumption is needed: in this case, the Arrow-Lind Theorem for firms holds when the return on firms' projects is uncorrelated with investors' wealth. Finally, when the firm is an insurance company with a diffuse ownership base, risk-neutrality implies that the minimum premium required by investors to consider selling insurance is the actuarially fair premium, i.e. the expected cover.

# Part III

# Research Papers

# Chapter 2

# Limited liability and the demand for insurance by individuals and corporations

ANDREA BERGESIO, PABLO KOCH-MEDINA AND COSIMO-ANDREA MUNARI[1]

**Abstract** Within the context of expected utility and in a discrete loss setting, we provide a complete account of the demand for coinsurance by strictly-risk averse agents and risk-neutral firms when they enjoy limited liability. When exposed to a bankrupting, binary loss and under actuarially fair prices, individuals and firms will either fully insure or not insure at all. The decision to insure will depend on whether the benefits the insuree derives from insurance after having compensated the damaged party are sufficiently attractive to justify the premium paid. When the loss is nonbinary, even when prices are actuarially fair, any amount of coinsurance can be optimal depending on the nature of the loss. Our results extend and partly rectify the results on insurance demand under limited liability encountered in the literature.

[1]A complementary version of this project, differing from this chapter in many parts, is available at SSRN: https://ssrn.com/abstract=3849959. Suggested citation: "Bergesio, Andrea and Koch-Medina, Pablo and Munari, Cosimo, Limited Liability and the Demand for Coinsurance by Individuals and Corporations (May 20, 2021). Swiss Finance Institute Research Paper No. 21-57." Partial support through the SNF, Switzerland project 189191 "Value-maximizing insurance companies: An empirical analysis of the cost of capital and investment policies" is gratefully acknowledged. Bergesio at University of Zurich and Swiss Finance Institute, Koch-Medina at University of Zurich and Swiss Finance Institute, Munari at University of Zurich and Swiss Finance Institute.

## 2.1 Introduction

Liability insurance protects the insuree against the claims from a third party who has suffered a loss for which the insuree is liable. Typical examples include general liability insurance, e.g., commercial general liability and business liability, professional liability insurance, providing coverage against, e.g., negligence or misrepresentation, and workers' compensation and employer liability insurance. When individuals and firms have limited liability, insurance may be the only means to ensure that the damaged party is compensated for the loss. Thus, the purpose of liability insurance is not only to protect the insuree, but also to protect the damaged party. As a result, understanding the demand for insurance by individuals and firms who enjoy limited liability is particularly important from a policymaking perspective. Indeed, as pointed out by Sinn (1982), if there is no intrinsic incentive to buy liability insurance, it may be desirable from a societal perspective to make it compulsory.[2] This paper revisits the impact of limited liability on the demand for coinsurance by strictly risk-averse individuals and by risk-neutral corporations. Within an expected utility framework and in a discrete setting, we provide a comprehensive analysis of the problem, thereby extending, unifying and clarifying results found in the literature.

#### Limited liability and strictly risk-averse individuals

The demand for coinsurance in an expected utility setting is well understood for a strictly risk-averse agent with deterministic initial wealth and an exposure to a random loss. In a competitive market, i.e., in a market where insurance sells at actuarially fair prices, such an agent will always opt for full insurance.[3] This result is often referred to as Mossin's theorem and goes back to Mossin (1968) and Smith (1968). Mossin's theorem has been extended in several directions ranging from allowing for random initial wealth (Doherty and Schlesinger (1983), Doherty

[2]Several papers have investigated the social and economic implications of limited liability and compulsory liability insurance. Early studies on the interaction between liability rules and insurance are Shavell (1982) and Shavell (1986). For further studies on the welfare impact of mandatory liability insurance for individuals and corporations, references include Keeton and Kwerel (1984), Han and MacMinn (1990), Jost (1996), and Boomhower (2019). For a general review of theoretical models and empirical applications related to personal and corporate liability insurance, see Baker and Siegelman (2013).

[3]If premia are actuarially unfavourable, strictly risk-averse agents will never opt for full insurance and may possibly not insure at all.

and Schlesinger (1985), Hong et al. (2011)), through considering other types of contract, e.g., deductible insurance (Schlesinger (1981) and Schlesinger (2000)) or upper-limit insurance (Schlesinger (2006)), to making behavioral assumptions beyond the expected-utility paradigm (e.g., Schlesinger (1997)). In this paper we are concerned with a different line of inquiry that was pursued in Sinn (1982), Shavell (1986), Konrad et al. (1993), and Gollier et al. (1997). In this literature, it is assumed that the agent has limited liability. This is a highly realistic feature since, typically, an individual cannot lose more than the value of what he or she owns, but may very well be exposed to "bankrupting" losses that exceed this value. Such losses may arise, for example, from claims due to damages to the property or health of third parties. The key question is whether agents have an intrinsic incentive to buy liability insurance. As observed in the cited literature, in the presence of a bankrupting loss agents may choose not to insure at all.

Sinn (1982), Shavell (1986), Konrad et al. (1993) work in a simple setting in which the individual is exposed to a binary loss, i.e., a loss of a fixed size either occurs or not. The main observation is that, even though prices are actuarially fair, the agent may choose not to insure at all. The reason for that is that having the third party absorb the loss may be more attractive than paying the premium to insure. However, if the agent chooses to insure, he or she will insure fully. The setting of Gollier et al. (1997) is more general and not limited to insurance.[4] In that paper, the main focus is on showing that an agent with limited liability will display greater willingness to accept risk than an agent with unlimited liability. Moreover, they show that if the wealth level is sufficiently low, the agent will choose a maximal risk exposure, i.e., the agent will not insure. Gollier et al. (1997) do not provide any results on whether it is sometimes optimal to have less than full risk exposure or even no risk exposure at all.

In this paper we provide a complete account of the demand for insurance when the loss is binary: under actuarially fair premia, the agent will insure fully if and only if full insurance increases expected utility. Otherwise, the agent will not insure at all. A similar result holds when premia are actuarially unfavorable but with partial insurance replacing full insurance. The cited literature contains part of these results,

[4] Gollier et al. (1997) address the problem from an optimal "risk taking" rather than an optimal "insurance" perspective. Moreover, translated into our setting, their proofs are given for losses whose distributions admit a continuous density function.

which, in particular, show that Mossin's theorem does not hold in the presence of limited liability. For a loss that is not binary, Mossin's theorem fails even more dramatically. As we show, even when prices are actuarially fair, the amount of insurance demanded by a strictly risk-averse agent may be anywhere between no insurance and full insurance and is case dependent. This invalidates the claim in Sinn (1982),[5] that all results obtained for binary losses remain true when losses are not binary. It also reinforces the need for caution when extrapolating from models with binary losses to models with more general types of losses. In any case, the main implication of the results for the binary loss model, namely that a strictly risk-averse agent may not have an incentive to protect damaged or injured third parties by insuring, remains true and the same policy implications remain valid.

**Limited liability and risk-neutral firms**

A good case can be made for assuming that publicly-held corporations are risk neutral because their owners can effectively diversify idiosyncratic risk; see, e.g., Mayers and Smith (1982). As a result, the firm's owners will maximize firm value. In a setting in which the firm is not exposed to financial market risk, this amounts to maximizing the expected cash flows to shareholders discounted at the risk-free rate. Similar to Froot et al. (1993), we think of firm value as being decomposed into two components. While *net asset value* corresponds to the value of the firm assuming it stops operating at the end of the period, *franchise value* corresponds to the value associated with the firm's ability to exploit future business opportunities in the subsequent periods.[6] We note that, in the corporate finance literature, Myers (1977) was among the first to explicitly recognize the value of future investment opportunities as the market value of the firm over and above the value of assets in place. Publicly-held firms always have limited liability, i.e., the firm's owners have the right to walk away from any obligations that exceed the value of the firm's assets. As in the case of individual agents, the firm may lose more than the value of the assets it owns when it is exposed to liability claims resulting from injuries or losses caused by the products or services it sells; see, e.g., Han and MacMinn (1990). The value of limited liability can be made explicit by decomposing net asset value into *default-free capital*, i.e., the value

[5] See footnote 5 in Sinn (1982).

[6] As in Froot et al. (1993) and Froot and Stein (1998), summarizing future expected profits in the end-of-period franchise value is a way to capture multi-periodicity in a single-period model.

of end-of-period capital assuming unlimited liability, and *default option value*, i.e., the value of the claims on the firm that need not be settled due to limited liability. Hence, we have

$$\text{Firm Value} = \text{ Default-Free Capital} + \text{Default Option Value} + \text{Franchise Value.}$$

To exploit its franchise, the firm is assumed to rely only on the available capital it has generated internally by the end of the period. The implication is that the firm cannot raise external capital, which is equivalent to consider a situation in which external financing is prohibitively costly due to extreme asymmetries of information or agency problems. This allows us to focus on insurance as the only source of financing when firms are prevented from accessing alternative financing channels.

The crucial question is: assuming the availability of insurance for the liability related losses at actuarially fair prices, will the firm purchase insurance and, if so, how much? The risk neutrality of the firm suggests that the firm should be indifferent between buying and not buying insurance. However, there are several factors that may induce a risk-neutral firm to behave *as if* it were risk averse. For example, as pointed out in Mayers and Smith (1982), the presence of frictions can make carrying risk costly and the firm may choose to insure to avoid incurring these costs.[7] When access to external capital is costless, a firm is not constrained in exploiting future business opportunities. However, when access to external financing is prevented, insurance can become attractive by providing a form of contingent capital that substitutes for external funds. This has been highlighted, for instance, in Froot et al. (1993) whose main focus is on discussing the benefits of hedging to protect the franchise value of the firm. While insurance may be beneficial to the firm either to avoid frictional costs or to protect its franchise, insurance also decreases the value of the option to default. As a result, retaining risk can have a positive impact on firm value; see, e.g., MacMinn (1987) and Han and MacMinn (1990). Hence, whether or not the firm chooses to buy insurance will ultimately depend on trade-offs that need to be made.

A critical trade-off is that between increasing the value of the default option and protecting the firm's franchise value in an environment where raising capital is costly or unfeasible. The relevance of this trade-off for value-maximizing firms has

[7]These frictions include taxation, costs of financial distress, agency costs, and regulatory costs; see Mayers and Smith (1982) for an overview.

been recognized and investigated extensively in other streams of literature. In the banking literature, Marcus (1984) and Keeley (1990) showed that in banks with a high franchise value,[8] the incentive to seek financial market risk induced by deposit insurance[9] was typically not exploited to the full because of the potential adverse effects on franchise value. In the context of insurance firms, Staking and Babbel (1995) and Babbel and Merrill (2005) argued that insurance firms with low franchise values, i.e., with poor prospects of attracting profitable new business, have an incentive to take excessive risk, whereas firms with a high franchise value to protect will tend to behave in a more risk-averse manner.

In this paper we provide a thorough analysis of how the "default option - franchise value" trade-off drives the demand for insurance of a risk-neutral limited-liability firm. To the best of our knowledge, this trade-off has not been systematically considered in the literature on the corporate demand for insurance. Several papers, e.g., MacMinn (1987), Mayers and Smith Jr (1987), Han and MacMinn (1990), and Garven and MacMinn (2013), have discussed the effect of limited liability, but have not considered the interplay with the firm's franchise value. The main reference for a discussion of the impact of risk on franchise value is Froot et al. (1993), who, however, ignore the impact on the default option. Hence, our paper complements the existent literature by addressing this trade-off in an explicit and rigorous manner.

Underlying the entire literature on why otherwise risk-neutral corporations may seek to reduce risk is the following key insight: taking into account the various frictions and the franchise value in a suitable manner, the firm "objective function" is formally equivalent to the expected utility of a strictly risk-averse agent. Hence, the behaviour of the firm is "observationally equivalent" to the behaviour of a strictly risk-averse agent. In the absence of limited liability, this means that the firm's demand for insurance is fully described by the classical version of Mossin's theorem for a strictly risk-averse individual. In the case of limited liability, one has to apply the corresponding results for strictly risk-averse agents with limited liability. Though corporate demand for insurance can thus be reduced to the problem of the demand for insurance by a strictly risk-averse individual, it is worthwhile to provide explicit results for firms. This is because the additional structure of the problem allows us to make plain the trade-offs that are at work behind the decision to insure or not.

[8] In the banking literature, franchise value is usually referred to as charter value.

[9] Which, in that context, plays a similar role to the firm's default option.

**Structure of the paper**

The paper is organized as follows. In Section 2.2, we present a brief account of the binary case that reproduces the results discussed above from a slightly different perspective. We show that, in this case, a version of Mossin's theorem under limited liability still holds. In Section 2.3, we highlight the trade-offs involved in a limited-liability firm's demand for insurance. Finally, in Section 2.4, we show that a limited-liability version of Mossin's theorem is not possible beyond the case of binary losses. Section 2.5 concludes.

## 2.2 The optimal insurance problem for a strictly risk-averse agent

In this section we provide a complete account of the demand for insurance by a strictly risk-averse agent with limited liability in the case of a binary loss and in the context of expected utility. The agent's utility function $u : \mathbb{R} \to \mathbb{R}$ is assumed to be differentiable, strictly increasing and strictly concave. Without loss of generality we assume that $u(0) = 0$. All random variables will be assumed to be defined on a nonatomic probability space $(\Omega, \mathcal{F}, \mathbb{P})$.[10] The expected value of a random variable $X$ is denoted by $\mathbb{E}[X]$. For every scalar $x \in \mathbb{R}$ we define $x_+ := \max\{x, 0\}$ and $x_- := \max\{-x, 0\}$. We start by recalling Mossin's theorem in the form that we need it.

### 2.2.1 Mossin's theorem

An agent with initial wealth $w > 0$ may suffer a random loss $L$. The agent can alter his or her terminal wealth by purchasing a coinsurance contract that we identify with the contractual level of coinsurance $\alpha \in [0, 1]$. The contractual indemnity for contract $\alpha$ is therefore given by $\alpha L$ and the associated premium equals

$$\pi_\lambda(\alpha) := (1 + \lambda)\alpha\mathbb{E}[L],$$

[10] If $(\Omega, \mathcal{F}, \mathbb{P})$ is nonatomic, we find for every $p \in [0, 1]$ an event $A \in \mathcal{F}$ such that $\mathbb{P}(A) = p$. We will need this flexibility when constructing a special random loss in Section 2.4.

where $\lambda \geq 0$ is the insurer's profit loading. The case where $\lambda = 0$, the "actuarially fair" case, corresponds to the case in which insurance markets are perfectly competitive and the insurer has zero expected profit. After purchasing insurance, the agent's terminal wealth is

$$W_\lambda(\alpha) := w - L - \pi_\lambda(\alpha) + \alpha L.$$

The agent's expected utility is given by

$$U_\lambda(\alpha) := \mathbb{E}[u(W_\lambda(\alpha))].$$

An optimal coinsurance level is a contract $\alpha^* \in [0, 1]$ that maximizes expected utility, i.e., such that

$$U_\lambda(\alpha^*) = \max_{\alpha \in [0,1]} U_\lambda(\alpha). \tag{2.1}$$

By our assumptions on $u$, the function $U_\lambda$ is strictly concave. Hence, problem (2.1) admits a unique optimal insurance level which we denote by $\alpha^*_\lambda \in [0, 1]$. The following result goes back to Mossin (1968) and Smith (1968), and describes the qualitative behaviour of the optimal insurance level in dependence of the profit loading.

**Theorem 2.2.1** (Mossin)**.** *There exists a critical insurance profit loading $\overline{\lambda} > 0$ such that:*

*(i)* $\alpha^*_\lambda = 1$ *for* $\lambda = 0$.

*(ii)* $0 < \alpha^*_\lambda < 1$ *for* $0 < \lambda < \overline{\lambda}$.

*(iii)* $\alpha^*_\lambda = 0$ *for* $\lambda \geq \overline{\lambda}$.

*Moreover,* $\{\alpha^*_\lambda\ ;\ 0 \leq \lambda \leq \overline{\lambda}\} = [0, 1]$.

### 2.2.2 Liability insurance with a binary loss

For the rest of this section we consider a binary loss $L$, i.e.

$$L = \begin{cases} \ell & \text{with probability } p, \\ 0 & \text{with probability } 1 - p \end{cases}$$

for some $\ell > 0$ and $0 < p < 1$. The premium for the contract $\alpha$ is then given by

$$\pi_\lambda(\alpha) = (1+\lambda)\alpha p \ell,$$

and the agent's terminal wealth by

$$W_\lambda(\alpha) := \begin{cases} w - \pi_\lambda(\alpha) & \text{in case of no loss,} \\ w - \pi_\lambda(\alpha) - (1-\alpha)\ell & \text{in case of loss.} \end{cases}$$

We assume that the loss arises from a claim for compensation against damage to a third party, so that $\alpha$ represents a liability insurance contract. Since the agent cannot lose more than his or her initial wealth, it is natural to assume limited liability, i.e., the agent will compensate the damaged party only up to the amount $w$. As a result, from the agent's perspective, the "relevant" wealth is not $W_\lambda(\alpha)$, which represents wealth "before limited liability", but wealth "after limited liability"

$$W_\lambda^+(\alpha) := [W_\lambda(\alpha)]_+ = \max\{W_\lambda(\alpha), 0\}.$$

The agent's expected utility "after limited liability" is then given by

$$V_\lambda(\alpha) := \mathbb{E}[u(W_\lambda^+(\alpha))].$$

In this case, an optimal level of insurance is any $\alpha^* \in [0,1]$ satisfying

$$V_\lambda(\alpha^*) = \max_{\alpha \in [0,1]} V_\lambda(\alpha). \tag{2.2}$$

Since $V_\lambda$ is continuous and $[0,1]$ is compact, the above maximization problem always admits a solution. However, in contrast to $U_\lambda$, the objective function $V_\lambda$ fails to be concave in general because terminal wealth enters as $W_\lambda^+(\alpha)$ and not as $W_\lambda(\alpha)$. As a result, it is unclear whether the optimal solution is unique or not. We are interested in understanding when optimal insurance levels are unique, when partial or full insurance are optimal, and, in particular, when full insurance is not optimal. As pointed out in the introduction, knowing when full insurance is not optimal is important because the loss primarily affects the damaged party rather than the prospective policyholder. As a result, if agents do not have an incentive to protect the damaged party, compulsory insurance may be desirable from a societal perspective.

### 2.2.3 Optimal coinsurance level

A first clear requirement for purchasing insurance to make sense is that the premium paid does not exceed the compensation received in the loss state, i.e. that $\pi_\lambda(\alpha) < \alpha\ell$ for some, and hence for every, $\alpha > 0$. This is easily seen to be equivalent to

$$0 \leq \lambda < \frac{1-p}{p}. \tag{2.3}$$

To analyze the maximization problem (2.2) it is useful to distinguish *bankrupting* losses ($\ell > w$) from *nonbankrupting* losses ($\ell \leq w$). When the loss is nonbankrupting and (2.3) holds, wealth after limited liability always coincides with wealth before limited liability, i.e., $W_\lambda^+(\alpha) = W_\lambda(\alpha)$ for every $\alpha \in [0,1]$. As a result, for nonbankrupting losses, the problems with and without limited liability coincide and the solution to the maximization problem (2.2) is already fully described by Theorem 2.2.1. We summarize the preceding discussion in the following proposition.

**Proposition 2.2.2.** *If $\ell$ is nonbankrupting, then $W_\lambda^+(\alpha) = W_\lambda(\alpha)$ for every $\alpha \in [0,1]$ and the optimization problems (2.2) and (2.1) coincide. In particular, the optimal level of coinsurance is fully described by Theorem 2.2.1.*

Assuming that the loss is bankrupting, we first note that if wealth *before* limited liability in the loss state is negative for some contract $\alpha \in (0,1]$, then the agent will not want to buy the contract because it is "ineffective". Indeed, wealth in the no-loss state would be strictly reduced by the premium and, in the loss state, any benefit obtained from insurance would be used to pay the claimants, so that wealth *after* limited liability would remain 0. Furthermore, if wealth before limited liability in the loss state is negative for every possible contract $\alpha \in (0,1]$, then the agent will never contemplate buying insurance. It is easy to see that this is the case if and only if $w \leq \pi_\lambda(1)$ or, equivalently, $\lambda \geq \frac{w-p\ell}{p\ell}$. Thus, we directly obtain the following result.

**Proposition 2.2.3.** *If $\ell$ is bankrupting and $\lambda \geq \frac{w-p\ell}{p\ell}$, then $\alpha = 0$ is the unique optimal coinsurance level.*

In light of Proposition 2.2.3, henceforth we may focus on bankrupting losses satisfying the following condition

$$0 \leq \lambda \leq \frac{w - p\ell}{p\ell}, \tag{2.4}$$

where, for convenience, we allow for $\lambda = \frac{w-p\ell}{p\ell}$ even though for this $\lambda$ insurance is not attractive. Note that, since the loss is bankrupting, this condition implies that for every contract the premium is strictly less than the compensation in case of loss, i.e., condition (2.4) implies condition (2.3).

Under condition (2.4), we define the *effectiveness threshold* by

$$\widetilde{\alpha}_\lambda := \min\{\alpha \in [0,1]\,;\ W_\lambda(\alpha) \geq 0\} = \frac{\ell - w}{\ell - (1+\lambda)p\ell} \in (0,1).$$

This threshold corresponds to the lowest coinsurance level such that wealth before limited liability in the loss state is positive. From the discussion above we infer that contracts in the interval $(0, \widetilde{\alpha}_\lambda]$ are ineffective and thus undesirable. Hence, the optimal contract on the interval $[0, \widetilde{\alpha}_\lambda]$ is $\alpha = 0$. By contrast, contracts in the interval $[\widetilde{\alpha}_\lambda, 1]$ are effective and we have $W_\lambda^+(\alpha) = W_\lambda(\alpha)$. This implies that the expected utility derived from contracts in that range is the same whether the agent has limited or unlimited liability. In particular, $V_\lambda$ is strictly concave and, hence, admits a unique maximum $\beta_\lambda^*$ on that sub-interval. As a result, the optimal solution of the optimization problem (2.2) under limited liability is either no insurance or coinsurance at the level $\beta_\lambda^*$. Noting that $V_\lambda(0) = V_0(0)$, we can establish which of these contracts is optimal by simply comparing $V_0(0)$ and $V_\lambda(\beta_\lambda^*)$. We record this in the following proposition.

**Proposition 2.2.4.** *If $\ell$ is bankrupting and $0 \leq \lambda \leq \frac{w-p\ell}{p\ell}$, then:*

*(i) If $V_0(0) > V_\lambda(\beta_\lambda^*)$, then $\alpha = 0$ is the unique optimal coinsurance level.*

*(ii) If $V_0(0) = V_\lambda(\beta_\lambda^*)$, then $\alpha = 0$ and $\alpha = \beta_\lambda^*$ are the only optimal coinsurance levels.*

*(iii) If $V_0(0) < V_\lambda(\beta_\lambda^*)$, then $\alpha = \beta_\lambda^*$ is the unique optimal coinsurance level.*

**Remark 2.2.5.** *Proposition 2.2.4 implies that the optimal level of coinsurance in the limited-liability case is always smaller than that under the unlimited-liability case.*

*Indeed, in both cases optimal coinsurance results from maximizing the same objective function, but in the limited liability case over a smaller interval. This point was also made in Gollier et al. (1997), who show that introducing limited liability increases the optimal exposure to risk of risk-averse agents.*

Together with Proposition 2.2.3 and Proposition 2.2.4, the following result provides a full characterization of the optimal demand for insurance when the agent has limited liability and is exposed to a binary loss. The theorem below distinguishes cases depending on the attractiveness of full insurance at actuarially fair prices.

**Theorem 2.2.6** (Individual demand with binary losses and limited liability)**.** *If $\ell$ is bankrupting and $w > \pi_0(1)$, then:*

*(i) If $V_0(1) < V_0(0)$, then $\alpha = 0$ is the unique optimal coinsurance level for every $\lambda \geq 0$.*

*(ii) If $V_0(1) = V_0(0)$, then $\alpha = 0$ and $\alpha = 1$ are the only optimal coinsurance levels for $\lambda = 0$ whereas $\alpha = 0$ is the unique optimal coinsurance level for $\lambda > 0$.*

*(iii) If $V_0(1) > V_0(0)$, then $\alpha = 1$ is the unique optimal coinsurance level for $\lambda = 0$ and there exists a critical insurance profit loading $0 < \overline{\lambda} < \frac{w-p\ell}{p\ell}$ such that:*

*(a) $\alpha = \beta^*_\lambda \in (0,1)$ is the unique optimal coinsurance level for $0 < \lambda < \overline{\lambda}$.*

*(b) $\alpha = 0$ and $\alpha = \beta^*_\lambda \in (0,1)$ are the only optimal coinsurance levels for $\lambda = \overline{\lambda}$.*

*(c) $\alpha = 0$ is the unique optimal coinsurance levels for $\lambda > \overline{\lambda}$.*

**Remark 2.2.7.** *It is clear how to generalize Theorem 2.2.6 to the case where the agent has a strictly positive minimum guaranteed wealth as in Sinn (1982). Note that our result encompasses the results in Sinn (1982), Shavell (1986), Konrad et al. (1993), who show that insurance may not be attractive to a risk-averse decision maker with limited liability even when prices are actuarially fair. However, these references provide neither a complete treatment nor a formal derivation of the agent's optimal demand for coinsurance. In addition, Sinn (1982) claims that his results can be generalized to settings with arbitrary losses. In Section 2.4 we show that this is, in fact, not possible.*

One of the main points highlighted in the literature on insurance demand under limited liability is the fact that, for small levels of wealth, a strictly risk-averse agent will choose to buy no insurance. Sinn (1982), Shavell (1986), and Konrad et al. (1993) show this in the case of binary losses and Gollier et al. (1997) for general loss profiles. This type of behavior is often referred to as "betting for resurrection".

We conclude this section by showing that in our setting such a behavior is implied by Theorem 2.2.6. We do this assuming a competitive insurance market, i.e., under actuarially fair prices. We complement the result by highlighting the impact of loss size on demand for insurance.

**Proposition 2.2.8.** *If $\lambda = 0$, then:*

*(i) There exists a critical wealth level $\widetilde{w} > 0$ such that:*

- *(a) $\alpha = 1$ is the unique optimal coinsurance level if $w > \widetilde{w}$.*
- *(b) $\alpha = 0$ and $\alpha = 1$ are the only optimal coinsurance levels if $w = \widetilde{w}$.*
- *(c) $\alpha = 0$ is the unique optimal coinsurance level if $w < \widetilde{w}$.*

*(ii) There exists a critical loss level $\widetilde{\ell} > 0$ such that:*

- *(a) $\alpha = 1$ is the unique optimal coinsurance level if $\ell < \widetilde{\ell}$.*
- *(b) $\alpha = 0$ and $\alpha = 1$ are the only optimal coinsurance levels if $\ell = \widetilde{\ell}$.*
- *(c) $\alpha = 0$ is the unique optimal coinsurance level if $\ell > \widetilde{\ell}$.*

Figure 2.1 illustrates part (i) of the preceding proposition by displaying the outcomes for high and low levels of wealth. The model used for both panels differs only in the amount of initial wealth. Panel (a), based on a larger wealth, shows a situation in which full insurance is optimal. In Panel (b), reduced wealth leads to a higher premium making full insurance less attractive than no insurance. Not shown in the figure but clear from continuity considerations is that for some intermediate level of wealth the agent will be indifferent between no insurance and full insurance.

In the next section, we turn our attention to the the optimal insurance problem of a firm, drawing a connection between our results for a strictly-risk averse agent and those we derive for the case of widely-held firms without access to capital markets.

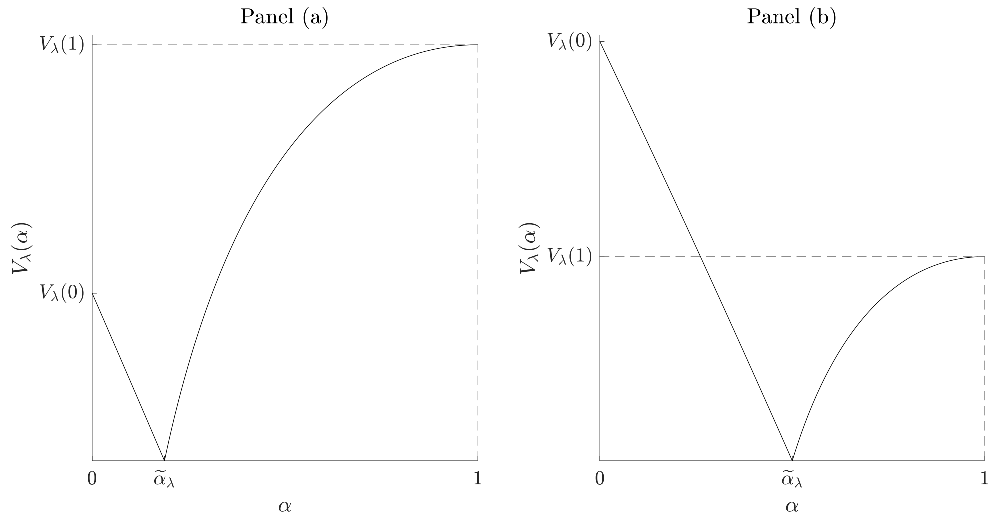



**Figure 2.1:** Expected utility $V_\lambda(\alpha)$ as a function of the coinsurance level $\alpha$ for actuarially fair prices ($\lambda = 0$). Panel (a) shows the case in which $V_\lambda(1) > V_\lambda(0)$, whereas Panel (b) shows the case in which $V_\lambda(1) < V_\lambda(0)$. The effectiveness threshold $\widetilde{\alpha}_\lambda$ represents the coinsurance level below which no insurance is optimal and above which full insurance is preferred. Both panels are generated using any utility function $u(x)$ such that $x \geq 0$ and liability loss $\ell = 10$ with probability $p = 0.2$. The expected utility in Panel (a) belongs to an agent with initial wealth $w = 8.5$, whereas the results in Panel(b) are obtained for $w = 6$.

## 2.3 The optimal insurance problem for a risk-neutral firm

We now investigate the demand for insurance by firms. On one extreme, when a firm is owned by a sole strictly risk-averse individual, the firm will purchase the level of insurance that maximizes the owner's expected utility, i.e., the problem fits into the setting of the previous section. When a firm is closely held by more than one owner, each with a different utility function, there will, in general, be no unanimity as to the objective of the firm so that the optimal insurance problem is not well posed. On the other extreme, when a firm is publicly traded with a broad ownership base and its profits are not exposed to financial market risk, its shareholders are assumed to hold well diversified portfolios that eliminate any exposure to idiosyncratic risk, so that the firm is considered to behave as if it were risk neutral; see, e.g., Mayers and Smith (1982). This is the case we focus on. As we point out in Remark 2.3.8, it is well known that when access to external capital is constrained, the costs of financing can induce the firm to act in a risk-averse manner in spite of the assumed risk neutrality.

In such cases, the insurance problem for a firm can be embedded into the model treated in the previous section.[11] Our aim, however, is to make explicit the trade-off that plays out between the impact of risk on the limited liability option, on the one side, and on the ability to exploit future business opportunities, on the other. To be able to focus on the main message, we deal exclusively with the case of competitive insurance markets, i.e., we assume that insurance is sold at actuarially fair rates.

We continue to work in a one-period model with two dates $t = 0$ and $t = 1$ and consider a risk-neutral firm with initial capital $\kappa > 0$, which it uses to generate a deterministic return $\rho > 0$. For convenience, we assume that the risk-free rate is equal to zero. At date $t = 1$, the firm is exposed to a random loss given for $\ell > 0$ and $0 < p < 1$ by

$$L = \begin{cases} \ell & \text{with probability } p, \\ 0 & \text{with probability } 1 - p. \end{cases}$$

As in the case of a strictly-risk averse individual, we allow for the loss to exceed the value of the firm's assets. Such a possibility arises, for instance, when a firm is exposed to claims resulting from injuries or losses caused by the products or services it sells; see, e.g., Han and MacMinn (1990). We assume that the firm is able to purchase insurance at actuarially fair prices, i.e., the premium for the contract $\alpha$ is

$$\pi(\alpha) := \alpha \mathbb{E}[L] = \alpha p \ell.$$

The firm's end-of-period capital after purchasing the contract $\alpha$ and before accounting for a possible default is given by the random variable

$$W(\alpha) := \begin{cases} \kappa(1 + \rho) - \pi(\alpha) & \text{in case of no loss,} \\ \kappa(1 + \rho) - \pi(\alpha) - (1 - \alpha)\ell & \text{in case of loss.} \end{cases}$$

In view of limited liability, end-of-period capital after buying insurance is given by

[11] Induced risk aversion refers to a behavior akin to those of risk-averse agents observed in otherwise risk-neutral individuals or firms. This phenomenon has been studied extensively within the literature on the determinants of corporate risk management, especially hedging and insurance policies. These studies include, but are not limited to, Smith and Stulz (1985), Nance et al. (1993), Froot et al. (1993), Froot and Stein (1998), and Purnanandam (2008).

$$W^+(\alpha) := [W(\alpha)]_+ = \max\{W(\alpha), 0\}.$$

As in the preceding section, we distinguish between *bankrupting* losses ($\ell > \kappa(1+\rho)$) and *nonbankrupting* losses ($\ell \leq \kappa(1+\rho)$).

### 2.3.1 Franchise value

In addition to terminal equity, there is another source of value at date $t = 1$ to which shareholders are entitled: the value of the firm's ability to exploit future business opportunities. To finance these future business opportunities, firms may rely in principle on a mix of internal capital and external funds raised at $t = 1$. On the one hand, when external capital is costless the firm can always fully exploit future growth opportunities regardless of the amount of internal funds that are available at time $t = 1$. In this case the firm has unconstrained access to funds and thus will be either indifferent towards insurance or will refrain from it entirely if priced at unfavourable rates. However, when external financing is costly due to market imperfections such as asymmetric information and agency problems, firms will be constrained in the amount of funds they can raise. Firms with costly financing may therefore find it attractive to purchase insurance coverage to reduce the expected costs that arise from relying on external financing (Froot et al. (1993)).

We consider the situation in which the firm can use only its available capital at the end of the period. We focus therefore on a fully constrained firm that cannot access external capital markets, or equivalently a firm that in principle has the ability to raise capital but is subject to extreme financing costs that make it economically unprofitable to resort to external funds. Since the firm cannot raise equity or debt capital, this setting allows us to draw attention to the case in which insurance represents the only source of financing.[12]

If, at date $t = 1$, the firm has at its disposal funds amounting to $d \geq 0$, it can exploit investment opportunities with present value $d + F(d)$, i.e., $F(d)$ represents the net present value, or the value added, when investing $d$. We shall call $F : [0, \infty) \to \mathbb{R}$ the firm's *growth-opportunities function* at date $t = 1$.

[12]Fully constrained firms have also been considered in the corporate finance literature to capture the maximal cost of financing frictions (see e.g. Almeida et al. (2004), Moyen (2004)).

We impose some fairly natural conditions on $F$, which are similar in spirit to those stipulated in Froot et al. (1993).[13]

**Assumption 2.3.1.** *The function* $F : [0, \infty) \to \mathbb{R}$ *is continuously differentiable, strictly increasing, strictly concave, and* $F(0) = 0$*. Moreover, we assume there exists* $F_\infty > 0$ *such that*

$$F(d) \to F_\infty \quad \textit{as } d \to \infty. \tag{2.5}$$

This assumption reflects the intuition that the firm cannot create value if no funds are available ($F(0) = 0$), that the net present value of future investments increases with the amount of capital invested ($F$ is strictly increasing), and that there are diminishing returns to new business opportunities as the firm grows in size ($F$ is strictly concave). In fact, we assume strongly diminishing returns, i.e., that for sufficiently large levels of capital, the marginal value from an additional unit of funds becomes negligible ($F(d) \to F_\infty$ as $d \to \infty$). Figure 2.2 shows the qualitative shape of $F$ under Assumption 2.3.1.

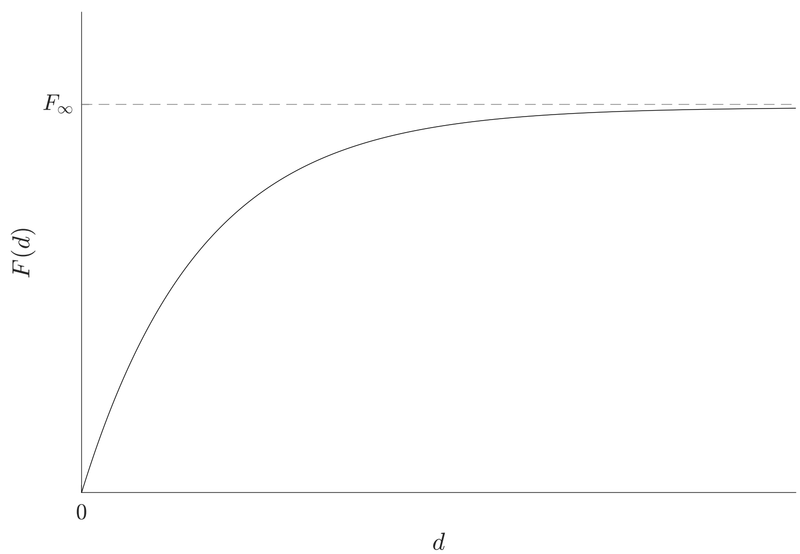



**Figure 2.2:** Qualitative shape of the growth-opportunities function $F$.

[13] Compared to Froot et al. (1993), we impose the additional intuitive condition that the function $F$ exhibits *strongly diminishing returns*, which requires that for sufficiently large levels of capital, the marginal value from an additional unit of funds becomes negligible; see, e.g., Morrison and Siegel (1997) and Basu and Fernald (1997).

When external capital is not available, another implication is that for fully constrained firms the default event is exogenous, i.e. it cannot be avoided by issuing capital. Specifically, default occurs whenever the capital $W(\alpha)$ available at the end of the period, after purchasing insurance, is negative. In a binary loss setting this implies that default is a direct consequence of *bankrupting* losses when the insurance indemnity is insufficient to restore the lost capital. As default prevents the firm from exploiting future business opportunities, insurance then becomes the only source of financing to which the firm can resort not only to relax its financing constraints but also to avoid the financial distress costs originated from lost growth opportunities.

### 2.3.2 Firm value and its decomposition

After purchasing the contract $\alpha$, the value of the firm at time $t = 1$ is given by the random variable

$$W^+(\alpha) + F(W^+(\alpha)),$$

which captures terminal equity and the value of future growth opportunities, provided funds are available at the end of the period. The risk neutrality of the firm implies that firm value at time $t = 0$ is determined by taking the expected value of firm value at time $t = 1$ and discounting it at the risk-free rate, which we have assumed to be 0, i.e. firm value at date $t = 0$ after buying the contract $\alpha$ is

$$V(\alpha) := \mathbb{E}[W^+(\alpha)] + FV(\alpha), \tag{2.6}$$

where $\mathbb{E}[W^+(\alpha)]$ is the value of terminal equity and

$$FV(\alpha) := \mathbb{E}[F(W^+(\alpha))]$$

is the *franchise value*.[14]

Note that we can rewrite terminal equity at date $t = 1$ as

[14]Note that terminology used to refer to the present value of future growth opportunities is not uniform across the literature. While the term "franchise value" is often used in the literature on insurance, see, e.g., Hancock et al. (2001) and Babbel and Merrill (2005), the banking literature typically favors the term "charter value", see, e.g., Keeley (1990). In the asset pricing and corporate finance literature the expressions "value of growth opportunities" and "value of the growth option" have become customary, see, e.g., Berk et al. (1999), Cooper (2006) and Da et al. (2012).

$$W^+(\alpha) = W(\alpha) + W^-(\alpha),$$

where $W(\alpha)$ corresponds to *default-free capital* at time $t = 1$, i.e., capital assuming unlimited liability, which is given as above by

$$W(\alpha) := \kappa(1 + \rho) - L - \pi(\alpha) + \alpha L,$$

and $W^-(\alpha)$ corresponds to the firm's *default option*, i.e., the amount by which the firm defaults at date $t = 1$, which is given by

$$W^-(\alpha) := [W(\alpha)]_- = \max\{-W(\alpha), 0\}.$$

It follows that firm value at date $t = 0$ can be expressed as

$$V(\alpha) = DFC(\alpha) + DO(\alpha) + FV(\alpha), \tag{2.7}$$

where

$$DFC(\alpha) := \mathbb{E}[W(\alpha)] \equiv \kappa(1 + \rho) - p\ell, \tag{2.8}$$

$$DO(\alpha) := \mathbb{E}[W^-(\alpha)] \tag{2.9}$$

are the value of *default-free capital* and the value of the *default option*, respectively.[15]

### 2.3.3 Optimal coinsurance level

We now study optimal levels of coinsurance, i.e. contracts $\alpha^* \in [0, 1]$ such that

$$V(\alpha^*) = \max_{\alpha \in [0,1]} V(\alpha). \tag{2.10}$$

As argued in the previous section, we can always ensure the existence of optimal solutions due to the continuity of $V$ but nothing can be said a priori on their multiplicity, because $V$ is not (strictly) concave.

[15] This is, in fact, a problem in Froot et al. (1993) because their mathematical model is based on normal returns for the risky asset, which means that internally generated capital can be negative in some states of the world. Hence, their growth-opportunities function cannot be a concave function of internally generated capital and their use of arguments invoking concavity is unjustified.

By (2.8), default-free capital $DFC(\alpha)$ does not depend on $\alpha \in [0,1]$. Hence, to understand optimal coinsurance levels it suffices to understand the impact of purchasing a contract $\alpha \in [0,1]$ on the value of the default option $DO(\alpha)$ and on the franchise value $FV(\alpha)$. We show next how the coinsurance level affects the default option.

**Lemma 2.3.2.** *The following statements hold:*

*(i) If $\ell$ is nonbankrupting, then $DO(\alpha) = 0$ for every $\alpha \in [0,1]$.*

*(ii) If $\ell$ is bankrupting and $\kappa(1+\rho) > p\ell$, then*

$$\widetilde{\alpha} := \frac{\ell - \kappa(1+\rho)}{(1-p)\ell} \in (0,1) \tag{2.11}$$

*and $DO$ is strictly decreasing on $[0,\widetilde{\alpha}]$ and constant and equal to $0$ on $[\widetilde{\alpha},1]$.*

*(iii) If $\ell$ is bankrupting and $\kappa(1+\rho) \leq p\ell$, then $DO$ is strictly decreasing on $[0,1]$.*

Note that, when the loss is bankrupting and $\kappa(1+\rho) > p\ell$, we have

$$\widetilde{\alpha} = \min\{\alpha \in [0,1]\,;\ \kappa(1+\rho) - \ell - \pi(\alpha) + \alpha\ell \geq 0\},$$

so that, similar to what we did in Section 2, $\widetilde{\alpha}$ corresponds to an *effectiveness threshold*, i.e. it is the threshold beyond which wealth after limited liability becomes strictly positive. Thus, this threshold corresponds to the lowest coinsurance level making the firm default free.

When the loss is nonbankrupting, the default option has no value regardless of the level of insurance bought. Hence, whether insurance increases firm value depends only on its impact on the firm's franchise value. The following result shows that when the loss is *nonbankrupting*, franchise value is maximized if full insurance is purchased.

**Proposition 2.3.3.** *If $\ell$ is nonbankrupting, then $FV$ is strictly increasing on $[0,1]$. In particular, $\alpha = 1$ is the unique optimal level of coinsurance.*

As was the case for the strictly risk-averse individual in Section 2, when the loss is bankrupting, it is disadvantageous to buy any insurance if terminal capital after limited liability remains equal to zero for all coinsurance levels, which is the case if and only if $\kappa(1+\rho) \leq p\ell$. We record this in the next proposition.

**Proposition 2.3.4.** *If $\ell$ is bankrupting and $\kappa(1+\rho) \le p\ell$, then FV is strictly decreasing on* $[0,1]$. *In particular,* $\alpha = 0$ *is the unique optimal coinsurance level.*

By the preceding proposition, from now on we may focus on bankrupting losses for which

$$\kappa(1+\rho) > p\ell,$$

in which case the effectiveness threshold $\widetilde{\alpha}$ defined in (2.11) is a well defined number in $(0,1)$. This threshold is useful to understand the impact of buying insurance on the value of the default option and franchise value. In fact, we already know from Lemma 2.3.2 that $DO$ is strictly decreasing on $[0,\widetilde{\alpha}]$ and constant and equal to 0 on $[\widetilde{\alpha},1]$. The following result describes the behaviour of the franchise value.

**Lemma 2.3.5.** *If $\ell$ is bankrupting and $\kappa(1+\rho) > p\ell$, then FV is strictly decreasing on* $[0,\widetilde{\alpha}]$ *and strictly increasing on* $[\widetilde{\alpha},1]$.

Thus, to determine the optimal coinsurance level, a trade-off between the loss in the value of the default option and the gain in franchise value needs to be resolved. After purchasing a contract $\alpha$, the *loss* in the value of the default option is given by

$$DO^{loss}(\alpha) := DO(0) - DO(\alpha),$$

which, by Lemma 2.3.5, is positive, strictly increasing on $[0,\widetilde{\alpha}]$, and equal to $DO(0)$ on $[\widetilde{\alpha},1]$. On the other hand, the *gain* in franchise value is equal to

$$FV^{gain}(\alpha) := FV(\alpha) - FV(0),$$

which, by Lemma 2.3.5, is positive, strictly decreasing on $[0,\widetilde{\alpha}]$, and strictly increasing on $[\widetilde{\alpha},1]$. It follows that on $[0,\widetilde{\alpha}]$ the default option value and the franchise value both decrease. Hence, it is never optimal to buy a contract in the interval $[0,\widetilde{\alpha}]$. On the other hand, $DO^{loss}(\alpha)$ is constant and $FV^{gain}(\alpha)$ is increasing on $[\widetilde{\alpha},1]$. Hence, it is always optimal to buy a contract in the interval $[\widetilde{\alpha},1)$. As a consequence, the only potential optimal coinsurance levels are $\alpha = 0$ and $\alpha = 1$, with full insurance being the optimal strategy if and only if

$$FV(1) - FV(0) = FV^{gain}(1) \ge DO^{loss}(1) = DO(0),$$

i.e., if and only if the loss in the value of the default option is offset by a sufficiently large gain in franchise value. The trade-offs we have just described are illustrated in Figure 2.3 for the case in which full insurance is optimal and Figure 2.4 for the case in which no insurance is optimal. The preceding discussion yields the following theorem, which is just a recasting of Theorem 2.2.6 into the corporate setting and, together with Proposition 2.3.3 and Proposition 2.3.4, completes the picture on the corporate demand for insurance by a risk-neutral firm under actuarially fair prices and no access to external capital markets.

**Theorem 2.3.6** (Trade-off between franchise value and default option)**.** *Consider a risk-neutral firm without access to external financing. If $\ell$ is bankrupting and $\kappa(1+\rho) > p\ell$, then:*

*(i) If $FV(1) - FV(0) > DO(0)$, then $\alpha = 1$ is the optimal coinsurance level.*

*(ii) If $FV(1) - FV(0) = DO(0)$, then $\alpha = 0$ and $\alpha = 1$ are the optimal coinsurance levels.*

*(iii) If $FV(1) - FV(0) < DO(0)$, then $\alpha = 0$ is the optimal coinsurance level.*

Our previous result can be cast in terms of critical levels for the firm's capitalization and for the size of the loss to which the firm's capital is exposed. As is to be expected, all else being equal, when either the initial capital is too low or similarly the loss size is too large, then the value of the default option will be too high to be offset by the gain in franchise when fully insuring. In the opposite situation, the value of the default option will be too low to be worth protecting by not insuring.

**Proposition 2.3.7.** *Consider a risk-neutral firm without access to external financing. Then:*

*(i) There exists a critical capital level $\widetilde{\kappa} > 0$ such that:*

*(a) $\alpha = 1$ is the optimal coinsurance level if $\kappa > \widetilde{\kappa}$.*

*(b) $\alpha = 0$ and $\alpha = 1$ are the optimal coinsurance levels if $\kappa = \widetilde{\kappa}$.*

*(c) $\alpha = 0$ is the optimal coinsurance level if $\kappa < \widetilde{\kappa}$.*

*(ii) There exists a critical loss level $\widetilde{\ell} > 0$ such that:*

(a) $\alpha = 1$ *is the optimal coinsurance level if* $\ell < \widetilde{\ell}$.

(b) $\alpha = 0$ *and* $\alpha = 1$ *are the optimal coinsurance levels if* $\ell = \widetilde{\ell}$.

(c) $\alpha = 0$ *is the optimal coinsurance level if* $\ell > \widetilde{\ell}$.

**Remark 2.3.8.** *Note that the firm value* $V$ *in Equation* (2.7) *can be written equivalently as*

$$V(\alpha) = \mathbb{E}[u([\kappa(1+\rho) - L - \pi(\alpha) + \alpha L]_+)],$$

*where* $u : \mathbb{R} \to \mathbb{R}$ *is any strictly-increasing and concave function satisfying*

$$u(x) = x + F(x), \ x \geq 0.$$

*If external capital can be raised at no cost, the firm has access to costless funds and* $u$ *can be taken to be affine, which implies the firm's problem is formally identical to the problem of a risk-neutral individual with limited liability maximizing expected*

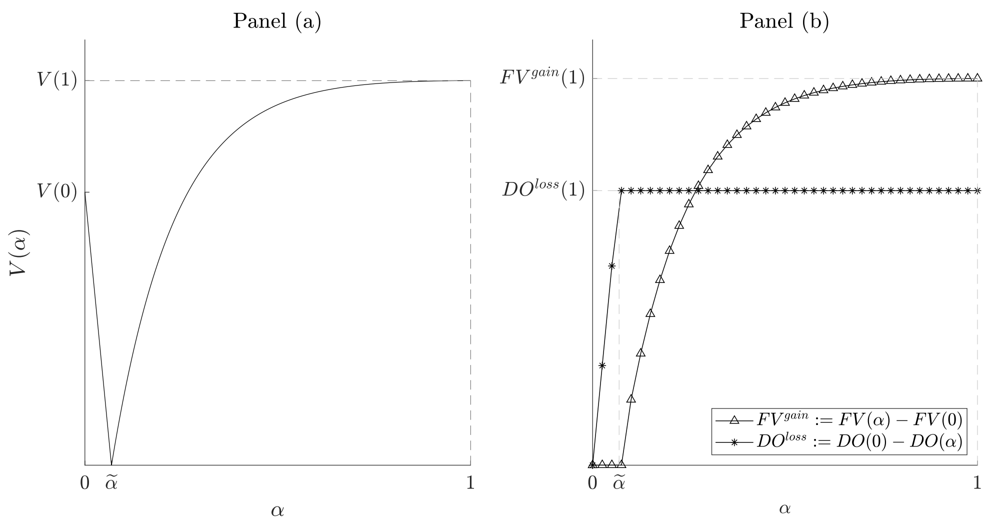



**Figure 2.3:** The figure plots the firm value $V$ and changes in the firm value components $DO$, $FV$ as a function of the insurance level $\alpha$. Panel (a) shows the firm value for the case in which full insurance is optimal, $V(1) > V(0)$. Panel (b) illustrates the *loss* in the default option value $DO^{loss}(\alpha)$ and the *gain* in franchise value $FV^{gain}(\alpha)$ corresponding to the value function in Panel (a). The x-axis shows the effectiveness threshold $\widetilde{\alpha}$, below which no insurance is optimal and above which *full* insurance is preferred. The figure is generated using the growth-opportunities function $F = 1 - e^{-\gamma x}$, $x \geq 0$ with $\gamma = 0.5$. Other model parameters are set as follows: $\ell = 12.7$, $p = 0.2$, $\kappa = 10$ and $\rho = 0.2$. The premium is actuarially fair, $\lambda = 0$.

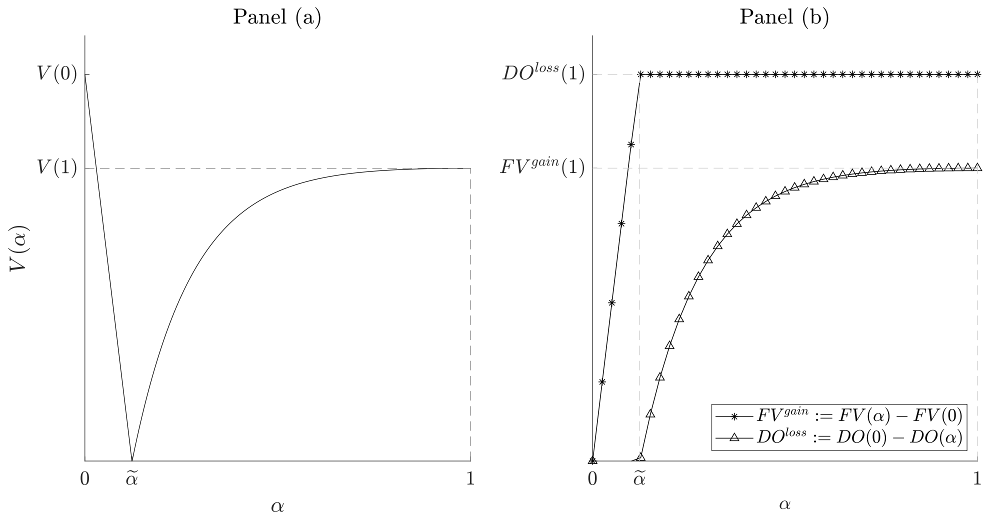



**Figure 2.4:** The figure plots the firm value $V$ and changes in the firm value components $DO$, $FV$ as a function of the insurance level $\alpha$. Panel (a) shows the firm value for the case in which no insurance is optimal, $V(1) < V(0)$. Panel (b) illustrates the *loss* in the default option value $DO^{loss}(\alpha)$ and the *gain* in franchise value $FV^{gain}(\alpha)$ corresponding to the value function in Panel (a). The x-axis shows the effectiveness threshold $\widetilde{\alpha}$, below which no insurance is optimal and above which *full* insurance is preferred. The figure is generated using the growth-opportunities function $F = 1 - e^{-\gamma x}$, $x \geq 0$ with $\gamma = 0.5$. Other model parameters are set as follows: $\ell = 13.3$, $p = 0.2$, $\kappa = 10$ and $\rho = 0.2$. The premium is actuarially fair, $\lambda = 0$.

