Optimal Contracting under Mean-Volatility Ambiguity Uncertainties

Jaeyoung Sung *

This version: March 10, 2018

Abstract

We present a continuous-time agency model under mean-volatility joint ambiguity uncertainties, where both the principal and agent exhibit Gilboa-Schmeidler's extreme ambiguity aversion. For this, we extend the martingale method well known in the agency literature, by allowing not only the mean but the volatility of the outcome process to be controlled in weak formulation. Unlike the existing literature, we distinguish between ex-post realized and ex-ante perceived volatilities. Then we argue that the second-best contract in general consists of two sharing rules: one for realized outcome and the other for realized volatility. The outcome sharing is for both uncertainty sharing and work incentives, and the volatility sharing is to align the agent's worst prior over ambiguity uncertainties with that of the principal. The optimal volatility sharing occurs when their worst priors become symmetrized. We show that the realized compensation is positively associated with the realized volatility, and that the sensitivity to the outcome is negatively related to the perceived volatility.

JEL classification: C61, C73, D81, D86, G34.
Keywords: mean-volatility ambiguity, moral hazard, optimal contract, stock option, managerial compensation, volatility control, weak formulation.

*Ajou University, School of Business Administration, Suwon, 443-749, South Korea. Phone: +82 31 219 2715. Fax: +82 31 219 1616. Email: jaeyoungsung1@yahoo.com. I would like to thank Xiaoyan Chen, Zengjing Chen, Jaksa Cvitanić, Shige Peng and Jianfeng Zhang for useful comments/discussions. All remaining errors of course are mine.

1 Introduction

We examine effects of ambiguity uncertainties on the optimal contract under moral hazard. It is well-accepted that economic future outcomes can be subject to 'risk' and 'ambiguity/Knightian' uncertainties. Under risk uncertainty, the true probability distribution of the uncertain outcome is known, whereas under ambiguity uncertainty, the true probability distribution is unknown.

In the presence of ambiguity uncertainty, the subjective expected utility paradigm pioneered by Savage (1954) and Anscombe and Aumann (1963) implies that a rational economic agent who adheres to a certain set of axioms, makes decisions maximizing his expected utility, as if he subjectively assigned probabilities to all ambiguous events. However, the famous Ellsberg (1961) paradox indicates that one of those axioms, namely the independence axiom, can be frequently violated by real-life thought experiments. Proposed as resolutions to this paradox in the literature are various modifications of the paradigm such as maxmin expected utility, Choquet expected utility, and smooth-ambiguity utility theories.[1]

We apply Gilboa and Schmeidler's (1989) maxmin (multiple-priors) utilities to a contracting problem where the principal (she) contracts the agent (he) to manage an asset under ambiguity uncertainties. Endowed with maxmin utilities, the two contracting parties perceive the ambiguity uncertainties as risk uncertainties under their subjective worst priors which have nothing to do with the unknown true distribution of the uncertain asset outcome. Moreover, since their worst priors can be different from each other, it is seemingly possible that the two different individuals behave as if they were given different information about the common uncertainties, and decided to 'agree to disagree.'[2]

[1]See Gilboa and Schmeidler (1989) for maxmin expected utility; Choquet (1953), Gilboa (1987) and Schmeidler (1989) for Choquet expected utility, and Klibanoff, Marinacci, and Mukerji (2005) for smooth-ambiguity utility theories.

[2]Epstein and Miao (2003) and Epstein (2001) consider two people with different levels of ambiguity. In this paper, two people, the principal and agent, face the common ambiguities.

In this paper, however, we show that the disagreement of their worst priors does not occur under the optimal contract. For this, we introduce to our agency model *joint* ambiguities in the mean and volatility of an Itô process-driven continuous-time outcome.[3]

It is well known from the existing literature that the volatility of the outcome importantly affects both risk-sharing and incentives in contracting. In the presence of volatility ambiguity, however, there are two different volatilities, ex-ante and ex-post volatilities: the former is subjectively ex-ante perceived and thus noncontractable, whereas the latter is objectively ex-post realized and thus contractable. Then, one can easily imagine that different subjective perceptions, if any, can cause inefficiency in contracting, just like information asymmetry (without learning) can. Thus, the principal may want to look for an extra contractual arrangement which can help reduce the perceptional difference.

