A Continuous-Time Optimal Insurance Design with Costly Monitoring

Hisashi Nakamura

Koichiro Takaoka

Working Paper Series No. 140

July 2012

A Continuous-Time Optimal Insurance Design with Costly Monitoring

Hisashi Nakamura and Koichiro Takaoka*
Hitotsubashi University

July 12, 2012

Abstract

We provide a theoretical and numerical framework to study optimal insurance properties for players' general utility forms. We consider a continuous-time model where neither the efforts nor the outcome of an insured firm are observable to an insurer. The insured may then cause two interconnected information problems: moral hazard and exaggerated claims. We show that, when costly monitoring is available, an optimal insurance contract distinguishes between the two information problems. Furthermore, if the insured's downward-risk aversion is weak and if the participation constraint is not too tight, then a higher level of the monitoring technology can mitigate both problems.

Keywords: Insurance, Costly monitoring, Moral hazard, Exaggerated claims.
JEL Classification: D82, D86, G22, G32.

---

*Address: Graduate School of Commerce and Management, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan. Corresponding author: Hisashi Nakamura. Phone: +81-42-580-8826. Fax: +81-42-580-8747. Email: hisashi.nakamura@r.hit-u.ac.jp. We would like to thank Takau Yoneyama for his constant encouragement and advice throughout the preparation of this work. We are also thankful to Ricard Gil, Chiaki Hara, Jean Kwon, Soichiro Moridaira, Mahito Okura, and all seminar participants at 2011 Nakanoshima Workshop (Osaka University), 12th SAET Conference on Current Trends in Economics (University of Queensland, Brisbane), Working Group on Globalization of Insurance Research (sponsored by Dai-ichi Life Insurance Company, Limited and Dai-ichi Life Research Institute Inc.), and Yonsei University. All errors are our own.

1 Introduction

As it is well known, insurers are often exposed to information problems in corporate insurance practices; see e.g. MacMinn and Garven [9]. Specifically, neither the efforts nor the outcome of insured firms are observable to insurers directly without a cost. The insureds may then cause purposely losses and/or exaggerate claims. Call the former the problem of moral hazard and the latter the problem of ex-post informational asymmetry (typically, insurance frauds).

The ex-post informational asymmetry problem distorts the insured's effort incentives in the moral hazard problem. In standard moral hazard models, it is often assumed that the insurer can observe the insured's outcome ex post; e.g., see Rogerson [14]. The insurer then faces only the incentive problem of inducing the insured to make desired efforts. By contrast, when there exists the ex-post informational asymmetry problem as well, the insurer needs to provide additionally the insured with an incentive to tell the truth ex post. When a claim is filed, the insurer does not find directly whether it is due to the insured's laziness, to the exaggeration, or neither. Much worse, when the insured has a chance to tell a lie about the ex-post outcome, he may have an incentive to be lazy in order to minimize his effort cost. Since the two information problems are interconnected in this way, it is very difficult to distinguish one from another directly. In practice, to overcome this difficulty, the insurer routinely investigates claims via a costly monitoring technology after they are filed (Harrington and Niehaus [6]).

The purpose of this paper is to provide a theoretical and numerical framework to study optimal insurance properties when the costly monitoring is available under the problems of moral hazard and ex-post informational asymmetry in an optimal contracting model with two players: an insurer (i.e., insurance company) and an insured (i.e., firm). We consider an environment in which neither the efforts nor the outcome of an insured firm are observable to an insurer, but the insurer can monitor the outcome by using a costly monitoring technology. The insurer writes an optimal contract to maximize her own expected utility, inducing the insured to make his optimal efforts and to report the truth while, at the same time, trying to reduce the expected monitoring cost. We solve the insurer's optimization problem with respect to the design of the optimal contract and the monitoring decision, subject to the insured's optimization with respect to the efforts and the reports. We then make clear the optimal insurance properties, especially dynamic equilibrium interaction between the two information problems.

In previous literatures, the paper of Cvitanić and Zhang [3] is related closely to our paper. Similarly to theirs, our paper looks at the optimal contracting problem in a continuous-time principal-agent model in which hidden actions and hidden information coexist. The continuous-time model is useful for studying dynamic, complex information problems due to its mathematical tractability. Specifically, fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ on a time interval $[0, T]$ for a finite time $T > 0$. Let $W = \{W_t\}_{0 \leq t \leq T}$ be a one-dimensional standard $\mathbb{F}$-Brownian motion on the probability space. Define $\mathbb{F}^W := \{\mathcal{F}_t^W\}_{0 \leq t \leq T}$ as the filtration generated by W up to T. Note that $\mathbb{F} \supseteq \mathbb{F}^W$. The insured produces the wealth X^u that is characterized by the stochastic differential equation: $\mathrm{d}X_t^u = v(u_t \, \mathrm{d}t + \mathrm{d}W_t)$, $X_0^u = 0$ where v is a positive constant. The insured controls u – call it effort – with a utility cost, i.e., a higher (lower) level of the costly effort leads to a higher (lower) level of the expected return of the wealth. The efforts u and some part of the wealth information X^u are hidden from the insurer.

Still, our paper departs from Cvitanić and Zhang [3] mainly in three respects. First, with regard to the problem of the hidden wealth information, our paper looks at ex-post informational asymmetry in an environment where the insurer cannot observe ex post the time path of the insured's outcome $X_\cdot^u$ except for its initial value X_0^u, whereas their paper looks at ex-ante adverse selection in an environment where the principal cannot observe the initial value X_0^u (i.e., the agent's ex-ante production ability) but can observe the ex-post trajectory $X_\cdot^u$.

Second, our model assumes that a costly monitoring technology is available. Cvitanić and Zhang [3] assume only costless reporting as a communication method, as usual in contract theory.[1] Their paper then shows that, in general, it is very hard to distinguish between the moral hazard problem and the hidden information problem not only statically but also dynamically, because, with only costless reporting, it is generally difficult to compute the Lagrangian multipliers associated with the two information problems.[2] On the other hand, in insurance practices, as it was mentioned

[1]In much of the literature on contract theory, communication games with costless reporting have been studied a lot in finite-horizon (typically, two or three period) discrete-time models. There are a few exceptions in a literature on insurance frauds; e.g., see Dionne, Giuliano and Picard [4].

[2]In each of the information problems, more comprehensive coverage is associated with high risk (Chiappori [2]). Accordingly, it is difficult to identify whether it is due to their ability, to their laziness, or neither. Much empirical insurance literature has differentiated moral hazard from ex-ante adverse selection by making use of some different dynamic properties of the incentive structures between the two information problems in insurance contracts for various *exogenous* cases (typically, a reform of regulatory framework); e.g., see Abbring et al. [1]. However, that has not used any dynamic optimal (i.e., *endogenous*) insurance properties.

above, insurance companies routinely verify the reports of insured firms via costly monitoring. In other words, the monitoring is another crucial communication method in insurance contracts. In contrast to the previous literature, in this paper, we examine the crucial role of monitoring to distinguish between the two information problems.

Third, technically, we solve the optimal contracting problem by using the standard method of stochastic optimal controls adapted to $\mathbb{F}$. On the other hand, by contrast, Cvitanić and Zhang [3] solve it by using the method of backward stochastic differential equations defined on the path space of X^u, in which all controls are adapted to $\mathbb{F}^{X^u}$. The information set generated by observing only X^u loses the information of the efforts u as compared to the information set generated by observing the true uncertainty W (and thus the efforts u as well). $\mathbb{F}^{X^u}$ is generally smaller than $\mathbb{F}^W$. The optimal control adapted to $\mathbb{F}^W$ is not necessarily in the set of the controls adapted to $\mathbb{F}^{X^u}$. Accordingly, our formulation defined on $\mathbb{F}^W$ is very natural and thus applicable to financial practices. In addition, we show that $\mathbb{F}^{X^u} = \mathbb{F}^W$ holds in equilibrium.[3]

Our main results are as follows. If optimal efforts are attained, the insurer can write the optimal insurance contract that distinguishes between the two information problems, by using the costly monitoring technology effectively. Specifically, when the monitoring technology is available, the ex-post informational asymmetry problem is reduced, although the insured can still enjoy an information advantage while in good shape. Meanwhile, if the insured's downward-risk aversion is weak and if the participation constraint of the insured is not too tight with respect to the monitoring cost, then a higher level of the monitoring technology (i.e., a smaller monitoring cost) can mitigate the problem of moral hazard.