*utility. However, when access to external capital is prohibited, u is strictly concave by Assumption (2.3.1), and the objective function of the risk-neutral firm turns out to be of the same form as the objective function of a strictly risk-averse individual with limited liability. Hence, all results of the previous section could be applied directly. Although this a straightforward translation, it is instructive to exploit the additional structure of the firm to disentangle the impact of insurance on the various components of firm value to expose the trade-offs driving the optimal insurance decision of the firm.*

## 2.4 The optimal insurance problem beyond binary losses

In the previous sections we have focused on binary losses and shown that, under limited liability, Mossin's theorem may fail to hold. Indeed, even at actuarially fair prices, it is possible for an agent to forego insurance if limited liability provides a sufficient form of "insurance" to her. In this section we explore the optimal insurance problem under limited liability in the presence of a general loss with discrete distribution. In a first step, we consider coinsurance contracts and show that, when losses are nonbinary, the failure of Mossin's theorem is even more pronounced. More precisely, we prove that a strictly risk-averse individual with limited liability may demand *any* level of coinsurance depending on the loss. In a second step, we widen the range of contracts and look for the optimal contract design under limited liability. Under suitable assumptions on the indemnity schedule, we establish the optimality of capped deductibles.

We consider an agent with utility function $u : \mathbb{R} \to \mathbb{R}$, which is assumed to be strictly increasing and strictly concave, and is normalized to satisfy $u(0) = 0$. We denote the agent's initial wealth by $w > 0$. The agent is exposed to a loss $L$ taking $m + 1$ different values

$$0 = \ell_0 < \ell_1 < \cdots < \ell_m,$$

with corresponding probabilities $p_0, \ldots, p_m \in (0, 1)$, which add up to 1.

### 2.4.1 Optimal coinsurance level

In this section we assume that the agent can alter his or her terminal wealth through the purchase of coinsurance at level $\alpha \in [0, 1]$. The contractual indemnity is given by $\alpha L$ and the associated premium, which we assume to be actuarially fair, by $\pi(\alpha) := \alpha \mathbb{E}[L]$. The agent's expected utility under limited liability after buying this contract equals

$$V(\alpha) := \mathbb{E}[u([w - \pi(\alpha) - L + \alpha L]_+)].$$

As before, an optimal insurance level for $L$ is any $\alpha^* \in [0, 1]$ such that

$$V(\alpha^*) = \max_{\alpha \in [0,1]} V(\alpha).$$

Recall from Theorem 2.2.6 that, if the loss $L$ is binary, then it is either optimal to buy no insurance or full insurance. The situation changes drastically if $L$ is nonbinary. The following result shows that, as soon as $L$ may take three different values, every partial coinsurance level can be optimal under limited liability. This emphasizes how special binary losses are.[16]

**Proposition 2.4.1.** *For every $\alpha \in (0,1)$ there exists a loss $L$ with $m = 2$ (i.e., $L$ takes three different values) such that $\alpha$ is the unique optimal coinsurance level.*

**Remark 2.4.2.** *The previous result shows that the claim in Sinn (1982), according to which his results in a binary loss setting can be generalized to settings with arbitrary losses, is not valid. As a result, not only Mossin's theorem fails when considering agents with limited liability, but the optimal coinsurance level beyond a binary loss can take any value depending on the loss distribution. This suggests that caution is needed when extrapolating results obtained in a binary setting to situations with more general losses.*

### 2.4.2 General indemnity schedules

The remainder of the section is devoted to studying the optimal insurance problem for a loss with discrete distribution under limited liability beyond the class of coinsurance contracts. This is important because it has been known since Huberman et al. (1983) that, in the presence of limited liability, coinsurance is not optimal for nonbinary losses. Moreover, Huberman et al. (1983) claim that "an optimal policy can contain an upper limit on coverage" when agents' wealth is protected by limited liability.[17] A direct implication of this statement is that Arrow's theorem (Arrow (1974)), according to which the optimal insurance contract for a strictly risk averse agent is of deductible type, fails when the agent enjoys limited liability.

Our setting is as follows: we consider an agent may alter terminal wealth by the purchase of insurance contracts characterized by an *indemnity (schedule)*

$$I : \{\ell_0, \dots, \ell_m\} \to [0, \infty)$$

assigning to each loss level a corresponding insurance cover. We always assume that $I(\ell_i) \leq \ell_i$ for every $i \in \{0, \dots, m\}$. This precludes the possibility of overinsurance, which would give rise to moral hazard. Incidentally, note that this implies $I(\ell_0) = I(0) = 0$. The class of all indemnity schedules avoiding overinsurance is denoted by

[16] Clearly, a corresponding version of Proposition 2.4.1 also holds for a limited-liability firm.
[17] See Proposition 2, p. 417 therein.

$\mathcal{I}$. We assume that an indemnity schedule $I \in \mathcal{I}$ can be purchased at the actuarially fair premium

$$\pi(I) := \mathbb{E}[I(L)].$$

The agent is assumed to have limited liability. In this situation, the agent's expected utility from terminal wealth after buying the indemnity $I \in \mathcal{I}$ is given by

$$V(I) := \mathbb{E}[u([w - \pi(I) - L + I(L)]_+)].$$

In the sequel, given a reference class $\mathcal{J} \subset \mathcal{I}$, we look for optimal indemnity schedules $I^* \in \mathcal{J}$ satisfying

$$V(I^*) = \max_{I \in \mathcal{J}} V(I). \tag{2.12}$$

Besides the class $\mathcal{I}$ of general indemnity schedules, we focus on two special classes, namely those of increasing and admissible indemnity schedules. An indemnity schedule $I \in \mathcal{I}$ is called *increasing* if

$$0 = I(\ell_0) \leq I(\ell_1) \leq \cdots \leq I(\ell_m).$$

The class of increasing indemnity schedules is denoted by $\mathcal{I}_{inc}$. The rationale for focusing on this type of contracts is well understood. If the insurance cover is an increasing function of the loss, then the insurer eliminates the risk of underreporting from the policyholder's side, i.e., of incurring a larger loss but reporting a smaller one to obtain a larger indemnization. An increasing indemnity schedule is called *admissible* if

$$0 = \ell_0 - I(\ell_0) \leq \ell_1 - I(\ell_1) \leq \cdots \leq \ell_m - I(\ell_m).$$

The class of admissible indemnity schedules is denoted by $\mathcal{I}_{adm}$. The interpretation of the condition above is also well discussed in the literature. By forcing the uninsured loss to be increasing as a function of the underlying loss, the condition eliminates the possibility that terminal wealth is higher for higher losses, which would result in moral hazard. It is easy to see that the optimal insurance problem (2.12) always admits a solution if $\mathcal{J}$ is taken to be either $\mathcal{I}$, or $\mathcal{I}_{inc}$, or $\mathcal{I}_{adm}$. This is because every $I \in \mathcal{I}$, respectively $I \in \mathcal{I}_{inc}$ or $I \in \mathcal{I}_{adm}$, can be identified with the vector $(I(\ell_1), \ldots, I(\ell_m))$ belonging to a compact subset of $\mathbb{R}^m$ and the objective function appearing in the problem is continuous.

It is immediate to see that coinsurance contracts are characterized by admissible indemnity schedules. The corresponding indemnity satisfies

$$I_\alpha(L) := \alpha L$$

for some $\alpha \in [0, 1]$. Another contract with admissible indemnity schedule are capped deductibles, which correspond to indemnities of the form

$$I_{d,c}(L) := (L - d)_+ - (L - c)_+$$

for $0 \leq d \leq c \leq \ell_m$. We refer to $d$ as the retention (or deductible) level and to $c$ as the cap level, or upper limit. Note that a capped deductible with zero deductible level coincides with an upper limit as

$$I_{0,c}(L) = L - (L - c)_+ = \min\{L, c\}.$$

Capped deductibles will play an important role in the sequel.

### 2.4.3 Preliminary considerations

In this section we collect some results and examples that help put our later results on optimal indemnity schedules into perspective. For our later analysis, it is useful to start defining for every indemnity $I \in \mathcal{I}$ the sets

$$D(I) := \{i \in \{0, \ldots, m\}\,;\ w - \mathbb{E}[I(L)] - \ell_i + I(\ell_i) < 0\},$$

$$S(I) := \{i \in \{0, \ldots, m\}\,;\ w - \mathbb{E}[I(L)] - \ell_i + I(\ell_i) > 0\}.$$

These sets capture the terminal states, through the corresponding indices, where wealth after buying insurance is strictly negative and strictly positive, respectively. In particular, the agent will benefit from limited liability in the "default" states belonging to $D(I)$ whereas he or she will have no need to receive that protection in the "surplus" states belonging to $S(I)$. The next result shows that, as soon as full insurance is not optimal, the presence of limited liability will always make acceptable for the agent to default in some states.

**Proposition 2.4.3.** *Let $I$ be optimal in $\mathcal{I}$. If full insurance is not optimal in $\mathcal{I}$, then $D(I) \neq \emptyset$. The same holds by replacing $\mathcal{I}$ with $\mathcal{I}_{inc}$ or $\mathcal{I}_{adm}$.*

As a first step in our study of the optimal insurance problem under general indemnity schedules, we focus on the situation where the underlying loss is binary. In this case, every indemnity can be trivially expressed as the indemnity schedule of a coinsurance contract (or equivalently of a capped deductible with maximum cap). Hence, the distinction among general, increasing, and admissible indemnity schedules is superfluous and it is optimal to either fully insure or not insure at all. This is consistent with Theorem 2.2.6 above. We provide a more direct proof based on Proposition 2.4.3.

**Theorem 2.4.4.** *If $m = 1$, then optimizing over $\mathcal{I}$, $\mathcal{I}_{inc}$, or $\mathcal{I}_{adm}$ is equivalent and the optimal indemnity schedule is either $I_0$ (no insurance) or $I_1$ (full insurance).*

As a next step, we focus on losses taking three values, or $m = 2$. The key observation in this case is recorded in the following simple but useful proposition showing that every admissible indemnity schedule can be expressed as the indemnity of a capped deductible.

**Proposition 2.4.5.** *Let $m = 2$. For every $I \in \mathcal{I}_{adm}$ there exist $0 \leq d \leq c \leq \ell_2$ such that $I = I_{d,c}$.*

The next result shows that, when the loss takes three values, the optimal admissible contract is a contract delivering full insurance or a capped deductible with zero deductible level and cap level that does not exceed the lower (strictly positive) loss.

**Theorem 2.4.6.** *Let $m = 2$ and let $I$ be an optimal indemnity schedule in $\mathcal{I}_{adm}$. Then, either $I = I_{0,c}$ for some $c \in [0, \ell_1]$ or $I = I_{0,\ell_2}$.*

The next example shows that, in a world with three states, optimizing over general indemnity schedules might deliver a strictly larger expected utility than optimizing over admissible ones. In addition, it complements Theorem 2.4.6 by showing that the optimal admissible indemnity may be given by a capped deductible with zero deductible level that provides only partial insurance against the lower loss.

**Example 2.4.7.** *Assume $L$ takes the values $(0, 1, 3)$ with equal probability $\frac{1}{3}$. Let $w = 2$ and $u(x) = 1 - e^{-x}$ for $x \in \mathbb{R}$. By Theorem 2.4.6, an optimal indemnity schedule $I \in \mathcal{I}_{adm}$ satisfies either $I = I_{0,3}$, in which case full insurance is optimal, or $I = I_{0,c}$ for some $c \in [0, 1]$. We claim that the optimal contract in $\mathcal{I}_{adm}$ is given by $I = I_{0,c}$ for some $c \in (0, 1)$. To see this, note first that*

$$V(I_{0,1}) = \frac{2}{3}u\left(\frac{4}{3}\right) \approx 0.49 > 0.48 \approx u\left(\frac{2}{3}\right) = V(I_{0,3}),$$

*showing that full insurance is not optimal. As a next step, define for $c \in [0, 1]$*

$$f(c) := V(I_{0,c}) = \frac{1}{3}u\left(2 - \frac{2}{3}c\right) + \frac{1}{3}u\left(1 + \frac{1}{3}\right).$$

*The left, respectively right, derivative of $f$ at $c = 0$, respectively $c = 1$, satisfies*

$$f'(0) = \frac{1}{9}u'(1) - \frac{2}{9}u'(2) \approx 0.01 > 0,$$

$$f'(1) = -\frac{1}{9}u'\left(\frac{4}{3}\right) \approx -0.03 < 0.$$

*This shows that neither $I_{0,0}$ nor $I_{0,1}$ are optimal. As a result, the optimal contract in $\mathcal{I}_{adm}$ is given by $I = I_{0,c}$ for some $c \in (0, 1)$ as claimed. We conclude by showing that such a contract is not optimal in $\mathcal{I}$. Indeed, define $J_c \in \mathcal{I} \backslash \mathcal{I}_{adm}$ by $J_c(0) := J_c(3) := 0$ and $J_c(1) := c$. Then*

$$V(J_c) = \frac{1}{3}u\left(2 - \frac{1}{3}c\right) + \frac{1}{3}u\left(1 + \frac{2}{3}c\right) > \frac{1}{3}u\left(2 - \frac{2}{3}c\right) + \frac{1}{3}u\left(1 + \frac{1}{3}c\right) = V(I_{0,c}).$$

### 2.4.4 Optimal insurance with general indemnity schedules

In this section we study the optimal insurance problem for general, increasing, and admissible indemnity schedules without imposing any condition on the discrete distribution of the underlying loss. We start by focusing on general indemnity schedules. In this case, we show that it is always optimal to fully insure some losses and leave the remaining losses uninsured.

**Theorem 2.4.8.** *Let $I \in \mathcal{I}$ be optimal. Then, there exists $A \subset \{1, \ldots, m\}$ such that $I(\ell_i) = \ell_i$ if $i \in A$ and $I(\ell_i) = 0$ if $i \notin A$.*

We collect a few observations on the previous result for general indemnity schedules in the next remark.

**Remark 2.4.9.** *We know from Example 2.4.7 that an optimal indemnity schedule in $\mathcal{I}$ need not be increasing. The previous theorem therefore implies that, in the pool of general indemnity schedules, it may be optimal to insure lower losses and leave higher losses completely uninsured. It follows from Example 2.4.7 that it may also be optimal to insure higher losses and leave lower losses uninsured.*

Next, we turn our attention to the class of increasing indemnity schedules. Before presenting the optimality result for this class, we make the following observation.

**Proposition 2.4.10.** *Let $I \in \mathcal{I}_{inc}$ be optimal. For every $i \in \{1, \ldots, m\}$ with $i \notin S(I)$ we have $I(\ell_{i-1}) = I(\ell_i) < \ell_i$.*

The next result shows that, within the class of increasing indemnity schedules, if the optimal contract provides insurance coverage to the insured, then the indemnity will be capped beyond some loss belonging to the "surplus" states in $S(I)$.

**Theorem 2.4.11.** *Let $I \in \mathcal{I}_{inc}$ be optimal. Then, either $I$ delivers no insurance or there exists $i \in \{1, \ldots, m\}$ such that $I(\ell_{i-1}) < I(\ell_i) = \cdots = I(\ell_m)$. Moreover, $i \in S(I)$.*

Finally, we address the optimal insurance problem with admissible indemnizations. We collect some properties of admissible schedules in the following proposition.

**Proposition 2.4.12.** *For every $I \in \mathcal{I}_{adm}$ the following statements hold:*

*(i) For every $i \in \{1, \ldots, m\}$ with $i \in S(I)$ we have $i - 1 \in S(I)$.*

*(ii) For every $i \in \{0, \ldots, m-1\}$ with $i \in D(I)$ we have $i + 1 \in D(I)$.*

*(iii) If $I$ is optimal in $\mathcal{I}_{adm}$ and neither full insurance nor no insurance is optimal in $\mathcal{I}_{adm}$, then there exist $1 \le h < k \le m$ such that $S(I) = \{0, \ldots, h\}$ and $D(I) = \{k, \ldots, m\}$. Moreover, $I(\ell_h) = \cdots = I(\ell_m)$.*

*(iv) If $w \geq \ell_m$, then full insurance is optimal in $\mathcal{I}_{adm}$.*

The next theorem, which is the key contribution of the paper, shows that an optimal admissible indemnity schedule is always a capped deductible.

**Theorem 2.4.13.** *Let $I \in \mathcal{I}_{adm}$ be optimal. Then, $I = I_{d,c}$ for some $0 \leq d \leq c \leq \ell_m$.*

The previous theorem derives a new type of optimal contracts, showing that Arrow's theorem fails under limited liability: any strictly risk averse agent will not prefer full insurance above a deductible, but the optimal insurance contract will contain a limit on coverage. The absence of limited liability is, however, only one of the assumptions under which Arrow's theorem was derived. The optimality of deductible contracts was established within the expected utility paradigm, with the premium principle based on the actuarial (or expected) value of the indemnity and assuming the insured and insurer share the same beliefs about the random loss.

Since its original formulation, the literature has extended Arrow's theorem in many directions.[18] For instance, the optimal indemnity schedule has been studied in settings beyond the expected utility framework (see e.g., Gollier and Schlesinger (1996), Bernard et al. (2015)). Furthermore, background risk was introduced in Dana and Scarsini (2007), Ping and Zanjani (2015) and Hofmann et al. (2019); alternative premium principles have been investigated in Carlier and Dana (2003) and Promislow and Young (2005) and heterogeneous beliefs between the parties were analyzed in Ghossoub (2017).

The result is that the optimal indemnity schedule varies considerably across models, ranging from contracts with a variable state-contingent deductible schedule under subjective beliefs (Ghossoub (2017)) to indemnities with upper limits on coverage when the *insurer*'s background risk is negatively correlated with the *insured* loss (Ping and Zanjani (2015)). Our result contributes to this literature by showing how Arrow's theorem changes for the case of a strictly risk averse agent with limited liability when available insurance contracts feature admissible indemnity schedules.

[18] For a recent and broad overview of this literature see Ghossoub (2017).

## 2.5 Conclusion

Understanding the insurance demand by individuals or firms exposed to liability claims is important, because whether or not they insure is a matter of public concern. Indeed, liability insurance protects not only the insurance taker but also the damaged party in case the insurance taker does not have sufficient wealth to fully meet obligations arising from a liability case. Within the context of expected utility and in a discrete loss setting, we have provided a complete account of the demand for insurance by strictly risk-averse agents and risk-neutral firms when they enjoy limited liability. When exposed to a bankrupting, binary loss and under actuarially fair prices, individuals and firms will either fully insure or not insure at all. The decision to insure will depend on whether the benefits the insuree derives from insurance after having compensated the damaged party are sufficiently attractive to justify the premium paid. In the case of firms, we provide a full analysis of the trade-off between the positive impact of risk on the value of the option to default and the negative impact on the firm's franchise value. When the loss is nonbinary, even when prices are actuarially fair, any amount of coinsurance can be optimal depending on the nature of the loss. In particular, this is a strong reminder of how special binary models are and calls for caution when extrapolating from these models to more general situations.

## 2.6 Appendix

This appendix collects the proofs of all results in the paper that are not immediate. Hence, we omit the proofs of Proposition 2.2.2, Proposition 2.2.3, Proposition 2.2.4 and Theorem 2.3.6.

### 2.6.1 Proof of Theorem 2.2.6

To prove Theorem 2.2.6 we require the following auxiliary result.

**Lemma 2.6.1.** *Let $p\ell \le w < \ell$. The function $\lambda \mapsto \beta^*_\lambda$ is continuous on $[0, \frac{w-p\ell}{p\ell}]$ and $\lambda \mapsto V_\lambda(\beta^*_\lambda)$ is continuous and strictly increasing on $[0, \frac{w-p\ell}{p\ell}]$. Moreover:*

*(i)* $\beta^*_0 = \beta^*_{\frac{w-p\ell}{p\ell}} = 1$.

*(ii)* $\beta^*_\lambda < 1$ *for* $0 < \lambda < \frac{w-p\ell}{p\ell}$.

*Proof.* Continuity follows from general theorems on the continuity of solutions of parametric optimization problems but we provide a direct proof for convenience. Let $(\lambda_n)$ be a sequence in $[0, \frac{w-p\ell}{p\ell}]$ such that $\lambda_n \to \lambda_0$. We have to show that $\beta^*_{\lambda_n} \to \beta^*_{\lambda_0}$. Note first that, since $(\beta^*_{\lambda_n})$ is a sequence in the compact interval $[0,1]$, we find a $\beta^* \in [0,1]$ such that $\beta^*_{\lambda_n} \to \beta^*$ (if necessary we can pass to a convenient subsequence). Observe that $\widetilde{\alpha}_{\lambda_n} \leq \beta^*_{\lambda_n} \leq 1$ for every $n$ implies that $\widetilde{\alpha}_{\lambda_0} \leq \beta^* \leq 1$. Moreover, since $(\lambda, \alpha) \mapsto V_\lambda(\alpha)$ is jointly continuous, we obtain that $V_{\lambda_n}(\beta^*_{\lambda_n}) \to V_{\lambda_0}(\beta^*)$. Assume that $\beta^* \neq \beta^*_{\lambda_0}$ so that $V_{\lambda_0}(\beta^*) < V_{\lambda_0}(\beta^*_{\lambda_0})$ by optimality of $\beta^*_{\lambda_0}$. It follows that for sufficiently large $n$ we have $V_{\lambda_n}(\beta^*_{\lambda_n}) < V_{\lambda_n}(\beta^*_{\lambda_0})$, contradicting the optimality of $\beta^*_{\lambda_n}$. We infer that $\beta^* = \beta^*_{\lambda_0}$ must hold. The fact that $\lambda \mapsto V_\lambda(\beta^*_\lambda)$ is continuous follows from the joint continuity of $(\lambda, \alpha) \mapsto V_\lambda(\alpha)$ and the continuity of $\lambda \mapsto \beta^*_\lambda$ we have just established. Monotonicity of $\lambda \mapsto V_\lambda(\beta^*_\lambda)$ is clear.

To prove (i), note first that $U_0$ attains its maximum over $[0,1]$ at $\alpha = 1$. Hence, the maximum of $V_0$ over $[\widetilde{\alpha}_0, 1]$ is also attained at $\alpha = 1$, i.e., $\beta^*_0 = 1$. Moreover, since $\widetilde{\alpha}_{\frac{w-p\ell}{p\ell}} = 1$, we have $\beta^*_{\frac{w-p\ell}{p\ell}} = 1$.

To prove (ii), let $0 < \lambda < \frac{w-p\ell}{p\ell}$. In this case, $U_0$ is decreasing in a neighborhood of $\alpha = 1$ and, hence, so is $V_\lambda$. It follows that $\alpha = 1$ cannot be a maximum, i.e., $\beta^*_\lambda < 1$. □

We now proceed with the proof of Theorem 2.2.6. If $V_0(1) < V_0(0)$, then Lemma 2.6.1 implies that $\alpha = 0$ is the only solution for $\lambda = 0$ and, hence, also for $\lambda > 0$. This proves (i). If $V_0(1) = V_0(0)$, then Lemma 2.6.1 implies that $\alpha = 0$ and $\alpha = 1$ are optimal solutions for $\lambda = 0$. Moreover, for $\lambda > 0$ we have $V_\lambda(\beta^*_\lambda) < V_0(1) = V_0(0)$, implying that $\alpha = 0$ is the optimal solution in this case. This establishes (ii). To conclude, observe that the function $\lambda \mapsto V_\lambda(\beta^*_\lambda)$ is continuous and strictly increasing by Lemma 2.6.1. Moreover, $V_0(\beta^*_0) = V_0(1) > V_0(0)$ and $V_{\frac{w-p\ell}{p\ell}}(\beta^*_{\frac{w-p\ell}{p\ell}}) = V_{\frac{w-p\ell}{p\ell}}(1) = 0 < V_0(0)$. It follows that there exists $0 < \tilde{\lambda} < \frac{w-p\ell}{p\ell}$ such that (a)-(c) are satisfied.

### 2.6.2 Proof of Proposition 2.2.8

Since $u$ is strictly concave and $u(0) = 0$, we have

$$(1-p)u(\ell) = (1-p)u(\ell) + pu(0) < u(\ell - p\ell).$$

First, recall that no insurance is optimal if $w < p\ell$ by Proposition 2.2.3 while full insurance is optimal if $w \geq \ell$ by Proposition 2.2.2. Now, let $p\ell \leq w < \ell$ and note that $V_0(0) = (1-p)u(w)$ and $V_0(1) = u(w - p\ell)$. For $w$ close to $\ell$ we have $(1-p)u(w) < u(w-p\ell)$. In this case, full insurance is optimal by Theorem 2.2.6. Moreover, as $w$ converges to $p\ell$ from above, $(1-p)u(w)$ converges to $(1-p)u(p\ell) > 0$ and $u(w-p\ell)$ converges to 0. As a result, $(1-p)u(w) > u(w-p\ell)$ for $w$ close to $p\ell$. In this case, no insurance is optimal again by Theorem 2.2.6. Since $u(w-p\ell)$ is increasing in $w$, there exists exactly one level $\widetilde{w}$ satisfying (i).

Similarly, note that full insurance is optimal if $\ell \leq w$ by Proposition 2.2.2 while no insurance is optimal if $\ell > \frac{w}{p}$ by Proposition 2.2.3. Now, let $w < \ell \leq \frac{w}{p}$ and note that $V_0(0) = (1-p)u(w)$ and $V_0(1) = u(w-p\ell)$. For $\ell$ close to $w$ we have $(1-p)u(w) < u(w-p\ell)$. In this case, full insurance is optimal by Theorem 2.2.6. Moreover, as $\ell$ converges to $\frac{w}{p}$ from below, $u(w-p\ell)$ converges to 0. As a result, $(1-p)u(w) > u(w-p\ell)$ for $w$ close to $p\ell$. In this case, no insurance is optimal again by Theorem 2.2.6. Since $u(w-p\ell)$ is decreasing in $\ell$, there exists exactly one level $\widetilde{\ell}$ satisfying (ii).

### 2.6.3 Proof of Lemma 2.3.2

Note first that $DO(\alpha) = \mathbb{E}[W^-(\alpha)] = p[\kappa(1+\rho) - \ell + (1-p)\alpha\ell]_-$ for every $\alpha \in [0,1]$.

*(i)* If $\kappa(1+\rho) \geq \ell$ and $\alpha \in [0,1]$, then $\kappa(1+\rho) - \ell + (1-p)\alpha\ell \geq 0$ and, thus, $DO(\alpha) = 0$.

*(ii)* If $\kappa(1+\rho) < \ell$ and $\kappa(1+\rho) > p\ell$, then clearly $\frac{\ell-\kappa(1+\rho)}{(1-p)\ell} \in (0,1)$ and

$$DO(\alpha) = \begin{cases} -p[\kappa(1+\rho) - \ell + (1-p)\alpha\ell] & \text{if } \alpha \in [0, \frac{\ell-\kappa(1+\rho)}{(1-p)\ell}], \\ 0 & \text{if } \alpha \in [\frac{\ell-\kappa(1+\rho)}{(1-p)\ell}, 1]. \end{cases}$$

In particular, $DO$ is strictly decreasing on $[0, \frac{\ell-\kappa(1+\rho)}{(1-p)\ell}]$.

*(iii)* If $\kappa(1+\rho) < \ell$ and $\kappa(1+\rho) \leq p\ell$, then $DO(\alpha) = -p[\kappa(1+\rho) - \ell + (1-p)\alpha\ell]$ for every $\alpha \in [0,1]$. In particular, $DO$ is strictly decreasing on $[0,1]$.

### 2.6.4 Proof of Proposition 2.3.3

If $\kappa(1+\rho) \geq \ell$ and $\alpha \in [0,1]$, then $\kappa(1+\rho) - \ell + (1-p)\alpha\ell \geq 0$ and, hence,

$$FV(\alpha) = pF(\kappa(1+\rho) - \ell + (1-p)\alpha\ell) + (1-p)F(\kappa(1+\rho) - \alpha p\ell).$$

Denote the left derivative of $FV$ at $\alpha = 1$ by $FV'(1)$ and note that

$$FV'(1) = p(1-p)\ell F'(\kappa(1+\rho) - p\ell) - (1-p)p\ell F'(\kappa(1+\rho) - p\ell) = 0.$$

Together with the strict concavity of $FV$, which follows from that of $F$, this implies that $FV$ is strictly increasing on $[0,1]$.

### 2.6.5 Proof of Proposition 2.3.4

By assumption, for every $0 < \alpha \leq 1$, we have $W^+(\alpha) = 0$ in the loss state. As a result, for every $0 < \alpha \leq 1$

$$FV(\alpha) = (1-p)F(\kappa(1+\rho) - \alpha p\ell).$$

Hence, $FV$ is clearly strictly decreasing on $[0,1]$. Since, by Lemma 2.3.2, $DO$ is also strictly decreasing on $[0,1]$, we immediately infer that $\alpha = 0$ is the unique optimal coinsurance level.

### 2.6.6 Proof of Lemma 2.3.5

By definition of $\widetilde{\alpha}$, we have

$$FV(\alpha) = \begin{cases} (1-p)F(\kappa(1+\rho) - \alpha p\ell) & \text{if } 0 \leq \alpha \leq \widetilde{\alpha}, \\ pF(\kappa(1+\rho) - \ell + (1-p)\alpha\ell) + & \\ \qquad\qquad +(1-p)F(\kappa(1+\rho) - \alpha p\ell) & \text{otherwise.} \end{cases}$$

It is clear that $FV$ is strictly decreasing on $[0,\widetilde{\alpha}]$. The proof that $FV$ is strictly increasing on $[\widetilde{\alpha},1]$ is analogous to the proof of Proposition 2.3.3.

### 2.6.7 Proof of Proposition 2.3.7

It follows from Proposition 2.3.4 that no insurance is optimal if $\kappa < \frac{p\ell}{1+\rho}$ or equivalently $\ell > \frac{\kappa(1+\rho)}{p}$. We also know from Theorem 2.3.6 that full insurance is optimal if $\kappa \geq \frac{\ell}{1+\rho}$ or equivalently $\ell \leq \kappa(1+\rho)$. Now, assume that $\frac{p\ell}{1+\rho} \leq \kappa < \frac{\ell}{1+\rho}$. In this case,

$$FV(1) - FV(0) = F(\kappa(1+\rho) - p\ell) + (1-p)F(\kappa(1+\rho)), \quad DO(0) = p(\ell - \kappa(1+\rho)).$$

If $\kappa$ tends to $\frac{p\ell}{1+\rho}$ from above, then $DO(0)$ converges to $p(1-p)\ell > 0$ while $FV(1) - FV(0)$ converges to $-(1-p)F(p\ell) < 0$. Hence, Theorem 2.3.6 implies that no insurance is optimal. Similarly, if $\kappa$ tends to $\frac{\ell}{1+\rho}$ from below, then $DO(0)$ converges to 0 while $FV(1) - FV(0)$ converges to $F((1-p)\ell) - (1-p)F(\ell)$. Observe that

$$F((1-p)\ell) - (1-p)F(\ell) > pF(0) + (1-p)F(\ell) - (1-p)F(\ell) = 0$$

by strict concavity of $F$. Hence, Theorem 2.3.6 implies that full insurance is optimal. Since, on the relevant interval, $DO(0)$ is a strictly-decreasing affine function of $\kappa$ and $FV(1) - FV(0)$ is a strictly-concave function of $\kappa$, we infer the existence of a unique point $\widetilde{\kappa}$ satisfying (i).

If $\ell$ tends to $\frac{\kappa(1+\rho)}{p}$ from below, then $DO(0)$ converges to $(1-p)\kappa(1+\rho) > 0$ while $FV(1) - FV(0)$ converges to $-(1-p)F(\kappa(1+\rho)) < 0$. Hence, Theorem 2.3.6 implies that no insurance is optimal. Similarly, if $\ell$ tends to $\kappa(1+\rho)$ from above, then $DO(0)$ converges to 0 while $FV(1)-FV(0)$ converges to $F((1-p)\kappa(1+\rho))-(1-p)F(\kappa(1+\rho))$. Observe that

$$\begin{aligned} &F((1-p)\kappa(1+\rho)) - (1-p)F(\kappa(1+\rho)) \\ &> pF(0) + (1-p)F(\kappa(1+\rho)) - (1-p)F(\kappa(1+\rho)) = 0 \end{aligned}$$

by strict concavity of $F$. Hence, Theorem 2.3.6 implies that full insurance is optimal. Since, on the relevant interval, $DO(0)$ is a strictly-increasing affine function of $\ell$ and $FV(1) - FV(0)$ is a strictly-concave function of $\ell$, we infer the existence of a unique point $\widetilde{\ell}$ satisfying (ii).

### 2.6.8 Proof of Proposition 2.4.1

Consider a loss $L$ taking three different values. Such a loss can always be described as

$$L = \begin{cases} 0 & \text{with probability } (1-p_2)(1-p_1) \\ \ell_1 & \text{with probability } (1-p_2)p_1 \\ \ell_2 & \text{with probability } p_2 \end{cases}$$

with $0 < \ell_1 < \ell_2$ and $p_1, p_2 \in (0,1)$. We fix $p_1$ and $\ell_1$ and view $p_2$ and $\ell_2$ as free parameters. Clearly, at least $\ell_2$ must be bankrupting since, otherwise, full insurance would be optimal by Theorem 2.2.1. We assume that only $\ell_2$ is bankrupting and that initial wealth is sufficiently large to buy full insurance. More explicitly, we assume that:

(A1) $0 < \ell_1 < w < \ell_2$.

(A2) $w > \mathbb{E}[L]$.

Similar to Section 2.2, we define the effectiveness threshold as

$$\widetilde{\alpha} := \inf\{\alpha \in [0,1]\,;\ w - \ell_2 - \alpha\mathbb{E}[L] + \alpha\ell_2 \geq 0\} = \frac{\ell_2 - w}{\ell_2 - \mathbb{E}[L]} \in (0,1).$$

We maximize $V$ on $[0, \widetilde{\alpha}]$ and $[\widetilde{\alpha}, 1]$ separately and obtain the solution to the global maximization over $[0,1]$ by comparing the respective solutions to these partial maximization problems. We start by maximizing $V$ over $[\widetilde{\alpha}, 1]$. As a first simple observation, note that $V$ coincides with the expected utility of the agent with unlimited liability on the interval $[\widetilde{\alpha}, 1]$. Hence, we know from Theorem 2.2.1 that the maximum of $V$ over $[\widetilde{\alpha}, 1]$ is attained at $\alpha = 1$. We proceed by maximizing $V$ over $[0, \widetilde{\alpha}]$. Denote by $\widetilde{L}$ a binary loss with loss size $\ell_1$ and loss probability $p_1$. For a contract $\alpha \in [0, \widetilde{\alpha}]$, wealth is always positive when $L = 0$ or $L = \ell_1$, but negative when $L = \ell_2$. Hence, as is easily seen,

$$V(\alpha) = (1-p_2)\mathbb{E}[u(w - \widetilde{L} - \alpha\mathbb{E}[L] + \alpha\widetilde{L})]. \tag{2.13}$$

Note that

$$\mathbb{E}[L] = (1-p_2)\mathbb{E}[\widetilde{L}] + p_2\ell_2 = (1 + \lambda(p_2, \ell_2))\mathbb{E}[\widetilde{L}], \tag{2.14}$$

where we have set

$$\lambda(p_2,\ell_2) := p_2\frac{\ell_2-\mathbb{E}[\widetilde{L}]}{\mathbb{E}[\widetilde{L}]} > 0.$$

Hence, we can rewrite (2.13) as

$$V(\alpha) = (1-p_2)\mathbb{E}[u(w-\widetilde{L}-\alpha(1+\lambda(p_2,\ell_2))\mathbb{E}[\widetilde{L}]+\alpha\widetilde{L})]. \tag{2.15}$$

It follows that, on the interval $[0,\widetilde{\alpha}]$, the function $V$ coincides up to multiplication by $1-p_2$ with the expected utility of an agent with utility function $u$, initial wealth $w$, exposed to the loss $\widetilde{L}$, and buying coinsurance priced with a loading $\lambda(p_2,\ell_2)$.

As a next step, we choose $p_2$ and $\ell_2$. Denote by $V_\lambda$ the expected utility for the strictly risk-averse agent with utility function $u$, initial wealth $w$, exposed to the loss $\widetilde{L}$, and buying coinsurance with a loading $\lambda$. Since $\ell_1$ is nonbankrupting, we know from Mossin's theorem (Theorem 2.2.1) that there exists a unique optimal coinsurance level $\alpha^*_\lambda$ for this agent. We now fix $\alpha^*\in(0,1)$ and denote by $\lambda>0$ the unique loading such that $\alpha^*_\lambda=\alpha^*$, which exists again by Theorem 2.2.1. Note that

$$w > (1+\lambda)p\ell \tag{2.16}$$

since otherwise we would have $\alpha^*=\alpha^*_\lambda=0$, which is not possible by assumption. Indeed, if $w\leq(1+\lambda)p\ell$, the function $\alpha\mapsto w-\ell_1-(1+\lambda)\mathbb{E}[\widetilde{L}]+\alpha\ell_1$, which corresponds to wealth in the loss state, would be decreasing implying that $\alpha^*_\lambda=0$. For every $\ell_2>\ell_1$ we set

$$p_2(\ell_2) := \frac{\lambda\mathbb{E}[\widetilde{L}]}{\ell_2-\mathbb{E}[\widetilde{L}]}\in(0,1)$$

and note that

$$\lambda(p_2(\ell_2),\ell_2)=\lambda. \tag{2.17}$$

Moreover, $p_2(\ell_2)\to 0$ as $\ell_2\to\infty$. From now on whenever we choose $\ell_2>\ell_1$ we will always implicitly assume that we choose $p_2=p_2(\ell_2)$. Note that, for every choice of $\ell_2>\ell_1$, we have by (2.14) that

$$\mathbb{E}[L]=(1+\lambda)\mathbb{E}[\widetilde{L}].$$

Hence, (2.16) implies that assumption (A2) is satisfied for every choice of $\ell_2>\ell_1$.

It remains to choose $\ell_2$ so that $\alpha^*$ is the global maximum of $V$. Choose $\ell_2>\ell_1$

sufficiently large so that

$$w - \ell_2 - \alpha^* \mathbb{E}[L] + \alpha^* \ell_2 < 0.$$

In particular, by definition of $\widetilde{\alpha}$, we have $0 < \alpha^* < \widetilde{\alpha}$. Using (2.15), it follows that the maximum of $V$ over $[0, \widetilde{\alpha}]$ is attained at $\alpha^*$. As a last step, we may enlarge $\ell_2 > \ell_1$, if necessary, to ensure that $\alpha^*$ is also the global maximum of $V$. To this effect, recall that $\alpha^* = \alpha^*_\lambda$ and note that

$$V_\lambda(\alpha^*) > V_\lambda(1) = u(w - (1+\lambda)\mathbb{E}[\widetilde{L}]) = u(w - \mathbb{E}[L]) = V(1)$$

by optimality of $\alpha^*_\lambda$ and by (2.14). Since $p_2(\ell_2) \to 0$ as $\ell_2 \to \infty$, we find $\ell_2$ large enough such that

$$V(\alpha^*) = (1 - p_2(\ell_2))V_\lambda(\alpha^*) > V(1)$$

by (2.15). For such $\ell_2$, the global maximum of $V$ is attained at $\alpha^*$ as desired.

### 2.6.9 Proof of Proposition 2.4.3

Assume that $D(I) = \emptyset$. In this case, limited liability is superfluous and we get

$$\begin{aligned} V(I) &= \mathbb{E}[u(w - \mathbb{E}[I(L)] - L + I(L))] \\ &\le u(\mathbb{E}[w - \mathbb{E}[I(L)] - L + I(L)]) = u(w - \mathbb{E}[L]) \end{aligned}$$

by Jensen's inequality. This would imply that full insurance is optimal, contradicting the assumption. As a result, $D(I) \neq \emptyset$ must hold.

### 2.6.10 Proof of Theorem 2.4.4

Let $I \in \mathcal{I}$ be optimal. It follows from Proposition 2.4.5 that $I = I_\alpha$ for some $\alpha \in [0,1]$. In particular, $I$ is admissible. If full insurance is not optimal, then $D(I)$ must be nonempty by Proposition 2.4.3. This implies that $1 \in D(I)$, whence

$$V(I) = p_0 u(w - \alpha \mathbb{E}[L]) \le p_0 u(w) \le V(I_0),$$

with strict inequality unless $\alpha = 0$. This shows that $I = I_0$ and concludes the proof.

### 2.6.11 Proof of Proposition 2.4.5

Take an arbitrary $I \in \mathcal{I}_{adm}$. A direct computation shows that $I = I_{d,c}$ for $d = \ell_1 - I(\ell_1) \in [0, \ell_1]$ and $c = I(\ell_2) + d \in [d, \ell_2]$. In particular, observe that $c \leq I(\ell_2) + \ell_1 - I(\ell_1) \leq \ell_2$ follows from admissibility of $I$.

### 2.6.12 Proof of Theorem 2.4.6

Let $I \in \mathcal{I}_{adm}$ be optimal. It follows from Proposition 2.4.5 that $I = I_{d,c}$ for some $0 \leq d \leq c \leq \ell_2$. If full insurance is not optimal, then $D(I)$ must be nonempty by Proposition 2.4.3. This implies that $2 \in D(I)$. First, suppose that $1 \notin S(I)$. In this case,

$$V(I) = p_0 u(w - \mathbb{E}[I(L)]) \leq p_0 u(w) \leq V(I_{0,0}),$$

with strict inequality unless $\mathbb{E}[I(L)] = 0$. This shows that $I = I_{0,0}$. Second, suppose that $1 \in S(I)$. We claim that $I(\ell_1) = I(\ell_2)$. If we otherwise have $I(\ell_1) < I(\ell_2)$, then take $\varepsilon \in (0, I(\ell_2) - I(\ell_1))$ and define an indemnity schedule $J \in \mathcal{I}_{adm}$ by setting

$$J(\ell_0) = 0, \quad J(\ell_1) = I(\ell_1), \quad J(\ell_2) = I(\ell_2) - \varepsilon.$$

Note that $\mathbb{E}[J(L)] < \mathbb{E}[I(L)]$. Moreover, $S(J) = S(I)$ provided $\varepsilon$ is small enough. As a result,

$$\begin{aligned} V(I) &= p_0 u(w - \mathbb{E}[I(L)]) + p_1 u(w - \mathbb{E}[I(L)] - \ell_1 + I(\ell_1)) \\ &< p_0 u(w - \mathbb{E}[J(L)]) + p_1 u(w - \mathbb{E}[J(L)] - \ell_1 + J(\ell_1)) = V(J). \end{aligned}$$

This contradicts the optimality of $I$ and shows that $I(\ell_1) = I(\ell_2)$ must hold. As a consequence, $I = I_{0,c}$ for $c = I(\ell_1)$. It remains to observe that $c = I(\ell_1) \leq \ell_1$.

### 2.6.13 Proof of Theorem 2.4.8

We prove the statement for $A = D(I)^c$. Fix $j \in \{1, \ldots, m\}$. First, assume that $j \in D(I)$. If $I(\ell_j) > 0$, then we define $J \in \mathcal{I}$ by setting for $i \in \{0, \ldots, m\}$

$$J(\ell_i) = \begin{cases} I(\ell_i) & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases}$$

Note that $\mathbb{E}[J(L)] < \mathbb{E}[I(L)]$ and $S(I) \subset S(J)$. As a result,

$$\begin{aligned} V(I) &= \sum_{i \in S(I)} p_i u(w - \mathbb{E}[I(L)] - \ell_i + I(\ell_i)) \\ &= \sum_{i \in S(I)} p_i u(w - \mathbb{E}[I(L)] - \ell_i + J(\ell_i)) \\ &< \sum_{i \in S(I)} p_i u(w - \mathbb{E}[J(L)] - \ell_i + J(\ell_i)) \\ &\leq \sum_{i \in S(J)} p_i u(w - \mathbb{E}[J(L)] - \ell_i + J(\ell_i)) = V(J), \end{aligned}$$

which is impossible, proving that $I(\ell_j) = 0$ must hold. Next, assume that $j \notin D(I)$. We claim that $I(\ell_j) = \ell_j$. If $I(\ell_j) < \ell_j$, then we define $J_\varepsilon \in \mathcal{I}$ by setting for $i \in \{0, \ldots, m\}$

$$J_\varepsilon(\ell_i) = \begin{cases} I(\ell_i) & \text{if } i \neq j, \\ I(\ell_j) + \varepsilon & \text{if } i = j, \end{cases}$$

where $\varepsilon \in [0, \overline{\varepsilon})$ for a suitable $0 < \overline{\varepsilon} < \ell_j - I(\ell_j)$. Note that $J_0 = I$. Note also that we may take $\overline{\varepsilon}$ small enough to have $D(J_\varepsilon) = D(I)$ for every $\varepsilon \in (0, \overline{\varepsilon})$. Define a function $f : [0, \overline{\varepsilon}) \to \mathbb{R}$ by

$$f(\varepsilon) := V(J_\varepsilon) = \sum_{i \in D(I)^c} p_i u(w - \mathbb{E}[J_\varepsilon(L)] - \ell_i + J_\varepsilon(\ell_i)).$$

Denote by $f'(0)$ the right derivative of $f$ at $\varepsilon = 0$. Then, setting $q := \sum_{i \in D(I)^c} p_i$, we obtain

$$\begin{aligned} f'(0) = &\sum_{i \in D(I)^c,\ i \neq j} p_i u'(w - \mathbb{E}[I(L)] - \ell_i + I(\ell_i))(-p_j) + \\ &\qquad + p_j u'(w - \mathbb{E}[I(L)] - \ell_j + I(\ell_j))(1 - p_j) \\ > &\sum_{i \in D(I)^c,\ i \neq j} p_i u'(w - \mathbb{E}[I(L)])(-p_j) + p_j u'(w - \mathbb{E}[I(L)])(1 - p_j) \\ = &\ p_j(1 - q) u'(w - \mathbb{E}[I(L)]) \geq 0. \end{aligned}$$

To obtain the strict inequality we used the fact that $u'$ is strictly decreasing and $I(\ell_j) < \ell_j$. As $f'(0) > 0$ contradicts the optimality of $I$, we infer that $I(\ell_j) = \ell_j$ must hold.

### 2.6.14 Proof of Proposition 2.4.10

Let $j \in \{1, \ldots, m\}$ and assume that $j \notin S(I)$ but $I(\ell_{j-1}) < I(\ell_j)$. Define $J \in \mathcal{I}_{inc}$ by

$$J(\ell_i) := \begin{cases} I(\ell_i) & \text{if } i \in \{0, \ldots, j-1\}, \\ I(\ell_{j-1}) & \text{if } i = j, \\ I(\ell_i) & \text{if } i \in \{j+1, \ldots, m\}. \end{cases}$$

Note that $\mathbb{E}[J(L)] < \mathbb{E}[I(L)]$. Note also that $S(I) \subset S(J)$. Indeed, for every $i \in S(I)$

$$\begin{aligned} w - \mathbb{E}[J(L)] - \ell_i + J(\ell_i) &> w - \mathbb{E}[I(L)] - \ell_i + J(\ell_i) \\ &= w - \mathbb{E}[I(L)] - \ell_i + J(\ell_i) > 0, \end{aligned}$$

showing that $i \in S(J)$. But then

$$\begin{aligned} V(J) &= \sum_{i \in S(J)} u(w - \mathbb{E}[J(L)] - \ell_i + J(\ell_i)) \\ &> \sum_{i \in S(I)} u(w - \mathbb{E}[I(L)] - \ell_i + I(\ell_i)) = V(I), \end{aligned}$$

contradicting the optimality of $I$. As a result, we must have $I(\ell_{j-1}) = I(\ell_j)$. Finally, observe that $I(\ell_j) < \ell_j$ for otherwise $\ell_j = I(\ell_{j-1}) \leq \ell_{j-1}$, which is impossible.

### 2.6.15 Proof of Theorem 2.4.11

Assume that no insurance is not optimal and define $j := \min\{i \in \{1, \ldots, m\}\,;\ I(\ell_i) = I(\ell_m)\}$. Since $I$ is increasing, it is clear that $I(\ell_{j-1}) < I(\ell_j) = \cdots = I(\ell_m)$. It follows from Proposition 2.4.10 that $j$ must belong to $S(I)$.

### 2.6.16 Proof of Proposition 2.4.12

To prove (i) and (ii), let $i \in \{1, \ldots, m\}$ satisfy $i \in S(I)$. Then, admissibility implies

$$w - \pi(I) - \ell_{i-1} + I(\ell_{i-1}) \geq w - \pi(I) - \ell_i + I(\ell_i) > 0,$$

showing that $i-1 \in S(I)$. Similarly, for $i \in \{0, \ldots, m-1\}$ with $i \in D(I)$, admissibility yields

$$w - \pi(I) - \ell_{i+1} + I(\ell_{i+1}) \leq w - \pi(I) - \ell_i + I(\ell_i) < 0,$$

showing that $i+1 \in D(I)$. To prove (iii), note first that, as no insurance is not optimal, we must have $w > \pi(I)$, for otherwise $V(I) = u(0) = 0 < p_0 u(w) \leq \mathbb{E}[u([w-L])_+]$ and no insurance would be strictly preferable to $I$. This shows that $0 \in S(I)$. As full insurance is not optimal, it follows from Proposition 2.4.3 that $D(I)$ is not empty. As a consequence of (i) and (ii), there exist $0 \leq h < k \leq m$ such that $S(I) = \{0, \ldots, h\}$ and $D(I) = \{k, \ldots, m\}$. Note that we must have $h > 0$, for otherwise $V(I) = p_0 u(w - \pi(I)) \leq p_0 u(w) \leq \mathbb{E}[u([w-L])_+]$ and no insurance would be optimal in $\mathcal{I}_{adm}$. Now, assume that $I(\ell_j) < I(\ell_{j+1})$ for some $j \in \{h, \ldots, m-1\}$. Let $\varepsilon \in (0, I(\ell_{j+1}) - I(\ell_j))$ and define an indemnity schedule $J \in \mathcal{I}_{adm}$ by

$$J(\ell_i) := \begin{cases} I(\ell_i) & \text{if } i \in \{0, \ldots, j\}, \\ I(\ell_i) - \varepsilon & \text{if } i \in \{j+1, \ldots, m\}. \end{cases}$$

Note that $\pi(J) < \pi(I)$ and $S(I) \subset S(J)$. Hence, we obtain

$$V(J) \geq \sum_{i=0}^{h} p_i u(w - \pi(J) - \ell_i + J(\ell_i)) > \sum_{i=0}^{h} p_i u(w - \pi(I) - \ell_i + I(\ell_i)) = V(I),$$

contradicting the optimality of $I$. As a result, we must have $I(\ell_h) = \cdots = I(\ell_m)$. Finally, to establish (iv), assume that $w \geq \ell_m$ and take any $I \in \mathcal{I}_{adm}$. Observe that $w - \pi(I) - \ell_m + I(\ell_m) \geq I(\ell_m) - \pi(I) > 0$ and, hence, $m \in S(I)$. By point (i), this implies that $S(I) = \{0, \ldots, m\}$. In other words, the value function $U$ coincides with $U_\infty$ on $\mathcal{I}_{adm}$. But then full insurance must be optimal in $\mathcal{I}_{adm}$.