We show that in the first-best contracting, such an extra contract is unnecessary because the usual standard outcome-based first-best contract alone automatically induces both the principal and agent to agree, in perception, on the volatility and mean of the ambiguous outcome. In the second best contracting, however, the outcome-based contract alone cannot optimally induce the agreement. As a result, the second best contract requires two distinct sharing rules: one for the realized outcome and the other for the realized volatility.[4] The outcome sharing is for uncertainty sharing and work incentives, and the volatility sharing for symmetrization of perceived risk burdens/premia between the two individuals. The two sharing rules are represented by their sensitivities, β_t and θ_t, to the outcome and the realized volatility, respectively.

[3]Our joint ambiguity assumption is necessary in order to derive a nontrivial resolution of the disagreement issue: it shall be seen later that the resolution requires a volatility-sharing scheme as an integral part of the optimal contract. Note that the issue does not arise in contracting if either the mean alone or the volatility alone of the outcome is ambiguous, because both the risk-and-ambiguity-averse principal and agent (automatically) agree on the most pessimistic probability measure for the outcome under the usual outcome-sharing contract. Chen and Epstein (2002) show that many agents with different concave utility functions can agree on the most pessimistic probability measure if the set of possible measures satisfies the rectangularity condition under the mean ambiguity. Epstein and Schneider (2008) consider a volatility ambiguity case.

[4]Cvitanić, Possamaï and Touzi (2014) also present a similar result from their principal-agent problem without ambiguity. Our result depends on a completely different economic reasoning from theirs.

We argue that volatility sharing is optimally achieved when the worst priors of the two contracting parties are symmetrized.[5] Differential priors about the volatility would result in differential risk premia: given an outcome sharing rule β_t, the agent may perceive excessively high risk from his ambiguity-uncertainty exposure, and demand too higher a risk premium than, the principal believes, he should. Then, the principal can improve the contract by using the volatility sharing rule θ_t to shift a part of the agent's perceived risk burden to herself, until their perceived risk premia are equalized, and so are their worst priors.[6] This intuition is reminiscent of *the law of one price* in financial markets where investors who find different prices for the same asset trade for profit until the price discrepancy disappears.

We further show that the optimal θ_t is positive: the greater the realized volatility, the more the realized compensation to the agent. We believe that this result is consistent, to some extent, with popular executive compensation practices of granting stock options, in the sense that stock options values are positively associated with realized volatilities. See Guay(1999), Core and Guay (2002), Coles, Daniel and Naveen (2006) and Murphy (2012) for positive empirical relation between pay and (total) option Vega, where the Vega measures the change of the option value per one percentage-point increase in volatility.

The two sensitivities β_t and θ_t are determined over time based on both static and dynamic tradeoffs. Static factors such as risk and work aversion, and ex-ante perceived outcome volatility are well known in the literature. Dynamic factors include not only the time and state, but *the principal's effective share* which we define (in Section 5.2) as the sum of the following two components: the nominal share $1 - \beta_t$, and an imaginary extra

[5] Note that a need for volatility sharing uniquely arises because of the presence of mean-volatility *joint* ambiguity uncertainties. Disagreement on the worst prior does not arise in contracting, if either the mean alone or the volatility alone of the outcome is ambiguous, because given the usual second-best outcome-based contract alone, both the risk-and-ambiguity-averse principal and agent automatically agree on the most pessimistic probability measure for the outcome. Chen and Epstein (2002) show that many agents with different concave utility functions agree on the most pessimistic probability measure under the mean ambiguity. Epstein and Schneider (2008) consider a volatility ambiguity case.