Based on the theoretical results, our model is tractable for numerical work. As a numerical example, we consider the case that the insured has a log utility and the insurer is risk-neutral. When the monitoring cost is an immediate level, the monitoring action is undertaken only for low wealth levels while the contract is deductible for very low wealth levels. Furthermore, the insurance premium is increasing in the monitoring cost. For very small monitoring costs, the monitoring action is undertaken for all nondeductible wealth levels (i.e., unless the contact is deductible for the wealth levels). The allocations are state-dependent. On the other hand, for very large monitoring costs, the monitoring action is necessarily avoided in equilibrium. The optimal contract is then of

[3]Note that our paper is not a generalization of Cvitanić and Zhang [3]. They assume that the contract can depend on the whole path of the wealth $X^u_\cdot$ whereas we assume that it depends only on its time-T value. Still, they show that the optimal contract depends only on the time-T value.

a state-independent debt type for all nondeductible wealth levels.

This paper is organized as follows. Next section defines an environment. Section 3 studies optimal insurance properties. Section 4 obtains numerical results. Final section concludes.

2 Environment

We consider an optimal contracting problem between two players: an insurer (i.e., insurance company) and an insured (i.e., firm) on a time interval $[0, T]$ for a finite time $T > 0$. For convenience, we will use female pronouns for the insurer, and male ones for the insured. Fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. Let $W = \{W_t\}_{0 \le t \le T}$ be a one-dimensional standard $\mathbb{F}$-Brownian motion on the probability space, i.e., for any t, s satisfying $0 \le t \le s$, $W_s - W_t$ is independent of $\mathcal{F}_t$. Define $\mathbb{F}^W := \{\mathcal{F}_t^W\}_{0 \le t \le T}$ as the filtration generated by W up to time T. Note that $\mathbb{F} \supseteq \mathbb{F}^W$.

The insured produces the wealth process X^u that is characterized by the following stochastic differential equation:[4]

$$\mathrm{d}X_t^u = v(u_t \, \mathrm{d}t + \mathrm{d}W_t), \quad X_0^u = 0 \tag{2.1}$$

where v is a positive constant and stands for riskiness of the wealth process. The insured can control the real-valued instantaneous expected wealth return process u – call it the insured's effort – with a utility cost $G_T(u) := \int_0^T g(u_t) \, \mathrm{d}t := \frac{1}{2} \int_0^T (u_t)^2 \, \mathrm{d}t$, i.e., a higher (lower) costly effort leads to a higher (lower) expected return of the wealth. Assume that the insured's effort process u is $\mathbb{F}$-adapted and bounded. For simplicity, assume $v = 1$ in the remainder of the theoretical sections. Obviously, any deterministic diffusion coefficients, including non-unity constant ones, would not change our theoretical results. Neither would do any non-zero initial value $X_0^u = x \neq 0$.

Let $U_i : \mathbf{R} \to \mathbf{R}$ denote player i's utility function of his or her own wealth, defined on $\mathbf{R}$, at time T $(i = 1, 2)$ where the index $i = 1$ denotes the insured and $i = 2$ denotes the insurer. The utility function U_i $(i = 1, 2)$ is three times continuously differentiable. In particular, the utility functions possess standard properties: $U_i' > 0$ for each $i = 1, 2$, and $U_1'' < 0$ and $U_2'' \le 0$. As usual in finance, both players are downward-risk averse: $U_1''' > 0$ and $U_2''' \ge 0$. Note that the insurer may

[4]Note that, in this model, the loss/gain process of the wealth is continuous. We may extend the process to have jumps under appropriate mathematical regularities. That will be our future work.

be risk-neutral.[5] In addition, we assume that the insured is risk-averse such that $-\frac{U_1''}{U_1'} \geq U_1'$.[6] This assumption is imposed for obtaining a necessary and sufficient condition for optimality below. The insured is exogenously given a reservation utility, denoted by a constant $r \in \mathbf{R}$, at time 0.

The information is asymmetric between the two players. We assume that $\{X^u, u\}$, except for the initial wealth $X_0^u = 0$, are the private information of the insured and are unobservable to the insurer. The players communicate the information of X^u with each other in the following two ways. First, the insured reports the trajectory of X^u at time T without a cost. The report may be a lie, i.e., the report is not necessarily equal to the true trajectory of X^u. Let the report be denoted by $\tilde{X}^u$. The reports $\tilde{X}^u$ will be mathematically specified shortly below. Second, a monitoring technology is available to the insurer at time T if she incurs a utility cost K_M. The technology is deterministic in the sense that, when demanded, it occurs with probability one, and delivers the true information of the time path of the wealth to the insurer with perfect accuracy.

The insured enters into a contract with the insurer and shares the time-T outcome X_T^u with the insurer according to terms of the contract for insuring against his wealth risk (typically, risk of property and liability). Specifically, the insurer offers a menu of contract payoffs C_T to the insured, and the insured decides whether or not to accept it. We assume that the insured's wealth allocation C_T takes the form of $C_T : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ as a functional of $(X_T^u, \tilde{X}_T^u)$, i.e., $C_T = C_T(X_T^u, \tilde{X}_T^u)$. Call $C_T(X_T^u, \tilde{X}_T^u)$ a contract.[7]

2.1 Sets of efforts and reports

Define the sets of the insured's controls. First, with regard to the efforts u,

Definition 1 $\mathcal{A}_1$ *is the set of the efforts u that are $\mathbb{F}$-adapted and bounded.*

For $u \in \mathcal{A}_1$, we define $\mathbb{P}^u$ as

$$\frac{\mathrm{d}\mathbb{P}^u}{\mathrm{d}\mathbb{P}} := \mathcal{E}\left(-\int u \,\mathrm{d}W\right)_T \tag{2.2}$$

[5]Most continuous-time optimal contracting models assume exponential utility forms; e.g., see Holmström and Milgrom [7], Schättler and Sung [15]. However, those forms are of limited use in financial practices. By contrast, we assume general utility forms.

[6]This is equivalent to the concavity of $e^{U_1(\cdot)}$ and is stronger than the concavity of U_1.

[7]The assumption of the dependence on the time-T values might look restrictive. As it is shown in Cvitanić and Zhang [3], it is not restrictive in equilibrium in Markovian settings.

where, for any $\mathbb{F}$-adapted real-valued processes θ,

$$\mathcal{E}\left(\int \theta \, \mathrm{d}W\right)_t := \exp\left(\int_0^t \theta_s \, \mathrm{d}W_s - \frac{1}{2}\int_0^t (\theta_s)^2 \, \mathrm{d}s\right).$$

Since u is bounded in $\mathcal{A}_1$, the Novikov condition is satisfied for u. Thus $\mathbb{P}^u$ is a probability measure. Accordingly, X^u is a driftless Brownian motion under the probability measure $\mathbb{P}^u$. Following notational conventions, $\mathbb{E}$ and $\mathbb{E}^u$ denote the expectation operators under $\mathbb{P}$ and $\mathbb{P}^u$, respectively.

Second, with regard to the reports $\tilde{X}^u$, assume that there exist $\mathbb{F}$-adapted real-valued processes $\tilde{u} \in \mathcal{A}_1$ such that, for each $\omega \in \Omega$,

$$\mathrm{d}\tilde{X}_t^u \quad = \quad (u_t - \tilde{u}_t) \, \mathrm{d}t + \mathrm{d}W_t, \quad \tilde{X}_0^u = X_0^u = 0.$$

Since the set $\mathcal{A}_1$ is a vector space, $u - \tilde{u}$ is in $\mathcal{A}_1$. For $u, \tilde{u} \in \mathcal{A}_1$, $\tilde{X}^u$ is a driftless Brownian motion under the probability $\mathbb{P}^{u-\tilde{u}}$ defined by:

$$\frac{\mathrm{d}\mathbb{P}^{u-\tilde{u}}}{\mathrm{d}\mathbb{P}} := \mathcal{E}\left(-\int (u - \tilde{u}) \, \mathrm{d}W\right)_T.$$

The term $\tilde{u}$ stands for the reported twist of the instantaneous expected return as compared to the truth, i.e., the insured could make the report that is lower than the truth X^u by $\int_0^T \tilde{u}_t \, \mathrm{d}t$. In addition, since u has been assumed to be unobservable to the insurer, she cannot distinguish between u and $\tilde{u}$ directly. This is why there exists dynamic interaction between the moral hazard problem and the ex-post informational asymmetry problem. Still, the insurer may distinguish between u and $\tilde{u}$ by designing the contract properly.