### 2.6.17 Proof of Theorem 2.4.13

If no insurance is optimal in $\mathcal{I}_{adm}$, then $I = I_{0,0}$ and we are done. Similarly, if full insurance is optimal in $\mathcal{I}_{adm}$, then $I = I_{0,\ell_m}$. Hence, assume that neither no insurance nor full insurance are optimal in $\mathcal{I}_{adm}$. Define the auxiliary indices $1 \le i_0 \le i_2 \le m$ by

$$i_0 := \max\{i \in \{1,\dots,m\}\,;\ I(\ell_{i-1}) = 0\}, \quad i_2 := \min\{i \in \{1,\dots,m\}\,;\ I(\ell_i) = I(\ell_m)\}.$$

Now, define a deductible level and a cap level $0 \le d \le c \le \ell_m$ by setting

$$d := \ell_{i_0} - I(\ell_{i_0}), \quad c := d + I(\ell_{i_2}).$$

Note that we indeed have $c \le \ell_m$ because $c \le \ell_{i_2} \le \ell_m$ by admissibility. The remainder of the proof is devoted to proving that $I = I_{d,c}$. As a first step, define the index

$$i_1 := \max\{i \in \{i_0,\dots,m\}\,;\ \ell_i - I(\ell_i) = d\}.$$

Note that for every $i \in \{i_0,\dots,i_1\}$ we must have $\ell_i - I(\ell_i) = d$. This is because

$$d = \ell_{i_0} - I(\ell_{i_0}) \le \ell_i - I(\ell_i) \le \ell_{i_1} - I(\ell_{i_1}) = d$$

by admissibility. Note also that $i_1 \le i_2$. Indeed, as $I(\ell_{i_1}) \le I(\ell_{i_2})$,

$$\ell_{i_1} \le \ell_{i_1} - I(\ell_{i_1}) + I(\ell_{i_2}) \le \ell_{i_2}$$

again by admissibility. We claim that

$$I(\ell_i) = I_{d,c}(\ell_i) \quad \text{for every } i \in \{0,\dots,i_1\} \cup \{i_2,\dots,m\}. \tag{2.18}$$

To establish this, take first $i \in \{0,\dots,i_0-1\}$. Then, $\ell_i \le \ell_{i_0} - I(\ell_{i_0}) = d$ by admissibility, yielding

$$I(\ell_i) = 0 = (\ell_i - d)_+ - (\ell_i - c)_+ = I_{d,c}(\ell_i).$$

Next, take $i \in \{i_0,\dots,i_1\}$. In this case, $d \le d + I(\ell_i) = \ell_i = d + I(\ell_i) \le d + I(\ell_{i_2}) = c$.

Hence,

$$I(\ell_i) = \ell_i - d = (\ell_i - d)_+ - (\ell_i - c)_+ = I_{d,c}(\ell_i).$$

Finally, if $i \in \{i_2, \ldots, m\}$, then $\ell_i \geq \ell_{i_2} \geq \ell_{i_0} - I(\ell_{i_0}) + I(\ell_{i_2}) = c$ by admissibility, implying

$$I(\ell_i) = I(\ell_{i_2}) = c - d = (\ell_i - d)_+ - (\ell_i - c)_+ = I_{d,c}(\ell_i).$$

This shows (2.18). The desired identity $I = I_{d,c}$ will therefore be established after we show that $i_1 \geq i_2 - 1$. To the contrary, suppose that $i_1 < i_2 - 1$. Define

$$\delta := \frac{\sum_{i=i_0}^{i_1} p_i}{\sum_{i=i_1+1}^{i_2-1} p_i} > 0.$$

Moreover, take any $\overline{\varepsilon} \in (0, \infty)$ satisfying

$$\overline{\varepsilon} < \min\left\{ I(\ell_0), \frac{\ell_{i_1+1} - I(\ell_{i_1+1}) - d}{1+\delta}, \frac{I(\ell_{i_2}) - I(\ell_{i_2-1})}{\delta} \right\}.$$

This is possible because all expressions in the minimum are strictly positive. For every $\varepsilon \in [0, \overline{\varepsilon})$ define an indemnity schedule $J_\varepsilon \in \mathcal{I}$ by setting

$$J_\varepsilon(\ell_i) := \begin{cases} I(\ell_i) & \text{if } i \in \{0, \ldots, i_0 - 1\}, \\ I(\ell_i) - \varepsilon & \text{if } i \in \{i_0, \ldots, i_1\}, \\ I(\ell_i) + \delta\varepsilon & \text{if } i \in \{i_1 + 1, \ldots, i_2 - 1\}, \\ I(\ell_i) & \text{if } i \in \{i_2, \ldots, m\}. \end{cases}$$

Note that $J_0 = I$. A direct verification shows that for every $\varepsilon \in [0, \overline{\varepsilon})$ we have $J_\varepsilon \in \mathcal{I}_{adm}$ (thanks to our choice of $\overline{\varepsilon}$), $\pi(J_\varepsilon) = \pi(I)$ (thanks to our choice of $\delta$), and $S(J_\varepsilon) = S(I)$. In particular, to verify that $S(J_\varepsilon) = S(I)$, use Proposition 2.4.12 to find $h \in \{1, \ldots, m-1\}$ such that $S(I) = \{0, \ldots, h\}$ and $I(\ell_h) = \cdots = I(\ell_m)$. Note that, by definition of $i_2$, we have $i_2 \leq h$. As a consequence,

$$w - \pi(J_\varepsilon) - \ell_i + J_\varepsilon(\ell_i) = w - \pi(I) - \ell_i + I(\ell_i) \begin{cases} > 0 & \text{if } i = h, \\ \leq 0 & \text{if } i = h + 1, \end{cases}$$

showing that $h \in S(J_\varepsilon)$ and $h+1 \notin S(J_\varepsilon)$ and implying that we indeed have $S(J_\varepsilon) = S(I)$ by Proposition 2.4.12. Now, define the function $f : [0, \overline{\varepsilon}) \to \mathbb{R}$ by

$$f(\varepsilon) := V(J_\varepsilon) = \sum_{i \in S(I)} u(w - \pi(I) - \ell_i + J_\varepsilon(\ell_i)).$$

The right derivative of $f$ at $\varepsilon = 0$ satisfies

$$\begin{aligned} f'_+(0) &= -\sum_{i=i_0}^{i_1} p_i u'(w - \pi(I) - \ell_i + I(\ell_i)) + \delta \sum_{i=i_1+1}^{i_2-1} p_i u'(w - \pi(I) - \ell_i + I(\ell_i)) \\ &\geq -\sum_{i=i_0}^{i_1} p_i u'(w - \pi(I) - \ell_{i_1} + I(\ell_{i_1})) + \delta \sum_{i=i_1+1}^{i_2-1} p_i u'(w - \pi(I) - \ell_{i_1+1} + I(\ell_{i_1+1})) \end{aligned}$$

by admissibility of $I$ and monotonicity of $u'$. Since

$$w - \pi(I) - \ell_{i_1} + I(\ell_{i_1}) = w - \pi(I) - d > w - \pi(I) - \ell_{i_1} + I(\ell_{i_1}),$$

the strict monotonicity of $u'$ eventually entails

$$f'_+(0) > \left( \delta \sum_{i=i_1+1}^{i_2-1} p_i - \sum_{i=i_0}^{i_1} p_i \right) u'(w - \pi(I) - \ell_{i_1} + I(\ell_{i_1})) = 0.$$

This contradicts the optimality of $I$ and shows that the assumption $i_1 < i_2 - 1$ is untenable. As a consequence, we must have $i_1 = i_2 - 1$ as claimed above. This concludes the proof.

# Chapter 3

# On corporate demand for insurance: A dynamic perspective on property insurance

Andrea Bergesio, Pablo Koch-Medina and Mario Šikić[1]

**Abstract** Why do firms purchase property insurance? If the firm has access to external capital and corporate liquidity, what is the effect of protecting productive capital on firm value? We address these questions in a dynamic model with endogenous investment, financing and demand for coverage against shocks to productive capital. We show that firms' demand for property insurance is driven by the relative size of the costs imposed by financial and investment constraints. Numerical results reveal that risk-neutral firms purchase property insurance mainly in response to inflexible productive capital, costly financing and, to a lesser extent, to increase debt capacity. We also study the impact of property insurance on firm value when the firm can choose among alternative financing options. Due to the complexity of the problem, we resort to Shapley values as the aggregate measure of the impact of property insurance on firm value relative to other corporate policies. Using a modified definition of Shapley values, we find that property insurance is more valuable for large firms, which have more productive capital at stake, and small firms with low productive capital exposed to heavy-tailed capital shocks, for which persistently low profit shocks increase the comparative advantage of property insurance.

[1]Partial support through the SNF, Switzerland project 189191 "Value-maximizing insurance companies: An empirical analysis of the cost of capital and investment policies" is gratefully acknowledged. Bergesio at University of Zurich and Swiss Finance Institute, Koch-Medina at University of Zurich and Swiss Finance Institute, Šikić at University of Zurich.

## 3.1 Introduction

It is well established that corporate risk management becomes relevant when firms are subject to various financial and investment frictions, which have been the focus of a long-standing literature since the seminal paper by Modigliani and Miller (1958). Research on this topic has developed theoretical models and produced empirical evidence to explain the economics of risk management, with a special focus on the incentives that lead firms to engage in financial hedging (Smith and Stulz (1985), Froot et al. (1993)) and purchase insurance (Mayers and Smith (1982), Mayers and Smith (1990), Schlesinger (2000)). In the insurance literature, the seminal paper by Mayers and Smith (1982) laid the trail for a stream of theoretical and empirical work that has investigated the drivers of demand for insurance by widely-held corporations. This research has shown that firms may purchase insurance to reduce different costs, including costs of external financing (Doherty (2000), Hau (2006a)), expected tax liabilities (Chen and PonArul (1989), Hoyt and Khang (2000)), financial distress cots (MacMinn (1987), Regan and Hur (2007)) and agency costs of debt (Schnabel and Roumi (1989), Garven and MacMinn (1993)), as well as to increase debt capacity (Caillaud et al. (2000)).

A narrower stream of theoretical research has shown how many of these factors can explain also why firms purchase protection against property losses, which can severely dampen their productive and financial capacity while tightening financial constraints (Hau (2006a)). In their empirical work, Aunon-Nerin and Ehling (2008) rely on existing models of corporate demand for insurance to investigate which drivers can explain observed levels of property coverage by firms, finding that property insurance helps firms to lower expected financial distress costs and reduce financial constraints. The authors, however, put forward the need for models that could provide further guidance on the mechanics behind firms' demand for property insurance:

> "*Ideally, we would like to test theories of corporate finance in a manner that permits unambiguous predictions. Current models, however, say very little about risk management through insurance* (Aunon-Nerin and Ehling (2008), p. 309)."

In particular, existing models of corporate insurance demand share two features: the use of static models, which ignore the dynamics of firms' decisions over time, and the focus on specific explanations for insurance purchase by firms, without exploring potential interactions among different frictions.

Following the call in Aunon-Nerin and Ehling (2008), we complement current literature by proposing a dynamic model of corporate demand for insurance and use it to answer two interrelated research questions: when firms have access to external and internal financial resources, what is the impact of insuring productive capital on firm value? Which are the main drivers of corporate demand for property insurance and what role is played by different financial and investment costs?

In order to address these questions while properly integrating property insurance within the set of financial decisions available to the firm, we develop a dynamic structural model. In general, dynamic models allow to understand how firms take marginal decisions differently based on their current state, which provides information on the their financial situation, productive stock and profitability. By construction, dynamic models account for the forward-looking nature of financial and investment decisions, which are taken according to their effect on future states and on the corresponding optimal decisions. Moreover, the inter-temporal nature of firms' investment and financial decisions becomes even the more relevant when firms are subject to frictional costs that constrain their decisions.

To investigate the drivers of corporate property insurance, we model in a parsimonious way a set of investment and financial costs that allow us to shed light on the relative importance of the frictions identified in the literature for the purchase of property coverage by different types of firm.

Our model is nested in dynamic discrete-time models of investment common in corporate finance (see e.g., Hennessy and Whited (2005), Gamba and Triantis (2008), DeAngelo et al. (2011)). Similar to these models, we consider a non-financial, value-maximizing firm that takes investment and financial decisions over time, given its current profitability and productive capital stock. The firm can invest in productive assets to generate operating profits that are ultimately subject to persistent profit shocks, which represent non-hedgeable background risk stemming from standard business risk. This feature, which is common across discrete-time models of investment, will enable us to investigate how standard results on insurance demand in the presence of background risk (or random initial wealth) change in a dynamic setting. In order to finance its investments, the firm can issue equity, raise short-term debt or change its available cash holdings. As in Hennessy and Whited (2005) and

DeAngelo et al. (2011), we consider firms that either hoard cash or raise debt, but never hold both simultaneously.[2]

The firm must incur costs when taking financial and investment decisions: external financing requires equity and debt issuance costs, whereas selling capital imposes liquidation and fire-sale costs, which capture the cost of financial distress in our model. Differently from Hennessy and Whited (2005) and Gamba and Triantis (2008), we allow for capital adjustment costs associated with building up capital stock. Cash holdings are also costly, due to taxes levied on corporate income and a rate of return shortfall related to agency costs of holding cash.[3] Moreover, collateral constraints effectively make debt default-free, thus limiting the firm's debt capacity by the value of the collateral the firm can pledge against debt (Hennessy and Whited (2005), Gamba and Triantis (2008), Livdan et al. (2009)).

We augment standard models of investment by assuming that not only productive assets are exposed to depreciation, but they are subject also to exogenous shocks generating losses that deplete the capital stock. Differently from non-hedgeable profits shocks, however, the firm can purchase protection against capital losses from insurance markets, which offer property coverage at prevailing market prices. In effect, property insurance represents a form of contingent capital that provides liquidity when capital shocks hit. Our model thus augments the workhorse investment model in corporate finance by introducing capital shocks as an additional source of uncertainty and extending the set of financial decisions to include property coverage.

Introducing property insurance in a dynamic model generates interesting interrelations among firm's decisions. While risk-free debt prevents us from studying the role of agency costs related to issuing risky debt (see e.g., Cooley and Quadrini (2001), Hennessy and Whited (2007)), collateral constraints introduce an additional way for insurance to create value by generating interdependence among debt, property insurance and investment. We show that not only firms with higher insurance coverage and capital stock have access to more debt capacity (Caillaud et al. (2000)), but also that

[2]Gamba and Triantis (2008) examine firms that resort to debt financing while simultaneously holding cash balances and show that different combinations of debt and cash positions generating the same net debt value might well deliver different firm values. In particular, their result is driven by the relative size of the cost of issuing long-term debt vis-a-vis taxes and financial distress costs.

[3]Cooley and Quadrini (2001) model the cost of holding cash by imposing a similar cost proportional to the rate of return on cash, while excluding taxes. Instead, Hennessy and Whited (2005) and Gamba and Triantis (2008) assume corporate taxes are levied on firm's profits, without imposing additional costs on the return on cash holdings.

expected capital liquidation and fire-sale costs can limit the degree to which property insurance contributes to the firm's collateral. Our results suggest that small firms subject to high transaction and financial distress costs will increase debt capacity by lowering the deductible level, whereas larger firms with more flexible capital stock will tend to raise the limit of coverage. Furthermore, we find that highly profitable firms in financial distress may be willing to increase their property coverage in order to relax collateral constraints, issue more debt and capture higher expected future growth opportunities.

In order to gauge the extent to which property insurance helps firms to restore their financial flexibility (Gamba and Triantis (2008)), we study the value that property insurance adds to an otherwise identical firm with no access to insurance for different firm sizes, profitability levels and frictional costs. We find that small, financially healthy firms subject to capital liquidation costs and costly external financing gain the largest benefits from property insurance, as it helps to protect capital with high marginal productivity. Moreover, the heterogeneity of insurance benefits across firms with different capital stocks is driven by capital liquidation costs: if firms can liquidate capital at no cost, then smaller firms have no comparative advantage from purchasing insurance over firms with a larger capital stock.

Interestingly, the model dynamics reveal that persistence in background risk may affect corporate demand for property insurance even if capital and profit shocks are independent processes. This effect arises because background risk is not i.i.d.: periods with high productivity are likely to persist, increasing the marginal expected productivity of capital in the next periods. The firm is thus induced to purchase property insurance to protect its productive capital and take advantage of potentially higher future productivity. In particular, property insurance is valuable to small firms with low productivity levels subject to capital liquidation costs, for which insurance indemnity acts as contingent capital that can partially offset expected shortfalls in operating profits.

Finally, we propose the use of Shapley values, appropriately modified to express the average marginal relative improvement in the value function for a certain financial policy, to study the relative importance of financing choices under different frictional costs. Modified Shapley values indicate that all financial decisions add the most value for firms with the least financial flexibility, specifically firms with a small capital stock and low profitability.

Furthermore, Modified Shapley values show that the expected marginal *relative* increase in firm value due to property insurance is higher for large firms, which have more productive capital at stake, and for firms exposed to capital losses with heavier tails, as the higher marginal expected cost of capital losses raises the marginal value that property insurance generates on average. The relevance of property insurance against heavy-tailed capital losses is especially high for small firms with low productive capital. In this case, the importance of property coverage is reversed as profitability rises: the exposure of small firms to heavy-tailed property losses induces a negative relation between profitability and corporate insurance, which is observationally equivalent to the results that would obtain when background risk produces *temporary* shocks that are negatively dependent on insurable shocks.

### Relation to the literature

We present next an account of the literature closest to our research, which lies at the intersection of two main streams of literature.

Our model borrows from the corporate finance literature and builds on existing dynamic models of investment in discrete time. Dynamic models of investment accommodate a wide range of firms' characteristics by allowing for a rich set of frictional costs associated with financing and investment decisions. For this reason, these models have been extensively used to generate testable predictions concerning firms' financial, investment and payout policies.

The models closest to ours are those developed in Hennessy and Whited (2005), Gamba and Triantis (2008) and DeAngelo et al. (2011). Similar to those, we allow for a rich set of controls that include (dis)investment, equity and debt issuance, cash holdings and dividends payout. Moreover, we model equity costs as in Hennessy and Whited (2005) and debt issuance costs as in (Gamba and Triantis (2008)), while also introducing collateral constraints on debt, corporate taxes, transaction costs and fire-sale costs from disinvestment. However, differently from Hennessy and Whited (2005) and Gamba and Triantis (2008) we abstract from the effects of personal taxes but allow for capital adjustment costs from increasing the capital stock, as in DeAngelo et al. (2011). The choice of ignoring personal taxes is motivated by the fact that corporate taxation is sufficient to capture the tax benefits of insurance in the form of tax deductible premiums. Capital adjustment costs, together with capital liquidation costs, represent frictions that make the stock of productive capital inflexible and thus motivate firms to manage their exposure to profit and capital shocks. A clear distinction between our model and other dynamic models in corporate finance is the presence of insurable capital shocks in addition to unhedgeable profit shocks. This feature extends standard dynamic models of investment in two directions: it introduces an additional source of uncertainty and enriches the set of financial decisions by allowing the firm to purchase property insurance against capital shocks. To help the reader trace how our modeling choices are related to the existing literature, we report the relevant references in a table at the beginning of the section on numerical results (see Table 3.1).

Our work is also related to the literature investigating the reasons for which firms purchase insurance. The seminal paper by Mayers and Smith (1982) clarified the need for distinction between the incentives to buy insurance for firms with closed and broad ownership. While the former should have the same preferences of their owner, firms owned by diversified investors should act *as if* they were risk neutral, as their owners can fully diversify away the firm's idiosyncratic risks. Mayers and Smith (1982) argued that firms with a diffuse ownership may find it optimal to purchase insurance coverage due to the existence of factors that make it costly to take risk, more specifically "because of (1) taxes, (2) contracting costs, or (3) the impact of financing policy on the firm's investment decision" (p. 282 therein). Their insights initiated a stream of theoretical and empirical work that has investigated further the drivers of the demand for insurance by risk-neutral firms.

Previous research has provided theoretical and empirical support for four main drivers of corporate demand: (i) relaxing financial constraints, (ii) lowering tax liabilities, (iii) reducing expected financial distress costs and (iv) avoiding expected agency costs of debt. The objective of relaxing the firm's' financial constraint is one shared with other risk management tools such as financial hedging (see e.g., Froot et al. (1993)). Relaxing financial constraints helps to increase firm value by securing valuable new financing sources and enhance investment. Doherty (2000) and Hau (2006a) posit that firms' main risk associated with property damage is a lack of liquidity, which forces them to sell their most liquid assets at a lower than desired price. Insurance becomes more attractive especially as the marginal cost of external financing increases with the amount raised. Caillaud et al. (2000) study how debt contracts with insurance clauses arise endogenously within a costly state verification setting in which a risk-neutral firm looks for funding from risk-neutral financiers.

Their contracting framework, which combines background risks and insurable liability risks, shows that firms' demand for insurance is driven by the effect of insurance on the face value of debt, i.e. it helps relaxing financial constraints. Nini (2020) finds that covenants requiring to purchase insurance are a common feature to most private credit agreements for publicly traded firms, which are required to use the insurance indemnity to repay the loan. Moreover, the presence and stringency of the insurance covenant are positively related with the existence of collateral used to secure the loan.

Another factor that has the potential to induce firms to buy insurance are corporate taxes characterized by a convex schedule. In this case, purchasing insurance can lower the expected tax liability by reducing the volatility of firm's taxable profits. Moreover, insurance can be an attractive option as the premium and indemnities are deductible for tax purposes (Main (1983), Chen and PonArul (1989)). Furthermore, the expected tax liability is affected by changes in the capital structure through debt interests. Rebello (1995) shows that a firm buying full insurance will prefer debt financing while a self-insuring firm will be indifferent between debt and equity.

Firms may also purchase insurance coverage in order to reduce expected financial distress costs, which include direct costs (e.g., legal fees) and indirect costs, such as lost future growth opportunities (MacMinn (1987), Garven and MacMinn (1993), Regan and Hur (2007)). Regan and Hur (2007) find that firms with a greater proportion of cumulative depreciation demand more insurance. Moreover, demand for insurance declines as liquidity declines and vice-versa; this might stem from the fact that firms facing a short-term cash crunch might reduce discretionary purchases, including insurance purchase. More recently, risk management in the presence of financial distress costs has been studied in the context of structured reinsurance products. For instance, Vincent et al. (2021) show that reinsurance deals whose indemnification is related to the performance of the cedent can represent an effective tool to reduce the costs of financial distress when the loss profiles of the insureds are correlated. In general, a natural analogy can be drawn between insurance demand by non-financial firms and demand for reinsurance by insurance firms, for which reinsurance acts as an alternative financing channel that provides contingent capital to the cedent (see Albrecher et al. (2017) for an overview of the roles of reinsurance). An example is given by excess-of-loss (XOL) treaties, which share many characteristics, including deductible levels and limit of coverage, with contracts available to non-financial firms to purchase property insurance (Aunon-Nerin and Ehling (2008)). In testing the demand-side theories of property insurance use, Aunon-Nerin and Ehling (2008) find a positive relation between expected distress costs and property insurance coverage. They also provide evidence for the so-called scale effect, i.e. the ratio of direct bankruptcy costs to firm size is larger for small firms than for large firms, and find that the dividend payout ratio is negatively associated with insurance coverage due to available cash on hand, easy access to capital markets, or both. In addition, they find that firms insure to increase debt capacity.

Several papers have focused on the effect of corporate insurance on the underinvestment problem associated with risky debt outstanding (Mayers and Smith Jr (1987), MacMinn (1987), Schnabel and Roumi (1989)). For instance, Mayers and Smith Jr (1987) study the underinvestment problem in a two-state, two-date model with risky debt and capital losses. In particular, they show that corporate insurance increases the incentive of the firm to replace the damaged assets at the end of the period by increasing the overall value of the firm, thus reducing the agency cost of risky debt. Note that in our model debt is risk-free, which eliminates the potential for underinvestment problems.

Our work relates also to the concept of Shapley value (Shapley (1953a)) and the literature related to it. The Shapley value is one of the key solution concepts in cooperative game theory, assigning to a cooperative game a unique distribution, among the players of that game, of the surplus generated by their cooperation, based on the average marginal contribution of each player to the surplus (see Hart (1989) for a review). Extensions of this concept in the game theory literature include Owen (1968), Owen (1972) and Kalai and Samet (1987), which proposed a modified version based on unequal weighting schemes to capture the potential heterogeneity across the players of a game. Shapley values have been adopted also in the machine learning community (Lundberg and Lee (2017)), which has employed such concept as it allows for the interpretation of otherwise too complex models.

**Structure of the paper**

The paper is structured as follows. Section 3.2 introduces our dynamic model, describing the firm's state, the set of feasible decisions and the dynamics of the firm's balance sheet. Section 3.3 investigates the relation between property insurance and collateral constraints under different frictional costs. Section 3.4 states the firm's problem formally and presents general results about the value function. Section 3.5

presents our numerical results, discussing the effect of property insurance on firm value and debt capacity under various frictional costs. We also compare corporate insurance to other financial policies using a modified definition of Shapley values based on relative increments of the value function. Section 3.6 concludes.

## 3.2 The model

We work in a discrete time economy extending for an infinite number of periods. We consider a non-financial, risk-neutral firm exposed to two exogenous shocks at the end of each period: a capital shock that can deplete its productive assets and a profit shock that can alter operating profits.

After the shocks realise, the firm takes operational and financial decisions, including (i) investment in productive capital, (ii) purchase of property insurance, (iii) changes in cash holdings, (iv) equity or debt issuance and (v) dividend distributions. Note that since expenditures and investment need to be financed by available funds, we can always express one of these controls as a function of the others.

Firm's decisions are constrained by frictions that are specific to each action. Changes to productive capital are costly: transaction and fire-sale costs lower the liquidation value of divested assets, while the firm must incur capital adjustment costs if productive capital is built up (Strebulaev (2007), Gamba and Triantis (2008), DeAngelo et al. (2011)). When purchasing property insurance, the firm must pay a premium that may include a positive profit loading, or mark-up, according to the degree of competitiveness of insurance markets (Rochet and Villeneuve (2011)).

If the firm hoards cash, the rate of return on cash holdings will be lower than the risk-free interest rate due to agency costs of holding cash (Stulz (1990), Bolton et al. (2011)). An alternative justification for cash growth rates below the risk-free interest rate is represented by differences between borrowing and lending rates (see e.g., Cooley and Quadrini (2001)). External financing, which includes equity and debt funds, is also costly: in this case the firm must incur issuance costs motivated by the presence of adverse selection or flotation costs (Myers and Majluf (1984), Hennessy and Whited (2007)).

### 3.2.1 The firm's balance sheet

At each date, information on the firm's balance sheet is encoded in the state $s = (c, k) \in \mathbb{R} \times \mathbb{R}_+$, which consists of the pair of cash balance $c$ and productive capital $k$; we will denote the set of states by $\mathcal{S} := \mathbb{R} \times \mathbb{R}_+$.[4] As in Strebulaev and Whited (2012), productive capital represents the *book value* of real assets and takes values on the positive realm, $k \geq 0$. On the contrary, the firm's cash balance can take either sign depending on the financial condition of the firm.

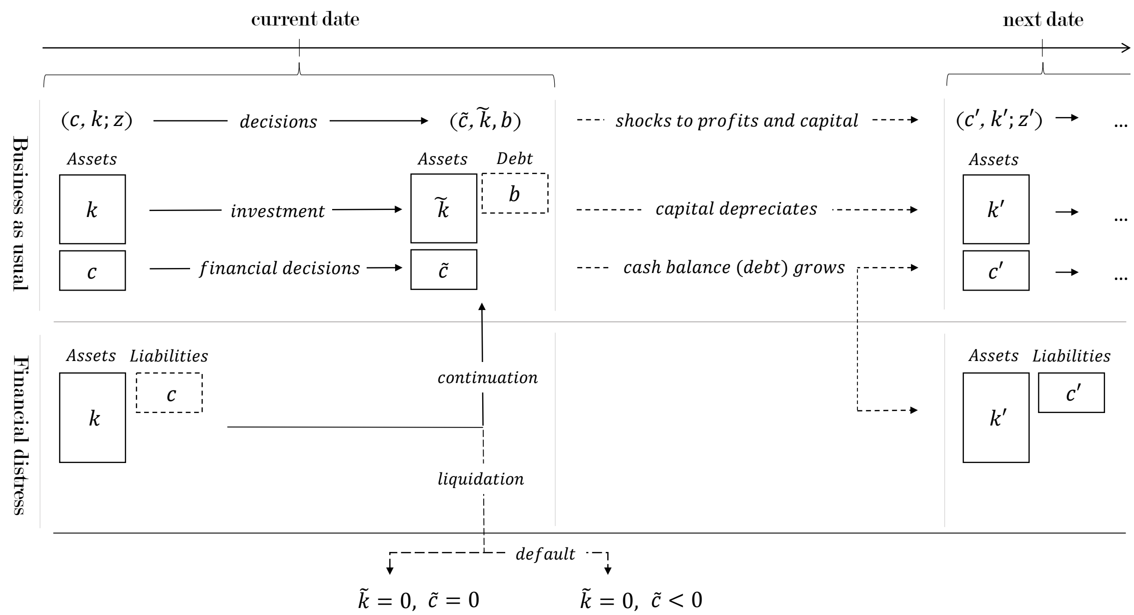



**Figure 3.1:** The figure shows the dynamics of the firm's balance sheet over one period, which is contained between the 'current date' and the 'next date'. The balance sheet comprises productive assets $k$ and cash holdings $c$. $b$ represents debt financing and $z$ is a bivariate random vector representing exogenous shocks to profits and capital. The balance sheet at any date can be 'business-as-usual' (BAU) or in 'financial distress' (FD). Note that liquidation leads to default only when the firm is in financial distress.

A positive cash balance indicates a 'business-as-usual' (BAU) balance sheet in which the firm is hoarding cash, while a negative cash balance represents outstanding liabilities resulting from obligations that could not be paid off in the previous period due to insufficient liquid resources.[5] Similar to Hennessy and Whited (2005), negative

[4]We use the terms productive capital, productive assets and real assets interchangeably. Moreover, we use $\mathbb{R}_+$ to denote the set of positive real numbers.

[5]Positive cash balances $c > 0$ are referred to in the text as cash holdings. We use outstanding liabilities to refer to negative cash balances $c < 0$.

cash balances identify states of financial distress (FD).[6]

Before taking financing and investment decisions, the firm decides whether to keep operating. If the business is closed, assets are liquidated and any proceeds remaining after paying outstanding liabilities are returned to shareholders. Moreover, liquidation is irreversible and is identified by the liquidation date $\tau$. If the firm remains active, it takes investment and financing decisions.

Figure 3.1 shows the firm's dynamics over a single period consistent with firm's decisions and the initial financial condition of firm.

### 3.2.2 Sources of Uncertainty and Capital dynamics

We start this section by discussing the source of uncertainty in our model. Uncertainty is modeled as an exogenous shock described by a bivariate random vector $z$. To identify its individual components we denote the shocks random vector as $z = (z^p, z^k)$, in which $z^p$ represents the shock to profits and $z^k$ the shock to productive capital. We assume that shocks to profits are independent of shocks to capital, as this assumption leads to a cleaner interpretation of the model and of the results.[7] The shock to profits represents pure unhedgeable business risk that may reflect shocks to demand, input prices and productivity. In line with the literature, we assume shocks to profits $z^p$ follow a Markov process (Whited (1992), Gomes (2001), Moyen (2004), Hennessy and Whited (2005)).[8] On the other hand, the shock to capital $z^k$ reflects exogenous events that cause damages to productive assets, such as natural catastrophes or operational failures, for which an insurance market exists. Given the unpredictable nature of

[6]The definition of financial distress and its consequences for the firm vary across the literature. In Gamba and Triantis (2008), financial distress occurs when liquidity at the end of a period is insufficient to cover the coupon payment. In that case, as in Hennessy and Whited (2005) but differently from our model, financial distress *requires* the firm to sell productive assets at a discount to pay off its debt obligations. Titman and Tsyplakov (2007) and Strebulaev (2007) specify a threshold for financial distress in terms of a cash flow-to-debt coupon coverage ratio and assume it leads to a loss in operating cash flows and to fire-sales on the market value of assets, respectively.

[7]The dependence between uninsurable shocks and shocks for which an active insurance market exists is the focus of the literature on insurance demand with background risk (see Hong (2018) for a review). Background risk is analogous to the risk underlying profit shocks in our model. In general, the literature has shown that the Bernoulli principle, according to which full insurance is optimal at actuarially fair prices for strictly risk averse agents, remains valid when background risk and insurable risk are independent (Schlesinger (2000), Hong et al. (2011), Hofmann et al. (2019)).

[8]This modeling assumption is made in order to capture the evolution of product market conditions over time.

these shocks, we assume that insurable capital shocks $z^k$ are identically distributed and independent over time. We collect the properties of the shock vector $z$ below.

**Assumption 3.2.1.** *The shock random process* $(z_t)$ *is a two dimensional process with components* $(z_t^p)$ *describing the profit shock and* $(z_t^k)$ *the capital shock. The bivariate shock takes values in the set* $Z \subseteq \mathbb{R} \times \mathbb{R}_+$. *The random variables* $\{z_t^k \,|\, t = 0, 1, \ldots\}$ *are independent and identically distributed, whereas* $\{z_t^p \,|\, t = 0, 1, \ldots\}$ *is a Markov process with transition kernel* $Q \colon \mathbb{R} \times \mathcal{Z} \to [0, 1]$, *with* $\mathcal{Z}$ *the canonical Borel sigma algebra on* $\mathbb{R}$. *We assume that the two stochastic processes are independent.*

**Investment and Operating profits.** At each date, the firm can alter its capital stock $k$ by investing $i \geq -k$, with positive values indicating an increase in productive assets and negative values representing divestment of capital. The new capital level after investment is then $\widetilde{k} = k + i$. Henceforth, we use $\widetilde{\cdot}$ to refer to a state variable *after* firm's decisions, but before shocks realise.[9]

Real assets $\widetilde{k}$ are put to productive use to generate operating profits $F(\widetilde{k}, z^p)$ at the end of the period after the profit shock $z^p$ realized. We assume the firm must incur fixed production costs $f > 0$ to run its business. To simplify notation, we will omit the superscript and write $F(\widetilde{k}, z)$. We impose standard conditions on the properties of the operating profit function, which we summarize below.[10]

**Assumption 3.2.2.** *Let* $K \subseteq \mathbb{R}_+$. *The operating profits function* $F \colon K \times Z \to \mathbb{R}_+$ *depends only on the* $z^p$ *component of the shock vector. Furthermore,*

1. *for every* $z \in Z$ *the operating profits are zero without capital, i.e.* $F(0, z) = 0$;
2. *for each realization of the shock* $z \in Z$, *the function* $k \mapsto F(k, z)$ *is strictly increasing and strictly concave; furthermore, it is continuous at* $k = 0$;
3. *for every constant* $k \geq 0$, $z \mapsto F(k, z)$ *is measurable.*

[9] We point out that our notation differs from that usually adopted in standard discrete-time models of investment. While the state variables refer to end-of-period quantities as in other dynamic models (e.g., Gomes (2001), Hennessy and Whited (2005)), we further distinguish between state variables *before* firm's decisions and variables *after* decisions have been implemented.

[10] In corporate finance, these are standard assumptions for the profit function. The monotonicity and concavity of $F$ reflect either a downward-sloping demand curve under imperfect competition (Hennessy and Whited (2005)) or the limited organizational resources available to firms (Lucas (1978)), which result in decreasing returns to scale.

*In addition, we need to impose a mild boundedness condition: there exists an increasing, concave function $c_p\colon K \to \mathbb{R}_+$, such that $F(k,z) \le c_p(k)$ for all $z \in Z$.*[11] *Furthermore, $c_p(k)/k \to 0$ as $k \to \infty$.*[12]

The last assumption ensures that expected profits from any fixed size of productive capital are well defined.[13] In order to undertake positive investments, the firm must incur capital adjustment costs $\phi\colon \mathbb{R}_+ \to \mathbb{R}_+$. On the other hand, disinvestment involves proportional transaction costs (Hennessy and Whited (2005), Strebulaev and Whited (2012)). In this case, the firm can sell its productive assets at a price $q \le 1$ per unit of capital, with strict inequality reflecting transaction costs. However, if disinvestment occurs in states of financial distress ($c < 0$), productive capital can be sold only at the fire-sale price $q_d \le q$, which captures financial distress costs that may stem from the time constraints faced by the firm when forced to settle outstanding liabilities (Hennessy and Whited (2005), Gamba and Triantis (2008)). Caillaud et al. (2000) also discuss the role of financial distress costs, which appear in their model as the cost borne by debtholders to verify the firm's cash flows. We enclose the costs attached to the (dis)investment decision in the cost function $C_i\colon \mathbb{R} \times \mathcal{S} \to \mathbb{R}_+$.

**Assumption 3.2.3.** *Let the capital adjustment costs be a continuous, increasing function $\phi\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that $\phi(0) = 0$. The investment cost function $C_i\colon \mathbb{R} \times \mathcal{S} \to \mathbb{R}_+$ is given by*

$$C_i(i,s) \coloneqq \begin{cases} \phi(i) & \text{if } i \ge 0, \\ i \cdot (q-1) & \text{if } i < 0 \text{ and } c \ge 0, \\ i \cdot (q_d - 1) & \text{if } i < 0 \text{ and } c < 0. \end{cases} \tag{3.1}$$

**Capital losses and Productive capital dynamics.** The dynamics of the firm's capital can be described as follows: after investment decisions are taken, productive capital $\widetilde{k}$ depreciates at the accounting and economic rate $\delta \in (0,1)$. The depreciated value $\widetilde{k}(1-\delta)$ is depleted based on the realized shock to capital, which gives rise

[11] This boundedness assumption is standard in the literature; see e.g., Gomes (2001), Gamba and Triantis (2008).

[12] There is no loss in generality in assuming that the function $c_p$ is differentiable. Then, the growth condition is equivalent to $c_p'(k) \to 0$ as $k \to \infty$. Furthermore, by the increasing, concave property, the derivative will be decreasing and positive where defined.

[13] Indeed, by the increasing property when $\widetilde{k} < 1$, then $F(\widetilde{k}, z) < F(1,z)$. Also, by concavity when $\widetilde{k} > 1$, then $F(\widetilde{k}, z) < \widetilde{k}\, F(1,z)$.

to a loss of productive capital. The properties of the capital loss function $\ell$, which depends on $z$ through $z^k$, are collected below.

**Assumption 3.2.4.** *The capital loss function $\ell\colon K \times Z \to \mathbb{R}_+$ depends only on the realization of the capital shock $z^k$, is continuous and is bounded above*

$$\ell(\widetilde{k}, z) \leq (1-\delta)\widetilde{k} \qquad \textit{for all } z \in Z.$$

*Furthermore, the function $\widetilde{k} \mapsto \ell(\widetilde{k}, z)$ is increasing for every $z \in Z$.*

Note that capital losses take zero as minimum value. Assumption 3.2.4 reflects the nature of capital losses: higher capital levels leave the firm exposed to larger capital losses, which however cannot exceed the amount of capital at stake. As a result, productive capital evolves according to the following dynamics

$$k' = \widetilde{k}(1-\delta) - \ell(\widetilde{k}, z), \tag{3.2}$$

in which we use $(\cdot)'$ to indicate state variables one period ahead.

### 3.2.3 Financing decisions

To finance its operations, the firm can choose from a set of financing sources that includes (1) equity issuance, (2) short-term debt, (3) internal cash holdings and (4) insurance coverage against property loss. The first three sources of funds are common to other investment models in corporate finance, whereas insurance coverage against property loss is specific to our model. As mentioned in the introduction to our model, we allow for a rich set of financial frictions that affect the firm's optimal mix of financing sources.

**Equity and Debt issuance.** The firm can finance its operations by raising equity $e \geq 0$ from existing shareholders, which must incur flotation costs denoted by $C_e\colon \mathbb{R}_+ \to \mathbb{R}_+$ (see e.g., Hennessy and Whited (2007)). As an alternative source of external financing, the firm can issue short-term debt with face value $b \geq 0$, which must be repaid at the end of the period. Issuing debt imposes two types of cost on the firm: a direct issuance cost $C_b\colon \mathbb{R}_+ \to \mathbb{R}_+$ and collateral constraints that limit the debt capacity of the firm.[14] In particular, collateral constraints make debt risk free by ensuring that the firm has always sufficient assets to pay back debt.[15] Since collateral constraints prevent the firm from defaulting on debt, in our model default can occur only on unsecured liabilities, which are represented by fixed costs of production $f$. The following assumption collects the properties of the cost functions associated with equity and debt financing.

**Assumption 3.2.5.** *The cost functions associated with external financing* $C_e, C_b\colon \mathbb{R}_+ \to \mathbb{R}_+$ *are increasing, lower-semicontinuous and satisfy* $C_e(0) = C_b(0) = 0$. *Moreover, debt issuance imposes collateral constraints that limit debt capacity, i.e.* $b \leq \overline{b}$ *for some* $\overline{b} < \infty$.

**Cash holdings.** In addition to external financing, the firm can alter its initial cash balance $c$. The new cash balance is denoted by $\widetilde{c}$ and grows at the interest rate $r_c \in [0, r]$, with $r_c \leq r$ representing the rate of return on cash holdings net of costs of hoarding cash.[16] If the firm is not liquidated, the new cash balance must satisfy $\widetilde{c} \geq 0$, which requires the firm to repay at the start of the period any outstanding liabilities $c < 0$ by divesting capital, raising external equity or rolling over debt in a suitable amount. In case the firm ceases operations, the cash balance becomes either zero or stays negative. A negative cash balance represents the portion of outstanding liabilities that cannot be paid off with the proceeds from liquidating the firm's assets, in which case liquidation leads to default. However, due to collateral constraints, note that the firm will default only if the proceeds from liquidations after debt repayment prove insufficient to cover the unsecured costs $f$. We formalize below the circumstances in which the firm defaults.

[14]We use borrowing constraints and collateral constraints as synonyms. Borrowing constraints have been shown to have their economic underpinning in the existence of asymmetric information between the firm and financiers (Clementi and Hopenhayn (2006)) and have been documented at large (see e.g., Stiglitz and Weiss (1981), Whited (1992)). The assumption of collateral constraints is widely present in the corporate finance literature (see e.g., Hennessy and Whited (2005), Hennessy and Whited (2007), Strebulaev and Whited (2012)).

[15]Since debt is risk-free, there is no agency problem associated with the firm increasing its leverage to expropriate wealth from existing creditors (Gamba and Triantis (2008)).

[16]As in Bolton et al. (2011), we introduce agency costs through $r_c < r$ as a form of carry cost of holding cash. An alternative source of carry cost of cash is represented by tax distortions, which part of the literature models by introducing a lower marginal tax rate on investors' interest income than the marginal tax rate on corporate income (see e.g., Graham and Harvey (2001), Faulkender and Wang (2006), Livdan et al. (2009)).

**Remark 3.2.6.** *If the firm is liquidated, the 'cash balance after liquidation' is either zero or negative, $\widetilde{c} \leq 0$. Moreover, liquidation coincides with default if and only if $c < 0$ and $\widetilde{c} < 0$.*

Since debt must be repaid at the end of the period out of available financial resources, there is no economic rationale for using debt financing to increase cash holdings, even more so given the cost of issuing debt and agency costs of holding cash. This observation can be used to show that the firm will never hold simultaneously debt and cash (DeAngelo et al. (2011), Strebulaev and Whited (2012)). This result, the proof of which is provided in the Appendix 3.7.2, is stated below.

**Proposition 3.2.7.** *Consider a policy where the firm is to hold cash and issue debt at the same time. This policy is (one-step) dominated by the policy such that the firm either holds cash or issues debt.*

**Property insurance.** The literature on corporate risk management provides arguments in favour of hedging as an alternative source of financing in the presence of frictions that make firm-specific risks costly to bear (Smith and Stulz (1985), Nance et al. (1993), Froot et al. (1993), Bolton et al. (2011)).

Our model incorporates an insurance market in which risk-neutral insurers compete to offer coverage against idiosyncratic shocks to real assets. Insurance firms propose a finite menu of insurance contracts based on accounting information provided by the firm about its productive capital, assuming the capital shock distribution is known to the insurers. Every contract $(I, \pi(I, \widetilde{k}))$ requires the firm to pay a premium $\pi(I, \widetilde{k})$ upfront in order to receive an indemnity $I$ at the end of the period according to the realized capital loss. Since property insurance is written on the assets' actual productive capacity, insurers set their offer according to information on the capital stock $\widetilde{k}$ available to the firm *after* (dis)investment decisions. It follows that in our model insurance and investment decisions are entangled, with the latter affecting the range of insurance prices and coverage available to the firm.

Insurance contracts include indemnity schedules that satisfy two standard properties in the insurance literature: i) the principle of indemnity (Gollier (1996)), according to which the indemnity is nonnegative and lower than the insurable loss and ii) the no-sabotage condition (Huberman et al. (1983)), which requires that the marginal indemnity is nonnegative and smaller than one.[17] The set of admissible insurance contracts is defined formally below.

**Definition 3.2.8.** *The set $\mathcal{I}_{adm}$ of admissible indemnity profiles is given by*

$$\mathcal{I}_{adm} := \big\{ I\colon \mathbb{R}_+ \to \mathbb{R}_+ \big| \, I(0) = 0,\ I(y) \leq I(x) \leq I(y) + (x-y) \text{ for all } 0 \leq y \leq x \big\}.$$

We denote the finite menu of insurance contracts offered to the firm by $\mathcal{I} \subset \mathcal{I}_{adm}$. The premium $\pi(I, \widetilde{k})$ represents the price of insurance coverage and is determined at the time of sale. Insurers set the premium based on the present value of the expected indemnity against capital losses next period under the constraint of nonnegative expected profits[18]

$$\pi(I, \widetilde{k}) = \frac{1+\lambda}{1+r}\, \mathbb{E}\big[I\big(\ell(\widetilde{k}, z)\big)\big], \tag{3.3}$$

in which $\lambda \geq 0$ is the insurer's profit loading, which reflects the competitiveness of insurance markets.[19] Markets are perfectly competitive if and only if insurance prices are so-called fair, i.e. $\lambda = 0$. Furthermore, note that the premia are independent of the profit shock $z^p$ by assumption of independence of the two shocks.

At this point, we need to discuss the type of indemnity payment received by the firm. In general, commercial property insurance provides indemnities based on either the replacement cost of damaged assets or their actual cash value. The replacement cost represents the cash amount needed to restore the full economic capacity of a damaged asset, independently of its depreciation. On the contrary, actual cash value captures the assets' remaining economic capacity, identified by the difference between

[17] The no-sabotage condition is required to reduce ex-post moral hazard (Huberman et al. (1983), Chi and Wei (2020)). The condition imposed on the marginal indemnity implies that larger loss realizations result in higher cash outflows for both the insurer and the insured. This ensures the insured has no incentives to overstate the actual loss.

[18] Let us point out that the premium is not a function of the current value of the shock. Indeed, the capital loss function depends only on past realizations of the capital shock $z^k$ which, by assumption, is an i.i.d. process. Note that Gollier (1996) has shown that when the insurance premium only depends on the expected indemnity, the optimal contract is deductible insurance if background risk and insurable risk are independent, although the result is obtained under unlimited liability.

[19] The condition $\lambda \geq 0$ ensures insurers' expected profits are nonnegative. We include a constant profit loading as a parsimonious way to capture competition in insurance markets. While this approach is standard in the insurance literature (e.g., Hau (2006b), Rochet and Villeneuve (2011)), alternatives include loadings related to risk, as it is typical in reinsurance premiums (Dionne et al. (2000)).

the replacement cost and the accumulated depreciation on the asset. In our model, all available information on capital at each date is given by its current book value, hence we cannot recover the replacement cost of lost capital. As a result, we consider insurance indemnities that return the actual cash value of real assets as captured by their book value, which is representative of those assets' actual economic capacity given that accounting and economic depreciation are the same.

The firm determines equity financing, debt issuance, cash holdings, and insurance purchase simultaneously with dividend distributions, $d \geq 0$. The latter are identified implicitly by the sources and uses of funds identity

$$\widetilde{c} = c - d + e + \left(\frac{b}{1+r} - C_b(b)\right) - \big(i + C_i(i,s)\big) - \pi(I, \widetilde{k}). \tag{3.4}$$

It is easy to show that the firm will never raise equity and pay out dividends simultaneously. The following results formalizes this statement. The proof of the statement can be found in Appendix 3.7.2.

**Proposition 3.2.9.** *If the firm issues equity, then dividends are not distributed.*

### 3.2.4 Profits, taxes and cash dynamics

The dynamics of the firm can be described as follows: after investment and financing decisions are implemented, the firm earns profits $P$ at the end of the period, after capital and profit shocks realise. Profits are the result of operating profits net of fixed production costs, minus the insurance premium, capital depreciation and uninsured capital losses, plus the net interest income on cash and debt

$$P = F(\widetilde{k}, z) - f - \pi(I, \widetilde{k}) - \delta\widetilde{k} - \ell(\widetilde{k}, z) + I\big(\ell(\widetilde{k}, z)\big) + r_c\widetilde{c} - \frac{rb}{1+r}. \tag{3.5}$$

The firm faces corporate taxes $T$ levied on taxable profits.[20] In our model, profits in Equation (3.5) provide the taxable income used to obtain the tax liability of the firm, $T(P)$. The corporate tax schedule shares some basic features with other models developed in previous studies (Hennessy and Whited (2005), Gamba and Triantis (2008)), as we describe below.

[20]In defining the taxable basis for corporate taxes, we exploit the distinct feature of most tax regimes that property losses and the insurance indemnity related to those losses are tax deductible (see e.g., Mayers and Smith (1982)).

**Assumption 3.2.10.** *The corporate tax function $T\colon \mathbb{R} \to \mathbb{R}_+$ is increasing, convex and satisfies*

$$T(p) = 0 \quad for \quad p \leq 0; \;\; and \quad \lim_{p\to\infty} \frac{T(p)}{p} = \overline{T} < 1. \tag{3.6}$$

The assumption of zero taxes when taxable profits are negative is equivalent to assuming that there are no losses carried back or losses carried forward.[21] Monotonicity and convexity of taxes reflect the fact that the marginal tax rate $T'(p)$ is increasing in taxable profits, where differentiable.[22]

The end-of-period cash balance $c'$ is obtained from the initial cash holdings $\widetilde{c}$ plus cash flows generated over the period. These are obtained from the firm's profits after adding back capital depreciation, capital losses and the insurance premium, less the debt principal $b/(1+r)$

$$c' = (1+r_c)\widetilde{c} + F(\widetilde{k}, z) - f + I\big(\ell(\widetilde{k}, z)\big) - T(P) - b. \tag{3.7}$$

We note here that, as suggested in the concluding remarks of Caillaud et al. (2000), our dynamic setting allows for endogenous investment decisions at the start of every period, thus leaving the firm with the choice between investing the insurance indemnity $I$ to replace damaged assets or use it otherwise.

### 3.2.5 Debt capacity and collateral constraints

In our model, the structure of the firm's liabilities is such that tax payments are senior to debt, which in turn is senior to fixed costs of production. As a result, collateral constraints require that cash flows after tax, but before any fixed cost of production, plus the liquidation value of capital at the end of the period are sufficient to pay back the face value of debt

$$b \leq F(\widetilde{k}, z) + I\big(\ell(\widetilde{k}, z)\big) - T(P) + k' - C_i(-k', s') \quad \text{with probability 1}, \tag{3.8}$$

[21]The corporate finance literature typically assumes a strictly convex tax schedule on the entire real line (see e.g., Hennessy and Whited (2005)). Therefore, instead of imposing the condition $T(p) = 0$ for $p \leq 0$, it is often assumed the tax function satisfies $\lim_{p\to-\infty} T'(p) = 0$.

[22]The corporate tax function is convex, hence locally Lipschitz. By the Rademacher theorem it is also differentiable almost everywhere.

where $k' - C_i(-k', s')$ represents the liquidation value of productive capital. The debt level itself affects the extent to which the collateral constraint binds: higher borrowing increases tax shields, but it also decreases the distance from financial distress. We define the debt capacity $\overline{b}$ by

$$\begin{aligned}\overline{b}(\widetilde{c}, \widetilde{k}, I) = \sup\big\{b \geq 0 \;\big|\; b \leq & F(\widetilde{k}, z) + I\big(\ell(\widetilde{k}, z)\big) - T(P) + \\ & + k' - C_i(-k', s') \;\text{ with probability } 1\big\},\end{aligned} \tag{3.9}$$

which represents the largest value of debt such that available funds at the end of the period are almost surely larger than the debt payment. Note that implicit in the collateral constraints (3.9) is that debt is expected to be repaid first out of after-tax cash flows. Only if available liquid resources are insufficient to pay off debt, which triggers financial distress, the firm may have to sell part of its productive assets. The firm, however, can make the decision that is best suited, in value maximizing terms, to repay liabilities: liquidate part of the capital, raise equity or roll over debt.