[6] One may alternatively view it as a social planner problem in a two-person world, where the planner uses the volatility sharing rule to allocate volatility-ambiguity uncertainties between the two parities, in order to maximize their perceived aggregate welfare.

share Z_t^P. This factor affects the two sensitivities over time: holding static factors fixed, the greater the principal's effective share, the higher-powered outcome-based incentives. In Section 6, we show that the imaginary share is positive (negative, zero), if the outcome exhibits increasing (decreasing, constant) returns to scale where the case of constant returns to scale corresponds to the Holmstrom and Milgrom (1987) stationary outcome. Effects of the imaginary share on the two sensitivities are greatest at the initial date, and converge to zero at the final date.

This paper is related to the recent literature on contracting under ambiguity uncertainties. Weinschenk (2010) examines linear contracts for a discrete-time contracting problem under ambiguity between the risk-ambiguity neutral principal and risk-ambiguity averse agent, but does not consider volatility-ambiguity sharing rules, because of the limitation of the discrete-time model. Szydlowski (2012) considers a dynamic contracting problem in the presence of the agent's limited liability and ambiguity about the agent's effort cost, and show that the optimal contract provides excessive incentives. Miao and Rivera (2015) also consider a continuous-time contracting case, and show that the optimal contract is a tradeoff between incentives and ambiguity sharing. Both Szydlowski and Miao and Rivera extend Sannikov (2008) by introducing ambiguity uncertainties into drifts of their outcomes, but not into volatilities, and that the two contracting parties have heterogeneous beliefs/perceptions: the agent faces no ambiguity, but the principal has to deal with ambiguity. In order to examine pure ambiguity effects on the optimal contract without being complicated by potential information-asymmetry issues, we assume, in this paper, that the two individuals share homogeneous beliefs about mean-volatility joint ambiguity uncertainty.[7]

[7] After the first version of this paper (2014) under a different title was circulated, Mastrolia and Possamaï's (2015) work came to my attention. The authors consider an extended version of Holmstrom-Milgrom (1987), allowing the volatility but not the mean to be ambiguous. They assume heterogeneous ambiguity beliefs between the principal and agent about the volatility of the outcome process. Then, information asymmetry and learning problems can simultaneously arise in their model, which can be potentially interesting issues for future research. In this paper, we avoid these issues by considering homogenous ambiguity beliefs.

This paper is also related to the volatility control literature. Volatility control problems in continuous-time contracting appear in early papers without ambiguity uncertainties, such as Sung (1995) and Ou-yang (2003). Both authors manage to solve their problems, because their problems can be transformed into Markovian volatility control problems as shown in Schättler and Sung (1993), and because the realized (ex post) volatility is a predictable process. We derive a Hamiltonian for mean-volatility control problems in the Appendix, directly extending Schättler and Sung. Recently, Cvitanić, Possamaï and Touzi (2014, 2015) examine volatility control issues in strong formulation to derive a path-dependent Hamilton-Jacobi-Bellman (HJB) equation for a continuous-time contracting problem without ambiguity uncertainties. Epstein and Ji (2013) consider their volatility control problems with singular measure changes in weak formulation utilizing Peng's (2006) G-Brownian Motion. For our mean-volatility control problem, we utilize the control-theoretic method which has recently been developed in the literature: see for example, Soner, Touzi, and Zhang (STZ, 2011a,b, 2012, 2013), Nutz and Soner (2012), Nutz (2012a,b), Bouchard and Nutz (2012), and Pham and Zhang (2014).

The rest of the paper is organized as follows. Section 2 describes the continuous-time contracting model with the outcome subject to mean-volatility joint ambiguity uncertainties. Before starting the analysis of the model, in Section 3, we narrow down, without loss of generality, the class of admissible contracts to a manageable subclass. Then, we start the analysis from Section 4, by presenting the first-best solution as a benchmark, and examine details of the second-best case in Section 5, where we show that the optimal contract depends not only on the outcome but on its quadratic variation, and that the most pessimistic priors of the two contracting parties are symmetrized. Section 6 provides an example using a linear-quadratic case with its joint-ambiguity parameters lying in a quadratic set. Finally, we summarize the paper in Section 7. The Appendix contains most of proofs and the martingale method for mean and volatility controls.