2.2 Set of contracts

Define f as a mapping (functional) from $\mathbf{R}$ to $\mathbf{R}$: for $u \in \mathcal{A}_1$,

$$f(X_T^u) := U_1\left(C_T(X_T^u, X_T^u)\right).$$

In other words, $f(X_T^u)$ denotes the insured's utility of the time-T wealth in case that the insured chooses $u \in \mathcal{A}_1$ and reports the truth $\tilde{X}_T^u = X_T^u$. Using this definition $f(X_T^u)$, define also F as a mapping (functional) from $\mathcal{A}_1$ to $\mathbf{R}$: for $u \in \mathcal{A}_1$, noting $X_T^u = W_T + \int_0^T u_t \, \mathrm{d}t$,

$$J(u) := \mathbb{E}\left[f\left(W_T + \int_0^T u_t \, \mathrm{d}t\right) - \int_0^T (u_t)^2 \, \mathrm{d}t\right].$$

This represents the insured's expected utility in case that the insured chooses $u \in \mathcal{A}_1$ and reports the truth $\tilde{X}_T^u = X_T^u$. Given the utility function U_1, the properties of f are linked to the ones of C_T. Note that C_T (and thus f) may not be differentiable everywhere.[8] Now, define mathematical regularities for the contracts C_T:

Definition 2 *Define the set $\mathcal{A}_2$ of the contracts $C_T(X_T^u, \tilde{X}_T^u)$ such that, for any $u, \tilde{u} \in \mathcal{A}_1$,*

(i) *$f(x)$ is continuous. In addition, $f(x)$ is differentiable except at a finite number of the points, say $\{x_1, x_2, \cdots, x_n\}$, and the derivative $f'(x)$, defined on $\mathbf{R} \backslash \{x_1, x_2, \cdots, x_n\}$, is bounded,*

(ii) *$\mathbb{E}\left[|U_2(X_T^u - C_T)|\right] < \infty$.*

Since $\sup_{x \in \mathbf{R} \backslash \{x_1, x_2, \cdots, x_n\}} |f'(x)| < \infty$, the linear growth condition is satisfied for f. Hence, for any $\alpha > 0$, $\mathbb{E}\left[e^{\alpha f(Z)}\right] < \infty$ where Z is a normal random variable with arbitrary parameters.

Next, as usual in financial contract theory, we restrict the contract space $\mathcal{A}_2$ to a further particular set, i.e., we impose the conditions of incentive compatibility and the insured's participation on the contract space. With regard to the incentive compatibility condition, we restrict the contracts to the ones that induce the insured to tell the truth (i.e., $\tilde{u} = 0$). Following standard discussions of costly monitoring, assume that $C_T(X_T^u, \tilde{X}_T^u)$ takes the form:

$$C_T(X_T^u, \tilde{X}_T^u) = \begin{cases} X_T^u - F = (X_T^u - \tilde{X}_T^u) + (\tilde{X}_T^u - F) & \text{if } \tilde{X}_T^u \geq b, \\ C_M(X_T^u) & \text{if } \tilde{X}_T^u < b \text{ and if } X_T^u = \tilde{X}_T^u, \\ 0 & \text{if } \tilde{X}_T^u < b \text{ and if } X_T^u \neq \tilde{X}_T^u \end{cases} \tag{2.3}$$

where $F \in \mathbf{R}$ and $b \in \mathbf{R}$ are constants and $C_M(X_T^u)$ is a function $\mathbf{R} \to \mathbf{R}$. We call this contract an insurance contract characterized by a triplet $\{F, b, C_M(X_T^u)\}$ in the following sense (as shown in Figure 1). First, F stands for the insurer's deterministic allocation when the monitoring is not undertaken and is interpreted as an insurance premium.[9]

[8]E.g., payoff functions are kinked in debt contracts and options.

[9]For example, consider the case that the insured could not commit to the contract at time T. In the case, the insurer could enforce the insured to pay F in advance (i.e., at time 0) in order to avoid failure to collect F at time T.

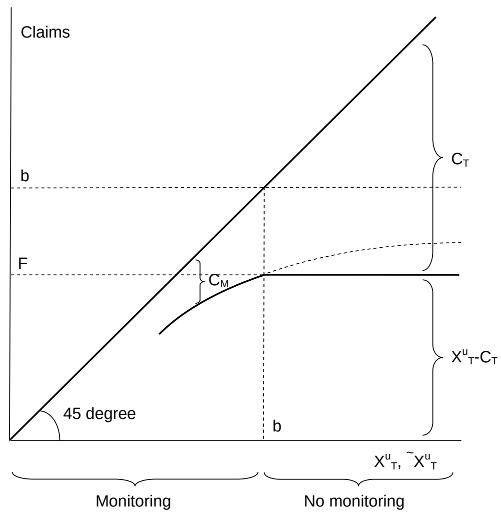

Figure 1: Insurance contract

Second, we assume that there exists a constant b such that, when $\tilde{X}_T^u$ is lower than b, a monitoring action is triggered. In other words, b stands for a threshold to trigger the monitoring. Note that we will verify the existence of such b below. The set $\{\tilde{X}_T^u < b\}$ is called the monitoring region on $\mathbf{R}$; its complement (i.e., $\{\tilde{X}_T^u \geq b\}$) the no-monitoring region. If the monitoring is undertaken and verifies the truth, the insurer provides the insured with the allocation $C_M(X_T^u)$ to insure against the low outcome. On the other hand, if a lie is verified, everything is confiscated from the insured and, in addition, the insured is penalized by the utility cost of $-\infty$.[10] Thus the verification of the false reporting would be out of equilibrium. Let $b = -\infty$ denote the case that only the no-monitoring region exists, and $b = +\infty$ denote the case that only the monitoring region exists. b is said to be feasible if $-\infty < b < +\infty$. Thus b is the threshold of insurance compensation in equilibrium.

Third, the function $C_M(X_T^u) : \mathbf{R} \to \mathbf{R}$ is the insured's time-T allocation when the monitoring is undertaken and verifies the truth. $C_M(X_T^u) - (X_T^u - F)$ is the insurance compensation. The compensation works as a put option. We impose two assumptions on $C_M(X_T^u)$. Firstly, $C_M(X_T^u)$ is assumed to be continuous. The continuity could be justified if the contract needs to be renegotiation-proof. Secondly, $C_M(X_T^u)$ is assumed to be non-decreasing. This assumption means a co-payment relationship, which looks relevant in practice.

Let us look at the incentive compatibility in the insurance contracts more specifically. When $X_T^u < b$, the insured could have an incentive to tell a lie if he is better off behaving as if he would be in the no-monitoring state. That would not be incentive compatible. Thus, to induce the insured

[10]The penalty can be interpreted as reputation loss and the cost of being imprisoned, for example.

to tell the truth, $C_M(X_T^u) \geq (X_T^u - F)$. On the other hand, when $X_T^u \geq b$, the monitoring action should not be undertaken. Therefore, C_M should be less than $X_T^u - F$. Also, as usual in contract theory, we assume that, when the insured is indifferent between two actions, he will choose the one that is better to the insurer. Thus the incentive compatibility condition is written as:

$$\Big(C_M(X_T^u) - (X_T^u - F)\Big)(X_T^u - b) \leq 0. \tag{2.4}$$

In previous standard costly monitoring models, due to the assumption of risk neutrality, minimizing the probability of undertaking a monitoring action is equivalent to maximizing the principal's expected wealth while providing the agent with no lower than his reservation utility; see e.g. Gale and Hellwig [5], Williamson [16]. Thus, to be incentive compatible, when being monitored, everything should be confiscated from the agent. Accordingly, that optimal contact form is called a simple debt. In contrast, in our present paper, since the utility functions are non-linear, the insured should receive positive allocation in the optimal risk-sharing arrangement when the monitoring action is made (i.e., $X_T^u < b$). Precisely, the insured receives the allocation even when X_T^u is larger than F, i.e., the monitoring action can be triggered optimally when the insured is liquid. This contract form is not the simple debt, but rather an insurance contract.