## 3.3 Debt capacity, firm size and insurance demand

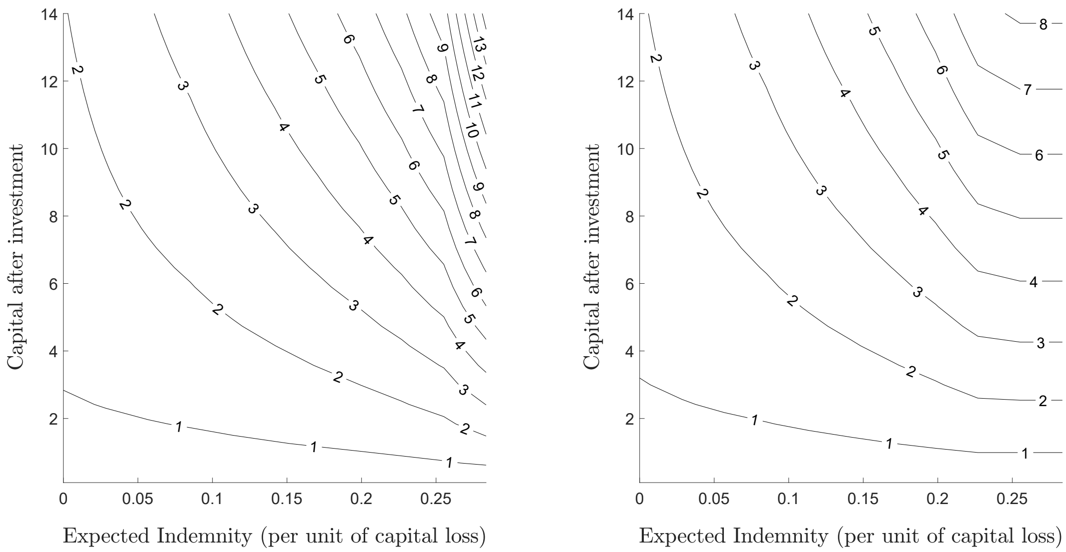
How does the firm's debt capacity depend on the firm's demand for property insurance? We report below an immediate observation, the proof of which is omitted, that shows how higher insurance coverage against capital losses relaxes collateral constraints on debt.

**Lemma 3.3.1.** *Consider two indemnity schedules $I_1 \leq I_2$, with insurance premiums $\pi(I_1, \widetilde{k})$ and $\pi(I_2, \widetilde{k})$. Moreover, let $\overline{b}_i$ be the debt capacity after purchasing the insurance contract $(I_i, \pi(I_i, \widetilde{k}))$, for $i \in \{1, 2\}$. Then $\pi(I_1, \widetilde{k}) \leq \pi(I_2, \widetilde{k})$ and $\overline{b}_1 \leq \overline{b}_2$.*

To gain intuition on how the firm's debt capacity depends on its capital stock and property insurance coverage, Figure 3.2 shows "iso-borrowing" curves (or contours) along which debt capacity remains constant, for different levels of capital after investment and expected insurance indemnity. For insurance contracts providing coverage over one period, Chapter 2 of this thesis has shown that capped-deductible contracts are optimal within the class of admissible insurance profiles for any strictly-risk averse agent with limited liability. As a result, we focus on capped-deductible contracts parametrized by a pair $(D, L)$, with $D$ the deductible level and $L$ the limit (or cap), that provide the following indemnity payment

$$I(D, L; \ell(\widetilde{k}, \cdot)) = \big(\ell(\widetilde{k}, z) - D\big)_+ - \big(\ell(\widetilde{k}, z) - (D + L)\big)_+, \tag{3.10}$$

where $(X)_+ = \max\{X, 0\}$. Figure 3.2 compares the firm's debt capacity when pro-




**(a)** Debt capacity with $q_d = q = 1$.

**(b)** Debt capacity with $q_d < q < 1$.

**Figure 3.2: Debt capacity, firm size and expected coverage.** The figure shows how debt capacity (contour lines) changes as a function of firm size after operations and of the expected indemnity payment per unit of capital loss. The underlying insurance contract is of capped-deductible type with varying limit and deductible level fixed and equal to zero. Panel (a) shows debt capacity for a firm with flexible capital ($q = q_d = 1$). Panel (b) illustrates debt capacity for a firm with inflexible capital in the base case ($q = 0.75, q_d = 0.5$). The remaining parameters are given in Table 3.1.

ductive capital is flexible, or costless (Panel (a)), to the debt capacity in the base case with transaction and fire-sale costs (Panel (b)). In either case debt capacity increases in the initial capital stock and insurance demand, with higher demand represented by a higher expected indemnity per unit of capital loss. Larger firms can pledge a higher collateral as more capital is available for liquidation to pay off debt, especially in the absence of divestment costs (see Panel (a)). Similarly, insurance coverage provides contingent capital that increases the firm's liquid funds in the event of a capital loss. Moreover, the impact of property insurance on debt capacity is stronger for larger capital levels, as the firm has more capital at stake. Comparing the two panels reveals not only that disinvestment costs lower the firm's debt capacity by decreasing the cash value of capital in the next period, but also that they may limit the debt-related benefit of purchasing property insurance. This latter effect occurs because disinvestment costs lower the impact of property losses on the cash value of capital and thus decrease the comparative advantage offered by insurance coverage. To gain intuition on this point, the next result shows under which instances each parameter of a capped-deductible contract is relevant for increasing the firm's borrowing capacity, assuming no taxes are levied on corporate profits.

**Proposition 3.3.2** (Debt capacity and capped deductibles)**.** *Consider a capped deductible contract. Denote by $D$ the deductible and by $L$ the upper limit, with $D, L \in [0, \overline{\ell}(\widetilde{k}, z)]$. Denote by $\overline{\ell}(\widetilde{k}, z)$ the maximum capital loss. Assume taxes are zero and let $q_d = q \leq 1$. Moreover, let $F(\widetilde{k}, z) = z^p \widetilde{k}^\theta$, with $\theta \in (0, 1)$, and $\ell(\widetilde{k}, z) = z^k \widetilde{k}(1-\delta)$. Then the firm's debt capacity is given by*

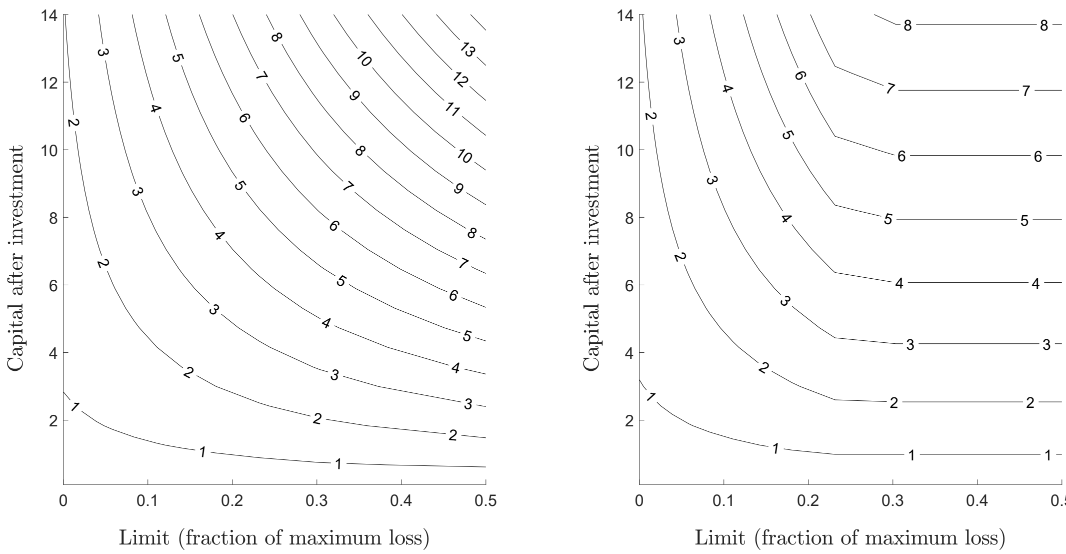
$$\overline{b}(\widetilde{c}, \widetilde{k}, D, L) = \begin{cases} \underline{z}^p \widetilde{k}^\theta + q(\widetilde{k}(1-\delta) - D), & \text{if } q < \frac{L}{\overline{z}^k \widetilde{k}(1-\delta) - D}, \\ \underline{z}^p \widetilde{k}^\theta + q\widetilde{k}(1-\delta)(1-\overline{z}^k) + L, & \text{otherwise}, \end{cases}$$

*where $\underline{z}^p$ is the smallest profit shock realization and $\overline{z}^k \leq 1$ is the largest loss per unit of capital.*[23]

The proof of the proposition is given in Appendix 3.7.3. We discuss the previous result using Figure 3.3, which illustrates contour lines for the debt capacity in the base case model as a function of firm size and insurance limit, assuming an insurance contract with zero deductible.

As before, Panel (a) shows debt capacity when capital is flexible, whereas Panel (b) presents how results change in the base case with inflexible capital. As Proposition 3.3.2 shows, the actual debt capacity of the firm depends on the deductible level, on the limit of coverage and on the flexibility of capital. If capital is perfectly flexible ($q = 1$), then the deductible level is irrelevant and the firm's borrowing capacity is increasing in the coverage limit (see Panel (a)). Intuitively, since borrowing constraints ensure debt repayment even when the largest loss occurs, only the choice of the limit has an impact on the value of the collateral and, thus, on the debt capacity. However, when capital is inflexible and cannot be sold at book value ($q < 1$), Proposition

[23]Formally, the maximum capital loss is given by $\overline{\ell}(\widetilde{k}, z) = \inf\{\alpha \,|\, \ell(\widetilde{k}, z) \leq \alpha \text{ with probability } 1\}$. Our specification of capital losses implies $\overline{\ell}(\widetilde{k}, z) = \overline{z}^k \widetilde{k}(1-\delta)$.




**(a)** Debt capacity with $q_d = q = 1$.

**(b)** Debt capacity with $q_d < q < 1$.

**Figure 3.3: Debt capacity, firm size and insurance limit.** The figures show how debt capacity (contour lines) changes as a function of capital after investment and of the limit of a capped-deductible contract. The limit is expressed as a fraction of the maximum capital loss. The deductible level is set to zero. Panel (a) shows debt capacity for a firm with flexible capital ($q = q_d = 1$). Panel (b) illustrates debt capacity for a firm with inflexible capital in the base case ($q = 0.75, q_d = 0.5$). Other parameters are as in Table 3.1.

3.3.2 shows that when disinvestment costs become sufficiently large given the chosen contract, increasing the limit of coverage does not increase borrowing capacity. This effect can be seen in Panel (b) of Figure 3.3, in which the contour lines flatten out when the limit exceeds a certain level. This shows not only that the limit becomes ineffective to increase debt capacity, but also that the thresholds beyond which such an effect occurs, in line with Proposition 3.3.2, depend on firm's size, disinvestment costs, insurance coverage and maximum capital loss.

**Remark 3.3.3.** *In line with Aunon-Nerin and Ehling (2008), our model suggests that firms are expected to use the deductible level and the limit as substitutes for the purpose of improving debt capacity. Moreover, for a given capped deductible contract, the debt capacity is expected to be more sensitive to the deductible level for small firms with highly inflexible capital and limited property loss exposure. On the other hand, the limit of coverage becomes more relevant for the borrowing capacity of larger firms with more flexible capital and higher loss exposure.*

## 3.4 The firm's problem

In this section, we discuss first the size of the firm's balance sheet and its implications for the state space. Then, we define the firm's problem.

### 3.4.1 On the size of the firm's balance sheet

Before proceeding with the analysis of the optimization problem, let us make a standard assumption in the literature. The proofs of this section are deferred to Appendix 3.7.4.

**Assumption 3.4.1.** *There is an upper bound $\overline{k}$ on capital, such that the firm would never invest to increase productive capacity beyond the limit $\overline{k}$. Furthermore, there is an upper bound $\overline{c}$ on cash reserves, such that the firm would never retain cash holdings above the level $\overline{c}$ on the books.*

This boundedness assumption is usually very difficult to prove, if at all possible, and depends closely on the fine structure of the model; see e.g., Koch-Medina et al. (2021). The rationale offered in the literature is as follows:[24] holding capital $k$ yields operating profits of $F(k;z) \leq c_p(k)$ with $c_p$ concave such that $c_p(k)/k \to 0$ as $k \to \infty$. However, productive capital depreciates over the period, so that a firm would never hold so much capital that depreciation overcomes operating profits, i.e. there is a level $\overline{k}$ such that

$$\frac{F(k;z)}{k} \leq \frac{c_p(k)}{k} \leq \delta,$$

for all $k \geq \overline{k}$. This argument thus points to the existence of an upper bound on capital, which we denote by $\overline{k}$. The existence of an upper bound on cash reserves can be argued following a similar line of reasoning. In particular, one can argue that a firm would never hold more cash than the level $\overline{c}$ that would allow the firm, given the return on cash holdings, to finance in each future period the maximal productive capital while maintaining its cash holdings to finance operations in subsequent periods. If we ignore taxes for simplicity, we obtain

$$\widetilde{c} \cdot (1 + r_c) \geq \overline{k} + C_i(\overline{k}, s) + \widetilde{c}, \tag{3.11}$$

[24]See Gomes (2001) and Hennessy and Whited (2005), among others.

which simplifies to

$$\widetilde{c} \cdot r_c \geq \overline{k} + C_i(\overline{k}, s), \tag{3.12}$$

for every $\widetilde{c} \geq \overline{c}$, given $s$ such that $c = 0$. Since hoarding cash above the level $\overline{c}$ is not economically profitable, this argument suggests the existence of an upper bound on cash holdings, denoted by $\overline{c}$.

Although the intuition above is convincing, it is difficult to turn it into a rigorous argument. For instance, consider the firm's capital. On the one side, capital depreciation is tax deductible; on the other side, if investment costs are too high, it might happen that capital depreciation is overcome by investment costs, which would make it more complex to deal with the standard argument reproduced above. As far as cash holdings are concerned, if equity and debt issuance are not available, it is easy to imagine that the firm would not distribute dividends if holding cash in the company allowed it to invest in productive capacity. The discussion above shows the challenges that underlie those arguments that suggest firms with sufficiently large capital and cash holdings will downsize their capital and pay out "excess cash" as dividends. The opposite question however could be asked, namely whether a firm would ever find it optimal to invest in productive capital or accumulate cash beyond the above specified limits. In the absence of a proof for the existence of such bounds in our model, we retain Assumption 3.4.1 to establish the compactness of the state space.[25]

### 3.4.2 Firm's decisions and firm value

The firm maximizes the net present value of the stream of expected future cash flows to shareholders. At each time step, the firm chooses among the set of feasible policies, given the current state $c, k$ and shock state $z$. We denote this set by $\mathcal{P}_0$ and collect the admissibility conditions here: in case the firm decides to stop operating, the set of available policies is

$$\mathcal{P}_0(c, k; z)^{stop} = \left\{(i, e, b, d, I) \,\middle|\, i = -k,\ e = 0,\ b = 0,\ d = (c + k - C_i(-k, s))_+,\ I \equiv 0\right\}$$

and in case the firm decides to continue operations, the set of admissible controls is

[25] The boundedness condition can be proved rigorously in the case where there is a cap on the costs functions, as in the case of fixed costs only. We do not reproduce this argument here, but it would follow the same lines as Koch-Medina et al. (2021) and yield explicit bounds on the capital and cash levels.

given by

$$\mathcal{P}_0(c,k;z)^{cont} = \left\{(i,e,b,d,I) \,\middle|\, 0 \le \widetilde{k} \le \overline{k},\ d,e \ge 0,\ I \in \mathcal{I},\ 0 \le b \le \overline{b},\ 0 \le \widetilde{c} \le \overline{c}\right\},$$

where $\mathcal{I} \subset \mathcal{I}_{adm}$, the upper bound $\overline{b}$ is given in equation (3.9) and $\widetilde{c}$ is defined by the identity (3.4). The set of strategies, where a strategy identifies a collection of admissible controls over time, is now given by $\mathcal{P}(c,k;z) = \{(p_t)_{t\in\mathbb{N}_0} \,|\, p_t \in \mathcal{P}_0(c_t,k_t;z_t)\}$. Given the initial state $(c,k) \in \mathcal{S}$ and shock $z \in Z$, the value of the firm to shareholders is given by

$$\begin{aligned} V(c,k;z) = \sup_{p\in\mathcal{P}(c,k;z)} \mathbb{E}\Bigg[\sum_{t=0}^{\tau-1} \frac{1}{(1+r)^t}(d_t - e_t - C_e(e_t)) + \\ + \frac{1}{(1+r)^\tau}\big(c_\tau + k_\tau - C_i(-k_\tau;s_\tau)\big)_+\Bigg], \end{aligned} \tag{3.13}$$

subject to the state dynamics (3.2), (3.7) and the accounting identity (3.4). The first sum in the above expression is the net present value of cash flows up to the liquidation time $\tau$. The second term represents the liquidation value of the firm at time $\tau$ subject to the limited liability of the company.

Our first result shows that, for every state $s$ and shock $z$, the value function $V(c,k;z)$ is well defined. This means, in particular, that we need to show that the sum in the above expectation is integrable, i.e. it admits an integrable upper bound; it is bounded below, as the company can always choose to liquidate, taking advantage of the limited liability.

**Theorem 3.4.2.** *The value function is well defined.*

Next, we show that the value function satisfies the dynamic programming equation. This yields, in particular, that the optimal policy is of feedback type, i.e. it depends only on the current state, but not on the past values of such.

**Theorem 3.4.3.** *The value function satisfies the dynamic programming equation*

$$\begin{aligned} V(c,k;z) = \sup_{p\in\mathcal{P}_0(c,k;z)} \max\Bigg\{d - e - C_e(e) + \frac{1}{1+r}\mathbb{E}\big[V(c',k';z')\big], \\ c + k - C_i(-k,s), 0\Bigg\}. \end{aligned} \tag{3.14}$$

*Furthermore, there exists a Borel measurable map*

$$f\colon (c, k; z) \mapsto \mathcal{P}_0(c, k; z)$$

*such that the optimal strategy is stationary with the stage policy $f(c, k; z)$, i.e. at each time $t$, the policy is given by $p_t = f(c_t, k_t; z_t)$.*

The first term in the dynamic programming equation is the continuation value. The second term represents the value of closing the company, whereas the third term represents the value of default and encodes the fact that we are working in a limited liability setting.

Note that the above theorem just guarantees the existence of the value function and of the optimal policy map. It does not imply any additional property of the value function. From the problem formulation, however, it is clear that additional regularity properties hold.

**Lemma 3.4.4.** *The value function is increasing in both cash reserves $c$ and productive capital $k$.*

## 3.5 Results

In this section we present and discuss our numerical results. We start by describing the methods used to implement the profit and capital shock processes, and to evaluate numerically the value function. After introducing the parameter set used in the base case model, we discuss the impact of frictional costs on firm value and how property insurance can help firms to manage these costs and restore financial flexibility.

We investigate also the interrelation between insurance and debt capacity and show that firms may actually purchase property insurance to relax their collateral constraints. Finally, we investigate the relevance of property insurance relative to other firm's financial policies by means of Shapley values expressed in terms of average *relative* improvements in the value function brought about by a given policy.

### 3.5.1 Implementation details

We look for the approximation of the value function in the class of piecewise affine functions. The state space for our optimization problem is three dimensional. The class of affine functions was, thus, selected by fixing a grid on the space and using all the functions obtained by bilinear interpolation of the function values on the grid points. This replaces the infinite dimensional problem of finding the value function with a finite dimensional problem of finding the function values on the grid points. The assumptions on the model imply that the state space is also bounded, thus no extrapolation is needed.

The benefit of using discretization, and linear approximation approach in general, over other functional approximators like neural networks is that one has convergence and approximation guarantees; see Sutton and Barto (2018), Chapter 9.4 and Bäuerle and Rieder (2011).

The above described procedure replaces the original optimization problem by a discrete one. The crux of the proof of the Bellman equation lies in showing that the operator on the right-hand side is a contraction; thus, simple iteration will yield the optimal policy and the initial approximation is irrelevant. We discuss this point further below.

**Profit shocks.** First, we assume that the log of the profit shock process is an AR(1) process

$$\log z^p_{t+1} = \mu + \rho \log z^p_t + \varepsilon_{t+1},$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ and $\{\varepsilon_i \,|\, i \geq 0\}$ are independent and $\mu$ is the expected value of the log-shock. In order to approximate the dynamics of the profit shock process, we apply the quadrature method of Tauchen (1986). This yields an approximation of the above AR(1) process by a Markov chain on a finite state space.

**Capital shocks.** Shocks to real assets may follow a variety of distributions depending on the nature of the shock itself (e.g., natural v.s. man-made). A useful parametrization of capital shocks is provided by the MBBEFD distribution function, see Bernegger (1997), which is flexible enough to allow for different loss behaviours.[26]

In particular, we assume that $z^k \sim F_{b,g}(x)$ with parameters $b, g \geq 0$ and cumulative distribution function given below

[26] The Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac (MBBEFD) distribution function has been applied in the reinsurance industry for calculating the expected indemnity payment for reinsured insurance contracts on property losses.

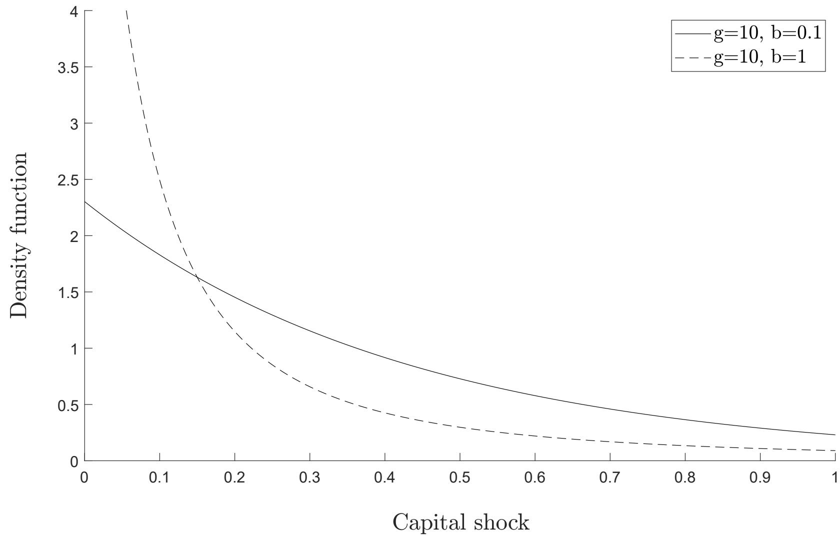



**Figure 3.4: Capital shock distribution.** The two curves illustrate the probability density functions for two parametrizations of the capital shock ($z^k$), with $z^k \sim F_{b,g}(x)$, $b, g \geq 0$ (see Equation (3.15)). Note that only the continuous part of the density functions is presented.

$$F_{b,g}(x) = \begin{cases} 1 & x \geq 1, \\ 0 & x < 1 \text{ and } (g = 1 \text{ or } b = 0), \\ 1 - \frac{1}{1+(1-g)x} & x < 1, b = 1, g > 1, \\ 1 - b^x & x < 1, bg = 1, g > 1, \\ 1 - \frac{1-b}{(g-1)b^{1-x}+(1-gb)} & x < 1, b > 0, bg \neq 1, g > 1. \end{cases} \tag{3.15}$$

Note, in particular, that the random variable $z^k$ has a point mass in 1. Moreover, $z^k$ takes values on the unit interval, which allows us to interpret capital shocks realizations as losses per unit of capital when the capital loss function is linear in $z^k$ (see Table 3.1). Figure (3.4) shows the probability density functions of the continuous part of the capital shock distribution, for two different parametrizations. The dotted line refers to a MBBEFD function such that $b = 1, g > 1$, which generates a loss distribution with more density mass concentrated on lower shock realizations. The continuous line is obtained by assuming $b = 0.1, g = 10$, such that $bg = 1, g > 1$, which describes the distribution of capital shocks with a heavier tail. In the numerical results we will investigate the impact of these capital shock distributions on the relative value added by property insurance.

**Bellman equation.** The piece-wise linear approximation to the value and policy function was obtained using the policy iteration approach; see Hernández-Lerma and Lasserre (2012), Chapter 8. The method is based on iterating the Equation (3.14) and evaluating the right-hand side

$$V_{n+1}(c,k;z) = \sup_{(i,I,e,b,d)\in\mathcal{P}_0(c,k;z)} \max\Big\{ d - e - C_e(e) + \frac{1}{1+r}\mathbb{E}\big[V_n(c',k';z')\big], \\ c + k - C_i(-k,s), 0\Big\}$$

on each point $(c,k;z)$ of the state space and all values of the shock, where $n \in \{1,2,\dots\}$ is the iteration step. The general theory implies that the value function is uniquely defined as a fixed point of the Bellman equation; see (3.14). Furthermore, the iteration above converges to the approximation of the true value function and the optimal policy is the policy that attains the maximum.

Although the value function is unique and the procedure above converges to the true value function, the optimal policy need not be. Indeed, a sufficient condition for uniqueness of the optimal policy would be strict convexity of the map

$$(i,(I,\pi(I,\widetilde{k})),e,b,d) \mapsto d - e - C_e(e) + \frac{1}{1+r}\mathbb{E}\big[V(c',k';z')\big].$$

Yet, this does not hold in general. In our implementation, if the policy is non unique, we always choose the smallest value of the control. The code was written in Julia.

**Parameter values and functional forms.** Table 3.1 provides the parameter values used in the base case model. The first column identifies the components of the model, grouped in different panels. The second and third columns show respectively the actual functional form used for the functions described in Section 3.2 and the values of the model parameters adopted in the base case. The last column specifies the relevant references that motivated the choices reported in columns two and three. The values chosen for the base case are supposed to refer to a firm that is constrained by various financial and investment costs (see e.g., Gamba and Triantis (2008)).

**Table 3.1:** Base-case parameter values and functional forms.

| Model component | Functional form | Parameter value | References |
|---|---|---|---|
| Operating profits & profit shocks | | | |
| *Operating profits* | $F(\widetilde{k}, z) = z^p \widetilde{k}^\theta$ | $\theta = 0.45$ | Hennessy and Whited (2005) |
| $z^p$ *dynamics* | $\log z^p_{t+1} = \mu + \rho \log z^p_t + \varepsilon_{t+1},\ \varepsilon \sim \mathcal{N}(0, \sigma^2)$ | $\mu = 0,\ \rho = 0.62,\ \sigma = 0.15$ | Gamba and Triantis (2008) |
| *Unsecured fixed cost of production* | | $f = 0.8$ | Gamba and Triantis (2008) |
| Capital dynamics | | | |
| *Capital losses* | $\ell(\widetilde{k}, z) = z^k \widetilde{k}(1-\delta)$ | | Hau (2006a) |
| $z^k$ *dynamics* | $z^k \sim F_g(x)$ with $g > 1$, $z^k$ is *i.i.d.*, $F_g(x) = \begin{cases} 1 & \text{if } x \geq 1, \\ 1 - \frac{1}{1+(g-1)x} & \text{if } x \in [0,1). \end{cases}$ | $g = 10$ | Bernegger (1997) |
| *Depreciation rate* | | $\delta = 0.1$ | Hennessy and Whited (2005) |
| Investment costs | | | |
| *Capital adjustment costs* | $\phi(i) = \phi_0 i^2/(2\widetilde{k}) + \phi_1 \widetilde{k} \mathbb{1}_{i \neq 0}$ | $\phi_0 = 0.01, \phi_1 = 0$ | Strebulaev and Whited (2012) |
| *Liquidation value for assets sale* | | $q = 0.75$ | Gamba and Triantis (2008) |
| *Fire-sale discount for assets sale* | | $q_d = 0.5$ | Gamba and Triantis (2008) |
| External issuance costs | | | |
| *Equity issuance costs* | $C_e(e) = \eta_0 e + \eta_1 e$ | $\eta_0 = 0.08, \eta_1 = 0.06$ | Gomes (2001), Gamba and Triantis (2008) |
| *Debt issuance costs* | $C_b(b) = \xi b$ | $\xi = 0.02$ | Gamba and Triantis (2008) |
| Corporate taxes | | | |
| *Tax function* | $T(p) = \begin{cases} \int_0^p T_m(\xi) d\xi & \text{if } p \geq 0, \\ 0 & \text{if } p < 0. \end{cases}$ | $\overline{T} = 0.39,\ \mu_T = 0,$ $\sigma_T = 2.3$ | *Parameter values*: own calculations |
| | $T_m(\xi) = \overline{T} \cdot \Phi(\xi, \mu_T, \sigma_T)$, $\Phi$: Normal c.d.f | | Hennessy and Whited (2007) |
| Cash holdings and insurance | | | |
| *Risk-free borrowing rate* | | $r = 0.05$ | Gamba and Triantis (2008) |
| *Return on cash holdings* | | $r_c = 0.03$ | Bolton et al. (2011) |
| *Insurance loading* | | $\lambda = 0.1$ | Zhang and Nielson (2009), Jiang et al. (2021) |

### 3.5.2 Simulation results

As our model does not have a closed-form solution, we present the predictions of the model based on numerical results. In order to understand the drivers of corporate demand for property insurance in a dynamic setting, we analyze three measures in steps: the value of financial flexibility (*VFF*), the added value of insurance (*AVI*) and a modified definition of Shapely values (*MSV*).

Financial flexibility refers to the degree to which firms can timely finance profitable business opportunities at reduced costs and avoid liquidation and financial distress costs when unexpected shocks occur (Gamba and Triantis (2008)). By limiting the actions a firm can attain, financing constraints and investment costs reduce firms' financial flexibility and thus their value, inducing firms to behave as risk averse agents. It follows that firms with larger frictional costs and lower financial flexibility have the potential to seek property insurance to relax their financing constraints (Mayers and Smith (1982)). To capture the value of financial flexibility and, in turn, firms' appetite for property insurance, we measure the value loss suffered by a firm subject to various frictional costs compared to the same firm under perfect markets ($VFF$), assuming no access to insurance markets.

When firms with costly financing can purchase property coverage, the insurance indemnity provides contingent capital that lowers the expected costs of financial distress (Garven and MacMinn (2013)), relaxes collateral constraints (Caillaud et al. (2000)) and reduces reliance on costly external financing (Hau (2006a)).

As a result, firms' financial flexibility increases and so does their value. To measure the value added by corporate insurance for firms subject to costly internal and external financing, capital liquidation costs and taxes, we analyze the percentage of firm value that is recovered by purchasing insurance relative to the value lost under different frictions ($AVI$). Finally, we are interested in how property insurance ranks among the set of financing choices that are available to the firm to improve its financial flexibility. In principle we could calculate the added value of every financial policy and use it to find how property insurance compares to other decisions. However, the added value ignores how the marginal impact of each policy on firm value depends on the interactions among other financing decisions available to the firm. Due to the complexity of the problem, we resort to a modified definition of the Shapley value (Shapley (1953a)) to gauge the average marginal relative improvement in firm value brought about by each financing decision. This measure allows us not only to capture the average marginal increase in firm value generated by property insurance, but also to determine the implied ranking among different financing channels based on their relative contribution to firm value.

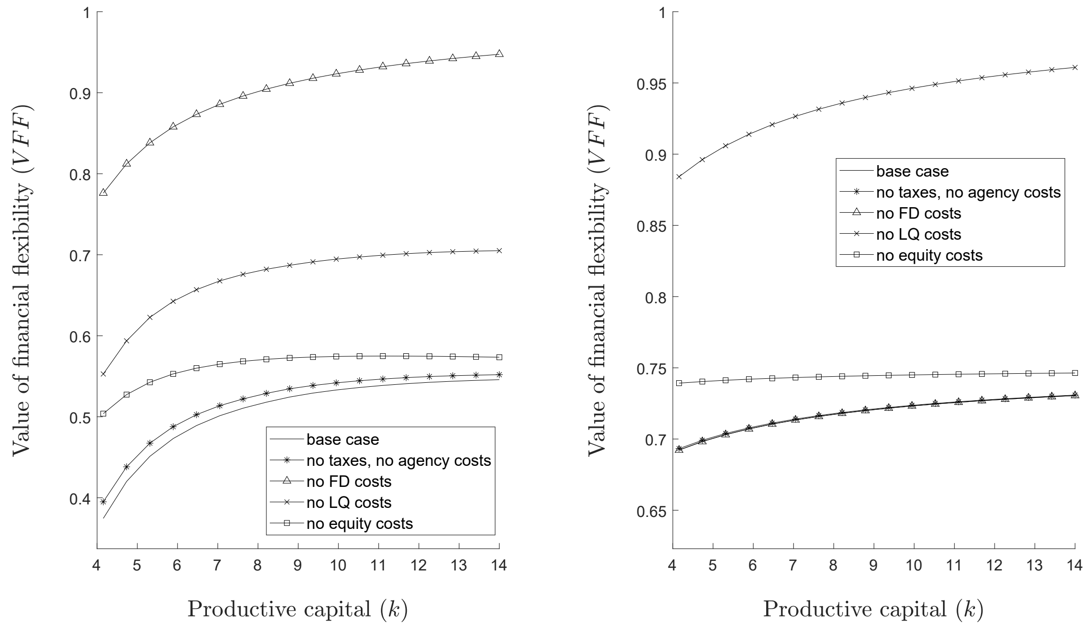
### 3.5.3 The cost of financial and investment frictions

To study the potential incentives for firms with costly financing to purchase property insurance, we investigate the effect of different frictional costs on their financial flexibility. To this aim, we consider firms without access to insurance markets, for which both profit and capital shocks represent pure background risks that cannot be hedged. We define the value of financial flexibility as the percentage loss in firm value due to taxes, financing and investment costs relative to an otherwise identical firm with no such costs (Gamba and Triantis (2008))

$$VFF = \frac{V_F}{V_0}. \tag{3.16}$$

$V_F$ is the value of a firm subject to frictional costs and no access to insurance markets, while $V_0$ is the value of a firm operating in perfect markets (*unconstrained* firm). As a consequence of the irrelevance result in Modigliani and Miller (1958), insurance matters for the value of a firm only if market imperfections exist, such as agency problems and asymmetric information. By comparing the value of a firm with and without market imperfections, $VFF$ is informative of the firm value lost due to frictional costs and thus of the potential relevance of insurance under various frictions.

Figure 3.5 and 3.6 show how the value of financial flexibility changes for firms with different capital levels, profitability and initial financial conditions. Focusing on firms without access to internal financing ($c \leq 0$), Figure 3.5 addresses the question: what is the loss in firm value due to financing costs for firms with different sizes of productive capital? Similar to the findings in Gamba and Triantis (2008), both panels show that small firms bear the largest loss of value, independently of the type of financing costs. This is because firms with low levels of productive capital are also those that are more constrained in their financing choices due to a lower cash value of capital and thus tighter collateral constraints. Panel (a), in which $c < 0$, shows that financing frictions are especially costly for small firms in financial distress: not only outstanding liabilities lower the value of the firm compared to an otherwise firm without liabilities (see Panel (b)), but they also induce a stronger aversion towards



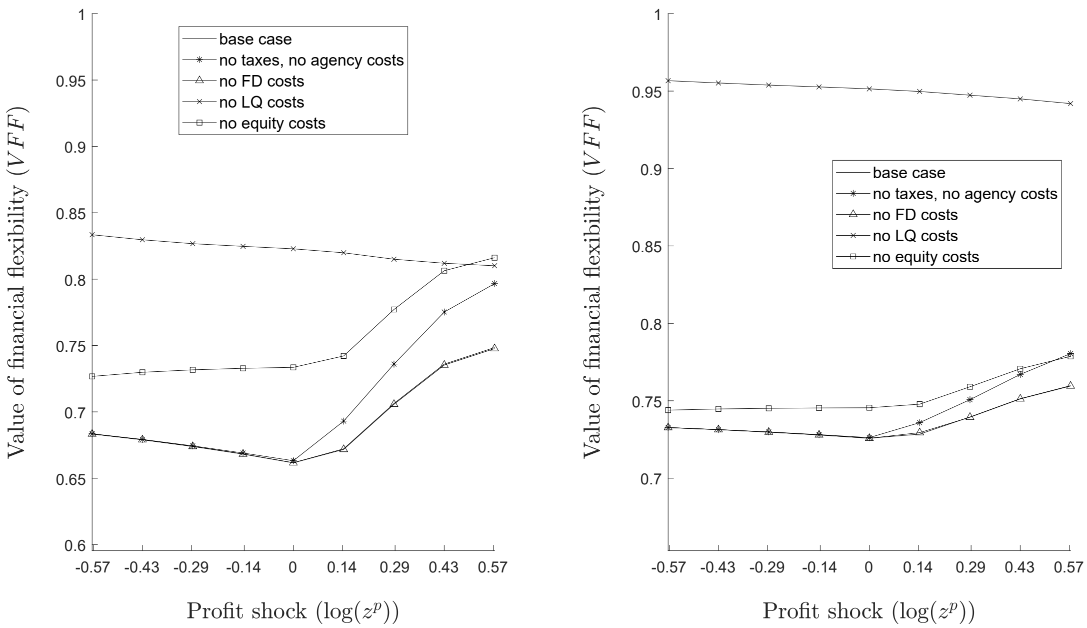

**(a)** Financial distress ($c < 0$).

**(b)** Business as usual ($c = 0$).

**Figure 3.5: Value of financial flexibility vs. productive capital.** The figure shows the value of a firm subject to costly financing relative to the value of the same firm with no financing costs, for different levels of productive capital ($k$). Each plot shows five cases: the *base case*, *no taxes nor agency costs* ($\overline{T} = 0, r_c = r$), *costless financial distress (FD)* ($q_d = 1$ but $q < 1$), *costless capital liquidation (LQ)* ($q_d < 1$ and $q = 1$) and *no equity costs* ($C_e(e) = 0, \forall e \geq 0$). The profit shock realization is fixed and set to $z^p = 1$, or $\log(z^p) = 0$. Panel (a) refers to firms with outstanding liabilities ($c = -2$). Panel (b) is based on a firm with $c = 0$. Base-case parameter values are as in Table 3.1.

risk as suggested by the steeper curvature of $VFF$ for low values of $k$.

The value of financial flexibility is also associated with the type of frictional costs. Naturally, the value of a firm in the base case is the most constrained compared to the other cases illustrated in Figure 3.5, in which the firm has unconstrained access to some forms of financing. For instance, the costs of internal financing due to taxes and agency costs or equity issuance costs account for a relatively small percentage of the firm value lost in the base case. On the contrary, firm value is highly impacted by the costs associated with the divestment of productive capital. If the firm incurs liquidation costs when divesting capital, then the lower cash value of capital will limit the firm's debt capacity and decrease the proceeds from liquidating capital in the next period. This in turn increases the firm's expected financial constraints that stem from relying on alternative costly financing resources. Interestingly, firms with outstanding liabilities (Panel (a)) will suffer a considerable loss in value from the current and expected fire-costs due to capital divestment in financial distress. On the other hand, 'business-as-usual' firms are barely affected by financial distress costs: their value is mostly affected by liquidation and equity issuance costs (Panel (b)).




**(a)** Small firm ($k = 2.4$). **(b)** Large firm ($k = 11.1$).

**Figure 3.6: Value of financial flexibility vs. profitability.** The figure shows the value of a firm subject to costly financing relative to the value of the same firm with no financing costs, for different levels of profitability ($z^p$). Each plot shows five cases: the *base case*, *no taxes nor agency costs* ($\overline{T} = 0, r_c = r$), *costless financial distress (FD)* ($q_d = 1$ but $q < 1$), *costless capital liquidation (LQ)* ($q_d < 1$ and $q = 1$) and *no equity costs* ($C_e(e) = 0, \forall e \geq 0$). Cash holdings are fixed and set to $c = 0$. Panel (a) refers to a small firm ($k = 2.4$); Panel (b) is for a firm with $k = 11.1$. Base-case parameter values are as in Table 3.1.

Figure 3.6 illustrates how the profitability level, or equivalently the realized profit shock, affects $VFF$ for firms with either small (Panel (a)) or large (Panel (b)) productive capital stocks. Intuitively, financial flexibility is more valuable for firms with a lower capital stock and in periods of persistent low productivity, in which cash flows from operating profits are modest and are expected to remain low. This is apparent in Panel (a), which shows that small and highly profitable firms are able to significantly improve their financial flexibility by benefiting from higher current and future productivity to lower the expected costs of equity issuance, cash holdings and capital divestment. Figure 3.6, however, shows that profitability is relevant for improving financial flexibility when firms are subject to liquidation costs: if capital is perfectly flexible ($q = 1$), then the loss in firm value is almost insensitive to current profitability, which suggests that current (and expected) profit shocks realizations become relevant for firms' financial flexibility via their mitigation of expected fire-sale costs.

**Remark 3.5.1.** *Figure 3.5 and 3.6 suggest that the value of financial flexibility is highest for firms in financial distress, with low profitability and exposed to fire-sale and capital liquidation costs. Moreover, small firms face tighter financial constraints due to their limited productive capacity. Hence, we expect property insurance to be more valuable precisely for these types of firms for which financial frictions are especially costly. We investigate the value added by property insurance in the next section.*

### 3.5.4 The value of corporate property insurance

Since firms with unhedgeable exposure to capital shocks lose much value due to different types of frictions, it is natural to investigate how property insurance helps to restore the lost firm value. In this section, we address the question of which frictions actually drive corporate demand for insurance, as well as to which degree property insurance improves firms' financial flexibility.

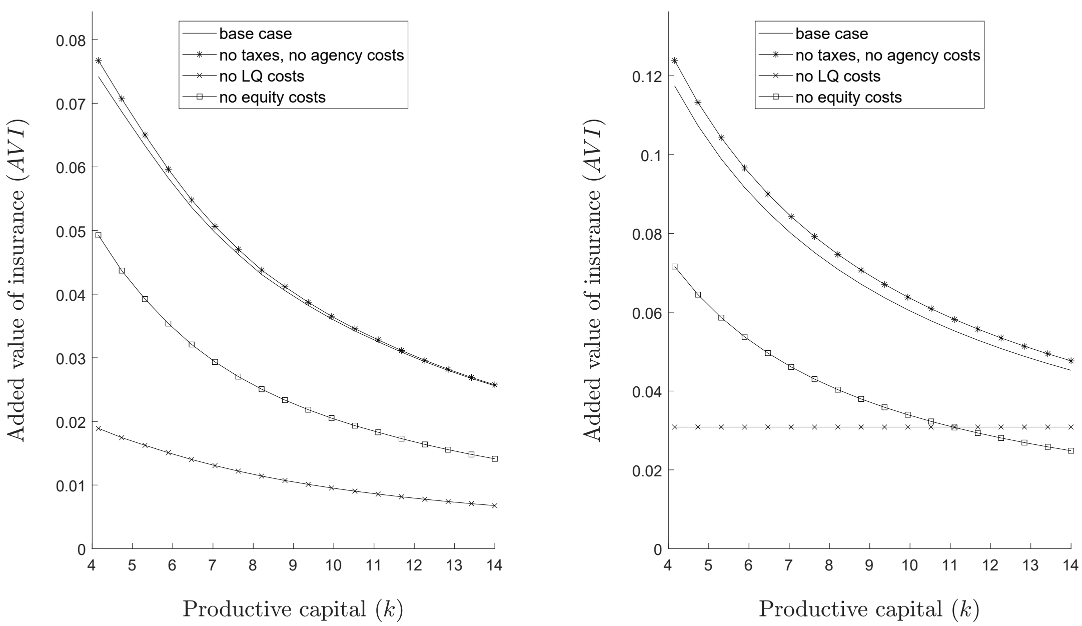
To find the value added by property insurance, we construct a measure that captures the proportion of firm value lost due to costly financing, capital costs and taxes that is recovered thanks to access to insurance. We define the added value of insurance as the value that property insurance adds to an otherwise identical firm with no access to insurance, divided by the loss in firm value due to frictional costs, namely

$$AVI = \frac{V_I - V_F}{V_0 - V_F} \geq 0, \tag{3.17}$$

where $V_I$ denotes the firm value when insurance is an available financing choice, $V_F$ is the firm value without access to insurance markets and $V_0$ is the value of an unconstrained firm. It is noteworthy that the added value of insurance can be related explicitly to the value of financial flexibility as follows

$$AVI = \frac{\frac{V_I}{V_0} - VFF}{1 - VFF}, \tag{3.18}$$

which is readily obtained by dividing each factor in Equation (3.17) by $V_0$. Intuitively, the previous expression shows that the ex-ante benefit of insurance as captured by $AVI$ is higher when the firm is subject to larger financing costs, or equivalently when financial flexibility is low (i.e., low $VFF$).



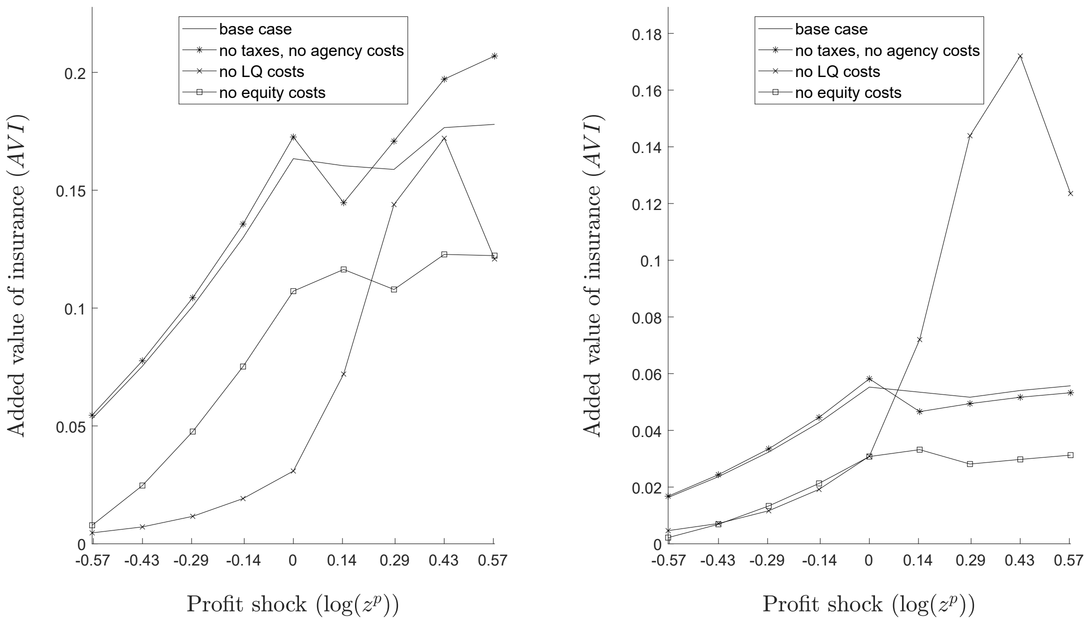

**(a)** Financial distress ($c < 0$). **(b)** Business as usual ($c = 0$).

**Figure 3.7: Added value of insurance vs. productive capital.** The figure shows the proportion of firm value lost due to costly financing that is recovered after purchasing property insurance, for different levels of productive capital ($k$). Each plot shows four cases: the *base case*, *no taxes nor agency costs* ($\overline{T} = 0, r_c = r$), *costless capital liquidation (LQ)* ($q_d = 0.75$ and $q = 1$) and *no equity costs* ($C_e(e) = 0, \forall e \geq 0$). The profit shock realization is fixed and set to $z^p = 1$, or $\log(z^p) = 0$. Panel (a) refers to firms with outstanding liabilities ($c = -2$). Panel (b) is based on a firm with $c = 0$. Base-case parameter values are as in Table 3.1.

Figure 3.7 shows the added value of property insurance for firms with different capital levels, given either outstanding liabilities (left-hand panel) or no cash holdings (right-hand panel). Despite the value of financial flexibility is higher in financial distress (Figure 3.5, Panel (a)), the value restored through insurance coverage is lower for firms with outstanding liabilities (Figure 3.7, Panel(a)) than for business-as-usual firms. Since firms with lower financial flexibility are expected to extract higher benefits from property insurance, this result suggests that firms in financial distress may retain higher exposure to property losses due to their limited funds compared to an otherwise financially healthy firm. Financial distress forces firms to devote part of their funds to repay outstanding liabilities and thus either reduces the resources available to purchase insurance coverage or increases the cost of affording property insurance.

Independently of the initial financial condition, property insurance generates more value for firms that are less financially flexible, in line with the intuition gained from Figure 3.5. The costs of liquidating productive capital and issuing equity drive most of the loss in firm value that property insurance helps to recover. Moreover, insurance coverage generates higher benefits for firms with a small productive capital stock, for which marginal productivity is higher. Panel (b) of Figure 3.7, however, shows that this observation is driven by firms with inflexible capital: when capital can be sold at book value ($q = 1$), property insurance restores the same fraction of lost value irrespective of firm size.

Remark 3.5.1 suggested that small firms with low productivity and inflexible capital should benefit more from property insurance. It is therefore natural to ask whether such firms actually recover, and to which degree, the value lost due to frictional costs through insurance. Figure 3.8 shows how productivity affects $AVI$ for small and large firms with no internal financing. Panel (a) indicates that, as expected, property insurance helps to restore a higher fraction of lost firm value for small firms with costly capital liquidation and equity issuance costs, especially for low productivity levels. Moreover, as in Figure 3.7, this result is explained by the presence of capital liquidation costs: if capital can be sold at no cost ($q = 1$), then firms benefit from property insurance to the same degree regardless of their size. It is well-known that insurance demand by strictly risk averse agents is unaffected by background risk when insurable and non-insurable risks are independent (Doherty and Schlesinger (983a)). Since in our model capital shocks and profit shocks are independent, the latter being the source of background risk, one might be tempted to conclude that productivity is relevant for corporate property insurance only through its effect on the value of financial flexibility. For example, Figure 3.6 would suggest that small firms with lower productive assets would benefit the most from purchasing insurance due to their greater loss in firm value. Nonetheless, Figure 3.8 shows that the value recovered through insurance is increasing in firms' productivity, i.e. highly productive firms recover a higher fraction of firm value due to property insurance coverage. This




**(a)** Small firm ($k = 2.4$). **(b)** Large firm ($k = 11.1$).

**Figure 3.8: Added value of insurance vs. profitability.** The figure shows the proportion of firm value lost due to costly financing that is recovered after purchasing property insurance, for different levels of profitability ($z^p$). Each plot shows four cases: the *base case*, *no taxes nor agency costs* ($\overline{T} = 0, r_c = r$), *costless capital liquidation (LQ)* ($q_d = 0.75$ and $q = 1$) and *no equity costs* ($C_e(e) = 0, \forall e \geq 0$). Cash holdings are fixed and set to $c = 0$. Panel (a) refers to small firms ($k = 2.4$) and Panel (b) to firms with a large capital stock ($k = 11.1$). Base-case parameter values are as in Table 3.1.

result is due to the persistence in the process driving the profit shocks, which can positively affect corporate demand for property insurance even if capital and profit shocks follow independent processes. This effect arises because background risk in our model is not i.i.d.: periods with high productivity are likely to persist, increasing the marginal expected productivity of capital in the next periods.

**Remark 3.5.2.** *Using the added value of insurance as a metric to gauge the benefits of property coverage, our model suggests that firms may be induced to purchase property insurance to protect their productive capital and take advantage of higher future expected operating profits. Moreover, inter-temporal dependence in the source of background risk produces an effect similar to assuming positive dependence between an i.i.d. background risk and the insurable risk (Doherty and Schlesinger (1985)); property insurance becomes more valuable as firms' profitability increases. Hence, the* *volume of property insurance underwritten is expected to rise during periods of economic expansion, when firms' growth opportunities are larger. Moreover, heterogeneity in the benefits of property insurance, as measured by AVI, is driven by inflexible capital: small firms will benefit more from insurance coverage only when subject to capital liquidation costs.*

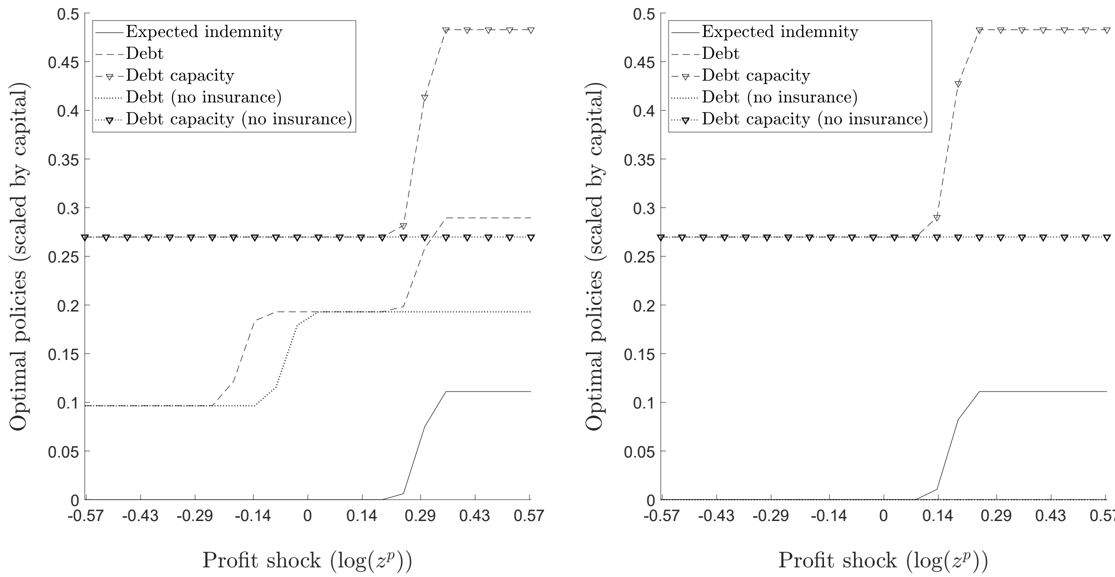
### 3.5.5 Debt capacity, financial slack and property insurance

The previous section has shown that property insurance helps financially constrained firms to restore part of their financial flexibility, especially when having access to costly external financing and inflexible productive assets. In this respect, property insurance acts as an additional financing channel providing contingent capital that lowers the firm's expected financing costs.