Notation.

Y**:** the outcome or coordinate process of $\Omega(\subset C[0,1])$, such that $Y_t(\omega) = \omega_t$, $\forall \omega \in \Omega$.

$\mathcal{P}$**:** the family of admissible singular and equivalent probability measures.

$\mathcal{P}^o$**:** the subfamily of amissible singular measures.

$\{\mathcal{F}_t\}, \{\hat{\mathcal{F}}_t\}$**:** resp., the natural filtration generated by Y, and the universal filtration defined in Section 2.

(u, v)**:** u and v, resp., are tuples of control and ambiguity parameter processes. In Sections A.1 and A.2, $u_t \in \mathcal{R}^n$ and $v_t \in \mathcal{R}^m$. In the text, $u_t = e_t \in \mathcal{R}$ and $v_t = (\mu_t, \nu_t) \in \mathcal{R}^2$.

(μ, ν)**,** $(\boldsymbol{\mu}, \boldsymbol{\nu})$**:** resp., *ex ante* perceived, and *ex post* realized/true ambiguity parameter processes.

U**,** D**:** subsets of $\mathcal{R}^n$ and $\mathcal{R}^m$, resp., for u_t and v_t for each t.

$U_t(Y)$**,** $D_t(Y)$**:** time-state dependent subsets of U and D, resp. In the text, $U_t(Y) \equiv U$ and $D_t(Y) = \{(\mu_t, \nu_t) \in D \,|\, \pi(\mu_t, \nu_t, t, Y) \geq 0\}$.

$\hat{\mathcal{U}}, \hat{\mathcal{D}}$**:** classes of $\hat{\mathcal{F}}^{\mathcal{P}}$-progressively measurable control and ambiguity-parameter processes, u and v, resp., with their values $(u_t, v_t) \in U \times D$ for each t.

$\mathcal{U}_t^1, \mathcal{D}_t^1$**:** subclasses of $\hat{\mathcal{U}}$ and $\hat{\mathcal{D}}$, resp., with $(u_s, v_s) \in U_s(Y) \times D_s(Y)$ where $s \in [t, 1]$. For brevity, $\mathcal{U}_0^1 \equiv \mathcal{U}$ and $\mathcal{D}_0^1 \equiv \mathcal{D}$.

$f(u, v, .)$**,** $\sigma(u, v, .)$**:** resp., drift and diffusion rates of Y under $P^{u,v}(\in \mathcal{P})$.

Φ, $\boldsymbol{\Sigma}$ **:** classes of admissible f and σ, resp.

$\Omega^\sigma(\subset \Omega)$**:** a collection of all sample paths of Y with $d\langle Y_t \rangle = \sigma_t^2 dt$.

H^A**,** H^P**,** H^o**:** resp., Hamiltonians for the agent, the principal, and a general case.

$\mathbb{L}^1_{\mathcal{P}}$: $L^1_{\mathcal{P}}$-closure of bounded uniformly continuous functions on Ω.

Ψ, $\bar{\Psi}$: classes of admissible contracts, defined in (2) and (9), resp.

β_t, θ_t: resp., outcome- and volatility-sharing sensitivities at time t.