With regard to the participation condition, the contracts are constrained to the ones that induce the insured to participate. Define the insured's optimal utility as $V_1 := \sup_{u \in \mathcal{A}_1} J(u)$ for each incentive compatible $C_T \in \mathcal{A}_2$. We assume that the insurer provides the insured with no lower utility than his reservation utility (i.e., $V_1 \geq r$), so as to make the insured enter into the contract. In particular, as usual in hidden action problems, we assume that the participation condition is binding:

$$V_1 = r. \tag{2.5}$$

In short, using the two constraints (2.4),(2.5), define the set $\mathcal{A}_2'$ of the contracts C_T, which is characterized by $\{F, b, C_M\}$ in Eq.(2.3), as follows.

Definition 3 *Define the set $\mathcal{A}_2'$ of the contracts $C_T \in \mathcal{A}_2$ satisfying*

(i) *C_M is continuous and non-decreasing.*

(ii) *C_T satisfies Conditions* (2.4) *and* (2.5)*.*

3 Optimal insurance design with costly monitoring

3.1 Insured's optimization

Proposition 3.1 *For $C_T \in \mathcal{A}_2'$, the insured's optimization problem $\sup_{u \in \mathcal{A}_1} J(u)$ has a unique solution u^* such that*

$$J(u^*) = \log \mathbb{E}[e^{f(Z)}], \quad \textit{where } \; Z \sim N(0, T).$$

Define

$$X_t^* := W_t + \int_0^t u_s^* \, \mathrm{d}s.$$

We then obtain $\mathbb{F}^{X^*} = \mathbb{F}^W$ and $u_t^* = h(X_t^*, t)$ where the deterministic function $h : \mathbf{R} \times [0, T] \to \mathbf{R}$ is defined as

$$h(x, t) := \frac{\mathbb{E}[e^{f(Z)} f'(Z)]}{\mathbb{E}[e^{f(Z)}]}, \quad \textit{where } \; Z \sim N(x, T - t). \tag{3.1}$$

Furthermore, u^* satisfies

$$e^{f(X_T^*)} = \mathbb{E}^*[e^{f(X_T^*)}] \, \mathcal{E}\Big(\int u^* \, \mathrm{d}X^* \Big)_T \tag{3.2}$$

where $\mathbb{E}^*$ denotes the expectation operator under the probability measure $\mathbb{P}^*$ defined by the Radon-Nikodym derivative:

$$\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}} := \mathcal{E}\Big(- \int u^* \, \mathrm{d}W \Big)_T = \frac{1}{\mathcal{E}\Big(\int u^* \, \mathrm{d}X^* \Big)_T}.$$

Proof: See appendix.

Since u^* is bounded, the Novikov condition is satisfied. By the Girsanov theorem, X^* is a driftless Brownian motion under the probability measure $\mathbb{P}^*$. Also, h is a bounded function, because, defining $K := \sup_{x \in \mathbf{R} \setminus \{x_1, x_2, \cdots, x_n\}} |f'(x)| < \infty$ as in Definition 2,

$$|h(x, t)| \leq \frac{\mathbb{E}[e^{f(Z)} |f'(Z)|]}{\mathbb{E}[e^{f(Z)}]} \leq \frac{\mathbb{E}[e^{f(Z)} K]}{\mathbb{E}[e^{f(Z)}]} = K.$$

Substituting the participation condition (2.5) into Eq.(3.2),

$$e^{-r}e^{f(X_T^*)} = \mathcal{E}\Big(\int u^*\,\mathrm{d}X^*\Big)_T = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{P}^*}. \tag{3.3}$$

This means that the choice of the probability measure associated with the optimal action u^* has an explicit functional relationship with the insured's allocation C_T. Thus, the insurer can implement the insured's optimal effort u^* by choosing the contract $C_T \in \mathcal{A}_2''$ in a way consistent with Eq.(3.3) subject to Conditions (2.4) and (2.5). Call Eq.(3.3) the implementability condition.

Using Condition (3.3), define the set $\mathcal{A}_2''$ of the contracts C_T as:

Definition 4 *Define the set $\mathcal{A}_2''$ of the contracts $C_T \in \mathcal{A}_2'$ satisfying Condition* (3.3).

3.2 Insurer's optimization

Now, we formulate the insurer's optimization problem:

$$\sup_{C_T\in\mathcal{A}_2''} \mathbb{E}\left[U_2\Big(X_T^* - C_T(X_T^*, X_T^*)\Big) - K_M\mathbf{1}_M\right] \tag{3.4}$$

where $\mathbf{1}_M$ is an indicator function such that $\mathbf{1}_M = 1$ when a monitoring action is undertaken (otherwise, 0). Although the insurer cannot observe the truth X^u directly, she can verify it by designing the optimal contract that implements the optimal effort u^* and is incentive compatible (i.e., $\tilde{u} = 0$). Accordingly, for $C_T \in \mathcal{A}_2''$, the insurer can take her expectation under $\mathbb{P}$ in Eq.(3.4). Due to Definition 2 (ii), the integrability is ensured in Eq.(3.4). Using the implementability condition (3.3), the optimization problem (3.4) is rewritten as

$$\begin{aligned}
&\sup_{C_T\in\mathcal{A}_2''} \mathbb{E}\left[U_2\Big(X_T^* - C_T(X_T^*, X_T^*)\Big) - K_M\mathbf{1}_M\right] \\
=\ &\sup_{C_T\in\mathcal{A}_2''} \mathbb{E}^*\left[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{P}^*}\Big(U_2\Big(X_T^* - C_T(X_T^*, X_T^*)\Big) - K_M\mathbf{1}_M\Big)\right] \\
=\ &\sup_{C_T\in\mathcal{A}_2''} \mathbb{E}^*\left[e^{-r}e^{f(X_T^*)}\Big(U_2\Big(X_T^* - C_T(X_T^*, X_T^*)\Big) - K_M\mathbf{1}_M\Big)\right] \\
=\ &\sup_{C_T\in\mathcal{A}_2''} \mathbb{E}^*\left[e^{-r}e^{U_1(C_T(X_T^*, X_T^*))}\Big(U_2\Big(X_T^* - C_T(X_T^*, X_T^*)\Big) - K_M\mathbf{1}_M\Big)\right].
\end{aligned} \tag{3.5}$$

Note that, in Eq.(3.5), we change the measure so as to have the driftless Brownian motion X^* under $\mathbb{P}^*$. Define the Lagrangian multipliers associated with Conditions (2.4) and (2.5) as μ and λ, respectively. Since $V_1 \geq r$ and, in particular, $V_1 = r$, we obtain $\lambda > 0$. Using Conditions (2.4) and

(3.3), the constrained optimization problem (3.5) is rewritten into:

$$\sup_{\{F,b,C_M\}} \left\{ \begin{array}{l} e^{-r}\,\mathbb{E}^*\left[e^{U_1(C_T(X_T^*,X_T^*))}\Big(\Big(U_2(X_T^*-C_T(X_T^*,X_T^*))-K_M\mathbf{1}_M\Big)+\lambda\Big)\right] \\ +\mu\left[\Big(C_M(X_T^*)-(X_T^*-F)\Big)(b-X_T^*)\right] \end{array} \right\}. \tag{3.6}$$

With regard to C_M, a necessary condition for optimality is:

$$e^{U_1(C_M)}U_1'(C_M)\left\{(\lambda-K_M)-\left(\frac{U_2'(X_T^*-C_M)}{U_1'(C_M)}-U_2(X_T^*-C_M)\right)\right\}+e^r\mu\,(b-X_T^*)=0. \tag{3.7}$$

On the assumption that C_M is continuous and non-decreasing, by the incentive compatibility condition (2.4), if b is feasible,

$$C_M(b)=b-F. \tag{3.8}$$

Accordingly, when $X_T^*\geq b$, the incentive compatibility condition is necessarily slack. Therefore, we can focus attention on the case of $X_T^*<b$ in the second term on the left-hand side of Eq.(3.7).

For notational convenience, regarding Eq.(3.7), define

$$\begin{aligned} H(y) &:= e^{U_1(y)}U_1'(y)\left\{(\lambda-K_M)-\left(\frac{U_2'(X_T^*-y)}{U_1'(y)}-U_2(X_T^*-y)\right)\right\}+e^r\mu\,(b-X_T^*)\,, \\ L(y) &:= \frac{U_2'(X_T^*-y)}{U_1'(y)}-U_2(X_T^*-y). \end{aligned}$$

Then,

$$\begin{aligned} H'(y) &= e^{U_1(y)}U_1'(y)\left\{U_1'(y)\Big((\lambda-K_M)-L(y)\Big)\left(1+\frac{U_1''(y)}{(U_1'(y))^2}\right)-L'(y)\right\} \\ &= e^{U_1(y)}U_1'(y)\left\{-\frac{\mu\,(b-X_T^*)\,e^{r-U_1(y)}}{U_1'(y)}\left(U_1'(y)+\frac{U_1''(y)}{U_1'(y)}\right)-L'(y)\right\}. \end{aligned}$$

Noting $-\frac{U_1''}{U_1'}\geq U_1'$, $\mu\leq 0$ and $L'(y)=\frac{-U_2''(X_T^*-y)U_1'(y)-U_2'(X_T^*-y)U_1''(y)}{(U_1'(y))^2}+U_2'(X_T^*-y)>0$, we obtain $H'(y)<0$. Therefore, Eq.(3.7) is the necessary and sufficient condition for optimality.