Another function of property insurance is related to collateral constraints: higher property coverage increases the value of the assets available to pay off debt, thus rising the debt capacity of the firm (see Figure 3.2). To what extent is property insurance purchased to relax collateral constraints?

Figure 3.9 shows how insurance demand, debt issuance and collateral constraints change for different profitability levels and initial financial conditions (FD v.s. BAU). Note that collateral constraints (lines with lower triangles) are independent of the firm's initial financial condition and thus are the same for both panels. As collateral constraints are also independent of current profitability, the firm's increased debt capacity at higher profitability levels is due to higher demand for property insurance and investment in the capital stock. Positive profit shocks increase current liquidity and improve the firm's expected profitability in the next periods, thus leading the firm to purchase more property insurance in periods of improved profitability.

However, Figure 3.9 suggests that the ultimate channels driving insurance demand might differ for firms with a different initial financial condition. Panel (b) shows that financially healthy firms may increase their property coverage without issuing debt. In this case the firm's debt capacity rises mechanically as a result of higher insurance coverage, which however is not purchased to improve the firm's collateral constraints. Instead, Panel (a) shows that firms in financial distress may be willing to buy property insurance, among other things, to increase their debt capacity. This claim follows from noting that, in addition to purchasing insurance, such firms issue more debt than it




**(a)** Financial distress ($c < 0$). **(b)** Business as usual ($c > 0$).

**Figure 3.9: Debt capacity, financial slack and property insurance.** The figure shows debt capacity and debt issuance (dashed lines) and expected indemnity (solid line) for a firm with access to insurance markets, as well as debt capacity and debt issuance (dotted lines) for an otherwise identical firm that cannot purchase property insurance. Financial slack is the difference between debt capacity and actual debt issued. Each variable is scaled by the capital stock and is plotted against the natural logarithm of the profit shock. The optimal policies and debt capacity are for a firm with capital $k = 4.14$. Panel (a) refers to a firm in financial distress ($c = -4$) and Panel (b) to a business-as-usual firm ($c = 8$). Base-case parameter values are given in Table 3.1.

would be allowed by the collateral constraints of an otherwise identical firm with analogous borrowing but no access to insurance markets.

Property insurance can thus help firms not only to harvest future expected business profits from improved profitability (see Figure 3.8), but also to relax current collateral constraints in times of financial distress. Parallel to this observation, it is worth noting that in line with the predictions of dynamic investment models (see e.g., Hennessy and Whited (2005), Strebulaev and Whited (2012)) firms tend to maintain some financial slack by issuing less debt than what is allowed by collateral constraints. In a dynamic setting this behavior is due to firms trying to preserve their financial flexibility by avoiding to exhaust their current debt capacity and lower the expected costs of financial distress. It follows that compared to a firm with no access to insurance markets, firms that purchase property insurance are able to increase their debt capacity, and potentially debt issuance, while retaining the desired financial slack.

### 3.5.6 Shapley values: ranking the firm's financing options

In this section we are interested in understanding how property insurance ranks among the set of financing choices available to the firm. More specifically, how valuable is property insurance to the firm relative to internal and external financing? We begin by noting that the marginal impact of each firm's decision on firm value depends on the actual number and types of financing channels available to the firm. For instance, our analysis in Section 3.5.4 relies on comparing the value of a firm with access to insurance markets to an otherwise identical firm without property insurance coverage. However, such a comparison ignores how the added value of insurance changes when the firm has access to larger or smaller subsets of alternative financing channels. As a result, we need to find another approach to measure how valuable each financing channel is to the firm. We propose the use of Shapley values, appropriately modified to express the average marginal *relative* improvement in the value function for a certain financial policy, to study the relative importance of firms' financing choices under different scenarios and frictional costs.

Since its first introduction, the Shapley value (Shapley (1953a)) has been applied in the game theory and economics literature as the main solution concept to cooperative games for problems of revenue sharing and cost allocations. Given a cooperative game in which the cooperation of all players (grand coalition) generates a total surplus, the Shapley value provides a unique distribution of the surplus among the players based on the average marginal contribution of each player to the surplus across all feasible coalitions (Hart (1989)). Among its many properties, the Shapley value is symmetric, namely it treats each player symmetrically by assigning the same value to players who provide the same contribution in every coalition. However, as players might differ in some characteristics (such as bargaining power or effort) that are not measurable through their individual contribution, the literature has developed a family of *weighted* Shapley values. Introduced in Shapley (1953b), a weighted Shapley value assigns to every player a positive weight that aims to capture the relative importance of that player to the total surplus of the cooperative game being modeled.[27] Typically the assignment of the weights is game specific, varying according to the features of the game (e.g., asymmetric information across players). In our setting, however, it is not clear a priori how to specify a distribution of weights across the firm's decisions; we thus leave this line of inquiry for future research.

Shapley values have been proposed in the context of explainability in machine learning in Lundberg and Lee (2017). The machine learning models often employed today are too complex to allow for an easy interpretation, so considerable effort is devoted to making interpretation of the models possible. Today there is a number of global approaches, in line with the corporate finance literature, where one studies the impact of introducing new controls, and local approaches, in the sense of comparative statics. The reason for proposing a new approach, the Shapley value, is that it has a rigorous game-theoretic interpretation. Beyond that, it is important to note that the importance of each individual control depends on the state. This implies that addressing the relevance of each decision by introducing new controls in a step-wise fashion, as usually applied in the corporate finance literature, would allow us only to qualitatively analyze the marginal impact of new controls. In our setting, Shapley values allow us to assign a numerical value to each control in every state. We start by presenting a modified definition of Shapley value as we use it in our analysis.

**Definition 3.5.3** (modified Shapley value, *MSV*)**.** *Denote by $\mathcal{C}$ the set of $n$ controls available to the firm. For $P \subset \mathcal{C}$ we denote the value function with only controls from the set $P$ by $V(c,k;z|P)$. The modified Shapley value $MSV_i(c,k;z)$ for control $i \in \mathcal{C}$ in the state $(c,k;z)$ is defined by*

$$MSV_i(c,k;z) = \sum_{P \subset \mathcal{C}\setminus\{i\}} \frac{|P|!(n-|P|-1)!}{n!}\left(\frac{V(c,k;z|P\cup\{i\})}{V(c,k;z|P)} - 1\right),$$

*where by $|P|$ we denote the cardinality of the set $P$, i.e. the number of its elements.*

Let us first note that in the above definition we define the Shapley values in the spirit of the original definition, though our metric of value change is not additive, i.e. $V(c,k;z|P\cup\{i\}) - V(c,k;z|P)$, as in the classical definition. This is because not

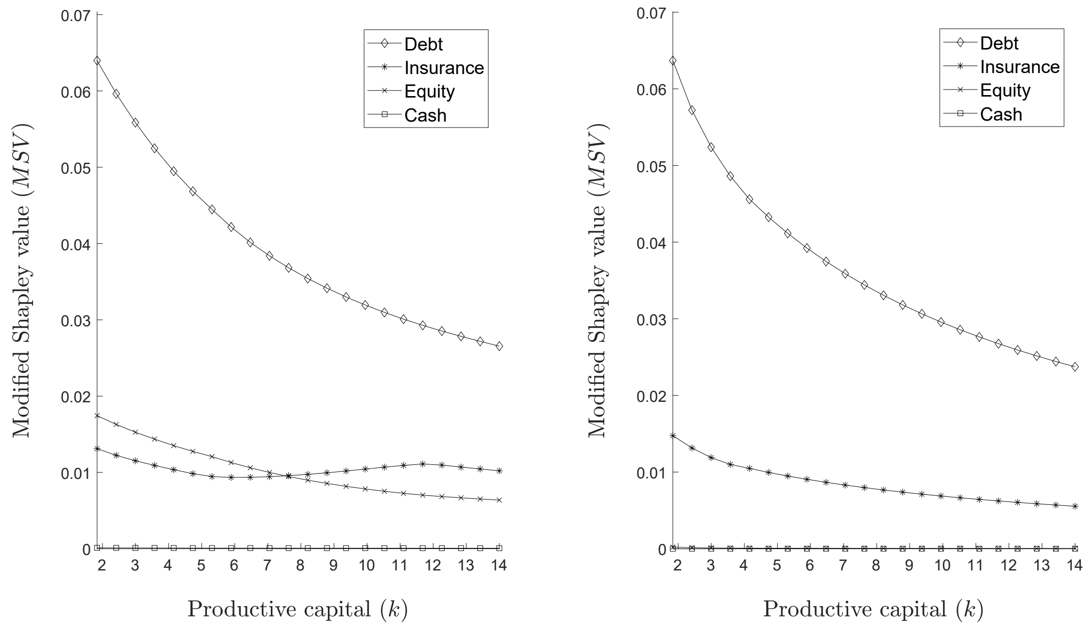
[27] A weighted Shapley value reverts to the standard Shapley value, in which every player is symmetric, when all the weights are the same (Kalai and Samet (1987)). Extensions of the concept of weighted Shapley value include Owen (1968) and Owen (1972), which provided alternative interpretations of the weights associated with the players, and Kalai and Samet (1987), in which some players can be assigned a zero weight.

only is the relative change in value the relevant metric, but our definition also gives results that are easier to analyze. Indeed, the modified Shapley value measures the *average marginal relative* contribution to firm value of each firm's financing choice relative to other policies. As a result, this metric allows us to establish a meaningful comparison among alternative financing channels by taking into account how their interaction changes across different sets of financing decisions.

One immediately notices that the major weakness of the (modified) Shapley value is that it requires recalculation of the value function for each subset of the set of controls. In our approach to calculating the value function, the computation of Shapley values is not a problem. This is because the computation time rises by an order of magnitude for each new control that is introduced. This means that if we are able to calculate the value function with the full set of controls, then we are able to calculate also the modified Shapley value for each control.

Our results on the value of financial flexibility (see Section 3.5.3) are informative of the qualitative properties that we expect for the modified Shapley value. The observation that the loss in firm value rises at a faster rate for lower levels of productive capital, driven by the induced concavity of the value function (see Figure 3.5), suggests $MSV$ will inherit such a behaviour. Specifically, we expect to observe that firms with less financial flexibility will exhibit ever higher modified Shapley values, or equivalently will benefit comparatively more from access to alternative financing channels.

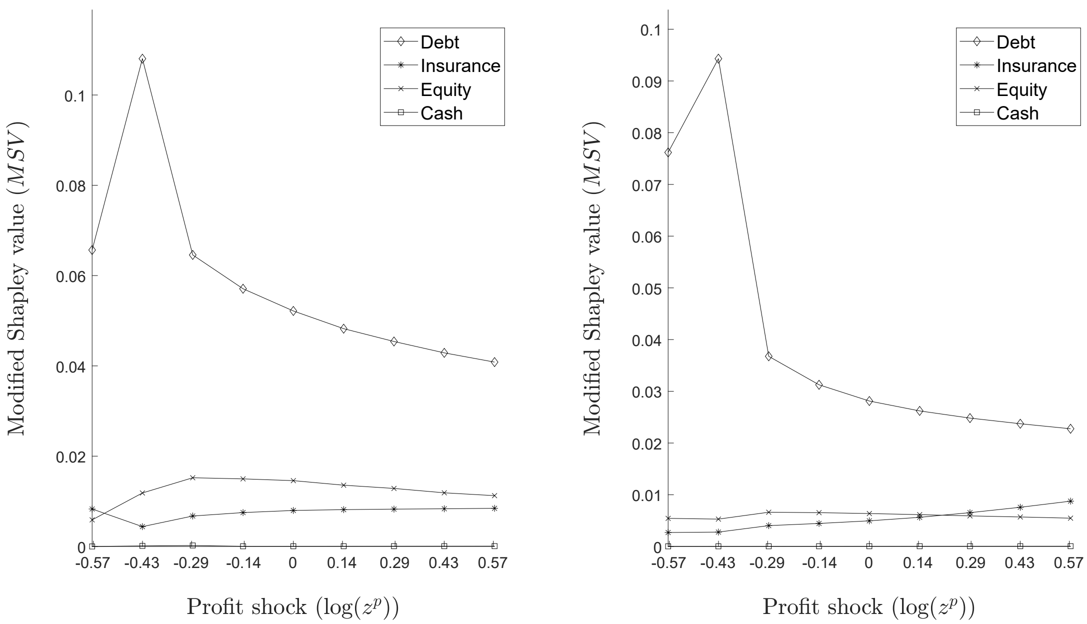
Figure 3.10 shows the effect of firm size on the modified Shapley values of debt issuance, external equity, cash holdings and property insurance for firms with profitable business opportunities but limited access to internal financing. A first, immediate remark is that modified Shapley values capture the fact that in our model a firm never finds it optimal to hold cash and borrow at the same time (see Proposition 3.2.7). In particular, throughout our numerical results we find that, independently of their size, firms gain little benefits on average from holding cash due to the presence of taxes and agency costs, while deriving much value from external financing by issuing debt (see e.g., Hennessy and Whited (2005), DeAngelo et al. (2011)). Furthermore, we confirm that the modified Shapley value mirrors the concavity of the value function: less financially flexible firms, which suffer greater losses in value due to frictional costs, are able to restore increasingly more value from access to additional financing channels than less constrained firms.




**(a)** Thin tailed distribution for the loss shock. **(b)** Heavy-tailed distribution for the loss shock.

**Figure 3.10: modified Shapley values vs. productive capital.** The figure shows the modified Shapley value in Definition 3.5.3 associated with different financing policies: debt issuance, property insurance, equity issuance and cash holdings. Each modified Shapley value is plotted as a function of the capital stock ($k$), given $c = 2$ and $\log(z^p) = 0.57$. Panel (a) refers to the base case model with $F_{b,g}(x)$ given in Table 3.1; Panel (b) shows $MSV$ for firms exposed to a loss with heavy-tailed distribution given by $F_{b,g}(x)$ with $g = 10, b = 0.1$. Base-case parameter values are as in Table 3.1.

Panel (a) in Figure 3.10 shows that the "pecking order" of a base-case firm, i.e. how its financing choices rank based on their $MSV$, depends on the size of productive capital. While debt adds the most value on average, the rank of equity and property insurance is reversed for small and large firms. We find that equity is more valuable to smaller firms than insurance, whereas for larger firms property insurance is more valuable than equity financing. While the literature has shown that equity financing is indeed a preferred choice for firms with limited sources of funds but high productivity shocks (Strebulaev and Whited (2012)), our model suggests that property insurance can provide an even more valuable financing channel for firms with much productive capital at stake. How does such ranking change when firms are exposed to capital losses with heavier tails? Panel (b) suggests that in this case all firms prefer to replace equity financing with property insurance and debt issuance, while the range of values attained by $MSV$ remains unaffected. The intuition for the increased importance of insurance among firms' financing policies is that the prospect of more extreme losses increases the marginal expected cost of capital losses, which in turn raises the marginal value that property insurance generates on average.




**(a)** Small firm ($k = 1.8$).

**(b)** Large firm ($k = 12.8$).

**Figure 3.11: modified Shapley values vs. profitability.** The figure shows the modified Shapley value in Definition 3.5.3 associated with different financing policies: debt issuance, property insurance, equity issuance and cash holdings. $MSV$ is shown as a function of firm's profitability ($z^p$), given $c = 8$. Panel (a) refers to a small firm ($k = 2.4$), whereas the Panel (b) to large firms ($k = 12.8$). Base-case parameter values are as in Table 3.1.

Figure 3.11 investigates the impact of background risk on the marginal value of financing policies when firms have access to sizable cash holdings, for firms with small and large capital levels. Panel (a) shows that small firms benefit from higher $MSV$ compared to larger firms, which is again a consequence of the induced concavity of the value function. Indeed, despite small firms having a highly productive capital, their limited financial flexibility prevents them from taking advantage of it, thus increasing the average value added by any financing channel that can help to relax their financial constraints. In general, both panels show that as firms become more profitable, the relevance of all financing policies decreases due to the likelihood of increased operating profits in the next periods. Moreover, debt adds the most value on average over the range of profitable profit shocks, though the comparative advantage of debt financing reaches its maximum when firms experience persistently low unexpected profits.

The effect of profitability on the value of equity financing and insurance is different for small and large firms. Panel (a) shows that for small firms equity financing remains more valuable than property insurance for any profitability level, despite property coverage gaining importance relative to equity issuance as small firms become more profitable. Interestingly, in Panel (b) we observe that the average marginal value contributed by equity finance and insurance is reversed for firms with larger productive assets. When firms with more capital at stake experience highly profitable shocks, property insurance is more valuable than equity issuance to capture future business profits while protecting productive capital against capital shocks. Overall, modified Shapley values suggest that the ranking of financing policies remains unchanged across productivity levels in spite of changes in their relative contribution.

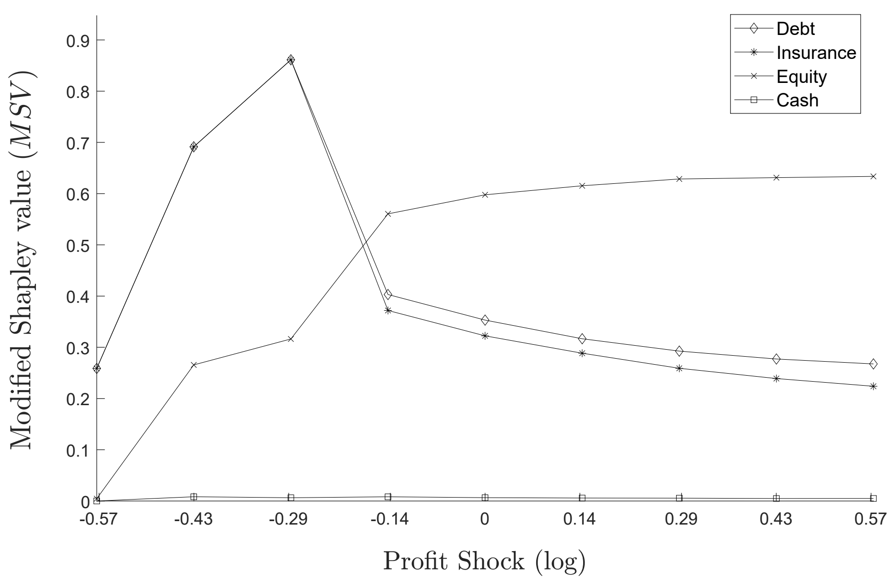



**Figure 3.12: modified Shapley values, fair premia and heavy-tailed capital losses.** The figure shows the modified Shapley value in Definition 3.5.3 associated with debt issuance, property insurance, equity issuance and cash holdings, when property insurance is priced at actuarially fair prices ($\lambda = 0$). The capital shock has a heavy-tailed distribution as in Figure 3.4 (solid line). $MSV$ is shown as a function of firm's profitability ($z^p$), for a firm with initial balance sheet $(c, k) = (0, 1.2)$. Other parameter values are as in Table 3.1.

As our main interest is the value contributed by property insurance, we study how the corresponding modified Shapley value changes when the firm has access to perfectly competitive insurance markets. This allows us to find the maximal contribution of insurance to firm value by increasing its comparative advantage relative to the rest of the financing channels. Since small firms suffer the most from the lack of financial flexibility, we report the average marginal value added by property insurance for small firms with no internal liquidity and exposed to heavy-tailed capital losses. Figure 3.12 shows that interesting dynamics emerge when highly constrained firms are exposed to capital shocks with a heavy-tailed distribution, for which the marginal expected value of property insurance is especially large. The pecking order of these firms depends on their productivity and is reversed when moving from low-productivity states to large profit shocks. Highly productive firms with a small capital stock prefer to finance their higher growth opportunities by issuing costly equity, while debt and insurance are less relevant. On the other hand, the prospect of periods with prolonged low-productivity raises the comparative advantage of insurance when property coverage is available at a fair premium. In this case, property insurance not only provides contingent cash flows that compensate for lower expected operating profits, but also relaxes the limited debt capacity of small firms. Effectively, the exposure of financially constrained firms to high-impact property losses induces a negative relation between property insurance and profitability that is observationally equivalent to the results obtained from models in which background risk and insurable risk are negatively dependent (Doherty and Schlesinger (1983), Schlesinger (2000)).

## 3.6 Conclusion

Our model complements the existing literature by introducing corporate demand for property insurance in a standard dynamic model of investment, which allows us to generate a wide range of predictions that account for the dependence of firms' decisions across time and over different sets of frictional costs.

We develop a dynamic model of investment in discrete time to investigate how property insurance helps to restore the loss in financial flexibility suffered by firms with costly financing, inflexible capital and taxes. We show that firms with low financial flexibility find it valuable to purchase property insurance to lower the future expected costs of financial distress and external financing, relax current collateral constraints and protect future growth opportunities. Our results suggest that highly profitable firms in financial distress may be willing to buy property insurance to expand their debt capacity and capture business opportunities.

We evaluate the model numerically and use a modified definition of Shapley value to rank property insurance among the set of financing channels available to the firm. Numerical results indicate that property insurance is especially attractive for small firms with limited access to external financing, even the more so when undergoing periods of low profitability and exposed to heavy-tailed capital losses. Moreover, despite profit and capital shocks being independent, we show that persistence in background risk can induce firms to purchase property insurance in a way that is observationally equivalent to the behavior suggested by static models in which background risk and insurable risk are dependent by assumption.

## 3.7 Appendix

### 3.7.1 On the state and action spaces

#### The firm's problem as a Markov Decision Problem

The optimization problem we are considering is an optimal control and stopping problem. In order to use the classical theory of Markov decision processes, we first need to expand the state space. Define $\mathcal{S}^*$ to be

$$\mathcal{S}^* := (\mathcal{S} \times Z) \cup \{*\},$$

where the state $*$ is an isolated state and represents the state of the firm after being closed. The decision to close the firm now becomes not a stopping time $\tau$ as a control, but a decision process $\{\xi_t\}$ that takes values in $\{\mathit{stop}, \mathit{cont}\}$. The value of the control $\xi = \mathit{cont}$ encodes the decision not to close the company and the value $\xi = \mathit{stop}$ encodes the decision to stop operations.

The set of actions available at each step to the manager is denoted by $\mathcal{P}_0(s^*)$ and depends only on the extended state $s^* \in \mathcal{S}^*$. In the case when the state is $*$, i.e. the firm has already closed, the firm has no controls $\mathcal{P}_0(*) = \varnothing$, but for simplicity of exposition, we will write instead $\mathcal{P}_0(*) = \{(i, e, b, d, I, \xi) = (0, 0, 0, 0, 0, \mathit{stop})\}$. In case the firm has not been liquidated yet, the state is given by the tuple $s^* = (c, k; z)$.

The set of actions are investment into the capital stock $i$, equity issuance $e$, debt issuance $b$, dividends $d$, insurance coverage $I$ and the liquidation decision $\xi$. If the firm decides to close, the only strategy to take is to liquidate the capital stock and distribute the liquid assets, if any, in the form of dividends. The available controls are, then, given by:

$$\mathcal{P}_0(c,k;z)^{stop} = \big\{(i,e,b,d,I)\,\big|\,i=-k,\ e=0,\ b=0,\ d=(c+k-C_i(-k,s))_+,\ I\equiv 0\big\}.$$

Thus, given the state $s$ and shock value $z$, if the manager decides to close the firm, there is a unique action to take. In the case when the manager decides not to close the firm, the set of available actions is given by

$$\mathcal{P}_0(c,k;z)^{cont} = \big\{(i,e,b,d,I)\,\big|\,0\le \widetilde{k}\le \overline{k},\ d,e\ge 0,\ I\in\mathcal{I},\ 0\le b\le \bar{b},\ 0\le \widetilde{c}\le \overline{c}\big\}.$$

Let us mention at this point a technical result concerning the set of policies that will become relevant for the proof of the dynamic programming equation.

**Lemma 3.7.1.** *The correspondence $\mathcal{P}_0$ is compact-valued.*

*Proof.* Because $*$ is an isolated point of $\mathcal{S}^*$ and the set $\mathcal{P}_0(*)$ is a singleton, there is nothing to prove; the graph is trivially closed. For $s^*\neq *$, the set of actions splits into a disjoint union of a finite number of sets

$$\mathcal{P}_0(s;z) = \bigcup_{\xi\in\{stop,cont\}}\ \bigcup_{I\in\mathcal{I}} \mathcal{P}_0(s;z)^{\xi}_{I=I},$$

thus, in the remainder of the proof, we will assume that the insurance indemnity is fixed, as well as the decision on whether to liquidate the firm.

In case the manager decides to liquidate the firm, there is a unique action available to the manager. Thus, the correspondence $(s;z)\mapsto \mathcal{P}_0(s;z)^{stop}$ is also trivially closed-valued.

The remaining case is that of fixed insurance indemnity and the manager deciding not to close the firm. In this case, it is easy to see that the set

$$\mathcal{P}_0(s)^{cont}_{I=I} = \big\{(i,e,b,d,I)\,\big|\,0\le \widetilde{k}\le \overline{k},\ d,e\ge 0,\ I\in\mathcal{I},\ 0\le b\le \bar{b},\ 0\le \widetilde{c}\le \overline{c}\big\}$$

is closed. □

**Remark 3.7.2.** *Note that $\mathcal{P}_0$ is never outer-semicontinuous.*[28] *Indeed, for that property to hold, the map $(c, k) \mapsto c+k-C_i(-k, s)$ — i.e. the dividend in the liquidation case — would need to be a continuous function. But this cannot be as long as we have disinvestment costs. The consequence is, thus, that the general theory would just guarantee Borel measurability of the value function, not its upper semicontinuity. However, the problem formulation gives us additional properties.*

On this, extended, state space we next need to describe the firm dynamics. Start from the observation that the state $*$ is absorbing, so if the current state is equal to $*$, it will remain in such state. Next, assume that the current state is $s \coloneqq (c, k; z) \in \mathcal{S}^*$. We have to consider two cases: If the firm decides to liquidate, the set of policies is equal to $\mathcal{P}_0(c, k; z)^{stop}$, specified above, but the next state $s' = *$. Lastly, if the firm decides not to liquidate, but chooses a policy from $\mathcal{P}_0(s)$, then the next state is given in the main text of the paper.

Finally, in this, extended, state and policy spaces, the value function is defined by

$$V(c, k; z) = \sup_{p \in \mathcal{P}(c,k;z)} \mathbb{E}\Bigg[\sum_{t=0}^{\tau-1} \frac{1}{(1+r)^t}(d_t - e_t - C_e(e_t)) + \\ + \frac{1}{(1+r)^\tau}\big(c_\tau + k_\tau - C_i(-k_\tau; s_\tau)\big)_+\Bigg].$$

### 3.7.2 Proofs of results from Section 3.2.3

*Proof of Proposition 3.2.7.* The proof of this proposition is a direct calculation. We will assume that the firm holds cash and issues debt at time 0; the general case follows by simple re-indexing. Denote by $(i, I, e, b, d)$ the firm policy at time 0. By assumption, $b > 0$ and the cash amount $\widetilde{c}$ after financial and investment decisions have been implemented is determined as follows

$$\widetilde{c}(b) = c - d + e + \frac{b}{1+r} - C_b(b) - i - C_i(i, s) - \pi(I, \widetilde{k}) > 0,$$

where we write the above 'uses and sources of funds' in the functional form as we

[28]The terminology surrounding set-valued maps is very non-uniform. We follow here the terminology of Rockafellar and Wets (2009). Outer-semicontinuous maps are called upper hemicontinuous by Aliprantis and Border (2013); upper-semicontinuous by Aubin and Frankowska (2009) and Hernández-Lerma and Lasserre (2012)

will consider what happens when we change the debt issuance policy. Consider the following perturbation of the debt issuance policy keeping everything else constant: Let $b > b^{\Delta} > 0$ be such that $\widetilde{c}(b^{\Delta}) > 0$; we can do this by the assumption on continuity of the function $C_b$. The value $b^{\Delta}$ presents a perturbation of the starting strategy $b$. We will show that the proposed perturbation is preferable to the original strategy.

From the properties of the debt issuance cost function $C_b$, we immediately see that
$$\widetilde{c}(b) - \widetilde{c}(b^{\Delta}) = \frac{b - b^{\Delta}}{1+r} - (C_b(b) - C_b(b^{\Delta})) \leq \frac{b - b^{\Delta}}{1+r}.$$
The latter estimate holds by the assumption that $C_b$ is increasing. At the end of the period, the taxable income of the firm is given by

$$P(b) = F(\widetilde{k}, z) - f - \pi(I, \widetilde{k}) - \delta\widetilde{k} - \ell(\widetilde{k}, z) + I\big(\ell(\widetilde{k}, z)\big) + r_c\widetilde{c}(b) - \frac{rb}{1+r},$$

with $b$ appearing only in the last two terms. Thus, at the end of the period, the capital balance of the two strategies is the same and the cash balance is given by

$$c'(b) = (1 + r_c)\widetilde{c}(b) + F(\widetilde{k}, z) - f + I\big(\ell(\widetilde{k}, z)\big) - T(P(b)) - b.$$

So,

$$\begin{aligned}
c'(b) - c'(b^{\Delta}) &= (1 + r_c)\big(\widetilde{c}(b) - \widetilde{c}(b^{\Delta})\big) - \big(T(P(b)) - T(P(b^{\Delta}))\big) - (b - b^{\Delta}) \\
&\leq \frac{1 + r_c}{1+r}(b - b^{\Delta}) - \big(T(P(b)) - T(P(b^{\Delta}))\big) - (b - b^{\Delta}) \\
&= \frac{r_c - r}{1+r}(b - b^{\Delta}) - \big(T(P(b)) - T(P(b^{\Delta}))\big) \\
&\leq \frac{r_c - r}{1+r}(b - b^{\Delta}) - \overline{T}\left(P(b) - P(b^{\Delta})\right) \\
&\leq 0.
\end{aligned}$$

The last estimate holds assuming

$$\beta r_c \overline{T} \leq \frac{r - r_c}{1+r}(1 - \overline{T}),$$

which would imply that if the taxes are high enough, one could issue more debt because of tax-deductible interests. The conclusion is that if a firm holds cash and issues debt at the same time, it is better to issue less debt. This proves the Proposition. □

*Proof of Proposition 3.2.9.* The claim of the Proposition is clear. Indeed, assume that the firm issues equity $e > 0$ and distributes dividends $d > 0$ simultaneously. Then, if the firm issued $d > d^{\Delta} > 0$ as dividends, such that also $e > d - d^{\Delta}$, then the firm dynamics would be precisely the same, but the cash flows to shareholders would be

$$d^{\Delta} - e^{\Delta} - C_e(e^{\Delta}) = d - e - C_e(e^{\Delta}) \geq d - e - C_e(e),$$

which is a higher value strategy. □

### 3.7.3 Proofs of results from Section 3.3

*Proof of Proposition 3.3.2.* We begin by recalling the debt capacity in Equation (3.8) for a firm that purchases a capped deductible contract $I(D, L; \ell(\widetilde{k}, z))$. For simplicity we denote debt capacity by $\bar{b}(D, L)$

$$\begin{aligned}\bar{b}(D, L) = \sup\{b \geq 0 \mid b \leq F(\widetilde{k}, z) + &I(D, L; \ell(\widetilde{k}, z)) - \\ &- T(P) + k' - C_i(-k', s') \;\text{ a.s.}\}.\end{aligned}$$

We derive $\bar{b}(D, L)$ explicitly based on the assumptions of Proposition 3.3.2. Given $q = q_d \leq 1$, the investment cost function $C_i(-k', s')$ simplifies as follows

$$\begin{aligned}C_i(-k', s') &= -k'\Big((q-1)\mathbf{1}_{c'\geq 0} + (q_d - 1)\mathbf{1}_{c'<0}\Big), \\ &= -k'(q\mathbf{1}_{c'\geq 0} + q_d\mathbf{1}_{c'<0} - 1), \\ &= k'(1-q),\end{aligned} \tag{3.19}$$

where $\mathbf{1}_{c'\geq 0}$ is an indicator function that takes value one if next-period cash $c' \geq 0$ and zero otherwise. Let $F(\widetilde{k}, z) = z^p\widetilde{k}^{\theta}$, with $\theta \in (0, 1)$, and $\ell(\widetilde{k}, z) = z^k\widetilde{k}(1-\delta)$. Moreover, taxes are zero by assumption, $T(P) = 0, \forall P \in \mathbb{R}$. We denote the resulting value of the collateral available to secure the debt by

$$\begin{aligned}f(z^p, z^k) &= F(\widetilde{k}, z) + I(D, L; \ell(\widetilde{k}, z)) - T(P) + k' - C_i(-k', s'), \\ &= F(\widetilde{k}, z) + I(D, L; \ell(\widetilde{k}, z)) + k' - k'(1-q), \\ &= F(\widetilde{k}, z) + \big(\ell(\widetilde{k}, z) - D\big)_+ - \big(\ell(\widetilde{k}, z) - (D+L)\big)_+ + \\ &\qquad + q\big(\widetilde{k}(1-\delta) - \ell(\widetilde{k}, z)\big),\end{aligned}$$

$$= z^p \widetilde{k}^\theta + \big(z^k \widetilde{k}(1-\delta) - D\big)_+ - \big(z^k \widetilde{k}(1-\delta) - (D+L)\big)_+ + \\ + q\widetilde{k}(1-\delta)(1-z^k).$$

The second equality follows from the assumption of zero taxes and Equation (3.19). The third equality uses the capital dynamics Equation (3.2) and the indemnity of a capped deductible contract in Equation (3.10). The last equality uses the functional form of the profit function and capital losses. The debt capacity $\overline{b}(D, L)$ is obtained by finding the smallest value of the collateral for any realization of the shocks $(z^p, z^k)$.

First, note that $f(z^p, z^k)$ depends on $z^p$ only through the profit function $F(\widetilde{k}, z)$, which is increasing in the profit shock $z^p$. Hence, $f(z^p, z^k)$ is also increasing in $z^p$, which implies the smallest value of $f(z^p, \cdot)$ is attained at the smallest profit shock realization, $\underline{z}^p$.

Second, $f(z^p, z^k)$ depends on $z^k$ only through the capital losses $\ell(\widetilde{k}, z)$, which are increasing in the capital shock $z^k$. However, $f(\cdot, z^k)$ is quasi-concave in the capital losses: $f(\cdot, z^k)$ is increasing for $\ell(\widetilde{k}, z) \in [D, D+L)$ and decreasing over $\ell(\widetilde{k}, z) \in [0, D)$ and $\ell(\widetilde{k}, z) \in [D+L, \overline{\ell}(\widetilde{k}, z)]$.[29] As a result, $f(\cdot, z^k)$ attains its minimum values at $D$ and $\overline{\ell}(\widetilde{k}, z)$.

Given how $f(z^p, z^k)$ depends on the shocks $(z^p, z^k)$, the debt capacity $\overline{b}(D, L)$ is obtained by retaining the smallest of the two values of the collateral at $D$ and $\overline{\ell}(\widetilde{k}, z)$

$$\overline{b}(D, L) = \begin{cases} \underline{z}^p \widetilde{k}^\theta + q(\widetilde{k}(1-\delta) - D), & \text{if } q < \frac{L}{\overline{z}^k \widetilde{k}(1-\delta) - D}, \\ \underline{z}^p \widetilde{k}^\theta + q\widetilde{k}(1-\delta)(1-\overline{z}^k) + L, & \textit{otherwise}, \end{cases} \tag{3.20}$$

where $\underline{z}^p$ is the smallest profit shock realization and $\overline{z}^k \leq 1$ is the largest loss per unit of capital. □

### 3.7.4 Proofs of results from Section 3.4

#### 3.7.4.1 Proof of Theorem 3.4.2

**Theorem 4.2. The value function is well defined.**

*Proof of Theorem 3.4.2.* What we need to show is that the infinite sum inside the expectation sign converges and is absolutely integrable.

[29] Using the functional form for the capital losses, we have $\overline{\ell}(\widetilde{k}, z) = \overline{z}^k \widetilde{k}(1-\delta)$, where $\overline{z}^k \leq 1$ is the largest loss per unit of capital (see Assumption 3.2.4).

We argued above that the manager always prefers staying in the bounded interval $0 \le k \le \overline{k}$ and $0 \le \widetilde{c} \le \overline{c}$ for some constants $\overline{k}, \overline{c}$. Thus, by Assumption 3.2.2, the profit of the firm is bounded above by $F(\overline{k}; z_t)$ at every time period $t$; we assumed the latter to be integrable.

The dividend issued will always be smaller than the liquidation value of the firm, which we can crudely estimate as

$$0 \le d_t \le (1+r)\overline{c} + \overline{k} + c_p(\overline{k}),$$

i.e. starting from an admissible position at time $t-1$, the largest possible liquidation value at time $t$ is estimated as above. On the other hand, the equity issued will always be smaller than is necessary to invest in the complete productive capacity and have the 'maximal' amount of liquid assets remaining, that is

$$0 \le e_t \le \overline{c} + f + \overline{k} + C_i(\overline{k}, s),$$

where the fixed costs $f$ denote the maximal loss of the firm.

Coming back to the main argument, irrespective of the admissible policy the following estimate holds

$$\begin{aligned} \mathbb{E}\left[\left|\sum_{t=0}^{\infty} \frac{1}{(1+r)^t}(d_t - e_t - C_e(e_t))\right|\right] &\le \mathbb{E}\left[\sum_{t=0}^{\infty} \frac{1}{(1+r)^t}\left|d_t - e_t - C_e(e_t)\right|\right] \\ &\le \mathbb{E}\left[\sum_{t=0}^{\infty} \frac{1}{(1+r)^t} const.\right] \\ &< \infty. \end{aligned}$$

Thus, the absolute value of the sum admits an integrable upper bound and the value function is well defined. □

#### 3.7.4.2 Proof of Theorem 3.4.3

*Proof of Theorem 3.4.3.* This follows from the main theorem in Hernández-Lerma and Lasserre (2012), Chapter 8; there it is shown that under certain conditions, the function $V$ is a fixed point of the operator

$$\mathcal{T}(f)(c,k;z) = \sup_{p\in\mathcal{P}(s)}\left\{d - e - C_e(e) + \frac{1}{1+r}\mathbb{E}\big[f(c',k';z')\big], c + k - C_i(-k,s), 0\right\}$$

on the Banach space of Borel measurable, essentially bounded random variables on $\mathcal{S}^*$.

Given that the stage cost is bounded above, as argued in the proof of Theorem 3.4.2, we only need to show that the problem satisfies Assumption 8.3.1. in Hernández-Lerma and Lasserre (2012); the weight function we will be working with is $w(s) = 1$ for all $s \in \mathcal{S}^*$ because the stage cost function is bounded. This amounts to two properties that need to be shown:

(1) the stage-cost function is lower semicontinuous, bounded below and inf-compact. The first condition is clear, the stage cost function being

$$(s, z; p) \mapsto d - e - C_e(e).$$

Indeed, this follows directly from Assumption 3.2.5. The lower bound was established in the proof of Theorem 3.4.2, whereas inf-compactness follows from the bounds on $\mathcal{P}_0$ and the bounds shown on the set of 'good' states.

(2) The following continuity property holds: For every bounded Borel measurable function $g\colon \mathcal{S}^* \to \mathbb{R}$, the function

$$a \mapsto \mathbb{E}[g(s'(a))]$$

is continuous; by $s'(a)$ is denoted the next state of the firm, given action $a \in \mathcal{P}_0(s)$, starting in state $s$. Continuity is clear for $s = *$, as this is an isolated point of the state space. Similarly in the case where the company decides to default, $\xi = \textit{stop}$; indeed, $s'(a) = *$ for the unique policy $a \in \mathcal{P}(s)^{\textit{stop}}$.

The only point that requires proof is when the manager decides not to close the firm, $\xi = \textit{cont}$. As in the proof of Lemma 3.7.1, we use the fact that the set $\mathcal{I}$ of admissible contracts is finite. Thus, for the proof, we only need to argue the situation with $\xi = \textit{cont}$ and indemnity $I \in \mathcal{I}$ fixed. The dynamics of the model is the following

$$\begin{aligned}
&\widetilde{k} = k + i,\\
&k' = \widetilde{k}(1-\delta) - \ell(\widetilde{k}, z'),\\
&\widetilde{c} = c - d + e + \frac{b}{1+r} - C_b(b) + i - C_i(i,s) - \pi(I, \tilde{k}),
\end{aligned}$$

$$P = F(\widetilde{k}, z') - f - \pi(I, \tilde{k}) - \delta\widetilde{k} - \ell(\widetilde{k}, z') + I(\ell(\widetilde{k}, z')) + r_c\widetilde{c} - \frac{rb}{1+r},$$
$$c' = (1 + r_c)\,\widetilde{c} + F(\widetilde{k}, z') - f + I(\ell(\widetilde{k}, z')) - T(P) - b.$$

Note that the conditions in Assumption 3.2.2 imply that the map $F\colon K \times Z \to \mathbb{R}_+$ is jointly measurable.[30] For a fixed initial state $s$, it is easy to see that all the above functions are continuous, but not continuous in the state $s$. Furthermore, all the random variables above have a density. This implies the result by continuity of functions $(k, z) \mapsto F(k, z)$ and $(k, z) \mapsto \ell(k, z)$. □

#### 3.7.4.3 Proof of Lemma 3.4.4

*Proof of Lemma 3.4.4.* Indeed, assume that it is optimal not to liquidate the firm in the state $(c, k; z)$. Liquidating capital and distributing the proceeds as dividends shows that $V(c, k + k^\delta; z) \geq V(c, k; z) + k^\delta q_d$. Similarly, distributing additional cash reserves shows that $V(c + c^\delta, k; z) \geq V(c, k; z) + c^\delta$. In particular, $V$ is strictly increasing in cash reserves and capital in the states where the function is not zero. □

[30]Indeed, the conditions imposed on the operating profit function above imply that it is a Carathéodory map; see Example 14.29 in Rockafellar and Wets (2009).

# Chapter 4

# The economics of the risk margin: Theory and evidence in a market-consistent framework

ANDREA BERGESIO, PAUL HUBER, PABLO KOCH-MEDINA AND LORIANO MANCINI[1]

**Abstract** What is the financial nature of the Solvency II risk margin and how material is it? We answer these questions in a market-consistent framework and show that the risk margin should represent a provision for the market-implied frictional costs of holding capital. The provision for frictional costs, which captures market expectations about taxes and other deadweight costs of capital, arises endogenously as part of the market-consistent value of insurance liabilities, in addition to the best estimate liability and default option value. We estimate the size of market-implied frictional costs based on observed market prices for publicly-held U.S. P&C insurers. Our results suggest that insurers are expected to carry significant deadweight costs due to agency problems, taxes and regulatory constraints, with estimates varying from 20% to over 80% of insurers' market capitalization. Despite these deadweight costs, we find that capital is necessary to support future business opportunities.

[1]Partial support through the SNF, Switzerland project 189191 "Value-maximizing insurance companies: An empirical analysis of the cost of capital and investment policies" is gratefully acknowledged. Bergesio at University of Zurich and Swiss Finance Institute, Huber at Swiss Re, Koch-Medina at University of Zurich and Swiss Finance Institute, Mancini at Università della Svizzera Italiana.

## 4.1 Introduction

Insurance reserves represent the main liability of insurance firms, providing for future realized claims and expenses that originate from their underwriting business. As a result, an appropriate valuation of insurance reserves is crucial to assess not only their underlying economic drivers, but also the economic net worth and financial strength of insurance firms (Hancock et al. (2001), Huber and Kinrade (2018)). Insurers calculate insurance reserves based on different sets of rules, which vary across countries and according to the relevant accounting frameworks, including regulatory, tax and financial reporting regimes (Nissim (2010), Sharara et al. (2010)). This heterogeneity complicates the valuation of insurers (Nissim (2013a,b)) and creates challenges for insurers that must comply with several reporting standards. If we denote by $J$ the set of existing valuation regimes, the value of insurance liabilities in any valuation regime $j \in J$ can be split in two components

$$\underbrace{VIL_j}_{\substack{\text{Value of insurance} \\ \text{reserves in regime } j}} = \underbrace{BEL}_{\substack{\text{Present value of} \\ \text{future expected claims}}} + \underbrace{(VIL_j - BEL)}_{\substack{\text{Margin that is} \\ \text{specific to regime } j}}.$$

As a matter of arithmetic, the value of insurance liabilities $VIL_j$ in any regime $j \in J$ can be seen as the sum of a best estimate of insurance liabilities $BEL$, which represents the present value of future expected claims from existing underwriting business, and a margin $VIL_j - BEL$, which is either defined explicitly or implicitly according to the valuation regime.

While regulatory regimes broadly agree on the definition of the best estimate of insurance reserves, there is a wide variety of approaches for determining deviations from the best estimate. Some regimes justify such a deviation as a "prudential margin" supposed to account for a certain level of unexpected deviations of insurance reserves from the best estimate (e.g., U.S. and Australian regulatory frameworks).[2] Other regulatory frameworks have moved recently towards economic solvency regimes

[2]The prudential perspective is endorsed by the U.S. Statutory Accounting Principles (SAP) and by the Australian Prudential Regulatory Authority (APRA). According to the U.S. SAP, P&C insurers are not allowed to recognize any discounting within the value of statutory insurance liabilities, implying that the prudential margin above the best estimate liability is identified implicitly by the value of the unrecognized discount (National Association of Insurance Commissioners (2015)). On the other hand, APRA statutory principles provide explicit requirements for the calculation of the margin in excess of the best estimate liability (APRA (2019)).

based on a cost-of-capital approach, including Solvency II in the European Union, the Swiss Solvency Test (SST) in Switzerland and the International Accounting Standards (IAS).[3] Under a cost-of-capital approach, deviations from the best estimate liability are defined as an explicit provision for the cost of holding the capital supporting the underwriting business over its outstanding term.[4]

The disagreement around the financial nature of the margin above the best estimate liability not only creates room for potential regulatory arbitrage, but leaves open the question of which margin is consistent with an economic valuation framework (see Thérond (2016) for an overview of market-consistent valuation frameworks). A regulatory framework that is not grounded on economic principles will introduce either unnecessary costs or unintended incentives for the supervised companies.[5] As a result, developing a consistent theoretical foundation for the economic valuation of insurance liabilities is a key step to design effective policy and regulatory measures.

We derive the economic value of insurance liabilities within a rigorous theoretical framework and develop an empirical strategy to investigate the size of the corresponding margin implied by observed market prices. Specifically, we ask the following research questions: (i) what is the financial nature of the margin embedded in the economic value of insurance liabilities? and (ii) given our answer to (i), what do financial markets tell us about the size of this margin? Note that the answer to the first research question is critical to address the second question: only by unveiling the economic forces producing the margin in a market-consistent framework we can identify observable variables useful to estimate the corresponding market-implied margin.

For insurers with a diffuse shareholder base, we show that the economic value of insurance liabilities can be separated in two major components based on market expectations:[6] i) a best estimate of insurance liabilities, adjusted for the value of shareholders' option to default, and ii) a provision for frictional costs of holding capital. In particular, we show that any deviation of the economic value of insurance

[3]The terms 'economic' and 'market-consistent' are used as synonyms.

[4]The margin for capital costs is known as risk margin (European Union (2009), IAIS (2020)) or market value margin (Federal Office Of Private Insurance (2006)).

[5]Recent research has shed light on the costs that arise when regulatory regimes deviate from economic principles, for instance concerning the valuation of statutory liabilities (Koijen and Yogo (2015, 2016), the classification of fixed-income investments (Ellul et al. (2015), Becker and Ivashina (2015)) and the accounting of financial hedging (Sen and Humphry (2018), Sen (2019)).

[6]Throughout this work we focus on widely-held insurers, which are insurance firms with a diffuse shareholder base, or equivalently insurers owned by well-diversified shareholders.

liabilities from the best estimate liability stems from the presence of market imperfections that generate deadweight costs of holding capital, such as taxes, agency costs and regulatory costs.[7] This result not only sheds light on the financial nature of the margin embedded in the economic value of insurance liabilities, but provides also crucial guidance for our empirical analysis. It is worth noting that differently from prior research on the economic value of insurance liabilities (see e.g., Möhr (2011)), which *assumes* the existence of a margin above the best estimate liability, we derive the provision for frictional costs endogenously from insurers' equity market price.[8]

Our theoretical approach is rooted in a market-consistent valuation framework (Koch-Medina et al. (2021), Artzner et al. (2020)) and relies on the decomposition of an insurer's market capitalization in its main economic drivers (Babbel and Merrill (1998), Hancock et al. (2001), Huber and Kinrade (2018)). Our starting point is to recognize that an insurer's market value of equity, defined as the stream of future expected cash flows to shareholders, provides an implicit equation for the economic value of insurance liabilities. To recover this implied value, we develop a simple model of insurance business that specifies how cash flows to shareholders evolve over time, similar to Koijen and Yogo (2015) and Sen and Humphry (2018). In addition, we value these cash flows using the *unique* economic valuation measure that ensures pricing is carried out consistently with financial market prices *and* with a broad shareholder base (Koch-Medina et al. (2021)). One major advantage of this approach is that the economic valuation measure can be applied regardless of the investment strategy of individual firms, instead of relying on risk-adjusted discount rates. Since the economic value of insurance liabilities should account only for existing business (Hancock et al. (2001), Babbel (1999)), in our theoretical framework we separate out the present value of future business opportunities, or franchise value, from the market value of equity (Babbel and Merrill (1998), Huber and Kinrade (2018)). In our empirical analysis, though, we recognize that part of the observed market capitalization may

[7]It follows that, in a Modigliani-Miller economy with perfect markets, the economic value of insurance liabilities coincides with the best estimate liability for widely-held insurance firms. See Doherty and Tinic (1981), Froot (2007) and Dhaene et al. (2017b) for an application and review of the Modigliani-Miller logic in insurance.

[8]The insurance literature tends to separate the economic value of insurance liabilities in the sum of a best estimate liability plus an additional liability (or margin), which is modeled according to existing regulatory regimes (Möhr (2011), Engsner et al. (2017)). Therefore, this literature remains silent as to the nature of the margin embedded in the economic value of insurance liabilities.

well capture insurers' future business opportunities and thus we allow for a non-zero franchise value.

We rely on the predictions of our market-consistent framework to estimate the provision for frictional costs implied by observed market prices. Using a sample of publicly-traded U.S. P&C insurers from 2009 to 2019, we estimate the value of the provision for frictional costs implied by their market capitalization. Our empirical strategy relies on the separation of the market value of equity into its observable and unobservable components. The main challenge is that the unobservable components of market capitalization include not only the market-implied provision for frictional costs, but also an intangible value given by the sum of the value of the option to default and the franchise value (Hancock et al. (2001), Babbel and Merrill (2005)). Since we cannot observe the intangible value separately, we develop an empirical strategy to separate out these unobservable components and recover the market-implied provision for frictional costs. For this purpose, we describe both unobservable components as parametric functions of observable firms' characteristics (see e.g., Korteweg (2010), Koijen and Yogo (2015)), which we use to explain the variation in the observable value of market capitalization. To identify the market-implied provision for frictional costs, we select observable variables recognized in the literature as proxies of insurers' deadweight costs. In addition, we interact these variables with insurers' capitalization to capture the dependence of deadweight costs on the capital level invested in an insurance firm (Zanjani (2002), Exley and Smith (2006)). Similarly, insurers' intangible value (franchise value and default option) is estimated as function of observed proxies of insurers' economic future business opportunities (Babbel and Merrill (1998)) and exposure to default risk (Gavira-Durón et al. (2020)).

Our results show that insurers' capital is expected to carry considerable frictional costs, with larger capital levels generating higher deadweight costs as a fraction of market capitalization (Zanjani (2002), Exley and Smith (2006) and Sherris and Van Der Hoek (2006)). In general, we find that the market-implied provision for frictional costs ranges from 20% to over 80% of the market capitalization for insurers with large capital levels. In particular, market-implied frictional costs are driven by taxes, agency costs and regulatory costs, as suggested by larger expected frictional costs for insurers with more dispersed analysts' forecasts, higher taxes, more reinsurance purchase and reduced dividends payouts. Our results suggest also that insurance firms with lower effective tax rates and reinsurance ceded bear substantially smaller frictional costs due to taxes and agency costs; in this case, the frictional margin ranges from 10% to 30% of the market capitalization for small insurers and between 30% and 70% for larger firms. Interestingly, our results show that even insurers with virtually no capital are expected to exhibit a strictly positive frictional margin, at around 20% of their market value of equity. This finding may well reflect the fact that markets expect these firms to recapitalize in the future and, thus, be poised to bear frictional costs. At the same time, we find that intangible value rises for firms with small and large capitalizations, reaching its minimum for insurers with capital between 2 and 3 $ billion. This suggests that despite holding capital is costly, insurers need their capital base to support the insurance business and capture future business opportunities. As a result, insurers are expected to face a trade-off between potentially higher future underwriting profits and larger deadweight costs when holding higher capital levels (Exley and Smith (2006)).

We also show that for most insurers in our sample the market-implied provision for frictional costs is larger than the corresponding margin implied in the U.S. statutory value of insurance reserves, which is given by the difference between the *un*discounted and discounted value of insurance liabilities. This suggests that should the U.S. regulatory framework move towards an economic solvency regime, U.S. P&C insurers with larger capitalization levels would find their statutory requirements increase substantially according to our estimation.