2 The Model

We introduce ambiguity uncertainty to the standard continuous-time agency problem with one principal and one agent, whose preferences are represented by constant absolute risk aversion with their CARA coefficients given by γ_P and γ_A, respectively, where $\gamma_P, \gamma_A \geq 0$. The time horizon of interest is the unit interval $[0, 1]$. The principal has an asset in place which will produce a cumulative monetary outcome of Y_t at time $t(\in [0, 1])$. The agent has expertise to manage the asset, and considers entering into a contract with the principal. His reservation utility is $-\exp(-\gamma_A \mathcal{W}_0)$, where $\mathcal{W}_0 \in \mathcal{R}$. At time 0, both the principal and agent sign a contract S subject to the participation constraint which requires S to guarantee, at the least, the agent's reservation utility. The contract specifies how the two contracting parties share the outcome, as a function of objectively verifiable information based on the whole history of the outcome, $Y \equiv \{Y_t; t \in [0, 1]\}$. After time 0, the agent exerts effort to improve the probability distribution of the outcome process Y. Both the principal and agent observe the whole history of Y as it realizes.

The uncertainty of the outcome Y is characterized by the filtered probability space $(\Omega, \hat{\mathcal{F}}, \{\hat{\mathcal{F}}_t\}, \{P \in \mathcal{P}\})$, where Ω is the space of all continuous functions on [0,1] with each function starting at zero, i.e., $\Omega := \{\omega \in C[0, 1] \,|\, \omega(0) = 0\}$; $\{\hat{\mathcal{F}}_t\}$ is the universal filtration, $\hat{\mathcal{F}}_1 = \hat{\mathcal{F}}$, and $\mathcal{P}$ is the admissible family of probability measures, including singular measures. The probability space is the common knowledge between the principal and agent, i.e., there is no asymmetric information.

We assume that the universal filtration satisfies the set of properties listed in STZ (2011a, Definition 2.2). The filtration is constructed as follows. Let $\{\mathcal{F}_t\}$ be the natural filtration generated by Y, and $\mathcal{F}_t^+ = \bigcap_{s>t} \mathcal{F}_s$. Also let $\mathcal{F}_t^P := \mathcal{F}_t^+ \vee \mathcal{N}^P(\mathcal{F}_t^+)$, for each $P \in \mathcal{P}$, where

$$\mathcal{N}^P(\mathcal{F}_t^+) = \left\{ A \in \Omega \,|\, \text{there exists } \tilde{A} \in \mathcal{F}_t^+ \text{ such that } A \subset \tilde{A} \text{and } P(\tilde{A}) = 0 \right\}.$$

Then the universal sigma algebra at time t is

$$\hat{\mathcal{F}}_t^{\mathcal{P}} = \bigcap_{P \in \mathcal{P}} \left(\mathcal{F}_t^P \vee \mathcal{N}_{\mathcal{P}} \right),$$

where $\mathcal{N}_{\mathcal{P}}(:= \bigcap_{P\in\mathcal{P}} \mathcal{N}^P(\mathcal{F}_1))$ is the collection of so-called $\mathcal{P}$-polar sets, each of which is a null set for all admissible probability measures. The universal filtration $\{\hat{\mathcal{F}}_t\}(\equiv \{\hat{\mathcal{F}}_t^{\mathcal{P}}\})$ is necessary for each admissible stochastic process to be well-defined not only on its own support of the probability measure, but on its null sets where other admissible (singular) probability measures live.

The family $\mathcal{P}$ is constructed by expanding the subfamily $\mathcal{P}^o$ of (partially or completely) singular measures, with absolutely continuous measures to each singular measure in the subfamily. Those singular measures are constructed by partitioning Ω by the quadratic variation (QV) of each sample path, and by endowing each partition Ω^σ with a Wiener measure P^σ as in Lemma A.1. Then, we use the Girsanov Theorem to generate absolutely continuous measures with respect to each $P^\sigma (\in \mathcal{P}^o)$. See the Appendix for details.