For the reference, similarly to Cvitanić and Zhang [3], we look at the case that there exists moral hazard without ex-post informational asymmetry. It means that the incentive compatibility condition is slack, i.e., $\mu=0$. The insurer does not need to monitor. Thus C_M can be replaced by C_T. Hence,

$$\frac{U_2'(X_T^* - C_T)}{U_1'(C_T)} - U_2(X_T^* - C_T) = \lambda. \tag{3.9}$$

And,

$$0 < \frac{\mathrm{d}C_T}{\mathrm{d}X_T^*} = 1 - \frac{U_2' U_1''}{U_2'' U_1' + U_2' U_1'' - U_2'(U_1')^2} < 1. \tag{3.10}$$

Furthermore, we examine the case that there are no moral hazard and no ex-post informational asymmetry. The standard Borch rule is then obtained:

$$\frac{U_2'(X_T^* - C_T)}{U_1'(C_T)} = \lambda, \tag{3.11}$$

$$0 \leq \frac{\mathrm{d}C_T}{\mathrm{d}X_T^*} = 1 - \frac{U_2' U_1''}{U_2'' U_1' + U_2' U_1''} < 1. \tag{3.12}$$

Accordingly, from Eq.(3.9) and Eq.(3.11), we see that the term $U_2(X_T^* - C_T)$ stands for the effect of moral hazard, whereas the term $\mu\,(b - X_T^*)$ stands for the effect of ex-post informational asymmetry.

From Eq.(3.10) and Eq.(3.12), $\frac{\mathrm{d}C_T}{\mathrm{d}X_T^*}$ is less than one, either with or without moral hazard. In addition, it is higher in Eq.(3.10) than in Eq.(3.12). I.e., when X_T^* gets higher, the larger compensation is required in the moral hazard case due to the necessity to induce the insured to make the optimal efforts. The effect of moral hazard on C_T is represented by the difference between Eq.(3.10) and Eq.(3.12):

$$\begin{aligned}
\Delta\left(\frac{\mathrm{d}C_T}{\mathrm{d}X_T^*}\right) &:= \left(1 - \frac{U_2' U_1''}{U_2'' U_1' + U_2' U_1'' - U_2'(U_1')^2}\right) - \left(1 - \frac{U_2' U_1''}{U_2'' U_1' + U_2' U_1''}\right) \\
&= \frac{-U_1''(U_1' U_2')^2}{(U_2'' U_1' + U_2' U_1'')\,(U_2'' U_1' + U_2' U_1'' - U_2'(U_1')^2)} > 0.
\end{aligned} \tag{3.13}$$

3.3 Characterization

We characterize the optimal C_M from Eq.(3.7). Suppose that, in the monitoring region, the optimal effort is attained. Then,

$$\frac{U_2'(X_T^* - C_M(X_T^*))}{U_1'(C_M(X_T^*))} - U_2(X_T^* - C_M(X_T^*)) = \lambda - K_M. \tag{3.14}$$

Noting strict concavity of U_1,

$$0 < \frac{\mathrm{d}C_M}{\mathrm{d}X_T^*} = 1 - \frac{U_2' U_1''}{U_2'' U_1' + U_2' U_1'' - U_2' (U_1')^2} < 1. \tag{3.15}$$

By Eq.(3.8) and Eq.(3.15), $C_M \geq X_T^* - F$ in the monitoring region, i.e, the insured is better off telling the truth when the payment rule C_M as well as (F, b) are given. Also, by construction, there is no informational asymmetry in the no-monitoring region. Hence, $\mu = 0$. Accordingly, if optimal efforts are attained as in Eq.(3.14), the problem of moral hazard is differentiated from the problem of ex-post informational asymmetry in the monitoring region. Furthermore, from Eq.(3.15), we can confirm that the threshold b is well-defined in Eq.(2.3) satisfying the incentive compatibility condition (2.4).

Implications are as follows. Similarly to Cvitanić and Zhang [3], C_T is non-linear in X_T^* in contrast to Holmström and Milgrom [7] and Schättler and Sung [15]. Furthermore, this paper draws richer implications of insurance than Cvitanić and Zhang [3]'s. First, if the optimal efforts are attained in Eq.(3.14), the insurer can write explicitly the insurance contract that distinguishes between the two information problems, by using the costly monitoring effectively. Note that Cvitanić and Zhang [3], by contrast, face difficulty with writing the optimal contract that differentiates between the informational problems, even by using dynamic data, because, with only costless reporting, it is very hard to compute the Lagrangian multipliers associated with the informational problems.

Second, we look at dynamic interaction between the problem of moral hazard and the problem of ex-post informational asymmetry in the optimal insurance contract. We focus on the case that b is feasible. We compare this model with the case of moral hazard without ex-post informational asymmetry characterized by Eq.(3.9). The ex-post informational asymmetry problem is reduced, i.e., the insured can still enjoy an information advantage while in good shape. On the other hand, in the monitoring region, there is no μ, i.e., informational asymmetry is removed there. Eq.(3.14) looks equivalent to Eq.(3.9), except for the existence of K_M on the right-hand side. Let us examine the effect of the monitoring on moral hazard, which is measured by the effect on $\frac{\mathrm{d}C_T}{\mathrm{d}X_T^*}$ like $\Delta \left(\frac{\mathrm{d}C_T}{\mathrm{d}X_T^*} \right)$

in Eq.(3.13). From Eq.(3.15), when $U_2'' < 0$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}K_M}\left(\frac{\mathrm{d}C_M}{\mathrm{d}X_T^*}\right) &= \frac{\mathrm{d}}{\mathrm{d}C_M}\left(\frac{\mathrm{d}C_M}{\mathrm{d}X_T^*}\right)\cdot\frac{\mathrm{d}C_M}{\mathrm{d}(\lambda-K_M)}\cdot\frac{\mathrm{d}(\lambda-K_M)}{\mathrm{d}K_M} \\
&= \underbrace{\frac{U_2'U_1'U_2''U_1''}{\left(U_2''U_1'+U_2'U_1''-U_2'(U_1')^2\right)^2}}_{\geq 0}\left(\begin{array}{l}\left(\left(-\frac{U_2'''}{U_2''}\right)-\left(-\frac{U_2''}{U_2'}\right)\right)\\ +\left(\left(-\frac{U_1'''}{U_1''}\right)-\left(-\frac{U_1''}{U_1'}\right)\right)\\ +\frac{U_1'}{\left(-\frac{U_2''}{U_2'}\right)}\left(\left(-\frac{U_1'''}{U_1''}\right)-2\left(-\frac{U_1''}{U_1'}\right)\right)\end{array}\right) \\
&\quad\cdot\underbrace{\frac{1}{\frac{-U_2''U_1'-U_2'U_1''}{(U_1')^2}+U_2'}}_{>0}\cdot\frac{\mathrm{d}(\lambda-K_M)}{\mathrm{d}K_M}. \qquad (3.16)
\end{aligned}$$

To draw economic implications of insurance from Eq.(3.16), we add several definitions as follows. In economics, $-\frac{U_i'''(x)}{U_i''(x)} > 0$ is called absolute prudence, which measures downward-risk aversion, i.e., the strength of the precautionary saving motive. Because $U_i'''(x) > 0$ for $i \in \{1,2\}$ here, both players are downward-risk averse in this model. We put a further classification of downward-risk aversion as follows. For $i \in \{1,2\}$, player i's downward-risk aversion is said to be weak if $-\frac{U_i'''}{U_i''} \leq -\frac{U_i''}{U_i'}$, in that $\frac{\mathrm{d}\left(-\frac{U_i''(x)}{U_i'(x)}\right)}{\mathrm{d}x} = \frac{U_i''}{U_i'}\left\{\left(-\frac{U_i'''}{U_i''}\right)-\left(-\frac{U_i''}{U_i'}\right)\right\} \geq 0$. I.e., the player has the precautionary saving motive in the sense of absolute prudence, but his or her absolute risk aversion $-\frac{U_i''(x)}{U_i'(x)}$ is increasing in wealth. On the other hand, for $i \in \{1,2\}$, player i's downward-risk aversion is said to be strong if $-\frac{U_i'''}{U_i''} > -\frac{U_i''}{U_i'}$. Note that, when player i's downward-risk aversion is weak, $-\frac{U_1'''}{U_1''} \leq -\frac{U_1''}{U_1'} < -2\frac{U_1''}{U_1'}$. Also, we can guess that larger K_M leads to higher λ due to the tighter participation constraint. Still, we are not certain whether $\frac{\mathrm{d}\lambda}{\mathrm{d}K_M} \geq 1$ or $\frac{\mathrm{d}\lambda}{\mathrm{d}K_M} < 1$. The participation constraint is said to be not too tight (with respect to K_M) if $\frac{\mathrm{d}\lambda}{\mathrm{d}K_M} < 1$.