Finally, we compare the estimated provision for frictional costs to the Solvency II risk margin of publicly-traded E.U. P&C insurers, which is closer in nature to the provision for frictional costs in our framework. We show that the E.U. risk margin can reach 40% of the market value of equity for insurers with large capital levels, which seems to suggest our estimates for U.S. insurers lie in a reasonable range. Nonetheless, the insurance literature has discussed many aspects in which Solvency II departs from a fully market-consistent framework (see e.g., Vedani et al. (2017)), which implies the risk margin provides only an approximate benchmark for our estimates.

### Relation to the literature

Our research is related to various streams of literature. The recent strand of the insurance literature on the "market-consistent" (also known as "fair" or "economic") valuation of insurance liabilities has typically defined the economic value of insurance liabilities as the sum of a best estimate liability and a risk margin, assuming a cost of capital approach for the latter (see for instance Möhr (2011), Wüthrich (2011), Dhaene et al. (2017a), Engsner et al. (2017), Delong et al. (2019), Barigou et al. (2022) and references therein).[9] However, by adopting such an approach these studies remain silent about the economic forces that can lead the economic value of insurance liabilities to deviate from their best estimate. Huber and Kinrade (2018) derive an economic representation of insurance liabilities starting from an equity valuation model and obtain a cost of capital based risk margin as a result, although no market-consistent valuation measure is specified.

Our study relates also to the line of research that has separated the market value of insurers into aggregate or less granular components that help understand the sources of value creation in the insurance industry. Studies include Babbel and Merrill (1998), Babbel (1999), Hancock et al. (2001), Babbel and Merrill (2005) and Koch-Medina et al. (2021). These works recognize that the value of insurers can be decomposed into economic net worth (i.e., the market value of assets less the default-free value of insurance liabilities), franchise value (i.e., the present value of future economic profits from new business) and the value of the shareholder's option to default on insurance liabilities. This decomposition proves useful not only to separate the market value of business in place (economic net worth and option to default) from the market value of future business, but also to identify the intangible portion of insurers' equity value (franchise value and option to default).[10]

Another important strand of literature has focused on the deadweight costs, also known as frictional costs, associated with investing capital in an insurance firm. An early attempt to include frictional costs in the (economic) valuation of insurance liabilities was undertaken in Hancock et al. (2001). A growing number of studies has since then recognized the importance to account for the role of frictional costs in insurance pricing (Zanjani (2002), Ng and Varnell (2003), Froot et al. (2004), Koijen and Yogo (2015)) and in the valuation of insurers' equity (Exley and Smith (2006),

[9]For a discussion of the conceptual issues relating to the cost of capital approach and its consistency with classical financial economics theory, see Floreani (2011). Recent reviews of potential issues related to the current specification of the Solvency II risk margin are provided by Pelkiewicz et al. (2020) and Albrecher et al. (2022).

[10]The inherent trade-off between franchise value and default option value, for which higher values of insurers' franchise are associated with a lower value of the default option and vice-versa, has been the focus of recent empirical studies (see e.g., Kirti and Peria (2017), Liu (2018) and Grace et al. (2019)). For a review of empirical measures of default risk and franchise value, see respectively Gredil et al. (2022) and Heidinger and Gatzert (2018).

Hitchcox et al. (2007), Bergesio et al. (2019)). This line of research has grouped the frictional costs faced by insurance companies into four categories: 1) agency costs of holding capital, 2) regulatory costs, 3) financial distress costs and 4) double taxation.[11]

Estimates of frictional costs in the insurance industry are scarce. Scotti (2005) provides an early attempt to quantify such costs, pointing out the intrinsic difficulty in estimating the different drivers of deadweight costs separately. Other studies have relied on historical insurance premiums to estimate the magnitude of the frictional costs implied by insurance product markets. Zanjani (2002) provides suggestive evidence of substantial frictions borne by U.S. P&C insurers for holding capital against catastrophe insurance. Yow and Sherris (2008) develop a one-period model and calibrate it to Australian insurers balance sheets to assess the relative importance of different frictional costs. The costs stemming from regulatory requirements have been the focus of recent studies. Koijen and Yogo (2015) and Koijen and Yogo (2016) estimate the costs of U.S. life insurance companies resulting from regulatory rules that do not align with economic principles. Becker and Ivashina (2015) show that the U.S. regulatory-defined buckets for fixed-income investments have created incentives for insurance companies to search for yield. Sen and Humphry (2018) and Sen (2019) show that inconsistencies in regulatory rules concerning insurers' hedging activity can distort the hedging choices of insurers. In contrast to these studies, however, we estimate the unobserved value of frictional costs as implied by *financial market* prices rather than product market prices. As a result, for publicly-held insurers with a broad shareholder base, we provide estimates of the deadweight costs consistent with financial assets prices and thus with market expectations. Studies relying on written insurance premiums, instead, capture not only information available to product markets (rather than financial markets) but also economic forces such as the elasticity of the demand for insurance (see e.g., Koijen and Yogo (2015)). We therefore present an approach to the estimation of the frictional costs for insurers that is complementary to the existing research.

[11]Research on the agency costs of holding capital includes Merton et al. (1993), Pottier and Sommer (2006), Zhang et al. (2009) and Chiang et al. (2019). Regulatory costs have been investigated more recently in Ellul et al. (2011), Ellul et al. (2015), Koijen and Yogo (2015, 2016) and Sen (2019). The role of financial distress costs has been analyzed among others in Staking and Babbel (1995), Smith et al. (2003) and Froot et al. (2004)). The effect of double taxation has been the focus of many papers, which include Myers and Cohn (1987), Cummins and Grace (1994), Derrig and Ostaszewski (1997) and Harrington and Niehaus (2003).

**Structure of the paper**

The remaining sections are organized as follows. Section 4.2 introduces our theoretical market-consistent framework. Section 4.3 derives the economic value of insurance liabilities. Section 4.4 develops an empirical strategy to identify and estimate the unobserved component of the economic value of insurance liabilities, namely the provision for frictional costs based on market expectations. Section 4.5 describes our sample of publicly-traded U.S. P&C insurance firms. Section 4.6 presents estimates of the size of the risk margin implied by observed market prices. Section 4.7 concludes.

## 4.2 Theoretical framework

In this section, we develop a simple theoretical framework that allows us to derive the economic value of insurance liabilities. This value originates from the economic balance sheet of an insurance firm in which assets, liabilities and equity are recorded at their economic values. The use of economic (or market-consistent) values ensures consistency with financial market prices, i.e. economic values and market prices will coincide for those securities traded on the financial market.

Figure 4.1 depicts our conceptual model of economic balance sheet, in which assets at market values back the economic value of insurance liabilities ($L$) plus the market value of equity ($V$).[12] The figure shows that our objective is equivalent to uncover the theoretical value of insurance liabilities that is *consistent* with the market value of equity of insurers.

Assets include the market value of investments ($A$), which constitute an insurer's tangible assets, and franchise value ($FV$), which represents the market value of future expected profits from new business. As such, franchise value represents an intangible asset that belongs to the firm and thus contributes to the market value of equity

$$V = ENW + FV, \tag{4.1}$$

where $ENW$ is the economic net worth, defined as the economic value of tangible assets in excess of the economic value of insurance liabilities, namely

[12]For simplicity, we ignore financial debt and defer the discussion of a firm with general capital structures to the end of the section.

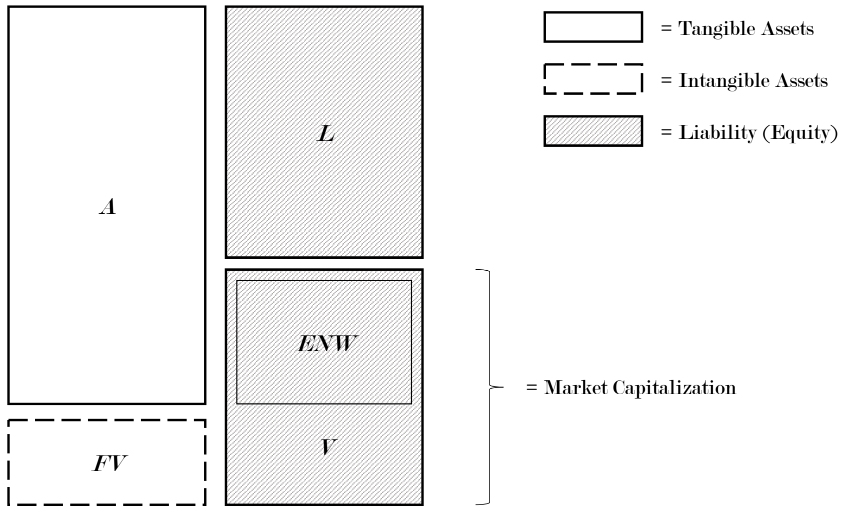



**Figure 4.1:** The figure offers a schematic representation of an insurer's economic balance sheet. Assets (white fill) include the economic value of tangible assets ($A$) and franchise value ($FV$), which represents an intangible asset (dashed line). The liability side (pattern fill) comprises the market value of equity ($V$) and the economic value of insurance liabilities ($L$), which includes the shareholders' value to default, as shown later. Economic net worth ($ENW$) is obtained from tangible assets in excess of the economic value of insurance liabilities. Financial debt is not considered here.

$$ENW = A - L. \tag{4.2}$$

The previous equations show that, provided the market value of investments $A$ is observable, we could recover the economic value of insurance liabilities if we observed either the economic net worth (Eq. (4.2)) or the franchise value implied in the market capitalization of publicly-traded insurers (Eq. (4.1)). However, neither of these can be observed directly and thus $L$ remains unobservable. Moreover, even if $L$ could be recovered, the market-consistent value of insurance liabilities *alone* would not provide insights into its economic drivers.

For these reasons, we introduce a market-consistent framework in which the economic value of insurance liabilities is decomposed into its basic components according to the expectations of well-diversified shareholders. Decomposing $L$ is valuable in two respects: i) it helps to shed light on the financial nature of the margin above the best estimate liability and ii) it provides theoretical guidance for the identification and estimation of its empirical counterpart. In order to recover the economic value of insurance liabilities consistent with market expectations, our starting point is the market value of equity of an insurance company, defined as the present value of future expected cash flows to shareholders. This definition requires to specify two elements:

1. the pricing measure with respect to which expectations are taken;
2. a workable definition of cash flows, for which we need a model of the insurance business.

In developing our framework, we assume that insurers have a diffuse shareholder base, or equivalently that investors are well-diversified. Either assumption implies that the firm can be valued *as if* the insurer were risk neutral (see e.g., Arrow and Lind (1970), Baumstark and Gollier (2014)). We discuss the corresponding pricing measure in Section 4.2.1, which shows that there exists a unique *economic* measure that is consistent with financial market prices, i.e. it replicates the price of traded securities, *and* with a broad ownership base. Section 4.2.2 introduces a simple model of insurance business that allows us to recover cash flows to shareholders.

### 4.2.1 Market-consistent framework

To value cash flows to shareholders and derive the resulting economic value of insurance liabilities, we need to work in a market-consistent valuation framework. Building upon the market-consistent framework in Koch-Medina et al. (2021), we extend the focus of their study to an insurance firm with multi-period insurance policies and subject to corporate taxation.

We work in discrete time and consider a limited-liability insurer with a diffuse equity ownership, selling at each date policies in insurance markets and investing the proceeds in financial markets. The relevant sources of uncertainty for the insurer are therefore underwriting risk, which stems from in-force insurance liabilities, and financial market risk associated with invested assets. It follows that to value cash flows to shareholders, which derive from underwriting and investment operations, any valuation framework needs to describe how pricing is carried out in financial and insurance markets. Hereafter we use the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to represent uncertainty in the economy, with $\mathbb{P}$ representing the "physical" probability measure.

**Market model.** Financial markets are assumed to trade continuously and are characterized by a money-market account and a risky security. The money-market account pays the instantaneous risk-free rate $\hat{r} > 0$, whereas the risky asset evolves according to a geometric Brownian motion with diffusion driven by a standard Brownian motion under the probability measure $\mathbb{P}$, which represents the only source of financial market risk.[13] The flow of information on market prices is described by the *market filtration* $(\mathcal{F}_t^M, t \geq 0)$. Provided the financial market is arbitrage-free and complete for any maturity $T > 0$, there exists a unique probability (pricing) measure $\mathbb{P}^*$ that prices at time $t \leq T$ any $\mathcal{F}_T^M$-measurable payoff $X_T$, with price given by

$$\pi_{t,T}(X_T) = \mathbb{E}^{\mathbb{P}^*}\left[ e^{-(T-t)\hat{r}} X_T \middle| \mathcal{F}_t^M \right]. \tag{4.3}$$

Thus, the market model is characterized by the unique pricing measure $\mathbb{P}^*$ that returns the market price of securities traded on financial markets.

**Insurance model.** The portfolio of insurance policies sold by the insurer gives rise to a sequence $(\ell_t, t \in \mathbb{N}\backslash\{0\})$ of i.i.d. non-negative random variables $\ell_t$ that represent aggregate losses at each date. Insurance losses are assumed to be independent of $\mathcal{F}_t^M$ for every $t \in \mathbb{N}\backslash\{0\}$, which implies that insurance risk

$$\epsilon_t = \mathbb{E}^{\mathbb{P}}[\ell_t] - \ell_t, \;\; t \in \mathbb{N}\backslash\{0\}$$

is independent of financial market risk. In addition to having access to information on financial market prices through $\mathcal{F}_t^M$, the insurer has information on realized insurance losses. As a result, the information set of the insurer at each date includes information about both financial markets *and* insurance losses. We characterize the insurance model by the larger *insurer's filtration* $(\mathcal{F}_t, t \geq 0)$, which represents the flow of information on financial market prices and insurance losses that is available to the insurer.

**The extended market.** When considering trading strategies with respect to the larger information set of the insurer, whose cash flows are adapted to the insurer's

[13]Note that from a modeling standpoint the difference in frequency between continuous trading in financial markets and discrete decision making at the firm level is consistent with the lower frequency at which insurers take decisions (Koch-Medina et al. (2021)). We assume the risky security evolves based on a geometric Brownian motion to ease the exposition. However, our results would follow through also if the risky asset followed an Itô diffusion, provided one takes the necessary precautions on the conditions imposed on the model parameters.

filtration, the extended market remains arbitrage free but becomes *incomplete*. Incompleteness implies that there are infinitely many possible risk-neutral probability measures consistent with market prices. However, *only one* of these measures is consistent with a diffuse shareholder base, which requires to value the insurance firm i) as if shareholders were risk-neutral and ii) in a market-consistent way. In particular, there exists a unique risk-neutral probability *economic* measure $\mathbb{Q}^*$ such that (Koch-Medina et al. (2021), Theorem 1.1):

(i) $\pi_{t,T}(X_T) = \mathbb{E}^{\mathbb{Q}^*}\left[e^{-(T-t)\hat{r}}X_T \middle| \mathcal{F}_t\right]$, for every $X_T$ in the marketed space at $T$, which is the set of all payoffs maturing at time $T$ that can be replicated in the market from the perspective of date 0 and, hence, date $t$ for any $t$;

(ii) $\mathbb{E}^{\mathbb{Q}^*}[X_T] = \mathbb{E}^{\mathbb{P}}[X_T] = 0$, for every $X_T$ orthogonal to the marketed space at $T$.

Moreover, the following relationship holds: $\mathbb{E}^{\mathbb{Q}^*}[X_T] = \mathbb{E}^{\mathbb{P}^*}[\mathbb{E}^{\mathbb{P}}[X_T|\mathcal{F}_t^M]]$.[14] Point (i) ensures the economic measure $\mathbb{Q}^*$ is market consistent, i.e. when applied to traded securities it replicates their market price. The second requirement stems from shareholders being well diversified: idiosyncratic risks such as insurance risks, which are driven by risk factors independent of financial market prices, can be diversified away and thus are not priced in by well-diversified investors. It follows that $\mathbb{Q}^*$ is the (unique) probability measure we need to value cash flows to shareholders. To simplify notation, we denote conditional expectations as $\mathbb{E}_t^{\mathbb{Q}^*}[X_T] := \mathbb{E}^{\mathbb{Q}^*}[X_T|\mathcal{F}_t]$.

### 4.2.2 A simple model of insurance firm

In this section, we introduce a simple model of an insurer's dynamics that allows us to derive cash flows to shareholders as a residual claim on the insurer's underwriting and investment results. The notation we adopt is such that stock items (e.g., assets, liabilities) are denoted by capital letters, whereas lowercase letters identify flow items (e.g., profits and insurance losses).

Absent agency problems, the insurer's management is assumed to take decisions optimally. This implies that cash flows used for the valuation of the insurance firm are assumed to derive from decisions taken consistently with the objective of shareholders'

[14] When working in a finite time horizon, $\mathbb{Q}^*$ coincides with the 'QP-rule' probability measure used in Artzner et al. (2020) for the valuation of insurance contracts.

value maximization. In the general model, however, we will allow for market imperfections that can generate costs related to sub-optimal decisions.

**Insurer's dynamics.** We consider an insurance firm that begins every period with a balance sheet in which assets at market value back the economic value of insurance liabilities plus the economic net worth (see Figure 4.1). The market value of assets at any date $t > 0$ is determined by investment and underwriting results over the previous period. Let $A_{t-1}$ denote the value of assets available at the beginning of period $t$. Investment returns and underwriting of new insurance policies determine the random return $\gamma_t$ on assets in period $t$, whereas policies in force generate insurance losses $\ell_t$. Assets at the end of period $t$ are

$$\gamma_t A_{t-1} - \ell_t,$$

which, if negative, indicate the insurer has insufficient assets to meet current insurance payments.[15]

Given the balance sheet at $t$, the insurer takes decisions concerning dividends payout and equity capital injections. The flow of dividends net of capital injections is denoted $z_t$ and identifies cash flows to shareholders at $t$. Positive cash flows $z_t \geq 0$ represent dividend payments and $z_t < 0$ capital injections.[16] The market value of assets after firm's decisions is given by $A_t$ and the dynamics of insurer's assets are

$$A_t = \gamma_t A_{t-1} - \ell_t - z_t. \tag{4.4}$$

Note that the previous equation provides an implicit expression for cash flows to shareholders, which can be obtained from the difference between assets before and after insurer's operations. In particular, Equation (4.4) represents the cash flow stream that we will use to derive the economic value of insurance liabilities in the baseline model (see Section 4.3.1).

[15]When the capital structure includes debt financing, the insurer must honour also debt payments. In this case, the assets available to pay debt obligations depend on the seniority structure of the insurer's liabilities. Section 4.3.4 shows the effect of capital structures that include debt funds on the economic value of insurance liabilities.

[16]Other firm's decisions such as investment in risky assets, reinsurance demand and premium price setting are not investigated in this paper. Previous papers have studied how an insurer takes these decisions optimally consistently with the objective of firm value maximization. For an overview of this stream of research, see e.g., Bi et al. (2019) and Koch-Medina et al. (2021).

As the insurance firm is exposed to both insurance and financial market risk, the insurer can default on insurance liabilities. However, default is endogenous as it depends on the willingness of the insurer to issue equity capital. We denote $\delta_t \in \{0, 1\}$ the endogenous liquidation decision, which takes value one if the firm continues operations at time $t$ and zero otherwise.[17]

**Regulatory constraint.** The insurer is subject to regulatory requirements that specify the minimum level of assets $\underline{A}_t$ necessary to operate the insurance business at $t$. In particular, $\underline{A}_t = \phi \underline{L}_t$ with $\phi \geq 1$ and $\underline{L}_t$ representing the statutory value of insurance liabilities. It follows that if $\phi > 1$ the insurer is required to hold a minimum level of statutory capital, $\underline{K}_t := (\phi - 1)\underline{L}_t$. An alternative formulation shows that the regulatory constraint is equivalent to a leverage constraint

$$A_t \geq \underline{A}_t \iff \frac{A_t}{\underline{L}_t} \geq \phi. \tag{4.5}$$

Moreover, regulatory requirements can be expressed as a constraint on the capital ratio (Koijen and Yogo (2015, 2016))

$$\frac{A_t - \underline{L}_t}{\rho \underline{L}_t} \geq \psi, \tag{4.6}$$

in which $\phi = 1 + \rho\psi$, with $\rho > 0$ being a risk weight and $\psi \geq 1$ the regulatory capital ratio. Empirical evidence shows that observed capital ratios typically satisfy $\psi > 1$, i.e. insurers hold more capital than the regulatory minimum, which suggests the existence of economic factors inducing insurers to hold capital buffers.[18]

If a regulatory breach occurs, the insurer is liquidated unless the minimum level of assets is restored by issuing new capital. In case the firm is liquidated, we assume that insurance liabilities get transferred to another insurer at their statutory value, whereas shareholders receive any assets in excess of the transfer price. In particular, we assume that the statutory value represents a sufficient transfer price for another

[17] We assume that liquidation is irreversible. As a result, $\delta_t = 0$ implies $\delta_{t+j} = 0$ for all $j > 0$.

[18] Evidence suggests this behaviour may stem from firms targeting ratios that meet the requirements set by credit rating agencies (Kirti and Peria (2017)). Moreover, in line with the sensitivity of insurance demand to insurers' financial strength, holding capital buffers helps insurers to protect market power (Koijen and Yogo (2015)) while avoiding costs associated with closer regulatory oversight and credit downgrades (Grace et al. (2019)).

insurance firm to take over the insurance liabilities of the liquidated insurer.[19] If assets are insufficient to meet insurance obligations, liquidation triggers the insurer's default and assets are seized by creditors.

### 4.2.3 Cash flows to shareholders and equity market value

Armed with a model of insurance firm and with a market-consistent pricing measure, we can derive explicitly the market capitalization of an insurer with a broad shareholder base. The market value of equity of an insurance firm at date 0 is obtained as the present value of the expected future stream of cash flows to shareholders

$$V_0 = \sum_{k=1}^{\infty} \frac{1}{(1+r)^k} \mathbb{E}_0^{\mathbb{Q}^*}[z_k], \tag{4.7}$$

in which $r := e^{\hat{r}} - 1$ is the one-period risk-free rate and $z_k$ are cash flows to shareholders at date $k$, valued according to the *economic* valuation measure $\mathbb{Q}^*$ that is consistent with market prices *and* with a diffuse ownership. The market value of equity provides the starting point to derive the market-consistent value of insurance liabilities.

## 4.3 The economic value of insurance liabilities

The value of insurance liabilities at any date should account only for in-force underwriting business. For this reason, we derive the economic value of insurance liabilities for an insurer with zero franchise value, i.e. *as if* no new business was written. In this case the market value of equity coincides with the insurer's economic net worth, $V_0 = ENW_0$, which is the present value of future expected cash flows to shareholders over the run-off period of insurance liabilities.

To convey the intuition, results will be presented for an insurer operating in a two-period economy and then extended to a generic horizon in Appendix 4.8.1. We

[19] For instance, under Solvency II the statutory value of insurance liabilities is supposed to ensure that '*insurance and reinsurance undertakings [...] take over and meet the insurance and reinsurance obligations*' being transferred (European Union (2009), Article 77(3)). In general, the transfer price would include the best estimate liability, a provision for the costs of holding capital and a mark-up based on the market power of the acquirer. For the sake of simplicity, we retain the assumption that the transfer price is equal to the statutory value of insurance liabilities.

start our discussion from a baseline model that is characterized by perfect markets. We will then use this model as a useful benchmark for the general model.

### 4.3.1 The baseline model

Consider an insurer that operates in an economy with two periods and three dates, $t = 0, 1, 2$. Financial markets are perfect, which implies there are no deadweight costs associated with investing in the insurance firm. To recover the economic value of insurance liabilities from the insurer's economic net worth, we need to identify cash flows to shareholders based on our model of the insurance firm. Table 4.1 shows cash flows to shareholders at date $t = 1, 2$ in the baseline model, in which the liquidation decision is taken at $t = 1$ and insurance liabilities run off at $t = 2$.

| | Cash flows to shareholders | |
|---|---|---|
| | $t = 1$ | $t = 2$ |
| $\delta_1 = 1$ | $\gamma_1 A_0 - \ell_1 - A_1$ | $(\gamma_2 A_1 - \ell_2)_+$ |
| $\delta_1 = 0$ | $(\gamma_1 A_0 - \ell_1 - \underline{L}_1)_+$ | |

**Table 4.1:** Cash flows to shareholders in the baseline model with two periods.

In the table we have used $(X)_+ = \max(X, 0)$. The table shows that limited liability benefits shareholders when insurance liabilities either run off or are transferred to another insurer: in these cases default is triggered if available assets are insufficient to cover insurance obligations.

Economic net worth is obtained as the present value of cash flows in Table 4.1

$$ENW_0^b = \frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}[\delta_1(\gamma_1 A_0 - \ell_1 - A_1)] + \frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}[(1-\delta_1)(\gamma_1 A_0 - \ell_1 - \underline{L}_1)_+] + \\ + \frac{1}{(1+r)^2}\mathbb{E}_0^{\mathbb{Q}^*}[\delta_1(\gamma_2 A_1 - \ell_2)_+],$$

in which we use the superscript $b$ to identify quantities specific to the baseline model. It is useful to decompose $ENW_0^b$ into a few drivers. We start by collecting cash flows to shareholders when the insurer is liquidated. The present value of the expected cash flows when the insurer is liquidated represents the value of the liquidation option

$$
\begin{aligned}
LO_0^b =& \frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}[(1-\delta_1)(\underline{L}_1+\ell_1-\gamma_1 A_0)_+] + \frac{1}{(1+r)^2}\mathbb{E}_0^{\mathbb{Q}^*}[(\delta_1(\ell_2-\gamma_2 A_1)_+]+ \\
&+ \frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}[(1-\delta_1)(BEL_1-\underline{L}_1)].
\end{aligned} \tag{4.8}
$$

The value of the liquidation option embeds the value of the option to default, which represents the present value of future expected shortfalls for which shareholders are not held accountable in the event of default due to limited liability. By indicating the default event at $t = 1$ with the indicator function $\Phi_1^b = \mathbb{1}_{\gamma_1 A_0 - \ell_1 - \underline{L}_1 < 0}$, which takes value one if insurance liabilities cannot be transferred at $t = 1$ and zero otherwise, the value of the option to default can be written as

$$
\begin{aligned}
DO_0^b =& \frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}[(1-\delta_1)(BEL_1+\ell_1-\gamma_1 A_0)\Phi_1^b] + \\
&+ \frac{1}{(1+r)^2}\mathbb{E}_0^{\mathbb{Q}^*}[(\delta_1(\ell_2-\gamma_2 A_1)_+],
\end{aligned} \tag{4.9}
$$

in which $BEL_1 = \frac{1}{1+r}\mathbb{E}_1^{\mathbb{Q}^*}[\ell_2]$ is the value at $t = 1$ of foregone contractual payments at $t = 2$. The difference $LO_0^b - DO_0^b$ is the value to shareholders when insurance liabilities can be transferred in full, which in the baseline case is given by

$$
LO_0^b - DO_0^b = \frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}[(1-\delta_1)(BEL_1-\underline{L}_1)(1-\Phi_1^b)].
$$

Note that the statutory value of insurance liabilities usually exceeds the best estimate liability, i.e. $\underline{L}_1 > BEL_0$.[20] As a result, in the baseline model the insurer will never find it optimal to liquidate unless default is triggered, i.e. $LO_0^b = DO_0^b$.[21] In general, liquidation and default coincide if statutory and economic balance sheets are the same.

Having accounted for the value of liquidation, we can decompose the economic net worth by rearranging cash flows when the insurer is not liquidated

[20] For instance, with Solvency II we have $\underline{L}_0 = BEL_0 + RM_0$, with $RM_0$ the risk margin. In this case $LO_0^b$ becomes

$$
LO_0^b = -\frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}[(1-\delta_1)RM_1] < 0.
$$

[21] When liquidation coincides with default, we can refine the equity value decomposition $V = ENW + FV$ (see Eq. (4.1)) and write the market value of equity as $V = NAV + DO + FV$ (see Babbel (1999) and Babbel and Merrill (2005)), with $NAV$ being the tangible net asset value.

$$
\begin{aligned}
ENW_0^b =& DO_0^b + \frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}[\gamma_1 A_0 - \ell_1] - \frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}[\delta_1 A_1] + \frac{1}{(1+r)^2}\mathbb{E}_0^{\mathbb{Q}^*}[\delta_1\gamma_2 A_1 - \ell_2],\\
=& DO_0^b + \underbrace{\frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}[\gamma_1 A_0]}_{=A_0} - \underbrace{\sum_{k=1}^{2}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\ell_k]}_{=BEL_0},\\
=& DO_0^b + A_0 - BEL_0.
\end{aligned}
$$

The previous equality shows that with perfect markets the economic net worth represents the value of assets, including current investments $A_0$ and the default option value $DO_0^b$, in excess of the best estimate liability $BEL_0$, which is the present value of future expected losses from in-force policies.[22] The best estimate liability net of shareholders' default option identifies the economic value of insurance liabilities in the baseline model

$$
L_0^b = BEL_0 - DO_0^b. \tag{4.10}
$$

Note that while the default option value is effectively an asset to shareholders that decreases the economic value of insurance liabilities, sometimes only the default-free best estimate liability is considered (see e.g., Albrecher et al. (2022)). Moreover, the default option value is systematically removed from the *statutory* value of insurance liabilities, as it would otherwise increase the regulatory value of the assets available to insurers precisely in times of distress.

### 4.3.2 The general model

The environment in which insurers operate is typically far from that of perfect markets. By investing in an insurance firm, investors are exposed to the effects of double taxation: not only are taxes levied on dividends paid out to investors, but also on the insurer's profits (Ke (2001), Feldblum (2007)). Insurance regulation generates substantial costs for insurers (Shim (2010), Koijen and Yogo (2015)), while the opaqueness of insurance liabilities (Chiang et al. (2019)) and illiquid investments (Ozdagli and Wang (2019)) create additional costs due to asymmetries of information between investors and the insurer's management (Dicks and Garven (2022)). Moreover, agency

[22] If insurance policies do not include savings components, we can simplify the best estimate liability further by replacing $\mathbb{Q}^*$ with the "physical" measure $\mathbb{P}$.

problems arise when the management has considerable discretion over the use of capital locked in the firm (Mayers and Smith Jr (2005), Laux and Muermann (2010)).

To gauge the effects of market imperfections, we model two types of costs that weigh on cash flows to shareholders. To capture the costs of taxation, we extend the baseline model by assuming that a constant corporate tax rate is levied on taxable profits based on a tax balance sheet. Furthermore, we allow for deadweight costs other than taxes to capture additional costs generated by market imperfections such as agency problems and asymmetric information.

Cash flows to shareholders in the general model are shown in Table 4.2.

| | Cash flows to shareholders | |
|---|---|---|
| | $t = 1$ | $t = 2$ |
| $\delta_1 = 1$ | $\gamma_1 A_0 - \ell_1 - \tau p_1^\tau - A_1 - c_1$ | $(\gamma_2 A_1 - \ell_2 - \tau p_2^\tau)_+$ |
| $\delta_1 = 0$ | $(\gamma_1 A_0 - \ell_1 - \tau p_1^\tau - \underline{L}_1)_+$ | |

**Table 4.2:** Cash flows to shareholders in a two-period economy with taxes ($\tau$) levied on taxable profits and deadweight costs ($c_1$) due to market imperfections.

At each date $t \in \{1, 2\}$, the insurer pays taxes according to the tax rate $\tau \in (0, 1)$, with taxable profits given by

$$p_t^\tau = A_{t-1}(\gamma_t - 1) - \ell_t - (T_t - T_{t-1}), \tag{4.11}$$

where $T_t$ is the value of insurance liabilities at $t$ based on the tax code.[23] Since taxable profits can take either sign, tax payments can be either positive or negative. In line with the corporate finance literature (see e.g., Hennessy and Whited (2005), DeAngelo et al. (2011)), we allow for negative taxable profits and interpret negative tax payments as losses carried forward that are generated in periods with negative taxable profits.[24] Deadweight costs $c_1$ decrease cash flows to shareholders when the

[23]While we model corporate taxes, personal taxes levied on dividends paid out to investors are not considered (similar to e.g., PonArul and Viswanath (1995) and Garven and Louberge (1996)). Niehaus (2019) discusses the implications of personal taxes on insurers' cost of equity capital and shows how tax costs for insurers have varied over time under different regimes. Prior research that incorporated personal taxation includes studies on the effect of taxes on insurers' supply of catastrophe insurance (Harrington and Niehaus (2003) and on managerial compensation (Ke (2001)).

[24]In order to preserve the simplicity of the model, we ignore the role of liquidation on losses carried forward, which would introduce additional complexity while not altering the intuition of the model.

firm continues operations ($t = 1, \delta_t = 1$) and capture costs associated with issuing costly external capital, agency costs of holding capital in the insurance firm, including possibly sub-optimal management decisions, and costs from regulatory actions.

Economic net worth in the general model is obtained as the present value of future expected cash flows in Table 4.2. Similar to the baseline model, the value of the liquidation option is obtained as the present value of cash flows to shareholders when the insurer is liquidated, which we denote $LO_0$. We comment on the liquidation option and the embedded default option later in this section.

The economic net worth in the general model can be decomposed by using $LO_0$, which collects cash flows to shareholders when the insurer is liquidated, and add any remaining cash flow that stem from continuing operations. The resulting economic net worth is given by

$$\begin{aligned} ENW_0 &= LO_0 + \frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}[\gamma_1 A_0 - \ell_1 - \tau p_1^\tau - c_1] - \frac{1}{(1+r)^2}\mathbb{E}_0^{\mathbb{Q}^*}[\ell_2 + \tau p_2^\tau], \\ &= LO_0 + A_0 - BEL_0 - \underbrace{\sum_{k=1}^{2} \frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\tau p_k^\tau]}_{=TL_0} - \underbrace{\frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}[c_1]}_{=DM_0}, \\ &= LO_0 + A_0 - BEL_0 - TL_0 - DM_0. \end{aligned}$$

Therefore, market imperfections generate taxes and other deadweight costs that give rise to additional liabilities identified by $TL_0$ and $DM_0$. In addition, these costs affect the value of the liquidation option $LO_0$ (see discussion below).

The tax liabilities $TL_0$ encapsulate the present value of future tax payments. When taxable profits differ from economic profits, tax liabilities can be further decomposed as the sum of a "tax margin" ($TM_0$) and deferred tax liabilities ($DTL_0$), namely $TL_0 = TM_0 + DTL_0$ (Huber and Kinrade (2018)). This allows us to distinguish $TM_0$, which identifies future expected taxes on investment income on assets backing capital in excess of the tax value of insurance liabilities, from $DTL_0$, which stems from temporary differences between the tax and economic balance sheets.[25]

[25] The tax margin $TM_0$ is the present value of future expected taxes levied on investment income earned on assets in excess of the *tax value* of insurance liabilities, formally

$$TM_0 = \sum_{k=1}^{2} \frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\tau \cdot r(A_{k-1} - T_{k-1})]. \tag{4.12}$$

The presence of market imperfections generates also a provision for deadweight costs, $DM_0$, which captures the present value of market expectations regarding the costs related to regulation (Koijen and Yogo (2015)), financial distress (Sherris and Van Der Hoek (2006)) and agency problems (Hancock et al. (2001)). We collect the liabilities due to taxes and other deadweight costs in a single provision for frictional costs, or frictional margin

$$FM_0 = TL_0 + DM_0 = TM_0 + DTL_0 + DM_0, \tag{4.13}$$

which represents market-implied frictional costs due to corporate taxes and other deadweight costs. Letting $DO_0$ denote the value of the default option under market imperfections, we obtain the economic value of insurance liabilities in the general model

$$L_0 = BEL_0 - DO_0 + FM_0. \tag{4.14}$$

Departing from perfect markets implies that the economic value of insurance liabilities, derived consistently with the market value of value-maximizing insurers, can be expressed as the sum of two components: a best estimate liability, net of the default option value adjusted for market imperfections, and a provision that captures market expectations about future frictional costs. Appendix 4.8.1 shows that this result holds for insurance liabilities extending over $T$ periods.

Our decomposition provides a financial interpretation for any deviation of the economic value of insurance liabilities from the best estimate liability. In a setting with market imperfections, the degree to which the (default-free) economic value of insurance liabilities exceeds their best estimate is explained by market expectations about the frictions that well-diversified shareholders incur to invest in the insurance business. In Section 4.4 we rely on this observation to develop an empirical strategy that allows us to estimate the market-implied value of insurers' frictional costs.

The economic value of insurance liabilities represents also the value of the so-

The second tax liability, $DTL_0$, identifies deferred tax liabilities and results from any difference between the present value of future tax payments and the tax margin. In particular, if the tax and economic balance sheet were the same (i.e., if $T_0 = BEL_0$), taxable and economic profits would coincide and so would the tax margin and the present value of future tax payments, implying $DTL_0$ would disappear. However, any difference in the valuation of insurance liabilities will be recognized in future expected taxable profits and taxed at rate $\tau$. The resulting present value of these temporary differences between the tax and economic balance sheet is then recognized in the value of deferred tax liabilities $DTL_0$.

called replicating portfolio (Hancock et al. (2001), Bergesio et al. (2019)), which consists of all insurer's investments used to meet the future expected costs, including claims and capital costs, generated by in-force insurance policies. In particular, the default-free best estimate liability represents the market value of a portfolio of liquid financial investments with cash flows that best match expected future insurance losses (Devineau and Chauvigny (2011), Möhr (2011)).

*Liquidation option and Default option.* It is instructive to split the value of the liquidation option $LO_0$, which is the present value of cash flows to shareholders when the insurer stops operating, into the value of the default option $DO_0$ and the value of liquidation $LO_0 - DO_0$ when no default occurs. Denote $\Phi_1 = \mathbb{1}_{\gamma_1 A_0 - \ell_1 - \tau p_1^\tau - \underline{L}_1 < 0}$ the indicator function that takes value one if insurance liabilities cannot be transferred at $t = 1$ and zero otherwise. Then, the value of the option to default can be written as

$$\begin{aligned} DO_0 =& \frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}\left[(1-\delta_1)\left(\frac{\ell_2 + \tau p_2^\tau}{1+r} + \ell_1 + \tau p_1^\tau - \gamma_1 A_0\right)\Phi_1\right] + \\ &+ \frac{1}{(1+r)^2}\mathbb{E}_0^{\mathbb{Q}^*}[\delta_1(\ell_2 + \tau p_2^\tau - \gamma_2 A_1)_+], \end{aligned} \tag{4.15}$$

which is the present value of the future expected benefits generated by limited liability in the event of default. Using Equations (4.14) and (4.15), we can gain insights on the reminder $LO_0 - DO_0$, which is the value of liquidation when insurance liabilities can be transferred in full to another insurer

$$\begin{aligned} LO_0 - DO_0 &= \frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}\left[(1-\delta_1)\left(c_1 + \left(\frac{\ell_2 + \tau p_2^\tau}{1+r} - \underline{L}_1\right)(1-\Phi_1)\right)\right], \\ &= \frac{1}{1+r}\mathbb{E}_0^{\mathbb{Q}^*}\left[(1-\delta_1)\left(c_1 + \left(BEL_1 + TL_1 - \underline{L}_1\right)(1-\Phi_1)\right)\right]. \end{aligned}$$

The previous equation shows that liquidation without default may be optimal only if the default-free economic value of insurance liabilities, comprising the best estimate liability, tax liabilities and deadweight costs, is greater than the transfer value of insurance liabilities, including foregone losses carried forward at $t = 1$.

In the next section, we discuss how the provision for frictional costs in our framework compares to the risk margin under Solvency II and other economic solvency regimes.

### 4.3.3 The Frictional margin and the Risk margin

Among the margins above the best estimate liability implied in the statutory value of insurance liabilities of existing regulatory regimes, the risk margin required by economic solvency regimes is the closest in nature to our provision for frictional costs.

Current economic solvency regimes specify the risk margin based on a cost of capital approach. In this approach, the risk margin is determined as the present value of future expected solvency capital requirements multiplied by a cost of capital rate, which is supposed to account for the costs of holding regulatory capital. In Solvency II and the SST, this cost of capital is set to 6%.[26] As a result, in our framework the frictional margin would coincide with the risk margin if one assumed that expected frictional costs were directly proportional to the minimum required regulatory capital

$$\mathbb{E}_t^{\mathbb{Q}^*}[\tau p_k^\tau + c_k] = r^c \cdot \mathbb{E}_t^{\mathbb{Q}^*}[\underline{K}_k], \quad \forall k > t, \tag{4.16}$$

with the constant of proportionality being the regulatory cost of capital equal to $r^c = 6\%$. Therefore, existing economic solvency regimes assume that even though insurers hold capital in excess of the minimum required level, the risk margin should account only for the cost of holding the required regulatory capital. An immediate implication is that under Solvency II, in which capital requirements are much sensitive to changes in interest rates, the risk margin inherits such sensitivity; on the other hand, deadweight costs need not be sensitive to interest rates. Furthermore, Equation (4.16) holds if the expected frictional costs are proportional to the cost of capital rate set by the regulator, which however corresponds to a total return on capital calibrated according to the CAPM model (CEIOPS (2009)). The CAPM cost of capital is problematic because without appropriate adjustments it includes return expectations on investments (base cost of capital), on future business (return on franchise), and an additional return expectation to compensate for the financial frictions to which insurance companies are subject.

[26] See European Union (2009), Art. 37 and Federal Office Of Private Insurance (2006), p. 92. In this context, it is worth noting the set of technical reports and consultation papers issued in the European Union related to the 'Solvency Capital Requirement review'. These include the second set for advice EIOPA (2018), which has confirmed the original calibration of the risk margin, and the consultation paper EIOPA (2019).

Based on our market-consistent framework, however, we argue that the risk margin should only allow for the frictional cost component, which can deviate substantially from the total return on capital.

### 4.3.4 The role of debt financing

Up to this point, we have considered insurers fully financed via equity capital. It is therefore natural to ask how the economic value of insurance liabilities varies when insurers are partially financed with debt. Here we discuss the impact of debt financing on the market-consistent value of insurance liabilities, while Appendix 4.8.1.4 shows how to adjust our model for generic capital structures that include financial debt. In general, we point out that introducing debt does not change our main result: as soon as market imperfections exist, market expectations about future deadweight costs will generate additional liabilities above the (frictionless) best estimate liability.

Part of the effects of financial leverage vary according to how debt ranks in the seniority hierarchy of liabilities (Hancock et al. (2001)). Senior debt cannot be used as part of the regulatory capital to support insurance risk, because it has priority over insurance obligations in case of default. As a result, senior debt financing is ineffective in decreasing the deadweight costs due to regulation. Subordinated debt is available in theory as regulatory capital, although its shorter duration compared to insurance liabilities creates challenges to its actual use as risk-bearing capital. Hybrid debt, which typically shares the tax advantages of debt and the risk characteristics of equity, can be used to free up locked-in equity and decrease the agency and tax costs of holding equity capital in the insurance firm (Zanjani (2002), Huber and Kinrade (2018)). Effectively, debt helps to reduce the capital base on which taxes and deadweight costs are incurred.

However, introducing debt in the capital structure increases the probability of distress states (Almagro and Ghezzi (1988), Cheng et al. (2012)). In corporate finance, the trade-off theory predicts that firms will choose the optimal level of financial debt to maximize the net benefits of leverage (Korteweg (2010)).[27]

[27]The corporate finance literature has studied extensively the trade-off implicit in the choice of firms' leverage ratios. The downside of debt financing is the introduction of a set of frictional costs (Dou et al. (2021)): in addition to the direct costs of bankruptcy (Warner (1977), Smith and Warner (1979), Müller (2022)), there exist indirect costs due to assets substitution (Jensen and Meckling (1976)) and debt overhang problems (Jensen and Meckling (1976)). More recently, empirical studies have investigated also indirect costs stemming from losses of human capital (Berk et al. (2010)) and

Since debt interests are tax deductible, market expectations about future tax costs will adjust downward to recognize the present value of the tax shield and the tax liability will decrease accordingly. On the other hand, to the extent that financial debt increases market expectations about future expected costs of financial distress, the provision for deadweight costs and thus the economic value of insurance liabilities will increase. Ultimately, the net effect of debt on the provision for frictional costs depends on the trade-off between the impact of tax-deductible interests on the tax liability and the higher expected costs of financial distress.

## 4.4 Empirical estimation

We explain our empirical strategy in steps. First, we discuss the challenges for the identification of the market-implied frictional costs of insurance. Then, we show how we overcome these difficulties and present our empirical specification.

### 4.4.1 The economic balance sheet revisited

Figure 4.2 shows an insurer's economic balance sheet in which the economic value of insurance liabilities $L$ is decomposed according to our market-consistent framework. From our former discussion, we know that the value of the option to liquidate the firm ($LO$) includes the value of shareholders' option to default on in-force liabilities. While the value of the default option effectively decreases the economic value of insurance liabilities and thus can be considered part of $L$ (Hancock et al. (2001)), we keep it separated from insurance liabilities (see e.g., Albrecher et al. (2022)).

The conceptual representation in Figure 4.2 is useful to illustrate the link between economic and regulatory (or statutory) balance sheets. As the main concern of regulators is the protection of policyholders, regulatory regimes typically prevent insurers from including intangible assets in their statutory balance sheet, as these assets are not available to back insurance liabilities. Moreover, despite statutory accounting rules varying across regulatory regimes, the statutory value of insurance liabilities can be decomposed in general into a best estimate liability plus a statutory margin

allocation of control rights (Franks and Loranth (2014)). Among the benefits of debt financing, prior research has studied not only tax-deductible interests (Kraus and Litzenberger (1973)), but also the lower agency costs stemming from lower free cash flows (Jensen (1986)). For an overview of the literature on optimal capital structure, see Graham and Leary (2011).

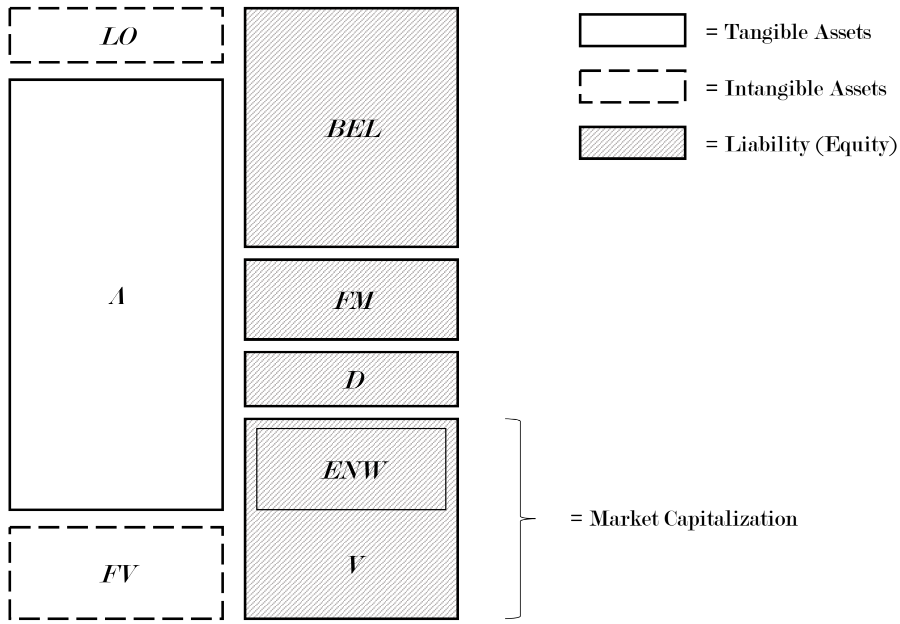



**Figure 4.2:** The figure provides a representation of the insurer's economic balance sheet in which the economic value of insurance liabilities is decomposed according to our market-consistent framework. Assets (white fill) include the value of shareholders' option to liquidate the insurance company ($LO$), the economic value of tangible assets ($A$) and the franchise value ($FV$). Intangible assets are represented by dashed lines. Insurer's assets back the best estimate liability ($BEL$), the present value of expected future frictional costs ($FM$) and the market value of equity ($V$). Economic net worth ($ENW$) represents the market value of equity net of franchise value. Financial debt is denoted by $D$.

that is specific to each regime. For instance, Solvency II in the European Union prescribes a best estimate liability plus the so-called risk margin, which is defined as 6% of the present value of future expected solvency capital requirements (European Union (2016), Art. 37). In the U.S. the NAIC requires non-life insurers to disclose their statutory insurance liabilities at nominal value, i.e. without discounting, which results in an implicit statutory margin equal to the difference between the nominal value of insurance liabilities and their best estimate (National Association of Insurance Commissioners (1998), Art. 11).

Statutory balance sheets differ from the economic balance sheet in Figure 4.2 due to intangible assets and any difference in the valuation of assets and liabilities. Crucially, changes in the valuation of insurance liabilities are driven by the way in which regulators define the statutory margin in excess of the best estimate liability. In addition, to the extent the best estimate liability can be observed or recovered reliably, the component of the economic value of insurance liabilities that is of interest for both economic and regulatory purposes is the provision for frictional costs, $FM$. Not only is its value relevant to assess the cost of frictions associated with in-force insurance liabilities, but it is also informative of the degree to which statutory margins are far off their economic counterpart. Our objective is thus to estimate the value of insurers' frictional costs implied by their market capitalization and their relation to statutory margins.

### 4.4.2 Identification assumptions

Based on our market-consistent framework, the market value of equity for an insurer with financial debt ($D$) can be written as follows

$$\underbrace{V_t}_{} = \underbrace{A_t - BEL_t - D_t}_{\substack{\text{Net asset}\\\text{value}}} - \underbrace{FM_t}_{\substack{\text{Provision for}\\\text{frictions}}} + \underbrace{LO_t + FV_t}_{\substack{\text{Intangible}\\\text{value}}}. \tag{4.17}$$

The previous equality shows that an insurer's market capitalization can be seen as the sum of three values. The net asset value, $A_t - BEL_t - D_t$, is the value of a default-free leveraged investment fund with total debt funding given by $BEL_t + D_t$. Market imperfections reduce shareholders' value by the market value of their expected costs, given by the provision for frictional costs $FM_t$. The remainder $IV_t = LO_t + FV_t$ captures an insurer's intangible value, which is the sum of the value of shareholders' liquidation option, including the value of the option to default, and franchise value.[28]

The separation of the intangible value from the net asset value, and thus from the value of insurance liabilities, is typical of a solvency perspective (Barth et al. (2008)). Regulatory regimes prevent insurers from recognizing intangible assets (e.g., default option and franchise value) for regulatory purposes, as these assets are not regarded as risk bearing (Floreani (2011)).[29] In our case, however, we adopt a similar separation in order to distinguish the observable portion of the market capitalization (the net asset value) from the *un*observable components ($FM_t$ and $IV_t$).

[28]Differently from Section 4.2, in which the economic value of insurance liabilities was derived ignoring new business, working with observed market values requires to account for the franchise value embedded in actual market prices.

[29]In the main regulatory frameworks the default option is referred to as "own credit risk", whereas the franchise value is captured by the goodwill recorded on the asset side of financial statements.

Our objective is to use the decomposition in Equation (4.17) to identify and estimate the market-implied value of $FM_t$, namely the value of the expected frictional costs implied by the market value of publicly-traded insurance companies. For this purpose it is convenient to write Equation (4.17) as

$$1 - \frac{A_t - BEL_t - D_t}{V_t} = \frac{IV_t}{V_t} - \frac{FM_t}{Vt}, \tag{4.18}$$

where we have divided each term by the market capitalization, $V_t$. Provided we can retrieve reliable data about the net asset value, the previous equality shows that for any publicly-traded insurer we can observe the difference between the intangible value and the provision for frictional costs.[30] However, as the intangible value of the firm is not observable, we cannot use Equation (4.18) to calculate directly the market-implied frictional costs. To overcome this issue, we explain how certain identification assumptions can be used to identify the market-implied frictional costs from the variation in insurers' balance sheets and market values.

Insurance research has identified and modeled different types of frictional costs, including regulatory costs, agency costs and financial distress costs. Prior research has described these costs as parametric functions of insurers' capitalization levels and their distance from minimum regulatory requirements (Ng and Varnell (2003), Froot et al. (2004), Chandra and Sherris (2006), Exley and Smith (2006), Sherris and Van Der Hoek (2006), Hitchcox et al. (2007)), Zhang and Nielson (2012)). Borrowing from this literature, we assume that frictional costs are a function of insurers' observable capital levels $K$, $FM(K)$. We define insurers' capital $K$ as the market value of assets in excess of the economic value of liabilities, which we recover from adjusted GAAP financial statements (see Appendix 4.8.2.2). Similarly, an insurer's capacity to attract new customers and generate future underwriting profits critically hinges on its available capital relative to the regulatory minimum (Ren and Schmit (2009), Koijen and Yogo (2015, 2016)). Therefore, also the intangible value is a function of insurers' capital, $IV(K)$.