Using these definitions, we characterize Eq.(3.16). When $U_2'' < 0$, if the insured's downward-risk aversion as well as the insurer's are weak and if the participation constraint is not too tight, then a higher level of the monitoring technology (i.e., smaller K_M) mitigates the problem of moral hazard. Also, even when the insurer's downward-risk aversion is strong, if the effect of the insured's weak downward-risk aversion overpowers the effect of the insurer's strong downward-risk aversion in Eq.(3.16) and if the participation constraint is not too tight, then a higher level of the monitoring technology (i.e., smaller K_M) can mitigate the problem of moral hazard. A logic behind this result is as follows. When the insured's downward-risk aversion is weak, he does not demand excessively high compensation. Also, when the tight participation constraint is not too tight, larger (smaller) K_M

does not tighten (loosen) the right-hand side of Eq.(3.14) much. In this case, it is not quite costly to induce the insured to make the optimal efforts. Thus, for larger (smaller) K_M, the monitoring technology is a less (more) useful devise to verify the truth. Therefore, a lower (higher) level of the monitoring technology (i.e., larger (smaller) K_M) then leads to more (less) moral hazard.

Also, we look at the case of $U_2'' = 0$:

$$\frac{\mathrm{d}}{\mathrm{d}K_M}\left(\frac{\mathrm{d}C_M}{\mathrm{d}X_T^*}\right) = \underbrace{\frac{-\left(U_2'\right)^2\left(U_1'\right)^2 U_1''}{\left(U_2'U_1''-U_2'(U_1')^2\right)^2}}_{\geq 0}\left(\left(-\frac{U_1'''}{U_1''}\right)-2\left(-\frac{U_1''}{U_1'}\right)\right)\cdot \underbrace{\frac{1}{\frac{-U_2'U_1''}{\left(U_1'\right)^2}+U_2'}}_{>0}\cdot\frac{\mathrm{d}(\lambda-K_M)}{\mathrm{d}K_M}. \tag{3.17}$$

We can draw the similar implications as in the above case of $U_2'' < 0$.

4 Numerical analysis of optimal insurance design

4.1 Numerical method

So far we have imposed several high-level assumptions on the set of the controls. Also, the effect of K_M on the endogenous variable λ has not been obtained explicitly in a closed form. Accordingly, we are not certain whether, for some practically relevant values of the structural parameters, we can obtain plausible solutions to the above optimal insurance design problem and can provide useful predictions regarding the insured's strategic behavior under moral hazard and ex-post informational asymmetry in the optimal insurance contract. To complete our study, we do numerical analyses and draw quantitative implications in this section.

Assume that v is a positive constant. Based on the results in Section 3, if b is feasible, a derivation method for the optimal values of b, F, C_M, u^*, λ consists of the following four steps.

(1) When Eq.(3.14) holds, $C_M(X_T^*)$ can be written as a function of λ, denoted by $C_M^\lambda(X_T^*)$.

(2) By Eq.(3.8), when b is feasible,

$$C_M^\lambda(b) = b - F. \tag{4.1}$$

Thus F can be written as a function of b and λ, denoted by $F(b, \lambda)$.

(3) By Condition (3.3), λ can be written as a function of b, denoted by $\lambda(b)$, satisfying:

$$\begin{aligned} e^r &= \mathbb{E}^*[e^{U_1(C_T)}] \\ &= \int_b^{+\infty} e^{U_1(X_T^*-F(b,\lambda))}\Phi(\mathrm{d}X_T^*) + \int_{-\infty}^b e^{U_1(C_M^\lambda(X_T^*))}\Phi(\mathrm{d}X_T^*) \end{aligned} \tag{4.2}$$

where Φ denotes the cumulative distribution function of X_T^* under $\mathbb{P}^*$ at time 0. Set $\Phi(x) = N\left(\frac{x}{v\sqrt{T}}\right)$ where $N(\cdot)$ denotes the standard normal cumulative distribution function, i.e., for $x \in \mathbf{R}$, $N(x) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x \exp\left(-\frac{u^2}{2}\right)\mathrm{d}u$; see e.g. Musiela and Rutkowski [10].

(4) By Eq.(3.6), with respect to b, the insurer optimizes her expected utility:

$$\begin{aligned} & e^{-r}\,\mathbb{E}^*\left[e^{U_1(C_T)}\Big((U_2(X_T^*-C_T)-K_M\mathbf{1}_M)+\lambda(b)\Big)\right] \\ = \quad & e^{-r}\left[\begin{array}{l} \Big(U_2(F(b,\lambda(b)))+\lambda(b)\Big)\int_b^{+\infty} e^{U_1(X_T^*-F(b,\lambda(b)))}\Phi(\mathrm{d}X_T^*) \\ +\int_{-\infty}^b e^{U_1(C_M^{\lambda(b)}(X_T^*))}\Big(U_2(X_T^*-C_M^{\lambda(b)}(X_T^*))-K_M+\lambda(b)\Big)\Phi(\mathrm{d}X_T^*) \end{array}\right]. \end{aligned} \tag{4.3}$$

4.2 Numerical example

As an example, let us look at the case that $U_1(x) = \log(x)$ and $U_2(x) = x$. There is one crucial caveat: this case does not fit the above theoretical formulation perfectly, in that, due to the log utility form, the insured's consumption should be strictly positive. Still, we examine this case for the following two reasons. First, the numerical formulation is simple and comprehensive clearly from a financial viewpoint. Second, we implicitly assume that, if any non-positive allocation is given to the insured, the players' contractual relationship is deductible, i.e., the insured and the insurer repudiate without a cost and live in autarky, which provides the players with zero utility, from the time onwards. If the probability of being deducted (as we will show below, $\mathbb{P}^*\Big(X_T^* < -(\lambda - K_M)\Big)$) is small, the problem is not serious.

From $-\frac{U_1''(x)}{U_1'(x)} = \frac{1}{x} = U_1'(x)$ and Eq.(3.14),

$$C_M = \frac{X_T^* + (\lambda - K_M)}{2}.$$

For the positivity of C_M, $X_T^* > -(\lambda - K_M)$. By Eq.(4.1), if b is feasible,

$$F = \frac{b - (\lambda - K_M)}{2}.$$

Note that the insurer's allocation can take negative values due to her risk neutrality, while the insured's one is positive due to his log utility. By Eq.(4.2),

$$\begin{aligned}
e^r &= \int_b^{+\infty} \left(x - \frac{b-(\lambda-K_M)}{2} \right) \Phi(\mathrm{d}x) + \int_{-(\lambda-K_M)}^{b} \left(\frac{x+(\lambda-K_M)}{2} \right) \Phi(\mathrm{d}x) \\
&= \int_b^{+\infty} \left(x - \frac{b}{2} \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2v^2T}}\,\mathrm{d}x + \int_{-(\lambda-K_M)}^{b} \left(\frac{x}{2} \right) \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2v^2T}}\,\mathrm{d}x \\
&\quad + \frac{(\lambda-K_M)}{2} \left(1 - N\left(\frac{-(\lambda-K_M)}{v\sqrt{T}} \right) \right)
\end{aligned} \tag{4.4}$$

By Eq.(4.3), the insurer optimizes her utility with respect to b and λ,

$$\begin{aligned}
&(F+\lambda) \int_b^{+\infty} (x-F)\Phi(\mathrm{d}x) + \int_{-(\lambda-K_M)}^{b} C_M \Big((x-C_M) + (\lambda-K_M) \Big) \Phi(\mathrm{d}x) \\
= &\left(\frac{b+\lambda+K_M}{2} \right) \int_b^{+\infty} \left(x - \frac{b-(\lambda-K_M)}{2} \right) \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2v^2T}}\,\mathrm{d}x \\
&+ \int_{-(\lambda-K_M)}^{b} \left(\frac{x+(\lambda-K_M)}{2} \right)^2 \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2v^2T}}\,\mathrm{d}x
\end{aligned}$$

subject to Eq.(4.4). Figure 2 illustrates the optimal insurance contract in this example.

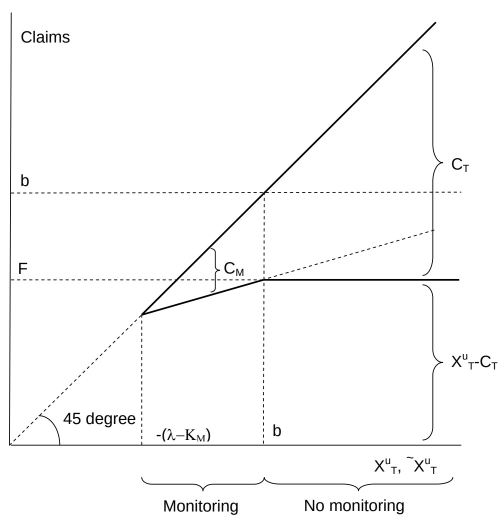

Figure 2: Optimal insurance contract

Finally, let us look at optimal insurance properties in the cases of $b=-\infty$ and $b=+\infty$. First, since the payment rule is continuous, we obtain $F=-(\lambda-K_M)$. From Eq.(4.4),

$$e^r = \int_{-(\lambda-K_M)}^{+\infty} (x+(\lambda-K_M))\,\Phi(\mathrm{d}x) = \int_0^{+\infty} z\,\Phi\Big(\mathrm{d}(z-(\lambda-K_M)) \Big).$$

Therefore, so long as $b=-\infty$ holds, $F=-(\lambda-K_M)$ remains constant as K_M is changed. Next, when $b = +\infty$, from Eq.(4.4),

$$e^r = \int_{-(\lambda - K_M)}^{+\infty} \frac{x + (\lambda - K_M)}{2}\, \Phi(\mathrm{d}x) = \int_0^{+\infty} \frac{z}{2}\, \Phi\Big(\mathrm{d}(z - (\lambda - K_M)) \Big).$$

Therefore, so long as $b = +\infty$ holds, $-(\lambda - K_M)$ remains constant as K_M is changed.

We set the wealth volatility to be a conventional level $v = 25\%$, based on previous empirical results. With regard to the estimates of monitoring cost K_M (relative to r), there is a controversy in previous empirical literatures. Thus we cover a wide range of K_M (relative to the insured's utility level r) in this numerical analysis. To determine the scale of this model, set $r = 1$. Parameterization and numerical results are shown in Table 1.

Table 1: Parameterization and numerical results

Parameters								
Wealth: Diffusion	v	25%	25%	25%	25%	25%	25%	25%
Reservation utility	r	1	1	1	1	1	1	1
Monitoring cost	K_M	$0.01 \sim 0.9$	1.0	1.6	2.2	2.8	5.0	10.0
Optimal results								
Insurance premium	F	n.a.	-1.09	-0.95	-0.94	-0.93	-1.06	-1.06
Compensation trigger	b	$+\infty$	-0.17	-0.32	-0.36	-0.45	$-\infty$	$-\infty$
Lagrangian multiplier	λ	$5.45 \sim 6.34$	3.00	3.19	3.72	4.21	6.06	11.06
Deductible threshold	$-(\lambda - K_M), F$	-5.44	-2.00	-1.59	-1.52	-1.41	-1.06	-1.06
Monitoring region	$b + (\lambda - K_M)$	> -5.44	1.83	1.27	1.16	0.96	0.00	0.00

The results are as follows. When the monitoring cost K_M is an immediate level, the monitoring action is undertaken only for low wealth levels $-(\lambda - K_M) < X_T^* < b$. The contract is deductible when the time-T wealth is less than $-(\lambda - K_M)$. This is a typical insurance contract with deductibles. Since $0 < \frac{\mathrm{d}\lambda}{\mathrm{d}K_M} < 1$, the participation constraint is not too tight. However, $-\frac{U_1'''}{U_1''} = -2\frac{U_1''}{U_1'} > -\frac{U_1''}{U_1'}$, i.e., the insured's downward-risk aversion is strong. From Eq.(3.17), we find that the problem of moral hazard is unchanged by a higher level of the monitoring technology (i.e., smaller K_M). As K_M gets larger, the monitoring region gets smaller. Also, the insurance premium F (the compensation trigger b, respectively) is slightly increasing (decreasing) in K_M. A logic behind these results is as follows. When K_M becomes larger, the whole pie to be shared becomes smaller. Thus the probability of monitoring is reduced. To make up for the high monitoring cost, the insurer demands the high insurance premium F and the low compensation trigger b.

For a very small K_M (i.e., $0.01 \leq K_M < 1$), the insurer necessarily prefers the truth, even by incurring the monitoring cost (i.e., $b = +\infty$). The monitoring action is then undertaken for all nondeductible X_T^* (i.e., unless the contract is deductible for X_T^*), and the allocations are state-dependent. As mentioned above, when $b = +\infty$ holds, $-(\lambda - K_M)$ remains constant as K_M is changed. On the other hand, for a very large K_M, the monitoring action would shrink largely the whole wealth to be shared. The monitoring action is necessarily avoided in equilibrium. The optimal contract is then of a state-independent debt-type for all nondeductible X_T^*, i.e., $b = -\infty$. Note that, in this case, F becomes a deductible threshold as well. Again, as mentioned above, when $b = -\infty$ holds, F and $-(\lambda - K_M)$ remain constant as K_M is changed.

5 Conclusion

We found the properties of optimal insurance in the model of continuous-time costly monitoring under moral hazard and ex-post informational asymmetry. For future work, we will extend the model to have (1) risk pooling in a model with multi-insureds and (2) securitized insurance contracts.

References

[1] Abbring, Jaap H., James J. Heckman, Pierre-André Chiappori, and Jean Pinquet (2003): "Adverse Selection and Moral Hazard In Insurance: Can Dynamic Data Help to Distinguish?" *Journal of the European Economic Association.*, 1, 512-521.

[2] Chiappori, Pierre-André (2000): "Econometric Models of Insurance under Asymmetric Information," in *Handbook of Insurance*, ed. by Georges Dionne. Kluwer Academic Publishers: Norwell, MA, 365-392.

[3] Cvitanić, Jakša, and Jianfeng Zhang (2007): "Optimal Compensation with Adverse Selection and Dynamic Actions," *Mathematics and Financial Economics*, 1, 21-55.

[4] Dionne, Georges, Florence Giuliano, and Pierre Picard (2009): "Optimal Auditing with Scoring: Theory and Application to Insurance Fraud," HEC Montreal, mimeo.

[5] Gale, Douglas, and Martin F. Hellwig (1985): "Incentive-Compatible Debt Contracts: The One-Period Problem, " *Review of Economic Studies*, 52, 647-663.

[6] Harrington, Scott E., and Gregory R. Niehaus (2004): *Risk Management and Insurance*, Second Edition, McGraw-Hill/Irwin: NY.

[7] Holmström, B., and P. Milgrom (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, 303-328.

[8] Karatzas, Ioannis, and Steve E. Shreve (1997): *Brownian Motion and Stochastic Calculus*, Second Edition, Springer.

[9] MacMinn, Ricard, and James Garven (2000): "On Corporate Insurance," in *Handbook of Insurance*, ed. by Georges Dionne. Kluwer Academic Publishers: Norwell, MA, 541-564.

[10] Musiela, Marek, and Marek Rutkowski (1997): *Martingale Methods in Financial Modelling*, New York, Springer-Verla.

[11] Ocone, Daniel L., and Ioannis Karatzas (1991): "A Generalized Clark Representation Formula, with Application to Optimal Portfolios," *Stochastics and Stochastics Reports*, 34, 187-220.

[12] Revuz, Daniel, and Marc Yor (1999): *Continuous Martingales and Brownian Motion*, Third Edition, Springer.