The literature on non-life insurance firms has argued that P&C insurers can share

[30]This observed difference is similar in nature to the franchise value defined in Exley and Smith (2006), in which the value of insurers' franchise is described as a hump-shaped function of leverage induced by the firm's growth opportunities and decreased by different types of frictions. Note that Exley and Smith (2006) consider *franchise value* net of frictional costs, thus excluding the value of shareholders' limited liability (see par. 5.4.7 therein).

a similar exposure to market forces, such as interest rates and inflation, and to market imperfections that originate frictional costs (Hill (1979), Hill and Modigliani (1987)).[31] As a result, we assume that the functions $IV(K)$ and $FM(K)$ can be used to describe the same relation among intangible value, frictional costs and capital levels across a sample of P&C insurers. Moreover, as insurers' capital is common to both functions, we can identify each function separately if either $IV$ and $FM$ have different functional forms or if each function is driven by the interaction of $K$ with variables that are specific to the unobserved quantity. We summarise our identification assumptions as follows:

(A1) Market-implied frictional costs of an insurer $i$ at time $t$ are a function of observable capital $K_{it}$ and observable variables $Z_{it}^{FM}$ associated with frictional costs: $FM_{it} = FM(K_{it}, Z_{it}^{FM})$. Similarly, the intangible value depends on $K_{it}$ and observable variables $Z_{it}^{IV}$ specific to the franchise value and to the liquidation option: $IV_{it} = IV(K_{it}, Z_{it}^{IV})$.

(A2) Functions $FM$ and $IV$ are common across firms. Moreover, they are either (i) dependent on capital $K_{it}$ only and described by different functional specifications or (ii) driven by the interaction of insurers' capital with different sets of observable variables $(Z_{it}^{FM}, Z_{it}^{IV})$ associated with frictional costs and with the intangible value, respectively.

Under these identification assumptions, the functional form of $FM$ and $IV$ fully describes how frictional costs and intangible value vary with insurers' capital levels. Furthermore, we can estimate the market-implied value of frictional costs from insurers' observed market capitalization, provided there exists sufficient heterogeneity in capital levels within the insurance industry.

### 4.4.3 Estimation

The insurance literature predicts that both $IV$ and $FM$ are non-linear functions of insurers' capital (Staking and Babbel (1995), Babbel (1999), Hancock et al. (2001), Froot et al. (2004), Exley and Smith (2006), Hitchcox et al. (2007)). Specifically, both functions vary in a convex way with the capital level, rising at an increasing

[31] Typically, these arguments are developed in relation to insurers' underwriting beta, which represents the exposure of the insurance business to systematic risk.

rate as insurers' capital becomes larger. For this reason we rely on assumptions (A1) and (A2) and describe the unobservable ratios via the same functional form, while allowing the capital level to interact with observable firms' characteristics that are specific to each function.

The shape of the intangible value is driven by the relation between insurers' capital, franchise value and the default option value. The two components of intangible value move in opposite directions as a function of insurers' capital, which makes $IV(K)$ a convex function of $K$. We describe the intangible value relative to market capitalization as a quadratic function of capital

$$\frac{IV_{it}}{V_{it}} = IV(K_{it}, \mathbf{Z}_{it}^{IV}) = \mathbf{Z}_{0it}^{IV}\gamma_{\mathbf{0}} + (\mathbf{Z}_{1it}^{IV} \cdot K_{it})\gamma_{\mathbf{1}} + (\mathbf{Z}_{2it}^{IV} \cdot K_{it}^2)\gamma_{\mathbf{2}}, \tag{4.19}$$

where $\gamma_0, \gamma_1$ and $\gamma_2$ are vectors of parameters common across insurers and $\mathbf{Z}_{0it}^{IV}, \mathbf{Z}_{1it}^{IV}$ and $\mathbf{Z}_{2it}^{IV}$ are vectors of observable firms' characteristics specific to the intangible value of insurer $i$ at time $t$. The quadratic term is meant to capture the non-linear effect of capital on the intangible value. For low enough values of capital, we expect function (4.19) to capture mostly an increased value of the default option, whereas the franchise value will predominate for large capital levels.

The degree of frictional costs borne by insurers is expected to rise at an increasing rate as their capitalization increases (Zanjani (2002), Ng and Varnell (2003), Exley and Smith (2006), Zhang (2008)).[32] To describe the hypothetical relation between expected deadweight costs and capital level, we consider the following specification

$$\frac{FM_{it}}{V_{it}} = FM(K_{it}, \mathbf{Z}_{it}^{FM}) = \mathbf{Z}_{0it}^{FM}\theta_{\mathbf{0}} + (\mathbf{Z}_{1it}^{FM} \cdot K_{it})\theta_{\mathbf{1}} + (\mathbf{Z}_{2it}^{FM} \cdot K_{it}^2)\theta_{\mathbf{2}}, \tag{4.20}$$

with $\theta_0, \theta_1$ and $\theta_2$ parameters common across firms and $\mathbf{Z}_{0it}^{FM}, \mathbf{Z}_{1it}^{FM}$ and $\mathbf{Z}_{2it}^{FM}$ vectors of observable variables expected to gauge the frictional costs faced by insurer $i$ at time $t$. The increasing growth in frictions at larger capital levels is captured by the squared capital term, which makes frictional costs increase or decline faster as insurers' capital changes. The previous specifications provide the building blocks for our estimation, which is carried out by means of the following linear regression

[32] Note that many works that have focused on the predicted relationship between insurers' capital, franchise value and frictional costs belong to research conducted by practitioners, which we rely upon to formulate part of our empirical strategy.

$$1 - \frac{A_{it} - BEL_{it} - D_{it}}{V_{it}} = IV(K_{it}, \mathbf{Z}_{it}^{IV}) - FM(K_{it}, \mathbf{Z}_{it}^{FM}) + u_{it}, \tag{4.21}$$

in which $u_{it} \sim \mathcal{N}(0, \sigma_u^2)$ accounts for measurement errors, which allow for temporary deviations from the theoretical relation in Equation (4.18). The error term is by assumption orthogonal to the explanatory variables $\mathbf{Z}_{it}^{IV}$, $\mathbf{Z}_{it}^{RV}$ and $K_{it}$, which eliminates potential simultaneity issues of $K_{it}$, $IV_{it}$ and $FM_{it}$ being jointly determined (Korteweg (2010)). We describe below how we choose the explanatory variables used to capture the intangible value and market-implied frictional costs, deferring their formal definition to Table 4.3.

### 4.4.4 Frictional costs, intangible value and insurers' characteristics

Insurance theory predicts that insurers' intangible value depends on their capital level in a convex fashion: at low capitalization levels the default option value increases the intangible value, whereas franchise value will grow and drive up most of the intangible value as insurers build up capital (Babbel (1999), Hancock et al. (2001)).

Insurers with low capital levels have limited market power (Zanjani (2002), Kirti and Peria (2017)) and are more likely to face tighter constraints from lower financial strength and closer regulatory supervision (Koijen and Yogo (2015), Sen (2019)), which reduce their franchise value. A natural candidate to capture franchise value is an insurer's market to book ratio (Ren and Schmit (2009)), which embeds the expected value of the future stream of profits. As demand for insurance is sensitive to credit quality (Koijen and Yogo (2015), Zanjani (2002)), an alternative measure of market power is provided by financial strength ratings, which affect franchise value through their impact on the elasticity of demand.[33]

An opposite argument holds for the default option, which provides value to shareholders when leverage increases and the insurer's distance to default decreases (Froot et al. (2004), Sherris (2006)). The risk of default and its associated value is cap-

[33] Moreover, Pottier and Sommer (2006) and Epermanis and Harrington (2006) find that growth rates are higher for insurers with higher ratings, and that the growth rate of a company moves up and down with rating changes.

tured by the Z-score based on return on equity (Gavira-Durón et al. (2020)) and the volatility in statutory capital changes (Sen (2019)). Both variables capture the higher default risk to which insurers are exposed when profitability and capital position deteriorate over time.

*Frictional costs of holding capital.* According to the insurance literature, frictional costs are expected to be directly related to the capital held by insurance firms, rising at an increasing rate for large capitalization levels (Zanjani (2002), Exley and Smith (2006)).

At low levels of capital, market prices embed the expected costs of increasingly higher financial frictions (e.g., higher equity issuance costs), larger financial distress costs (Smith et al. (2003)) and regulatory costs (Lorson et al. (2012), Koijen and Yogo (2015)) associated with a higher probability of regulatory intervention (e.g., restrictions on dividend payments). Financial distress costs are reflected indirectly in the degree to which insurers engage in risk transfer to financial and reinsurance markets. Reinsurance ceded can be used effectively whenever the cost of taking risk and issuance costs are higher than the cost of transferring risk (Hancock et al. (2001), Pottier and Sommer (2006)). The cost of regulation can be gauged by the degree of assets invested in illiquid assets, which may increase assets risk and thus the required statutory capital (Babbel and Merrill (2005), Ellul et al. (2015)). Moreover, dividend cuts can signal regulatory costs if limitations to dividend distributions are imposed by regulatory authorities (Scotti (2005), Hitchcox et al. (2007)).

Expected deadweight costs rise at an increasing rate for large capital holdings, reflecting the additional taxes levied on invested assets (Myers and Cohn (1987), Cummins and Grace (1994) and Derrig and Ostaszewski (1997)) and potentially larger agency costs (Merton et al. (1993), Morgan (2002) and Zhang (2008)). The effect of taxes can be measured by the effective corporate tax rate, which is given by the ratio of profits after tax to profits before tax. This is a measure that captures the degree to which double taxation through corporate taxes levies on shareholders' returns. Agency costs have been measured using several methods, including errors in reserves estimates and dispersion in analysts' forecasts as close proxies for the opaqueness of insurers and, by extension, of their agency costs (Hitchcox et al. (2007), Pottier and Sommer (2006) and Zhang et al. (2009)). Analysts' forecasts dispersion provides information similar to dispersion in financial strength ratings (Morgan (2002)), which however are issued and reviewed at a lower frequency than the former.[34] In this case, it is worth noting that identification of market-implied frictions due to agency costs is achieved by exploiting the second moment of variables that are typically used as controls of insurers' franchise (via their effect on financial strength). Table 4.3 reports the variables chosen to describe the unobservable components of insurers' market capitalization.

## 4.5 Data

We construct a sample that consists of quarterly observations for U.S. publicly-traded P&C insurance companies over the period 2009-2019. We choose this time period as the financial accounting standard 157 (FAS 157), which establishes a framework for measuring fair value, became effective on November 15, 2007 (as reported in WRDS COMPUSTAT). In Appendix 4.8.2.2 we show how we construct an economic balance sheet starting from fair value accounting.

The data required for our analysis is collected from several sources. Information on the shareholder base of insurers is from the SEC Form 13F, which requires investment managers with assets under management larger than $100mln to disclose their U.S. equity holdings. U.S. GAAP financial statements are from WRDS COMPUSTAT, whereas statutory statements are from S&P SNL Financial, which provides access to the regulatory schedules filed by U.S. insurers to the NAIC. Data on analysts' activity is from Thomson Reuters I/B/E/S; market prices are from CRSP (equity securities) and TRACE (debt securities).

We retrieve data on insurers' ownership via the 13F reports filed quarterly to the SEC, as these allow us to construct measures of (institutional) ownership concentration and thus assess whether insurers in our sample have a diffuse ownership.

From COMPUSTAT and SNL Financial we retrieve quarterly financial and statutory accounting data necessary to obtain the insurers' balance sheets at economic

[34] Indicators of agency costs for insurers are typically connected to informational asymmetries between managers and shareholders. These include transparency of pricing and reserving aspects of the business and the extent to which managers' decisions are taken in the best interest of investors (Hitchcox et al. (2007) p. 62). Following Armstrong et al. (2011), one can distinguish broadly two approaches to gauge the degree of information asymmetry: using market-based measures (e.g., bid-ask spread and its adverse selection component) or accounting-based variables (e.g., errors in reserves estimates). Additional measures of information asymmetry are provided by variables related to analysts' coverage.

| Intangible variables, $\mathbf{Z}_{i,t}^{IV}$ | | |
|---|---|---|
| Observable variable | Definition | Reference |
| ROE | *Return on Equity.* The ratio of (1) net income to (2) the average of shareholders' equity over the previous quarter. | Koijen and Yogo (2016) |
| $Z_{\text{ROE}}$ | *ROE Z-score.* The sum of one and ROE divided by the standard deviation of ROE over a four-quarter rolling window. | Lepetit and Strobel (2015), Gavira-Durón et al. (2020) |
| **Frictional variables, $\mathbf{Z}_{i,t}^{FM}$** | | |
| Observable variable | Definition | Reference |
| AFD | *Analysts' forecast dispersion.* The standard deviation of EPS analysts' forecast estimates over the stock market price at the end of the previous period. | Pottier and Sommer (2006), Zhang et al. (2009) |
| REINS | *Reinsurance ceded.* The ratio of (1) Net Premiums Written to (2) Gross Premiums Written. | Koijen and Yogo (2016), Irresberger and Peng (2017) |
| ETR | *Effective tax rate.* The ratio of (1) the federal and foreign income taxes incurred to (2) net income. | Derrig and Ostaszewski (1997), Harrington and Niehaus (2003), Irresberger and Peng (2017) |
| DIVCUT | *Dividend Cut.* A dummy variable indicating dividends paid are below the level paid in the prior period, conditional on being positive in the prior period. | Berry-Stölzle et al. (2014) |

**Table 4.3:** The table shows the observable variables chosen to explain the unobservable ratios $IV(K_{it}, \mathbf{Z}_{it}^{IV})$ and $FM(K_{it}, \mathbf{Z}_{it}^{FM})$. For each variable, the table provides how the variable is defined and the main references we rely on for its definition.

value. These databases provide also information on insurers' financial and underwriting performance, reinsurance activity, dividend policy and tax payments, which represent observable variables used to explain market expectations about insurers' intangible value and frictional margin (see Table 4.3). Statutory reporting data from SNL Financial is also useful to retrieve insurers' premiums by line of business and their geographic location by state. Details on the adjustments applied to the GAAP balance sheet to recover the corresponding economic value of assets and liabilities are provided in Appendix 4.8.2.2.

We complement the set of firms' characteristics by including data on each insurer's analysts' activity from Thomson Reuters I/B/E/S. This allows us to calculate for each quarter a measure of dispersion of analysts' forecasts, which provides information on the degree of disagreement among analysts and thus the perceived opacity about insurers' business.

We supplement the sample with quarterly market values of equity from CRSP and market values of debt from TRACE. Whenever debt is not traded or its market value cannot be reliably obtained, we use its book value. Market data is matched to the latest accounting information available to investors, corresponding to the latest earnings release date in WRDS COMPUSTAT.

The final sample includes 191 firm-quarter observations, for which we provide summary statistics in Table 4.4. Details on the sample selection are provided in Appendix 4.8.2.1.

As discussed in Section 4.4.2, it is important to observe a wide range of capital levels to be able to identify how frictional costs are related to insurers' capital. The standard deviation of capital is 2.91, with an average capital of \$2.15bln, hence we observe substantial variation of capital across insurers. Another remark is that our theoretical framework produces predictions that are valid for insurers with a diffuse shareholder base.[35] To check the validity of this assumption, Table 4.4 shows the proportion of insurers' shares owned by institutional investors ($IOR$) and the concentration of institutional ownership measured by the Herfindahl–Hirschman index ($HHI_{IO}$).[36] The

[35] Another requirement for insurers to behave as risk-neutral institutions is that the idiosyncratic risks of the insurance firm are uncorrelated with investors' wealth (Arrow and Lind (1970), Baumstark and Gollier (2014)). Knowledge of investors' positions and their exposure to aggregate market risk, however, is harder to verify.

[36] These measures have been employed in many research areas; among others, an important strand has focused on the relation between institutional ownership and stock returns (see e.g., Gompers and Metrick (2001), Sias et al. (2006) and Lewellen (2011)).

| Panel A: Summary Statistics | | | | | |
|---|---|---|---|---|---|
| | | Percentile | | | |
| | Mean | 10 | 50 | 90 | SD |
| AFD | 0.131 | 0.015 | 0.048 | 0.188 | 0.539 |
| ETR | 0.219 | 0 | 0.255 | 0.336 | 0.142 |
| REINS | 0.117 | 0.006 | 0.093 | 0.361 | 0.129 |
| DIVCUT | 0.235 | 0 | 0 | 1 | 0.424 |
| ROE | 0.098 | -0.031 | 0.109 | 0.234 | 0.154 |
| $Z_{ROE}$ | 3.668 | 1.975 | 3.766 | 5.005 | 1.149 |
| K ($bln) | 2.155 | 0.105 | 1.129 | 6.248 | 2.912 |
| IOR | 0.624 | 0.2 | 0.72 | 0.95 | 0.273 |
| $HHI_{IO}$ | 0.076 | 0.03 | 0.05 | 0.13 | 0.084 |
| DISCOUNT | 0.1 | 0.036 | 0.083 | 0.19 | 0.067 |

| Panel B: Correlation Matrix | | | | | | |
|---|---|---|---|---|---|---|
| | ETR | REINS | DIVCUT | ROE | $Z_{ROE}$ | K |
| AFD | -0.228 | -0.035 | -0.056 | -0.21 | -0.142 | -0.045 |
| ETR | | -0.075 | 0.079 | 0.403 | 0.391 | 0.053 |
| REINS | | | -0.091 | -0.052 | -0.125 | -0.104 |
| DIVCUT | | | | 0.048 | 0.123 | 0.076 |
| ROE | | | | | 0.381 | 0.142 |
| $Z_{ROE}$ | | | | | | 0.023 |

**Table 4.4:** This table presents summary statistics for the 191 firm-quarter observations in the sample. The variables are defined as follows: $AFD$ is the standard deviation of EPS analysts' forecast estimates over the stock market price at the end of the previous period; $ETR$ is the ratio of federal and foreign income taxes incurred to net income; $REINS$ is the ratio of net premiums written to gross premiums written; $DIVCUT$ is a dummy variable indicating that dividends paid are below the level paid in the prior period, conditional on being positive in the prior period; $ROE$ is the ratio of net income to the average GAAP shareholders' equity capital over the previous quarter; $Z_{ROE}$ is the sum of one and ROE divided by the standard deviation of ROE over a four-quarter rolling window; $K$ (in U.S.$ billion) is the market value of assets in excess of insurance, financial and other liabilities at market value; $IOR$ is the fraction of shares owned by institutional investors; $HHI_{IO}$ is the Herfindahl-Hirschman index based on institutional investors' shares; $DISCOUNT$ is the implicit statutory discount used to obtain discounted insurance reserves (see Appendix 4.8.2.2). For further details on the construction of the economic balance sheet please refer to Appendix 4.8.2.2. SD indicates the standard deviation. The data sources include COMPUSTAT, Thomson Reuters I/B/E/S, SNL Financial and 13F reports.

institutional ownership varies considerably, from less than 20% to more than 95% of insurers' capital, with an average of 62% of capital owned by institutional investors. Despite such heterogeneity in institutional shares, the ownership appears dispersed across institutional investors as shown by the Herfindahl–Hirschman index, with an average concentration of 0.08. We consider these statistics as providing evidence of insurers with a sufficiently broad shareholder base.

We observe considerable heterogeneity in firms' characteristics used to estimate the unobservable values $FM$ and $IV$. In particular, our sample is characterized by substantial variation in analysts' forecasts, effective tax rate and reinsurance use. Moreover, insurers are profitable on average based on their return on equity ($ROE$), with a mean ROE of 10%. The average implied discount in statutory insurance reserves is 10%, which can exceed 20% for long-tailed lines of business such as workers' compensation.

## 4.6 Results

In this section, we discuss the estimates of the intangible value and frictional costs implied by insurers' market capitalization, investigating the drivers of the market-implied deadweight costs and their relation to statutory requirements.

### 4.6.1 Intangible value and Market-implied Frictional costs

The estimates for our specification (4.21) are presented in Table 4.5 (column I), while Figure 4.3 illustrates the corresponding predicted relationships between intangible value, frictional costs and insurers' capital. Following our discussion of the explanatory variables in Section 4.4.4, capital $K_{it}$ interacts differently with the observable variables in $IV(K_{it}, \mathbf{Z}_{it}^{IV})$ and $IV(K_{it}, \mathbf{Z}_{it}^{FM})$. Observable variables related to agency costs ($AFD$), taxes ($ETR$) and the franchise value ($ROE$), which become relevant as insurers' capital increases, are grouped in $\mathbf{Z}_{1it}^{IV}$ and $\mathbf{Z}_{1it}^{FM}$, respectively. Firms' characteristics expected to capture regulatory costs, financial distress costs ($REINS$, $DIVCUT$) and the value of the option to default ($Z_{\text{ROE}}$) are likely to grow or decline at a faster rate as capital changes and are thus grouped respectively in $\mathbf{Z}_{2it}^{IV}$ and $\mathbf{Z}_{2it}^{FM}$.

Table 4.5 shows that profitable firms and insurers with highly volatile returns on equity experience a larger intangible value, with the value of the default option driving the increase in intangible value at low capitalization levels. This can be seen also in Figure 4.3 (top-right panel), which shows that the intangible value rises for insurers with small capitalization levels, for which the probability of default is higher, and when capital levels increase to the point where insurers can benefit from an improved franchise value.

Market-implied frictional costs increase as agency costs rise, in line with evidence on insurers' opaqueness (Morgan (2002), Chiang et al. (2019)), and when taxes and regulatory costs become larger. Furthermore, taxes, cuts in dividend payouts and the proportion of reinsurance purchase are sensitive to changes in insurers' capital. When coupled with the capitalization level, dividend cuts capture the costs for insurers to adjust their capital due to regulatory requirements and costly equity issuance, which become especially relevant at lower levels of capitalization (Koijen and Yogo (2015)). The results also suggest that higher reinsurance purchase may signal the attempt of insurers to reduce the costs of holding capital (Lorson et al. (2012)), either regulatory and issuance costs when capital is low (Froot et al. (1993), Hancock et al. (2001)) or agency costs weighing on higher capitalization levels (Pottier and Sommer (2006), Zhang et al. (2009)).

The result of our identification strategy for the functions $IV$ and $FM$ under Assumptions (A1) and (A2)(ii) is shown in Figure 4.3. For each estimated function we report the values predicted by our model and a polynomial approximation that is a best fit for the predicted data. The top-right panel confirms the convex relation of the intangible value with insurers' capital and shows that firms with higher capitalization levels benefit from a higher intangible value than otherwise smaller firms. Moreover, because the intangible value represents a positive part of the market capitalization, the estimated value of $IV/V$ is between zero and one.

Figure 4.3 shows that market-implied frictional costs not only increase as insurers build up capital, but can represent a large fraction of market capitalization for highly-capitalized insurers. Because the relationship between frictional costs and capital appears almost linear, we test an alternative specification in which we drop the quadratic terms in $FM(K_{it}, \mathbf{Z}_{it}^{FM})$ and show the results in Table 4.5 (II). To test whether the nested model (II) provides a better fit, we conduct a likelihood ratio test and reject the null hypothesis that the nested model is preferred to the original specification at the 1% significance level. The same conclusion holds when comparing

| Dependent variable: $1 - \frac{A_{it}}{V_{it}} + \frac{BEL_{it}}{V_{it}} + \frac{D_{it}}{V_{it}}$ | | |
|---|---|---|
| | I | II |
| constant | -0.04 | -0.452** |
| | (0.188) | (0.233) |
| constant (*K) | -0.222*** | 0.067 |
| | (0.059) | (0.081) |
| constant (*K$^2$) | 0.064*** | -0.001 |
| | (0.012) | (0.015) |
| ROE | 1.189** | 0.993* |
| | (0.665) | (0.715) |
| Z$_{\text{ROE}}$ | 0.58** | -0.456 |
| | (0.342) | (0.466) |
| ROE (*K) | -0.246* | -0.157 |
| | (0.166) | (0.235) |
| Z$_{\text{ROE}}$ (*K$^2$) | -0.036** | 0.011 |
| | (0.021) | (0.028) |
| AFD | 1.091** | -1.289*** |
| | (0.489) | (0.506) |
| ETR | 0.045* | -0.096** |
| | (0.033) | (0.047) |
| REINS | 0.049* | -0.024 |
| | (0.038) | (0.045) |
| DIVCUT | 0.004 | 0.01* |
| | (0.005) | (0.006) |
| AFD (*K) | 0.047 | 0.514*** |
| | (0.158) | (0.149) |
| ETR (*K) | 0.018* | 0.015 |
| | (0.013) | (0.021) |
| REINS (*K$^2$) | 0.01*** | |
| | (0.003) | |
| DIVCUT (*K$^2$) | 0.068* | |
| | (0.048) | |
| $R^2$ | 0.608 | 0.217 |
| N | 191 | 191 |

**Table 4.5:** This table reports the coefficient estimates from ordinary least squares regressions over the 2009 to 2019 sample of 22 insurers. Our observable variable (Eq. (4.21)) is regressed on the intangible value and frictional costs relative to firm value, $IV(K_{it}, \mathbf{Z}_{it}^{IV})$ and $FM(K_{it}, \mathbf{Z}_{it}^{FM})$, specified as a function of firm characteristics (see Equations (4.19) and (4.20), respectively). Explanatory variables are as in Table 4.3. Heteroscedasticity robust standard errors are reported in parenthesis. *, **, *** correspond to p-values lower than 10%, 5% and 1% respectively. $N$ is the number of firm-quarters.

models (I) and (II) using standard information criteria.

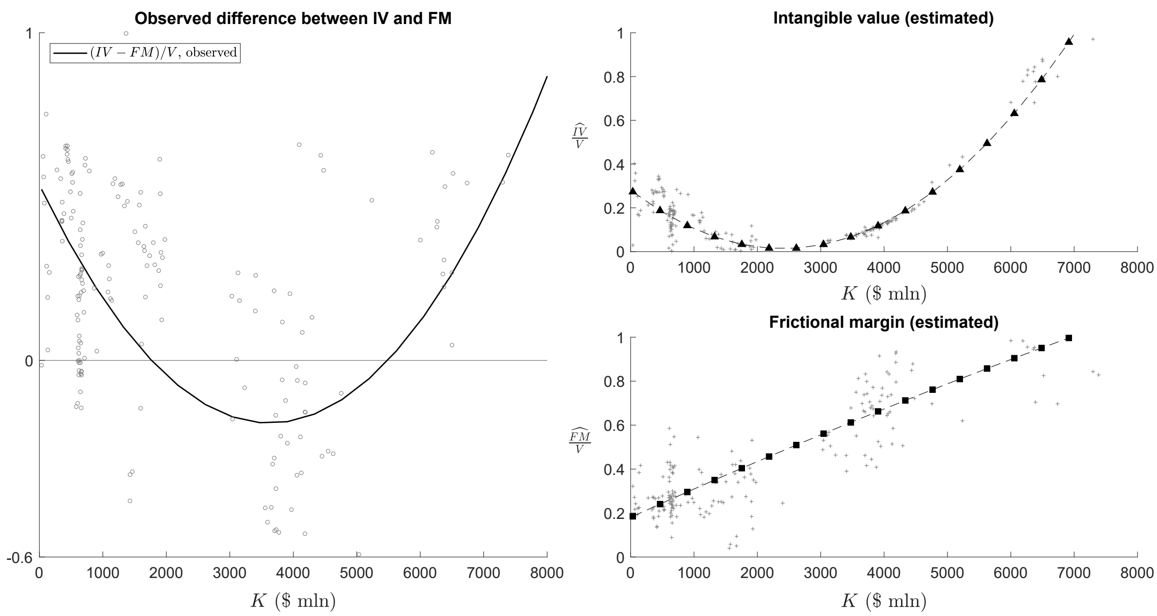



**Figure 4.3:** This figure shows how the intangible value and market-implied frictional costs, expressed as a fraction of market capitalization, vary as a function of insurers' capital according to our estimation (see specification (4.21)). The left-hand side plot shows the *observed* difference between $IV/V$ and $FM/V$. The right-hand side plots show the *estimated* intangible value $IV/V$ (top) and frictional margin $FM/V$ (bottom).

### 4.6.2 The drivers of Market-implied Frictional Costs

The relationship in Figure 4.3 between total estimated frictional costs and firm capital is the result of the combined effect of different firm characteristics and their interactions with insurers' capitalization. To investigate the effect of each characteristic on the market-implied frictional costs, Figure 4.4 plots how the estimated frictional costs vary with insurers' capital for different types of frictions. For each firm characteristic we take the 10%, 50% and 90% percentile while keeping the remaining characteristics at their median values.

All plots in Figure 4.4 show that as insurers' capitalization increases, frictional costs increase as a fraction of market capitalization. Consistent with the interpretation of firm characteristics in $Z_{it}^{FM}$ as proxies for different types of frictions, the plots show that insurers with more dispersed analysts' forecasts, higher taxes, more reinsurance ceded and larger reductions of dividends payouts have higher market-implied

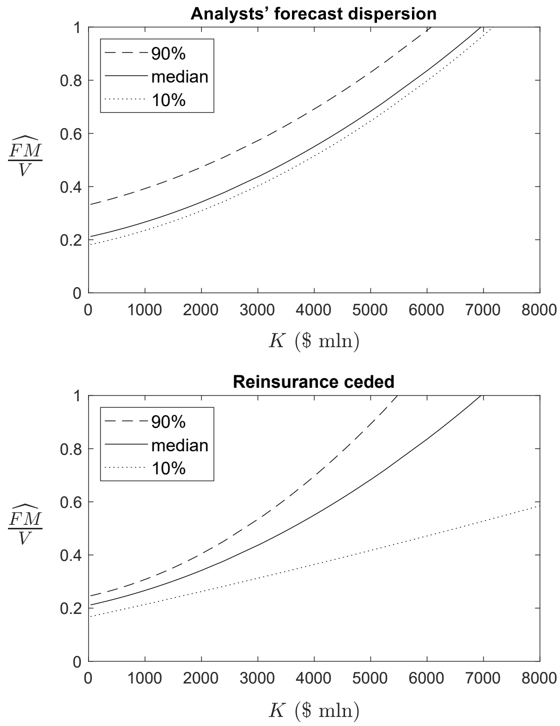



**Figure 4.4:** This figure shows how market-implied frictional costs, as a fraction of market capitalization ($FM/V$), vary with insurers' capital levels ($K$, on the horizontal axis). The plots are based on the parameter estimates of the full model (specification I, Table 4.5). Each graph compares the median firm to an otherwise equivalent firm with one frictional variable at its 10th and 90th percentile of the sample distribution. Each frictional variable is as defined in Table 4.3.

frictional costs. Moreover, according to our model the market expects insurers with higher reinsurance purchase and corporate taxes to experience ever increasing frictional costs at large capitalization levels. These results suggest that reinsurance is perceived by investors as a strategy to reduce frictional costs, especially agency costs associated with holding capital (Hancock et al. (2001), Pottier and Sommer (2006)).

Prior research has found a similar relation between different types of frictional costs and insurers' capital. Using data on observed insurance premiums for U.S. P&C insurers, Zanjani (2002) concludes that the impact of capital costs on premiums can be higher than 70% for insurers with large capital-to-premium ratios, while Koijen and Yogo (2015) show that the regulatory and financial costs for life insurers were as high as \$0.96 per dollar of statutory capital during the financial crisis in 2008. For catastrophe insurance, Harrington and Niehaus (2003) suggest that U.S. corporate income tax can produce tax costs on insurers' equity capital that exceed 100% of the present value of the expected claims.

In contrast to previous studies on insurers' frictional costs, however, we estimate their unobserved value as implied by *financial market* prices rather than *product market* prices. This implies that, for insurers with a diffuse shareholder base, our results provide an estimate of the value of frictional costs consistent with equilibrium assets prices formed according to information in financial markets. On the other hand, estimates based on written insurance premiums (Zanjani (2002), Koijen and Yogo (2015, 2016)) reflect information available to product markets, which is only partly accessible to financial markets through corporate reporting. Therefore, the two approaches are complementary because they rely on different information sets, providing estimates that are informative about expectations in the respective markets.

### 4.6.3 Market-implied frictions and statutory margins

Given our estimates of the frictional costs implied by insurers' market capitalization, it is natural to ask how the economic value of insurance liabilities compares to the value recorded under the relevant statutory regime. As argued in the introduction to this chapter, the value of insurance liabilities in any regime can be expressed as the sum of a best estimate liability plus an excess over the best estimate. It follows that variation in the value of insurance liabilities across regimes is explained by the regime-specific margin above the best estimate liability. We thus compare our estimates of the market-implied frictional costs, which capture the *economic* value of insurance liabilities in excess of the best estimate, to the statutory margin implied in U.S. statutory insurance reserves and to the size of the Solvency II risk margin.

According to the U.S. regulatory requirements, P&C insurers must report the statutory value of insurance reserves on an undiscounted basis and disclose any discount they might have applied when calculating the present value of future expected claims (National Association of Insurance Commissioners (1998)). It follows that the implied U.S. statutory margin above the best estimate is given by the discount that insurers are not allowed to record for statutory purposes.

In Figure 4.5, the top-left plot shows the average observed U.S. statutory discount as a fraction of the market value of equity for each insurer.[37] The average discount

[37]Not all insurers in our sample report an implicit discount, known as non-tabular discount, in

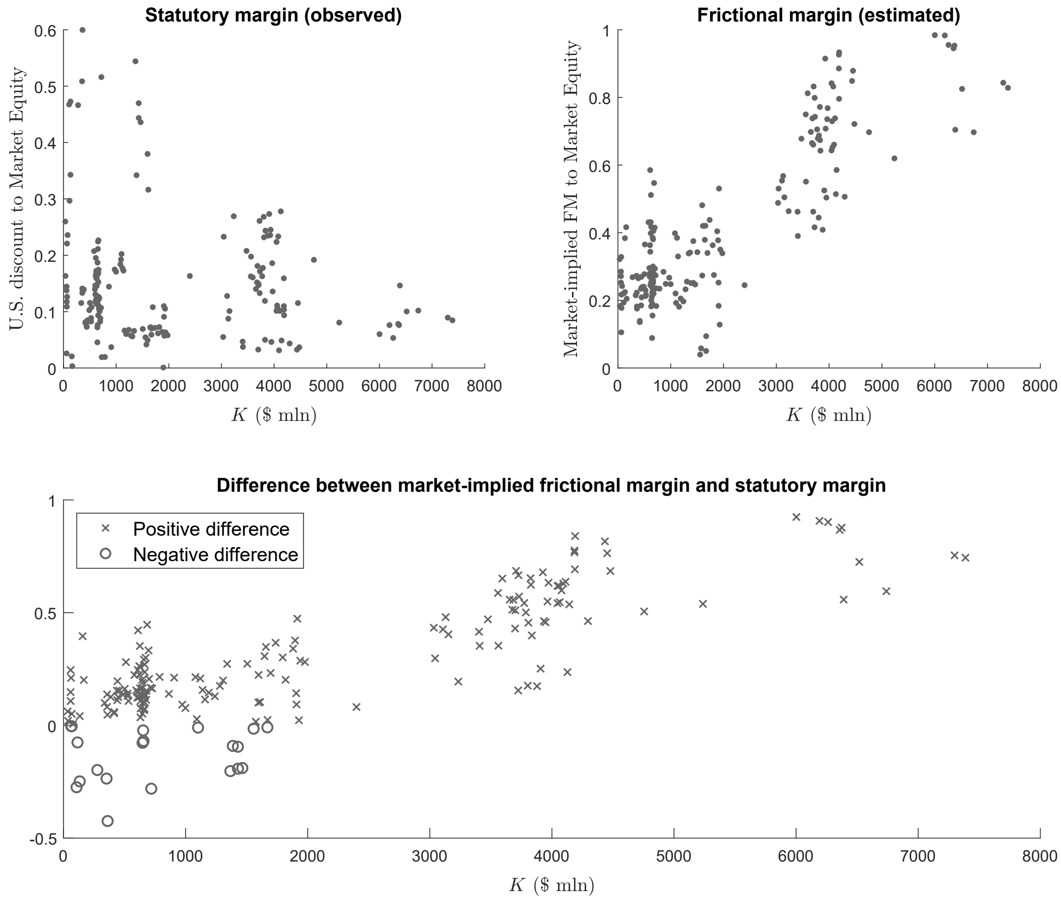



**Figure 4.5:** This figure shows how the average observed U.S. statutory discount (top-left plot) and the estimated frictional margin (top-right plot), both expressed as a fraction of market capitalization, vary with insurers' capital levels ($K$, on the horizontal axis). The graph at the bottom shows the difference between the estimated frictional margin (top-right plot) and the observed statutory margin (top-left plot) as a function of insurers' capital levels. The average observed U.S. statutory discount for each insurer is calculated based on the disclosed non-tabular discount in SAP statements (see Appendix 4.8.2.2 for details).

implied in the statutory value of insurance liabilities, expressed as a percent of market capitalization, decreases with capital. This suggests that the U.S. statutory regime, by requiring to record insurance reserves at their *undiscounted* value, can impose

their statutory statements. To obtain the statutory discount in these cases (see Appendix 4.8.2.2 for more details), we collect all the statutory discounts disclosed for each line of business by U.S. P&C insurers, expressed as a fraction of the undiscounted statutory value of insurance reserves, and calculate the weighted average discount for each insurer in our sample, with weights corresponding to the relative value of insurance reserves associated with each line of business.

implicitly higher requirements to insurers with lower capital levels.

To understand how the U.S. statutory requirements compare to a regime based on a market-consistent framework, the bottom plot in Figure 4.5 shows the difference between the estimated frictional margin and the implied statutory margin as a function of insurers' capital. This allows us not only to show how distant statutory liabilities are from their estimated economic value, but also to examine the change in the statutory value of insurance reserves that insurers would experience if the current statutory regime were to move to a framework based on economic principles.

The resulting difference in the value of insurance liabilities shows the consequences of adopting a market-based regulatory framework. Because some insurers with low capital levels bear proportionally smaller frictional costs compared to the required implicit statutory discount, Figure 4.5 shows that these insurers are also those that would benefit from recording their insurance liabilities at economic value. On the other hand, most insurers in our sample would find their capital positions worsening after passing to a market-consistent regime, as current statutory requirements ignore the frictional costs of holding equity capital invested in insurance firms.

The Solvency II regime in the European Union and the Swiss Solvency Test in Switzerland represent attempts to develop regulatory frameworks based on economic principles. Solvency II requires E.U. insurers to calculate their insurance liabilities as the sum of a best estimate liability plus a risk margin, given by the present value of future expected capital solvency requirements times a cost of capital rate equal to 6% (European Union (2013)). This approach is supposed to capture the market-consistent value of the capital costs that would be required on average by an insurance firm to take over the insurance liabilities of another insurer.

The similarity between the risk margin and the provision for frictional costs makes it instructive to relate our estimates to the actual size of the risk margin. Figure 4.6 shows how the risk margin for E.U. publicly-traded P&C insurers varies with their capitalization. Despite the 6% cost of equity rate is crucial to determine the relative size of the risk margin, Figure 4.6 is useful as it provides two pieces of evidence.[38] In line with our results, insurers with larger capital levels have a higher risk margin

[38] As pointed out in Bergesio et al. (2019), the capital cost under Solvency II, which has resorted to capital markets models for its calibration, may capture components other than the frictional costs of holding capital. These include the opportunity cost from investing directly in financial markets (base cost of capital) and market expectations about future returns on the franchise value (return on franchise, see e.g., Huber and Kinrade (2018)).

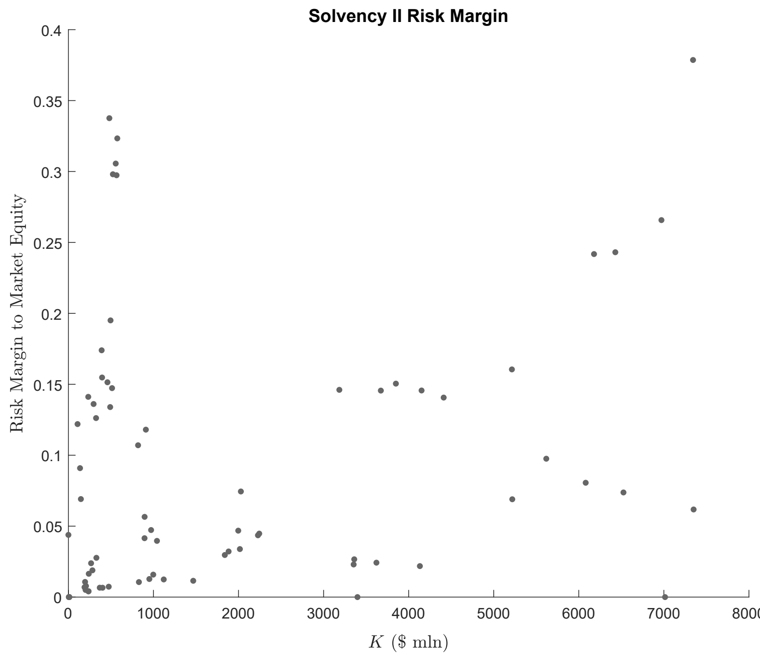



**Figure 4.6:** The figure shows how the Solvency II risk margin for publicly-traded E.U. P&C insurers, as a fraction of their market capitalization, varies with insurers' capital levels ($K$, on the horizontal axis). Insurers' capital is obtained from Quantitative Reporting Templates (QRT) as the difference between the Solvency II value of own funds.

relative to their market capitalization. Moreover, E.U. insurers can have a risk margin as high as 30% to 40% of their market capitalization, which suggests that our estimates may be reasonable when applied to P&C insurance companies in the U.S.[39] Given this evidence, it is natural to ask how our model would perform when applied to the sample of insurers in Figure 4.6. However, it is worth noting that while Solvency II provides publicly-available data on insurers' economic balance sheets, or close approximations of it, reliable data is only available at an annual frequency and from the date Solvency

[39]We point out here that the evidence provided in Figure 4.6 calls for caution when it comes to drawing definitive conclusions based on our estimates. Our analysis indicates that some U.S. insurers should bear a frictional margin well above the maximal risk margin of E.U. insurers. Moreover, the estimated functional forms in Figure 4.3 are different from those suggested by some authors, despite such predictions can only be partially related to our outcome variables. As a result, it would be desirable to test alternative hypotheses concerning the qualitative properties of the frictional margin and intangible value. In addition, variables that have the potential to improve on our estimates include capital ratios and leverage ratios, the growth in insurers' net income and the nature and duration of insurance liabilities. Hence, we recognize that our research is far from complete and further investigations are needed.

II entered into force (January 1st, 2016).[40]

## 4.7 Conclusion

We derive the economic value of insurance liabilities within a market-consistent framework, in which an insurer with a diffuse shareholder base values cash flows to shareholders consistently with financial market prices *and* with a broad ownership base. The resulting economic value of insurance liabilities is the sum of a best estimate liability plus a provision for future expected frictional costs due to market imperfections. We estimate the size of the provision for frictional costs implied in the market capitalization of publicly-traded U.S. P&C insurers and show that it increases with insurers' capital and vary from 20% to over 80% of their market capitalization. Moreover, the estimated market-implied provision for frictional costs is larger than the statutory provision for most U.S. insurers in our sample. When we compare our results to the risk margins of E.U. insurers, evidence suggests that our estimates for U.S. insurers lie in a reasonable range. We conclude that despite holding capital is costly due to frictional costs, insurers need to hold capital to support future business opportunities and balance the inherent trade-off between the franchise value and the frictional costs associated with any given capitalization.

## 4.8 Appendix

### 4.8.1 Economic value of insurance liabilities

This section of the appendix extends the results obtained within a two-period economy and derives the economic value of insurance liabilities for an insurer with in-force insurance policies providing coverage over $T > 0$ periods. Subsection 4.8.1.3 presents results for a fully equity-financed insurer, while Subsection 4.8.1.4 shows the impact of debt financing on the economic value of insurance liabilities. In line with our economic valuation framework, the extended results will be derived under the economic valuation measure $\mathbb{Q}^*$.

[40] While it would be interesting to conduct a similar analysis for E.U. insurance companies, we leave this study to future research.

#### 4.8.1.1 Definitions

We define below the relevant economic quantities that will be derived endogenously as part of the market-consistent value of insurance liabilities. Here, we retain the assumption of a fully equity-financed insurer.

For the purpose of the definitions, let $(\ell_t)_{t=1}^T$ be a sequence of i.i.d. non-negative random variables $\ell_t$ that represent aggregate losses at time $t$. Moreover, let $(\delta_t)_{t=0}^T$ be the sequence of liquidation decisions of the firm, with $\delta_t = 1$ indicating the insurer continues operations at $t$. To simplify the exposition, we set $\delta_0 = \delta_T = 1$. The sequence of assets at market value is denoted $(A_t)_{t=0}^T$, with $A_T = 0$, whereas $(\underline{L}_t)_{t=0}^T$ represents the sequence of statutory values of insurance liabilities. The sequence of taxable profits and the sequence of deadweight costs are given respectively by $(p_t^\tau)_{t=1}^T$ and $(c_t)_{t=1}^T$, with $c_T = 0$.

**Definition 4.8.1** (Best estimate liability)**.** *The best estimate liability at $t = 0$ is*

$$BEL_0 = \sum_{k=1}^{T} \frac{1}{(1+r)^k} \mathbb{E}_0^{\mathbb{Q}^*}[\ell_k].$$

**Definition 4.8.2** (Provision for frictional costs)**.** *Let $\tau \in (0,1)$ be the corporate tax rate. Moreover, consider the tax liability $TL_0$ and the provision for deadweight costs $DM_0$, given respectively by*

$$TL_0 = \sum_{k=1}^{T} \frac{1}{(1+r)^k} \mathbb{E}_0^{\mathbb{Q}^*}[\tau p_k^\tau], \qquad DM_0 = \sum_{k=1}^{T} \frac{1}{(1+r)^k} \mathbb{E}_0^{\mathbb{Q}^*}[c_k],$$

*where $TL_0$ identifies the present value of future expected tax payments and $DM_0$ represents the present value of future expected deadweight costs.*
*The provision for frictional costs is equal to the sum of a margin for taxes and a margin for deadweight costs*

$$FM_0 = TL_0 + DM_0.$$

**Definition 4.8.3** (Liquidation option value)**.** *The value of shareholders' option to liquidate the insurance firm at $t = 0$ is*

$$LO_0 = \sum_{k\in\{1,...,T-1\},T>1} \frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}\left[(BEL_k + FM_k - (\underline{L}_k - c_k)(1-\delta_k)\Pi_{i=0}^{k-1}\delta_i\right] +$$
$$+ \sum_{k\in\{1,...,T-1\},T>1} \frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)(\underline{L}_k + \ell_k + \tau p_k^\tau - \gamma_k A_{k-1})_+] +$$
$$+ \frac{1}{(1+r)^T}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{T}\delta_j(\ell_T + \tau p_T^\tau - \gamma_T A_{T-1})_+],$$

*where the first term represents the present value of the net benefits from transferring insurance obligations at their statutory value, whereas the second and third terms identify the present value of future expected cash flows from liquidating the insurer over the run-off period.*

#### 4.8.1.2 Equity value in run-off

We summarize our approach to derive the market-consistent value of insurance liabilities. We start from the market value of equity of an insurer, defined as the present value of future expected cash flows to shareholders over the run-off period of insurance liabilities (see Equation (4.7)), and derive the market-implied value of insurance liabilities by unfolding the present-value relation. As in Section 4.3, this approach is carried out by considering the market value of an insurer *as if* franchise value was zero. When the run-off period extends for $T > 0$ periods, Equation (4.7) implies that the market value of equity at $t = 0$ is

$$ENW_0 = \sum_{k=1}^{T} \frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k}\delta_j(\gamma_k A_{k-1} - \ell_k - \tau p_k^\tau - A_k - c_k)] + \tag{4.22}$$
$$+ \sum_{k\in\{1,..,T-1\},T>1} \mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)(\gamma_k A_{k-1} - \ell_k - \tau p_k^\tau - \underline{L}_k)_+] + \tag{4.23}$$
$$+ \frac{1}{(1+r)^T}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{T}\delta_j(\ell_T + \tau p_T^\tau - \gamma_T A_{T-1})_+], \tag{4.24}$$

where $r > 0$ is the one-period risk free rate of return. The previous equation is the natural extension of our two-period model to $T$ periods: the first summation (4.22) is the present value of future expected (default-free) cash flows from continuing operations until the T-th period; the second and third summations (4.23)-(4.24) represent the present value of future expected cash flows after taking into account the liquidation decision and limited liability of the insurer at each future date.

#### 4.8.1.3 Economic value of insurance liabilities with equity financing

Equipped with definitions (4.8.1)-(4.8.3), the next proposition shows the economic net worth, and the resulting economic value of insurance liabilities, when in-force policies provide coverage over $T$ periods.

**Proposition 4.8.4.** *Consider a limited-liability, risk-neutral insurer with economic balance sheet described by assets $A_0$ and in-force insurance liabilities $L_0$. Let the policies underlying insurance liabilities have coverage length extending for $T$ periods. Furthermore, assume the insurer has no financial leverage. Then, the insurer's economic net worth at time $t = 0$ is*

$$ENW_0 = A_0 - BEL_0 - FM_0 + LO_0 \tag{4.25}$$

*and the default-free market-consistent value of insurance liabilities is*

$$BEL_0 + FM_0, \tag{4.26}$$

*where $BEL_0$ is the default-free best estimate liability and $FM_0$ is the provision for frictional costs.*

*Proof.* We derive the result starting from the equity value $ENW_0$ in run-off with $T$ periods. We begin by considering summation (4.22), which we write equivalently as follows

$$\begin{aligned}
& \sum_{k=1}^{T} \frac{1}{(1+r)^k} \mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k} \delta_j (\gamma_k A_{k-1} - \ell_k - \tau p_k^\tau - A_k - c_k)], \\
= \quad & \sum_{k=1}^{T} \frac{1}{(1+r)^k} \mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1} \delta_j \gamma_k A_{k-1}] - \sum_{k=1}^{T} \frac{1}{(1+r)^k} \mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1} \delta_j (1-\delta_k) \gamma_k A_{k-1}] - \\
& - \sum_{k=1}^{T} \frac{1}{(1+r)^k} \mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k} \delta_j (\ell_k + \tau p_k^\tau + A_k + c_k)], \\
= \quad & \underbrace{\sum_{k=1}^{T} \frac{1}{(1+r)^k} \mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1} \delta_j \gamma_k A_{k-1}] - \sum_{k=1}^{T} \frac{1}{(1+r)^k} \mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k} \delta_j A_k)]}_{=A_0} -
\end{aligned}$$

$$
\begin{aligned}
& -\underbrace{\sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\ell_k]}_{=BEL_0}-\underbrace{\left(\sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\tau p_k^\tau]+\sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[c_k]\right)}_{=FM_0} - \\
& -\sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)\gamma_k A_{k-1}] + \\
& +\sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[(1-\Pi_{j=0}^{k}\delta_j)(\ell_k+\tau p_k^\tau+c_k)], \\
= \quad & A_0 - BEL_0 - FM_0 - \sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)\gamma_k A_{k-1}] + \\
& +\sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[(1-\Pi_{j=0}^{k}\delta_j)(\ell_k+\tau p_k^\tau+c_k)].
\end{aligned}
$$

We can re-write the last summation in the previous equality by using the identity

$$
1-\Pi_{j=0}^{k}\delta_j=\sum_{j=0}^{k}(1-\delta_j)\Pi_{i=0}^{j-1}\delta_i, \tag{4.27}
$$

to obtain

$$
\begin{aligned}
& \sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[(1-\Pi_{j=0}^{k}\delta_j)(\ell_k+\tau p_k^\tau+c_k)], \\
= & \sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}\left[\sum_{j=1}^{k}(1-\delta_j)\Pi_{i=0}^{j-1}\delta_i(\ell_k+\tau p_k^\tau+c_k)\right], \\
= & \sum_{k\in\{1,\ldots,T-1\},T>1}\frac{1}{(1+r)^{k+1}}\mathbb{E}_0^{\mathbb{Q}^*}\left[(\ell_{k+1}+\tau p_{k+1}^\tau+c_{k+1})\sum_{j=1}^{k}(1-\delta_j)\Pi_{i=0}^{j-1}\delta_i\right] + \\
& +\sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}\left[(\ell_k+\tau p_k^\tau+c_k)\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)\right].
\end{aligned}
$$

Moreover, one can express the first summation in the last equality as

$$
\sum_{k\in\{1,\ldots,T-1\},T>1}\frac{1}{(1+r)^{k+1}}\mathbb{E}_0^{\mathbb{Q}^*}\left[(\ell_{k+1}+\tau p_{k+1}^\tau+c_{k+1})\sum_{j=1}^{k}(1-\delta_j)\Pi_{i=0}^{j-1}\delta_i\right],
$$

$$= \sum_{k\in\{1,\ldots,T-1\},T>1} \frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}\left[(BEL_k + FM_k)(1-\delta_k)\Pi_{i=0}^{k-1}\delta_i\right].$$

Using the previous equations, (4.22) becomes

$$\begin{aligned}
&\sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k}\delta_j(\gamma_k A_{k-1} - \ell_k - \tau p_k^\tau - A_k - c_k)],\\
&= A_0 - BEL_0 - FM_0 -\\
&\quad - \sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)(\gamma_k A_{k-1} - \ell_k - \tau p_k^\tau - c_k)] +\\
&\quad + \sum_{k\in\{1,\ldots,T-1\},T>1}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}\left[(BEL_k + FM_k)(1-\delta_k)\Pi_{i=0}^{k-1}\delta_i\right].
\end{aligned}$$

Next, we consider the summation (4.23), which can be written as

$$\begin{aligned}
&\sum_{k\in\{1,\ldots,T-1\},T>1}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)(\gamma_k A_{k-1} - \ell_k - \tau p_k^\tau - \underline{L}_k)_+],\\
&= \sum_{k\in\{1,\ldots,T-1\},T>1}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)(\underline{L}_k + \ell_k + \tau p_k^\tau - \gamma_k A_{k-1})_+] +\\
&\quad + \sum_{k\in\{1,\ldots,T-1\},T>1}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)(\gamma_k A_{k-1} - \ell_k - \tau p_k^\tau - \underline{L}_k)].
\end{aligned}$$

Putting together (4.22), (4.23) and (4.24), we obtain

$$\begin{aligned}
ENW_0 &= A_0 - BEL_0 - FM_0 -\\
&\quad - \sum_{k=1}^{T}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)(\gamma_k A_{k-1} - \ell_k - \tau p_k^\tau - c_k)] +\\
&\quad + \sum_{k\in\{1,\ldots,T-1\},T>1}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}\left[(BEL_k + FM_k)(1-\delta_k)\Pi_{i=0}^{k-1}\delta_i\right] +\\
&\quad + \sum_{k\in\{1,\ldots,T-1\},T>1}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)(\underline{L}_k + \ell_k + \tau p_k^\tau - \gamma_k A_{k-1})_+] +\\
&\quad + \sum_{k\in\{1,\ldots,T-1\},T>1}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)(\gamma_k A_{k-1} - \ell_k - \tau p_k^\tau - \underline{L}_k)] +
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{(1+r)^T}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^T\delta_j(\ell_T+\tau p_T^\tau-\gamma_T A_{T-1})_+],\\
&= A_0 - BEL_0 - FM_0 + \\
&+\sum_{k\in\{1,\dots,T-1\},T>1}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}\left[(BEL_k+FM_k-(\underline{L}_k-c_k)(1-\delta_k)\Pi_{i=0}^{k-1}\delta_i\right]+\\
&+\sum_{k\in\{1,\dots,T-1\},T>1}\frac{1}{(1+r)^k}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^{k-1}\delta_j(1-\delta_k)(\underline{L}_k+\ell_k+\tau p_k^\tau-\gamma_k A_{k-1})_+]+\\
&+\frac{1}{(1+r)^T}\mathbb{E}_0^{\mathbb{Q}^*}[\Pi_{j=0}^T\delta_j(\ell_T+\tau p_T^\tau-\gamma_T A_{T-1})_+],\\
&= A_0 - BEL_0 - FM_0 + LO_0,
\end{aligned}
$$

where the last equality follows from Definition 4.8.3. Hence, the default-free market-consistent value of insurance liabilities is given by the sum of the best estimate liability ($BEL_0$) plus the margin for expected future frictional costs ($FM_0$), i.e. $BEL_0 + FM_0$. □

#### 4.8.1.4 Allowing for debt financing

In this section we show how to adjust our results when the capital structure includes debt financing. Consider an insurer issuing debt that is senior to insurance liabilities, i.e. it takes precedence over the insurer's assets in case of liquidation. To simplify exposition, assume the insurer issues one-period zero coupon bonds with face value $D_t$. Moreover, the insurer is subject to collateral constraints that ensure assets after tax payments are always sufficient to repay debt obligations, which effectively make debt default free.[41] In particular, the collateral constraints require that

$$A_t^\star \coloneqq \gamma_t A_{t-1} - \tau p_t^\tau - D_t \geq 0, \quad \text{a.s.} \tag{4.28}$$

where taxable profits account for the deduction of interest payments on debt

$$p_t^\tau = A_{t-1}(\gamma_t - 1) - \ell_t - r\frac{D_t}{(1+r)} - T_t + T_{t-1}. \tag{4.29}$$

Note that issuing debt increases the risk of the firm, as only assets after paying off

[41]Since debt is risk free, there are no agency costs associated with debt financing (Gamba and Triantis (2008)) and shareholders cannot benefit from their limited liability on debt obligations.

debt are available to honor insurance obligations. As a result, the minimum level of regulatory assets is likely to increase when leverage due to senior debt is introduced. Moreover, since part of the assets acts as collateral, the insurer will issue debt only when the firm has assets in excess of the regulatory minimum, because the firm would not otherwise satisfy the regulatory constraints and could not keep operating.