[13] Rogers, L. C. G., and David Williams (2000): *Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus*, Second Edition, Cambridge University Press.

[14] Rogerson, William (1985): "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 53, 1357-1368.

[15] Schättler, H., and J. Sung (1993): "The First-Order Approach to the Continuous-Time Principal-Agent Problem with Exponential Utility", *Journal of Economic Theory*, 61, 331-371.

[16] Williamson, Stephen D. (1987): "Costly Monitoring, Loan Contracts, and Equilibrium Credit Rationing," *Quarterly Journal of Economics*, 102, 135-146.

[17] Zvonkin, A. K. (1974): "A Transformation of The State Space of A Diffusion Process That Removes The Drift," *Math. USSR (Sbornik)*, 22, 129-149.

Appendix

A Proof of Proposition 3.1

Defined in Eq.(2.1), for $u \in \mathcal{A}_1$, $X^u_t = W_t + \int_0^t u_s \,\mathrm{d}s$. As it was shown above, X^u is a driftless Brownian motion under the probability measure $\mathbb{P}^u$ characterized by the Radon-Nikodym derivative:

$$\frac{\mathrm{d}\mathbb{P}^u}{\mathrm{d}\mathbb{P}} = \mathcal{E}\Big(- \int u \,\mathrm{d}W \Big)_T.$$

By the Martingale Representation Theorem, there exists an $\mathbb{F}^{X^u}$-adapted process H^u_s such that

$$\mathbb{E}^u[\, e^{f(X^u_T)} \,|\, \mathcal{F}^{X^u}_t] = \mathbb{E}^u[\, e^{f(X^u_T)}] + \int_0^t H^u_s \,\mathrm{d}X^u_s.$$

Define γ^u as

$$\gamma^u_t := \frac{H^u_t}{\mathbb{E}^u[\, e^{f(X^u_T)} \,|\, \mathcal{F}^{X^u}_t]}.$$

We then obtain

$$e^{f(X^u_T)} = \mathbb{E}^u[e^{f(X^u_T)}]\, \mathcal{E}\Big(\int \gamma^u \,\mathrm{d}X^u \Big)_T. \tag{A.1}$$

Since the Clark-Ocone formula is applicable to $e^{f(\cdot)}$,[11]

$$H^u_t = \mathbb{E}^u[\, e^{f(X^u_T)} f'(X^u_T) \,|\, \mathcal{F}^{X^u}_t]. \tag{A.2}$$

[11] As to the Clark-Ocone formula, see Appendix B.

Since X^u is the driftless Brownian motion under $\mathbb{P}^u$, $X^u_T \sim N(X^u_t, T-t)$ conditional on $\mathcal{F}^{X^u}_t$. Hence,

$$\gamma^u_t = \frac{\mathbb{E}^u[\, e^{f(X^u_T)} f'(X^u_T) \mid \mathcal{F}^{X^u}_t]}{\mathbb{E}^u[\, e^{f(X^u_T)} \mid \mathcal{F}^{X^u}_t]} = h(X^u_t, t) \tag{A.3}$$

as in Eq.(3.1). Therefore, γ^u is bounded. Taking logarithms on both sides of Eq.(A.1),

$$\begin{aligned} f(X^u_T) &= \log \mathbb{E}^u[e^{f(X^u_T)}] + \int_0^T \gamma^u_s \, \mathrm{d}X^u_s - \frac{1}{2}\int_0^T (\gamma^u_s)^2 \, \mathrm{d}s \\ &= \log \mathbb{E}^u[e^{f(X^u_T)}] + \int_0^T \gamma^u_s \, \mathrm{d}W_s - \frac{1}{2}\int_0^T \left\{(\gamma^u_s)^2 - 2u_s\gamma^u_s\right\} \mathrm{d}s. \end{aligned}$$

Subtracting $\frac{1}{2}\int_0^T u_s^2$ from both sides and taking expectations under the probability measure $\mathbb{P}$,

$$\begin{aligned} \mathbb{E}\left[f\Big(W_T + \int_0^T u_s \, \mathrm{d}s\Big) - \frac{1}{2}\int_0^T u_s^2 \, \mathrm{d}s\right] &= \log \mathbb{E}^u[e^{f(X^u_T)}] - \frac{1}{2}\,\mathbb{E}\left[\int_0^T (\gamma^u_s - u_s)^2 \, \mathrm{d}s\right] \\ &\le \log \mathbb{E}^u[e^{f(X^u_T)}]. \end{aligned} \tag{A.4}$$

Note that $\log \mathbb{E}^u[e^{f(X^u_T)}]$ on the right-hand side of Eq.(A.4) is independent of u. If $\gamma^u = u$, i.e., if there exists some $u \in \mathcal{A}_1$ that satisfies the implementability condition (3.3), it is an optimal solution. Moreover,

$$\begin{aligned} \gamma^u = u \quad &\Longleftrightarrow \quad u_\cdot = h(X^u_\cdot, \cdot) \qquad \because \text{Eq.(A.3)} \\ &\Longleftrightarrow \quad X^u_\cdot = W_\cdot + \int_0^\cdot h(X^u_s, s) \, \mathrm{d}s. \end{aligned} \tag{A.5}$$

It is well known that, by Zvonkin [17] (see also Karatzas and Shreve [8], Chapter 5 Notes, p.396), the last Markovian stochastic differential equation (A.5):

$$\mathrm{d}X^u_t = \mathrm{d}W_t + h(X^u_t, t)\, \mathrm{d}t, \qquad X^u_0 = 0$$

has a unique strong solution X^*. Since it is the strong solution, $\mathbb{F}^{X^*} = \mathbb{F}^W$. $\square$

B Supplementary note on the Clark-Ocone formula

The Clark-Ocone formula for general functionals is stated in terms of the Malliavin-Fréchet derivative (see e.g. Ocone and Karatzas [11], Revuz and Yor [12] and Rogers and Williams [13]), but, in order to show our formula (A.2), the following simple argument is sufficient.

Let us define the function $g : \mathbf{R} \times [0, T) \to \mathbf{R}$ by

$$g(x,t) := \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{y^2}{2(T-t)}} e^{f(x+y)} dy,$$

which is a C^2 function since it is the convolution of $e^{f(x)}$ and $\frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{x^2}{2(T-t)}}$. Also, we have

$$\mathbb{E}\big[e^{f(W_T)} \big| \mathcal{F}_t\big] = g(W_t, t) \quad \text{a.s.,} \ 0 \le t < T$$

and thus the process $\{g(W_t, t)\}_{0 \le t < T}$ is a martingale. It then follows from Itô's formula that

$$g(W_t, t) = g(0,0) + \int_0^t \frac{\partial g}{\partial x}(W_s, s) dW_s, \quad 0 \le t < T, \ \text{a.s.}$$

It remains to prove that

$$\begin{aligned} \frac{\partial g(x,t)}{\partial x} &= \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{y^2}{2(T-t)}} \frac{e^{f(x+y+\epsilon)} - e^{f(x+y)}}{\epsilon} \, \mathrm{d}y \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{y^2}{2(T-t)}} f'(x+y) e^{f(x+y)} \, \mathrm{d}y. \end{aligned}$$

This can be shown by the following two properties combined with Lebesgue's dominated convergence theorem:

- For almost every $x \in \mathbf{R}$, $\lim_{\epsilon \to 0} \frac{e^{f(x+\epsilon)} - e^{f(x)}}{\epsilon} = f'(x) e^{f(x)}$;
- Since the Lipschitz constant of f is $K := \sup_{x \in \mathbf{R} \setminus \{x_1, x_2, \cdots, x_n\}} |f'(x)| < \infty$, we have, for every $x \in \mathbf{R}$,

$$\begin{aligned} \sup_{0<|\epsilon|<1} \left| \frac{e^{f(x+\epsilon)} - e^{f(x)}}{\epsilon} \right| &= e^{f(x)} \sup_{0<|\epsilon|<1} \left| \frac{e^{f(x+\epsilon) - f(x)} - 1}{\epsilon} \right| \\ &\le e^{f(x)} \sup_{0<|\epsilon|<1} \frac{e^{|f(x+\epsilon) - f(x)|} - 1}{|\epsilon|} \\ &\le e^{f(x)} \sup_{0<|\epsilon|<1} \frac{e^{K|\epsilon|} - 1}{|\epsilon|} \\ &= (e^K - 1) e^{f(x)}. \end{aligned}$$