In the full model with debt financing and in-force insurance policies extending for $T$ periods, cash flows at date $t \in \{1, 2, ..., T\}$ become (assuming liquidation did not occur before $t$)

$$z_t = \begin{cases} A_t^\star - \ell_t - A_t + \frac{1}{1+r} D_{t+1} - c_t^D & \text{for } t \in \{1, 2, ..., T-1\} \text{ and } \delta_t = 1, \\ (A_t^\star - \ell_t + \tau(p_t^\tau)_- - \underline{L}_t)_+ & \text{for } t \in \{1, 2, ..., T-1\} \text{ and } \delta_t = 0, \\ (A_t^\star - \ell_t + \tau p_t^\tau)_+ & \text{for } t = T, \end{cases} \tag{4.30}$$

where $c_t^D \geq 0$ represents deadweight costs when the capital structure includes debt financing, which may introduce additional costs of financial distress compared to a fully equity-financed firm. Equation (4.30) can be used to derive the economic net worth, and thus the economic value of insurance liabilities, for insurers that resort to financial debt. Following a derivation analogous to Subsection 4.8.1.3, the economic net worth for an insurer with debt financing can be written as

$$ENW_0^D = A_0 - D_0 - BEL_0 - FM_0^D + LO_0^D, \tag{4.31}$$

where the superscript $D$ identifies the components of the economic net worth in Equation (4.25) for the model with financial debt. Equation (4.31) shows that financial debt affects the economic value of insurance liabilities via the liquidation option value $LO_0^D$ and the frictional margin $FM_0^D$.

With debt financing, Equation (4.28) implies that the value of the liquidation option, and in turn the value of shareholders' option to default, will embed the impact of collateral constraints on the value of the assets available for insurance payments. Specifically, when the capital structure includes debt, the option to default reflects the fact that only assets after debt payments are available to honour insurance obligations.

Debt financing generates also net benefits given by the tax shield from deductible debt interests minus the costs of financial distress associated with financial leverage.[42]

[42]The balance between tax benefits and costs of financial distress from debt funding is at the core

As interest payments on debt are tax deductible, the tax margin is reduced by the tax value of investment income on assets backing debt capital. As a result, debt financing effectively lowers the expected tax liability of the insurer and thus the resulting tax margin ($TM_t^D \leq TM_t$). However, the insurer's financial leverage may well increase market expectations concerning future expected financial distress costs, in which case the provision for deadweight costs ($DM_t^D \geq DM_t$) will increase. Hence, the net benefits to leverage depend on the trade-off between $TM_t^D$ and $DM_t^D$.

Another benefit of debt financing is obtained when the capital structure includes subordinated or hybrid debt, which in theory can be used to support risk taking (Hancock et al. (2001)). These forms of debt can help financial institutions to meet regulatory requirements and free up locked-in equity (Huber and Kinrade (2018)). When deadweight costs are associated with the level of equity capital held in the insurance firm, debt financing can be effective in decreasing these costs by lowering the capital base on which they are borne. This is particularly relevant when asymmetric information introduces agency costs that increase rapidly for large capital holdings, or when double taxation expose shareholders to increasingly high costs due to convex corporate tax schedules (Zanjani (2002)). It follows that debt funding affects markets expectations concerning the amount of deadweight costs (and taxes) that weigh on shareholders' capital. Note that assessing these benefits requires to model frictional costs explicitly, specifying the functional relation between frictional costs and the insurer's capital base on which those costs are borne.

### 4.8.2 Appendix: database construction

We construct our dataset by collecting information from several sources. U.S. GAAP financial statements are from COMPUSTAT. Statutory statements are from S&P SNL Financial, which provides access to the regulatory schedules filed by U.S. insurers to the NAIC. Data on analysts' activity is from Thomson Reuters I/B/E/S. Market prices are from CRSP (equity securities) and TRACE (debt securities).[43] In the following two subsections, we provide details on the criteria adopted and adjustments

of the trade-off theory, which predicts an optimal capital structure exists that maximises net benefits to leverage (see Korteweg (2010) for a review).

[43] SNL Financial, Wharton Research Data Services (WRDS) and databases available therein (COMPUSTAT, Thomson Reuters I/B/E/S, CRSP, TRACE) own the copyright of their respective data.

applied to arrive at our final sample. First, we present the criteria used to select our data. Then, we describe how we have adjusted U.S. GAAP financial statements to recover the economic balance sheet of the insurers in our sample.

#### 4.8.2.1 Sample selection

Following Scotti (2005), we retain those companies for which divergence between assets and insurance liabilities under statutory and U.S. GAAP principles does not exceed ±20% of the U.S. GAAP balance sheet. Moreover, we drop any observations for which the accounting or market value of assets, liabilities or equity are either negative or missing. We select insurers classified as groups in SNL Financial and exclude groups domiciled in Bermuda. Because the strategy we adopt to recover the best estimate liability relies on rules specific to P&C insurers, we include only groups with at least 90% of insurance liabilities related to the P&C business. Observable variables in Table 4.3 and our measure of insurers' capital are trimmed at the 5% and 95% percentile to ensure results are not driven by extreme observations. Furthermore, we retain observations for which our dependent variable is between -2 and 1. The former bound excludes 9 outliers with excessively low market capitalization compared to the corresponding market value of assets, whereas the upper bound keeps out 16 observations with assets lower than insurance and financial liabilities.

#### 4.8.2.2 Adjustments to U.S. GAAP balance sheet

Given U.S. insurers do not disclose their balance sheet at economic values, we apply a set of adjustments to their U.S. GAAP financial statements. For this purpose, we follow the approach in Hancock et al. (2001) and Scotti (2005), as well as the principles underlying the Swiss Re's Economic Value Management (EVM) framework. Table 4.6 lists the adjustments to selected balance sheet items of our P&C insurance sample.

Over our sample period, most of insurers' invested assets under U.S. GAAP are reported at fair value. However, a few asset classes are held at historical cost, including fixed income classified as held-to-maturity and real estate investments. In order to obtain a proxy for the economic value of these assets, we use Schedule D from NAIC statutory statements, which contains detailed information on the fair value of insurers' investments in bonds, stocks, real estate and other short term investments. If we cannot adjust the value of an asset, we retain its disclosed value under U.S. GAAP.

Insurers' investments are complemented with the remainder of insurers' assets, from which we exclude the value of goodwill and other intangible assets.[44]

| ITEM | ADJUSTMENT |
|---|---|
| *Assets* | |
| Fixed income securities | If there are fixed income investments that are not reported at fair value under U.S. GAAP, we retrieve the fair value of fixed income securities from NAIC Schedule D, Part 1. |
| Real Estate | The fair value of real assets investments is obtained from NAIC Schedule A, Part 1. |
| Goodwill and other intangible assets | These assets are not included in economic assets. |
| *Liabilities* | |
| Reserves for Unpaid losses and Loss adjustment expenses | To adjust U.S. GAAP loss reserves for discounting, we proceed in steps. First, we discount the insurer's statutory loss reserves for each business line and fiscal period by applying the median non-tabular discount disclosed for that business line and fiscal period by SNL P&C insurers. Second, we calculate an implied discount ratio for each firm and fiscal period as the ratio of total discounted statutory loss reserves to undiscounted statutory loss reserves. Third, we apply the resulting implied discount to U.S. GAAP loss reserves. |
| Financial debt | If the fair value of debt is not disclosed under U.S. GAAP, we resort to TRACE for the market value of the traded portion of financial debt. Otherwise, we consider the disclosed value of debt. |

**Table 4.6:** The table shows the adjustments applied to selected items of the U.S. GAAP balance sheet of P&C insurers in order to obtain estimates of their economic values. Adjustments are based on accounting rules in place over the sample period 2009-2019.

Insurance liabilities of P&C insurers typically are not discounted under U.S. GAAP, whereas our economic framework requires to discount future expected insurance cash flows to arrive at a best estimate. In order to obtain the discounted value of P&C reserves, we rely on statutory statements.[45] The statutory value of P&C insurance

[44]Non-invested assets include, among others, cash and cash equivalents, financial derivatives and reinsurance receivables. In principle, we might want to account for counterparty credit risk when valuing insurance-related assets (Swiss Re's EVM). However, due to lack of information, we use the reported value of these assets.

[45]The total value of statutory P&C insurance reserves results from i) loss and loss adjustment expense reserves (case reserves), ii) incurred but not yet reported reserves (IBNR reserves) and iii)

reserves, which is used to calculate insurers' statutory capital, does not allow for discounting.[46] For this reason, insurers are required to disclose in the NAIC Schedule P, Part 1 any discount implicit in the reported cases reserves, known as nontabular discount. We collect all nontabular discounts disclosed by SNL P&C insurers over our sample period and calculate the median discount for each line of business (LOB) and fiscal period. These discounts are then used to adjust the corresponding statutory (undiscounted) case reserves and IBNR reserves of those firms in our sample for which a discount is not available. The total discounted statutory loss reserves for each firm and fiscal period is obtained as the sum of discounted loss reserves across business lines. The last step consists in applying the resulting implied discount ratio, which we define as the ratio of discounted to non-discounted statutory loss reserves, to U.S. GAAP loss reserves for each firm and fiscal period. The remainder of insurance liabilities, which include unearned premiums reserves and other insurance liabilities, is taken at reported value. While the resulting best estimate of insurance liabilities is not free of potential distortions arising from management assumptions underlying disclosed discounts, we believe it represents a direct and transparent method to account for discounting of insurance liabilities.[47] The best estimate of insurance liabilities is complemented by the remainder of U.S. GAAP liabilities, which include among others derivative financial instruments, pension benefits and payables related to investing activities.

The reported value of financial debt is adjusted to fair values whenever possible. If fair values are not disclosed in U.S. GAAP financial statements, we retrieve the traded portion of financial debt from TRACE. If no adjustment can be applied, we retain the disclosed value of debt under U.S. GAAP.

---

unearned premiums reserves (UPR).

[46] Loss reserves can be discounted if arising from business lines with fixed and determinable payments, e.g. workers' compensation and long-term disability (SSAP n. 65 - Property and Casualty contracts).

[47] An alternative approach would consist in using information on historical loss experience, which is provided in loss triangles from the NAIC Schedule P (Part 2,3,4), to estimate projected future losses and obtain their discounted values. According to Nelson (2000), an advantage of using actual paid loss data is that errors in management estimates, whether intentional or unintentional, do not affect the projections. The downside, however, which prevents us from relying on loss triangles, is that such an approach requires reliable data on historical losses for each line of business. In our sample, most of the insurers operate in lines of business with long tails, for which loss triangles do not provide sufficiently long historical data to produce reliable estimates. Moreover, we are unable to obtain projected future claims for many lines of business with incomplete loss triangles.

# Part III

# Bibliography

# Bibliography


Albrecher, H., Beirlant, J., and Teugels, J. L. (2017). *Reinsurance: actuarial and statistical aspects*. John Wiley & Sons.

Albrecher, H., Eisele, K.-T., Steffensen, M., and Wüthrich, M. V. (2022). On the cost-of-capital rate under incomplete market valuation. *Working Paper*.

Aliprantis, C. D. and Border, K. C. (2013). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer-Verlag Berlin and Heidelberg GmbH & Company KG.

Almagro, M. and Ghezzi, T. L. (1988). Federal income taxes—provisions affecting property/casualty insurers. In *Proceedings of the Casualty Actuarial Society*, volume 75, pages 95–161.

Almeida, H., Campello, M., and Weisbach, M. S. (2004). The cash flow sensitivity of cash. *Journal of Finance*, 59(4):1777–1804.

APRA (2019). Insurance Liability Valuation. *Australian Prudential Regulation Authority (APRA), Prudential Standard GPS 340*.

Armstrong, C. S., Core, J. E., Taylor, D. J., and Verrecchia, R. E. (2011). When does information asymmetry affect the cost of capital? *Journal of Accounting Research*, 49(1):1–40.

Arrow, K. J. (1974). *Essays in the theory of risk-bearing*, volume 121. North-Holland Amsterdam.

Arrow, K. J. and Lind, R. C. (1970). Uncertainty and the evaluation of public investment decisions. *American Economic Review*, 60(3):364–378.

Artzner, P., Eisele, K.-T., and Schmidt, T. (2020). No arbitrage in insurance and the QP-rule. *Available at SSRN 3607708*.

Asmussen, S. and Steffensen, M. (2020). *Risk and insurance*. Springer.

Aubin, J.-P. and Frankowska, H. (2009). *Set-valued analysis*. Springer Science & Business Media.

Aunon-Nerin, D. and Ehling, P. (2008). Why firms purchase property insurance. *Journal of Financial Economics*, 90(3):298–312.

Babbel, D. F. (1999). Components of Insurance Firm Value, and the Present Value of Liabilities. In *Investment Management for Insurers*, chapter 2. Frank J. Fabozzi Associates, Publishers.

Babbel, D. F. and Merrill, C. (1998). Economic valuation models for insurers. *North American actuarial journal*, 2(3):1–15.

Babbel, D. F. and Merrill, C. (2005). Real and illusory value creation by insurance companies. *Journal of Risk and Insurance*, 72(1):1–21.

Baker, T. and Siegelman, P. (2013). The law and economics of liability insurance: A theoretical and empirical review. *Research handbook on the economics of torts*.

Barigou, K., Bignozzi, V., Tsanakas, A., et al. (2022). Insurance valuation: A two-step generalised regression approach. *ASTIN Bulletin*, 52(1):211–245.

Barth, M. E., Hodder, L. D., and Stubben, S. R. (2008). Fair value accounting for liabilities and own credit risk. *The Accounting Review*, 83(3):629–664.

Basu, S. and Fernald, J. G. (1997). Returns to scale in us production: Estimates and implications. *Journal of Political Economy*, 105(2):249–283.

Bäuerle, N. and Rieder, U. (2011). *Markov decision processes with applications to finance*. Springer Science & Business Media.

Baumstark, L. and Gollier, C. (2014). The relevance and the limits of the Arrow-Lind theorem. *Journal of Natural Resources Policy Research*, 6(1):45–49.

Becker, B. and Ivashina, V. (2015). Reaching for yield in the bond market. *Journal of Finance*, 70(5):1863–1902.

Bednarek, Z. and Moszoro, M. (2014). The Arrow–Lind theorem revisited: Ownership concentration and valuation. *Applied Financial Economics*, 24(5):357–375.

Bergesio, A., Huber, P., Koch-Medina, P., and Wilhelmy, L. (2019). The valuation of insurance liabilities: A framework based on first principles. *Available at SSRN 3386182*.

Berk, J. B., Green, R. C., and Naik, V. (1999). Optimal investment, growth options, and security returns. *Journal of Finance*, 54(5):1553–1607.

Berk, J. B., Stanton, R., and Zechner, J. (2010). Human capital, bankruptcy, and capital structure. *Journal of Finance*, 65(3):891–926.

Bernard, C., He, X., Yan, J.-A., and Zhou, X. Y. (2015). Optimal insurance design under rank-dependent expected utility. *Mathematical Finance*, 25(1):154–186.

Bernegger, S. (1997). The Swiss Re Exposure Curves and the MBBEFD Distribution Class. *ASTIN Bulletin: The Journal of the IAA*, 27(1):99–111.

Berry-Stölzle, T. R., Nini, G. P., and Wende, S. (2014). External financing in the life insurance industry: Evidence from the financial crisis. *Journal of Risk and Insurance*, 81(3):529–562.

Bi, J., Cai, J., et al. (2019). Optimal investment–reinsurance strategies with state dependent risk aversion and VaR constraints in correlated markets. *Insurance: Mathematics and Economics*, 85(C):1–14.

Bolton, P., Chen, H., and Wang, N. (2011). A unified theory of tobin's q, corporate investment, financing, and risk management. *Journal of Finance*, 66(5):1545–1578.

Boomhower, J. (2019). Drilling like there's no tomorrow: Bankruptcy, insurance, and environmental risk. *American Economic Review*, 109(2):391–426.

Brealey, R. A., Cooper, I. A., and Habib, M. A. (1997). Investment appraisal in the public sector. *Oxford Review of Economic Policy*, 13(4):12–28.

Caillaud, B., Dionne, G., and Jullien, B. (2000). Corporate insurance with optimal financial contracting. *Economic Theory*, 16(1):77–105.

Carlier, G. and Dana, R.-A. (2003). Pareto efficient insurance contracts when the insurer's cost function is discontinuous. *Economic Theory*, 21(4):871–893.

CEIOPS (2009). Final CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Technical Provisions – Article 86(d) Calculation of the Risk Margin. *Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS)*, DOC-36/09.

Chandra, V. and Sherris, M. (2006). Capital management and frictional costs in insurance. *Australian Actuarial Journal*, 12(4):399.

Chen, C. Y. and PonArul, R. (1989). On the tax incentive for corporate insurance purchase. *Journal of Risk and Insurance*, 56(2):306–311.

Cheng, J., Weiss, M. A., et al. (2012). Capital structure in the property-liability insurance industry: Tests of the tradeoff and pecking order theories. *Journal of Insurance Issues*, 35(1):1–43.

Cherbonnier, F. and Gollier, C. (2022). Risk-adjusted social discount rates. *The Energy Journal*, 43(4).

Chi, Y. and Wei, W. (2020). Optimal insurance with background risk: An analysis of general dependence structures. *Finance and Stochastics*, 24(4):903–937.

Chiang, C.-C., Kim, H. H., and Niehaus, G. (2019). Opaque liabilities, learning, and the cost of equity capital for insurers. *Journal of Risk and Insurance, Forthcoming*.

Clementi, G. L. and Hopenhayn, H. A. (2006). A theory of financing constraints and firm dynamics. *Quarterly Journal of Economics*, 121(1):229–265.

Cooley, T. F. and Quadrini, V. (2001). Financial markets and firm dynamics. *American Economic Review*, 91(5):1286–1310.

Cooper, I. (2006). Asset pricing implications of nonconvex adjustment costs and irreversibility of investment. *Journal of Finance*, 61(1):139–170.

Cummins, J. D. and Grace, E. (1994). Tax management and investment strategies of property-liability insurers. *Journal of Banking & Finance*, 18(1):43–72.

Da, Z., Guo, R.-J., and Jagannathan, R. (2012). CAPM for estimating the cost of equity capital: Interpreting the empirical evidence. *Journal of Financial Economics*, 103(1):204–220.

Dana, R.-A. and Scarsini, M. (2007). Optimal risk sharing with background risk. *Journal of Economic Theory*, 133(1):152–176.

DeAngelo, H., DeAngelo, L., and Whited, T. M. (2011). Capital structure dynamics and transitory debt. *Journal of Financial Economics*, 99(2):235–261.

Delong, L., Dhaene, J., Barigou, K., et al. (2019). Fair valuation of insurance liability cash-flow streams in continuous time: Theory. *Insurance: Mathematics and Economics*, 88(C):196–208.

DeMarzo, P. M. (1993). Majority voting and corporate control: the rule of the dominant shareholder. *The Review of Economic Studies*, 60(3):713–734.

Derrig, R. A. and Ostaszewski, K. M. (1997). Managing the tax liability of a property-liability insurance company. *Journal of Risk and Insurance*, 64(4):695–711.

Devineau, L. and Chauvigny, M. (2011). Replicating portfolios: calibration techniques for the calculation of the Solvency II economic capital. *Bulletin Français d'Actuariat*, 21:59–97.

Dhaene, J., Stassen, B., Barigou, K., Linders, D., Chen, Z., et al. (2017a). Fair valuation of insurance liabilities: Merging actuarial judgement and market-consistency. *Insurance: Mathematics and Economics*, 76(C):14–27.

Dhaene, J., Van Hulle, C., Wuyts, G., Schoubben, F., and Schoutens, W. (2017b). Is the capital structure logic of corporate finance applicable to insurers? Review and analysis. *Journal of Economic Surveys*, 31(1):169–189.

Dicks, D. L. and Garven, J. R. (2022). Asymmetric information and insurance cycles. *Journal of Risk and Insurance*, 89(2):449–474.

Dionne, G. et al. (2000). *Handbook of insurance*. Springer.

Doherty, N. A. (2000). Innovation in corporate risk management: the case of catastrophe risk. In *Handbook of Insurance*, pages 503–539. Springer.

Doherty, N. A. and Schlesinger, H. (1983). The optimal deductible for an insurance policy when initial wealth is random. *Journal of Business*, 56(4):555–565.

Doherty, N. A. and Schlesinger, H. (1983a). Optimal insurance in incomplete markets. *Journal of Political Economy*, 91(6):1045–1054.

Doherty, N. A. and Schlesinger, H. (1985). Incomplete markets for insurance: An overview. *Journal of Risk and Insurance*, 52(3):402–423.

Doherty, N. A. and Tinic, S. M. (1981). Reinsurance under conditions of capital market equilibrium: A note. *Journal of Finance*, 36(4):949–953.

Dou, W. W., Taylor, L. A., Wang, W., and Wang, W. (2021). Dissecting bankruptcy frictions. *Journal of Financial Economics*, 142(3):975–1000.

EIOPA (2018). EIOPA's second set of advice to the European Commission on specific items in the Solvency II Delegated Regulation. *European Insurance and Occupational Pensions Authority (EIOPA) Guidelines*, BoS-18/075.

EIOPA (2019). Consultation Paper on the Opinion on the 2020 review of Solvency II. *European Insurance and Occupational Pensions Authority (EIOPA) Guidelines*, BoS-19/465.

Ekern, S. and Wilson, R. (1974). On the theory of the firm in an economy with incomplete markets. *The Bell Journal of Economics and Management Science*, pages 171–180.

Ellul, A., Jotikasthira, C., and Lundblad, C. T. (2011). Regulatory pressure and fire sales in the corporate bond market. *Journal of Financial Economics*, 101(3):596–620.

Ellul, A., Jotikasthira, C., Lundblad, C. T., and Wang, Y. (2015). Is historical cost accounting a panacea? Market stress, incentive distortions, and gains trading. *Journal of Finance*, 70(6):2489–2538.

Engsner, H., Lindholm, M., and Lindskog, F. (2017). Insurance valuation: A computable multi-period cost-of-capital approach. *Insurance: Mathematics and Economics*, 72(C):250–264.

Epermanis, K. and Harrington, S. E. (2006). Market discipline in property/casualty insurance: Evidence from premium growth surrounding changes in financial strength ratings. *Journal of Money, Credit and Banking*, pages 1515–1544.

European Union (2009). Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II). *Official Journal of the European Union*, 52:1–155.

European Union (2013). Regulation (EU) No 462/2013 of the European Parliament and of the Council of 21 May 2013 amending Regulation (EC) No 1060/2009 on credit rating agencies Text with EEA relevance. *Official Journal of the European Union*, 56:1–33.

European Union (2016). Commission Delegated Regulation (EU) 2016/467 of 30 September 2015 amending Commission Delegated Regulation (EU) 2015/35 concerning the calculation of regulatory capital requirements for several categories of assets held by insurance and reinsurance undertakings (Text with EEA relevance). *Official Journal of the European Union*, 85:6–19.

Exley, C. and Smith, A. (2006). The cost of capital for financial firms. *British Actuarial Journal*, 12(1):229.

Faulkender, M. and Wang, R. (2006). Corporate financial policy and the value of cash. *Journal of Finance*, 61(4):1957–1990.

Federal Office Of Private Insurance (2006). Technical document on the Swiss Solvency Test. *Swiss Confederation*.

Feldblum, S. (2007). Capital structure, solvency regulation, and federal income taxes for property-casualty insurance companies. *Variance*, 1(2):157–172.

Fisher, A. C. (1973). Environmental externalities and the Arrow-Lind public investment theorem. *American Economic Review*, 63(4):722–725.

Fisher, I. (1930). *Theory of interest: as determined by impatience to spend income and opportunity to invest it.* Augustusm Kelly Publishers, Clifton.

Floreani, A. (2011). Risk margin estimation through the cost of capital approach: Some conceptual issues. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 36(2):226–253.

Foldes, L. P. and Rees, R. (1977). A note on the Arrow-Lind theorem. *American Economic Review*, 67(2):188–193.

Franks, J. and Loranth, G. (2014). A study of bankruptcy costs and the allocation of control. *Review of Finance*, 18(3):961–997.

Froot, K., Venter, G., and Major, J. (2004). Capital and value of risk transfer. In *14th Annual International AFIR Colloquium*, pages 181–195.

Froot, K. A. (2007). Risk management, capital budgeting, and capital structure policy for insurers and reinsurers. *Journal of Risk and Insurance*, 74(2):273–299.

Froot, K. A., Scharfstein, D. S., and Stein, J. C. (1993). Risk management: Coordinating corporate investment and financing policies. *Journal of Finance*, 48(5):1629–1658.

Froot, K. A. and Stein, J. C. (1998). Risk management, capital budgeting, and capital structure policy for financial institutions: an integrated approach. *Journal of Financial Economics*, 47(1):55–82.

Gamba, A. and Triantis, A. (2008). The value of financial flexibility. *Journal of Finance*, 63(5):2263–2296.

Gardner, R. (1979). The Arrow-Lind theorem in a continuum economy. *American Economic Review*, 69(3):420–422.

Garven, J. and MacMinn, R. D. (2013). On the demand for corporate insurance: Creating value. In *Handbook of Insurance*, pages 487–516. Springer.

Garven, J. R. and Louberge, H. (1996). Reinsurance, taxes, and efficiency: a contingent claims model of insurance market equilibrium. *Journal of Financial Intermediation*, 5(1):74–93.

Garven, J. R. and MacMinn, R. D. (1993). The underinvestment problem, bond covenants, and insurance. *Journal of Risk and Insurance*, 60(4):635–647.

Gavira-Durón, N., Mayorga-Serna, D., and Bagatella-Osorio, A. (2020). The financial impact of the implementation of Solvency II on the Mexican insurance sector. *The Geneva Papers on Risk and Insurance-Issues and Practice*, pages 1–26.

Geddes, R. R. and Goldman, J. (2022). Institutional economics and the cost of capital for infrastructure projects. *Journal of Economic Policy Reform*, 25(2):85–102.

Ghesquiere, F. and Mahul, O. (2010). Financial protection of the state against natural disasters: A primer. *Policy Research working paper ; no. WPS 5429. World Bank.*

Ghossoub, M. (2017). Arrow's theorem of the deductible with heterogeneous beliefs. *North American Actuarial Journal*, 21(1):15–35.

Gollier, C. (1996). Optimum insurance of approximate losses. *Journal of Risk & Insurance*, 63(3):369–380.

Gollier, C. (2011). Discounting and risk adjusting non-marginal investment projects. *European Review of Agricultural Economics*, 38(3):325–334.

Gollier, C. (2021). The welfare cost of ignoring the beta. Technical report, Working Paper.

Gollier, C., Koehl, P.-F., and Rochet, J.-C. (1997). Risk-taking behavior with limited liability and risk aversion. *Journal of Risk and Insurance*, 64(2):347–371.

Gollier, C. and Schlesinger, H. (1996). Arrow's theorem on the optimality of deductibles: a stochastic dominance approach. *Economic Theory*, 7(2):359–363.

Gomes, J. F. (2001). Financing investment. *American Economic Review*, 91(5):1263–1285.

Gompers, P. A. and Metrick, A. (2001). Institutional investors and equity prices. *The quarterly journal of Economics*, 116(1):229–259.

Grace, M. F., Kamiya, S., Klein, R. W., and Zanjani, G. H. (2019). Market discipline and guaranty funds in life insurance. *Available at SSRN*, 2597052.

Graham, J. R. and Harvey, C. R. (2001). The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics*, 60(2-3):187–243.

Graham, J. R. and Leary, M. T. (2011). A review of empirical capital structure research and directions for the future. *Annu. Rev. Financ. Econ.*, 3(1):309–345.

Grant, S. and Quiggin, J. (2003). Public investment and the risk premium for equity. *Economica*, 70(277):1–18.

Gravelle, H. and Rees, R. (2004). *Microeconomics.* Pearson education.

Gredil, O. R., Kapadia, N., and Lee, J. H. (2022). On the information content of credit ratings and market-based measures of default risk. *Journal of Financial Economics*, 146(1):172–204.

Grout, P. A. (2003). Public and private sector discount rates in public-private partnerships. *The Economic Journal*, 113(486):62–62.

Han, L.-M. and MacMinn, R. D. (1990). Limited liability, corporate value, and the demand for liability insurance. *Journal of Risk and Insurance*, 57(4):581–608.

Hancock, J., Huber, P., and Koch, P. (2001). Value creation in the insurance industry. *Risk Management and Insurance Review*, 4(2):1–9.

Harrington, S. E. and Niehaus, G. (2003). Capital, corporate income taxes, and catastrophe insurance. *Journal of Financial Intermediation*, 12(4):365–389.

Harstad, B. (2005). Majority rules and incentives. *The Quarterly Journal of Economics*, 120(4):1535–1568.

Hart, S. (1989). Shapley value. In *Game theory*, pages 210–216. Springer.

Hau, A. (2006a). The liquidity demand for corporate property insurance. *Journal of Risk and Insurance*, 73(2):261–278.

Hau, A. (2006b). Production under uncertainty with insurance or hedging. *Insurance: Mathematics and Economics*, 38(2):347–359.

Heidinger, D. and Gatzert, N. (2018). Awareness, determinants and value of reputation risk management: Empirical evidence from the banking and insurance industry. *Journal of banking & finance*, 91(6):106–118.

Hennessy, C. A. and Whited, T. M. (2005). Debt dynamics. *Journal of Finance*, 60(3):1129–1165.

Hennessy, C. A. and Whited, T. M. (2007). How Costly Is External Financing? Evidence from a Structural Estimation. *Journal of Finance*, 62(4):1705–1745.

Hernández-Lerma, O. and Lasserre, J. B. (2012). *Further topics on discrete-time Markov control processes*, volume 42. Springer Science & Business Media.

Hill, R. D. (1979). Profit regulation in property-liability insurance. *Bell Journal of Economics*, 10(1):172–191.

Hill, R. D. and Modigliani, F. (1987). The Massachusetts model of profit regulation in nonlife insurance: an appraisal and extensions. In *Fair rate of return in property-liability insurance*, pages 27–53. Springer.

Hirshleifer, J. (1965). Investment decision under uncertainty: choice—theoretic approaches. *Quarterly Journal of Economics*, 79(4):509–536.

Hitchcox, A., Hinder, I., Kaufman, A., Maynard, T., Smith, A., White, M., et al. (2007). Assessment of target capital for general insurance firms. *British Actuarial Journal*, 13(1):81–168.

Hofmann, A., Häfen, O. V., and Nell, M. (2019). Optimal insurance policy indemnity schedules with policyholders' limited liability and background risk. *Journal of Risk and Insurance*, 86(4):973–988.

Hong, L. (2018). A note on Mossin's theorem for deductible insurance given random initial wealth. *Scandinavian Actuarial Journal*, 2018(5):404–411.

Hong, S. K., Lew, K. O., MacMinn, R. D., and Brockett, P. (2011). Mossin's theorem given random initial wealth. *Journal of Risk and Insurance*, 78(2):309–324.

Hoyt, R. E. and Khang, H. (2000). On the demand for corporate property insurance. *Journal of Risk and Insurance*, 67(1):91.

Huber, P. and Kinrade, N. (2018). A Generic Framework for the Economic Valuation of Insurance Liabilities. *Working Paper*.

Huberman, G., Mayers, D., Smith Jr, C. W., et al. (1983). Optimal insurance policy indemnity schedules. *Bell Journal of Economics*, 14(2):415–426.

IAIS (2020). Instructions for the April 2020 Insurance Capital Standard (ICS) Data Collection Exercise of the Monitoring Period Project. *IAIS, International Association of Insurance Supervisors*.

Irresberger, F. and Peng, Y. (2017). Shadow insurance usage and capital management in life insurance groups. *Fox School of Business Research Paper*, (16-022).

Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review*, 76(2):323–329.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305 – 360.

Jiang, S.-j., Tsai, J. T.-H., and Xiang, F. (2021). Dynamics of underwriting profits in the US market: Payout patterns and regulation effects. *International Journal of Finance & Economics*, pages 1–9.

Jost, P.-J. (1996). Limited liability and the requirement to purchase insurance. *International Review of Law and Economics*, 16(2):259–276.

Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008). *Modern actuarial risk theory: using R*, volume 128. Springer Science & Business Media.

Kalai, E. and Samet, D. (1987). On weighted shapley values. *International Journal of Game Theory*, 16(3):205–222.

Ke, B. (2001). Taxes as a determinant of managerial compensation in privately held insurance companies. *The Accounting Review*, 76(4):655–674.

Keeley, M. C. (1990). Deposit insurance, risk, and market power in banking. *American Economic Review*, 80(5):1183–1200.

Keeton, W. R. and Kwerel, E. (1984). Externalities in automobile insurance and the underinsured driver problem. *Journal of Law and Economics*, 27(1):149–179.

Kelsey, D. and Milne, F. (1996). The existence of equilibrium in incomplete markets and the objective function of the firm. *Journal of Mathematical Economics*, 25(2):229–245.

Kirti, D. and Peria, M. S. M. (2017). When gambling for resurrection is too risky. *IMF Working Papers*, 2017(180).

Klein, M. (1997). The risk premium for evaluating public projects. *Oxford Review of Economic Policy*, 13(4):29–42.

Koch-Medina, P., Moreno-Bromberg, S., Ravanelli, C., and Šikić, M. (2021). Revisiting optimal investment strategies of value-maximizing insurance firms. *Insurance: Mathematics and Economics*, 99:131–151.

Koijen, R. S. and Yogo, M. (2015). The cost of financial frictions for life insurers. *American Economic Review*, 105(1):445–75.

Koijen, R. S. and Yogo, M. (2016). Shadow insurance. *Econometrica*, 84(3):1265–1287.

Konrad, K. A., Skaperdas, S., et al. (1993). Self-insurance and self-protection: A nonexpected utility analysis. *The Geneva Risk and Insurance Review*, 18(2):131–146.

Korteweg, A. (2010). The net benefits to leverage. *Journal of Finance*, 65(6):2137–2170.

Kraus, A. and Litzenberger, R. H. (1973). A state-preference model of optimal financial leverage. *Journal of Finance*, 28(4):911–922.

Laux, C. and Muermann, A. (2010). Financing risk transfer under governance problems: Mutual versus stock insurers. *Journal of Financial Intermediation*, 19(3):333–354.

Lepetit, L. and Strobel, F. (2015). Bank insolvency risk and z-score measures: A refinement. *Finance Research Letters*, 13:214–224.

Lewellen, J. (2011). Institutional investors and the limits of arbitrage. *Journal of Financial Economics*, 102(1):62–80.

Liu, W. S. (2018). Regulatory capital and portfolio investments: Evidence from life insurance companies. *Available at SSRN*, 3274791.

Livdan, D., Sapriza, H., and Zhang, L. (2009). Financially constrained stock returns. *Journal of Finance*, 64(4):1827–1862.

Lorson, J., Schmeiser, H., and Wagner, J. (2012). Evaluation of benefits and costs of insurance regulation–a conceptual model for Solvency II. *Journal of Insurance Regulation*, 31(1):125–156.

Lucas, R. E. (1978). Asset prices in an exchange economy. *Econometrica*, 46(6):1429–1445.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

MacMinn, R. D. (1987). Insurance and corporate risk management. *Journal of Risk and Insurance*, 54(4):658–677.

MacMinn, R. D. (2005). *The Fisher Model and Financial Markets*. World Scientific.

Magill, M. J. (1984). On the Arrow-Lind theorem. *Journal of Mathematical Analysis and Applications*, 102(1):211–219.

Main, B. G. (1983). Corporate insurance purchases and taxes. *Journal of Risk and Insurance*, 50(2):197–223.

Marcus, A. J. (1984). Deregulation and bank financial policy. *Journal of Banking & Finance*, 8(4):557–565.

Mas-Colell, A., Whinston, M. D., Green, J. R., et al. (1995). *Microeconomic theory*, volume 1. Oxford university press New York.

Mayers, D. and Smith, C. W. (1990). On the corporate demand for insurance: evidence from the reinsurance market. *Journal of Business*, pages 19–40.

Mayers, D. and Smith, C. W. J. (1982). On the corporate demand for insurance. *Journal of Business*, 55(2):281–296.

Mayers, D. and Smith Jr, C. W. (1987). Corporate insurance and the underinvestment problem. *Journal of Risk and Insurance*, 54(1):45–54.

Mayers, D. and Smith Jr, C. W. (2005). Agency problems and the corporate charter. *Journal of Law, Economics, and Organization*, 21(2):417–440.

McKean, R. N. and Moore, J. H. (1972). Uncertainty and the evaluation of public investment decisions: comment. *American Economic Review*, 62(1/2):165–167.

Mechler, R. and Hochrainer-Stigler, S. (2014). Revisiting Arrow-Lind: managing sovereign disaster risk. *Journal of Natural Resources Policy Research*, 6(1):93–100.

Merton, R. C., Perold, A., et al. (1993). Theory of risk capital in financial firms. *Journal of Applied Corporate Finance*, 6(3):16–32.

Modigliani, F. and Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review*, 48(3):261–297.

Möhr, C. (2011). Market-consistent valuation of insurance liabilities by cost of capital. *ASTIN Bulletin: The Journal of the IAA*, 41(2):315–341.

Morgan, D. P. (2002). Rating banks: Risk and uncertainty in an opaque industry. *American Economic Review*, 92(4):874–888.

Morrison, C. J. and Siegel, D. (1997). External capital factors and increasing returns in US manufacturing. *Review of Economics and Statistics*, 79(4):647–654.

Mossin, J. (1968). Aspects of rational insurance purchasing. *Journal of Political Economy*, 76(1):553–568.

Moszoro, M. (2014). Efficient public-private capital structures. *Annals of Public and Cooperative Economics*, 85(1):103–126.

Moyen, N. (2004). Investment–cash flow sensitivities: Constrained versus unconstrained firms. *Journal of Finance*, 59(5):2061–2092.

Müller, K. (2022). Busy bankruptcy courts and the cost of credit. *Journal of Financial Economics*, 143(2):824–845.

Myers, S. C. (1977). Determinants of corporate borrowing. *Journal of Financial Economics*, 5(2):147–175.

Myers, S. C. and Cohn, R. A. (1987). A discounted cash flow approach to property-liability insurance rate regulation. In *Fair Rate of Return in Property-Liability Insurance*, pages 55–78. Springer.

Myers, S. C. and Majluf, N. S. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of financial economics*, 13(2):187–221.

Nance, D. R., Smith Jr, C. W., and Smithson, C. W. (1993). On the determinants of corporate hedging. *Journal of Finance*, 48(1):267–284.

National Association of Insurance Commissioners (1998). SSAP No. 65 - Property and Casualty Contracts. *NAIC Annual Statement Instructions, Statutory Issue Paper No. 65.*

National Association of Insurance Commissioners (2015). Annual Financial Reporting Model Regulation. *Official Journal of the European Communities.*

Nelson, K. K. (2000). Rate regulation, competition, and loss reserve discounting by property-casualty insurers. *The Accounting Review*, 75(1):115–138.

Ng, H. M. and Varnell, E. (2003). Frictional costs. *Staple Inn Actuarial Society*.

Niehaus, G. (2019). Personal taxes, cost of insurer equity capital, and the case of offshore hedge fund reinsurers. *Journal of Risk and Insurance*.

Nini, G. (2020). Insurance covenants in corporate credit agreements. *Journal of Risk and Insurance*, 87(1):95–115.

Nissim, D. (2010). Analysis and valuation of insurance companies. *CE/ ASA (Center for Excellence in Accounting and Security Analysis) Industry Study 2*.

Nissim, D. (2013a). Implied cost of equity capital in the US insurance industry. *Journal of Financial Perspectives*, 1(1):87–103.

Nissim, D. (2013b). Relative valuation of US insurance companies. *Review of Accounting Studies*, 18(2):324–359.

Owen, G. (1968). Communications to the editor—a note on the Shapley value. *Management Science*, 14(11):731–731.

Owen, G. (1972). Multilinear extensions of games. *Management Science*, 18(5-part-2):64–79.

Ozdagli, A. K. and Wang, Z. K. (2019). Interest rates and insurance company investment behavior. *Available at SSRN 3479663*.

Pelkiewicz, A., Ahmed, S., Fulcher, P., Johnson, K., Reynolds, S., Schneider, R., and Scott, A. (2020). A review of the risk margin–Solvency II and beyond. *British Actuarial Journal*, 25.

Ping, X. and Zanjani, G. (2015). Optimal insurance contracts with insurer's background risk. *Working Paper*.

PonArul, R. and Viswanath, P. (1995). Solvency risk and the tax sheltering behavior of property-liability insurers. *Journal of Risk & Insurance*, 62(3):575–584.

Pottier, S. W. and Sommer, D. W. (2006). Opaqueness in the insurance industry: Why are some insurers harder to evaluate than others? *Risk Management and Insurance Review*, 9(2):149–163.

Promislow, S. D. and Young, V. R. (2005). Unifying framework for optimal insurance. *Insurance: Mathematics and Economics*, 36(3):347–364.

Purnanandam, A. (2008). Financial distress and corporate risk management: Theory and evidence. *Journal of Financial Economics*, 87(3):706–739.

Rebello, M. J. (1995). Adverse selection costs and the firm's financing and insurance decisions. *Journal of Financial Intermediation*, 4(1):21–47.

Regan, L. and Hur, Y. (2007). On the corporate demand for insurance: The case of korean nonfinancial firms. *Journal of Risk and Insurance*, 74(4):829–850.

Ren, Y. and Schmit, J. T. (2009). Franchise value, competition and insurer risk taking. *Available at SSRN 1230422*.

Rochet, J.-C. and Villeneuve, S. (2011). Liquidity management and corporate demand for hedging and insurance. *Journal of Financial Intermediation*, 20(3):303–323.

Rockafellar, R. T. and Wets, R. J.-B. (2009). *Variational analysis*, volume 317. Springer Science & Business Media.

Schlesinger, H. (1981). The optimal level of deductibility in insurance contracts. *Journal of Risk and Insurance (pre-1986)*, 48(3):465.

Schlesinger, H. (1997). Insurance demand without the expected-utility paradigm. *Journal of Risk and Insurance*, 64(1):19–40.

Schlesinger, H. (2000). *The Theory of Insurance Demand, in: G. Dionne, ed., The Handbook of Insurance, 131–151.* Boston, Kluwer Academic Publishers.

Schlesinger, H. (2006). Mossin's theorem for upper-limit insurance policies. *Journal of Risk and Insurance*, 73(2):297–301.

Schnabel, J. A. and Roumi, E. (1989). Corporate insurance and the underinvestment problem: An extension. *Journal of Risk and Insurance*, 56(1):155–159.

Scotti, V. (2005). Insurer's cost of capital and economic value creation: Principles and practical implications. *Swiss Re Technical Publishing Sigma*, 3.

Sen, I. (2019). Regulatory limits to risk management. *The Review of Financial Studies*. Available at SSRN: $https://ssrn.com/abstract = 3345880$.

Sen, I. and Humphry, D. (2018). Capital regulation and product market outcomes. *Bank of England Working Paper*.

Shapley, L. S. (1953a). A Value for n-person Games. Contributions to the Theory of Games II, Kuhn, H., Tucker, A.

Shapley, L. S. (1953b). *Additive and non-additive set functions*. Princeton University.

Sharara, I., Hardy, M., and Saunders, D. (2010). Regulatory capital standards for property and casualty insurers under the us, canadian and proposed Solvency II (standard) formulas. *CAS, CIA, and SOA Joint Risk Management Section*.

Shavell, S. (1982). On liability and insurance. *Bell Journal of Economics*, 13(1):120–132.

Shavell, S. (1986). The judgment proof problem. *International review of law and economics*, 6(1):45–58.

Sherris, M. (2006). Solvency, capital allocation, and fair rate of return in insurance. *Journal of Risk and Insurance*, 73(1):71–96.

Sherris, M. and 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.

Shim, J. (2010). Capital-based regulation, portfolio risk and capital determination: Empirical evidence from the us property–liability insurers. *Journal of Banking & Finance*, 34(10):2450–2461.

Sias, R. W., Starks, L. T., and Titman, S. (2006). Changes in institutional ownership and stock returns: Assessment and methodology. *Journal of Business*, 79(6):2869–2910.

Sinn, H.-W. (1982). Kinked utility and the demand for human wealth and liability insurance. *European Economic Review*, 17(2):149–162.

Smith, A., Moran, I., and Walczak, D. (2003). Why can financial firms charge for diversifiable risk. In *2003 Bowles Symposium Fair Valuation of Contingent Claims and Benchmark Cost of Capital–Casualty Actuarial Society*.

Smith, C. W. and Stulz, R. M. (1985). The determinants of firms' hedging policies. *Journal of financial and quantitative analysis*, 20(4):391–405.

Smith, C. W. and Warner, J. B. (1979). Bankruptcy, secured debt, and optimal capital structure: Comment. *Journal of Finance*, 34(1):247–251.

Smith, V. L. (1968). Optimal insurance coverage. *Journal of Political Economy*, 76(1):68–77.

Spackman, M. (2002). Public–private partnerships: lessons from the British approach. *Economic Systems*, 26(3):283–301.

Staking, K. B. and Babbel, D. F. (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.

Stiglitz, J. E. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Economic Review*, 71(3):393–410.

Strebulaev, I. A. (2007). Do tests of capital structure theory mean what they say? *Journal of Finance*, 62(4):1747–1787.

Strebulaev, I. A. and Whited, T. M. (2012). Dynamic models and structural estimation in corporate finance. *Foundations and Trends® in Finance*, 6(1–2):1–163.

Stulz, R. (1990). Managerial discretion and optimal financing policies. *Journal of financial Economics*, 26(1):3–27.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics letters*, 20(2):177–181.

Thérond, P.-E. (2016). About market consistent valuation in insurance. In *Modelling in Life Insurance–A Management Perspective*, pages 43–60. Springer.

Titman, S. and Tsyplakov, S. (2007). A dynamic model of optimal capital structure. *Review of Finance*, 11(3):401–451.

Vedani, J., El Karoui, N., Loisel, S., and Prigent, J.-L. (2017). Market inconsistencies of market-consistent European life insurance economic valuations: pitfalls and practical solutions. *European Actuarial Journal*, 7(1):1–28.

Vincent, L., Albrecher, H., Krvavych, Y., et al. (2021). Structured reinsurance deals with reference to relative market performance. *Insurance: Mathematics and Economics*, 101(PB):125–139.

Warner, J. B. (1977). Bankruptcy costs: Some evidence. *Journal of Finance*, 32(2):337–347.

Whited, T. M. (1992). Debt, liquidity constraints, and corporate investment: Evidence from panel data. *Journal of Finance*, 47(4):1425–1460.

Wüthrich, M. V. (2011). An academic view on the illiquidity premium and market-consistent valuation in insurance. *European Actuarial Journal*, 1(1):93–105.

Yow, S. and Sherris, M. (2008). Enterprise risk management, insurer value maximisation, and market frictions. *ASTIN Bulletin: The Journal of the IAA*, 38(1):293–339.

Zanjani, G. (2002). Pricing and capital allocation in catastrophe insurance. *Journal of Financial Economics*, 65(2):283 – 305.

Zhang, L. and Nielson, N. (2009). The pricing of multiple line P&C insurance based on the full information underwriting beta. *Assurances et Gestion des Risques*, 77(3-4):237–64.

Zhang, L. and Nielson, N. (2012). Pricing for multiline insurer: Frictional costs, insolvency, and asset allocation. *Risk Management and Insurance Review*, 15(2):129–152.

Zhang, T., Cox, L. A., and Van Ness, R. A. (2009). Adverse selection and the opaqueness of insurers. *Journal of Risk and Insurance*, 76(2):295–321.

Zhang, Y. (2008). Economic impact of capital level in an insurance company. *Variance*, 2(1):39–51.

# Part IV

# Curriculum Vitae

# Curriculum Vitae

## Personal Details

| | |
|---|---|
| Name: | Andrea Bergesio |
| Date of birth: | 02 January 1991 |
| Nationality: | Italian |

## Education

| | |
|---|---|
| 09/2016 – 02/2023 | **PhD candidate in Banking and Finance**<br>University of Zurich, Switzerland<br>Advisors: Prof. Dr. Cosimo-Andrea Munari<br>Prof. Dr. Pablo Koch-Medina |
| 09/2013 – 03/2016 | **Master of Science in Finance and Risk Management**<br>Università degli Studi di Firenze, Italy<br>Goethe-Universität Frankfurt, Germany (Erasums+) |
| 09/2010 – 10/2013 | **Bachelor of Science in Economics**<br>Università degli Studi di Firenze, Italy |

## Professional Experience

| | |
|---|---|
| 09/2017 – 12/2022 | Teaching and research assistant<br>Department of Banking and Finance,<br>University of Zurich, Switzerland |
| 09/2015 – 06/2016 | Consultant<br>Financial Risk Management - Credit Risk,<br>KPMG Advisroy S.p.A., Italy |