Each admissible measure $P^{e,v}(\in \mathcal{P})$ is indexed by the agent's effort process e and the pair v of uncertain/ambiguious parameter processes (μ, ν). Let $\mathcal{U}$ and $\hat{\mathcal{D}}$, respectively, be the classes of all $\hat{\mathcal{F}}^{\mathcal{P}}$-progressively measurable control and ambiguity-parameter processes such that $e(\in \mathcal{U}) : [0,1] \times \Omega \to U(\subset \mathcal{R})$ and $v(= (\mu,\nu) \in \mathcal{D}) : [0,1] \times \Omega \to D(\subset \mathcal{R} \times \mathcal{R}_+)$, respectively, where $\mathcal{R}_+$ is the strictly positive part of the real line. Let $D_t(Y) := \{(\mu,\nu) \in D \,|\, \pi(\mu,\nu,t,Y) \geq 0\}$, where $\pi : D \times [0,1] \times \Omega \to \mathcal{R}$ is continuously differentiable in (μ,ν). Also let $\mathcal{D} := \{v \in \hat{\mathcal{D}} \,|\, v(t,Y) \in D_t(Y), t \in [0,1]\}$. Then $\mathcal{D}$ is a subclass of $\hat{\mathcal{D}}$ constrained by $D_t(Y)$ for all t. We assume that both U and $D_t(Y)$ for all t are compact and convex with nonempty interiors.[8] Throughout the paper, the pair of arguments (t, Y) denotes $(t, \{Y_s, 0 \le s \le t\})$.

Under each $P^{e,v} \in \mathcal{P}$, the outcome (or the coordinate) process evolves according to the following scalar-process dynamics:

$$dY_t = f(e_t, \mu_t, \nu_t, t, Y)dt + \sigma(\nu_t, t, Y)dB_t^{e,v}, \tag{1}$$

with $Y_0 = 0$, where $B_t^{e,v}$ is a $P^{e,v}$-standard Brownian Motion (BM). The drift and diffusion rates, f and σ, respectively, are members of classes Φ and $\mathbf{\Sigma}$ of nonanticipative functionals, which are defined in Assumptions A.1 and A.2 with $u = e$ and $v = (\mu, \nu)$. In particular, the class $\mathbf{\Sigma}$ satisfies the uniform Lipschitz continuity property. As can be seen in the Appendix, admissible conditional probability measures, $P_t^{e,v}$ and $P_t^{e',v'}$, are mutually absolutely continuous if $\sigma_s = \sigma'_s$ for all $s \in (t, 1]$, and mutually singular if $\sigma_s \neq \sigma'_s$ all $s \in (t, 1]$.

The agent is allowed to privately choose effort $e(s, Y)$, incurring a cumulative private monetary cost of $\int_0^t c(e_s, s, Y)ds$ up to time t, where c is also a nonanticipative functional. Furthermore, we assume that f, σ and c, are continuously differentiable in (e, μ, ν), ν, and e, respectively, that both f and c are strictly increasing in e, and that σ is strictly increasing in ν. We let c_e and f_e denote partial derivatives of c and f, respectively, with respect to e.

Also let $\boldsymbol{v}(\in \mathcal{D})$ be the pair of the true parameter processes $(\boldsymbol{\mu},\boldsymbol{\nu})$. Given (e, S), if $P^{e,\boldsymbol{v}}$ were known, then the agent's expected utility would be

$$\begin{aligned} & E^{e,\boldsymbol{v}}\left[-\exp\left\{-\gamma_A\left(S - \int_0^1 c(e_t, t, Y)dt\right)\right\}\right] \\ & \text{s.t.} \quad dY_t = f(e_t, \boldsymbol{\mu}_t, \boldsymbol{\nu}_t, t, Y)dt + \sigma(\boldsymbol{\nu}_t, t, Y)dB_t^{e,\boldsymbol{v}}, \end{aligned}$$

where $E^{e,\boldsymbol{v}}$ is the expectation operator under probability measure $P^{e,\boldsymbol{v}}$. However $P^{e,\boldsymbol{v}}$ is unknown/ambiguous, except that $P^{e,\boldsymbol{v}} \in \mathcal{P}$.

[8]This condition is to satisfy the Karush-Kuhn-Tucker (KKT) constraint qualification conditions for almost all $t \in [0, 1]$. See, for instance, Takayama (1985) or Bazaraa, Sherali and Shetty (2006) for the constraint qualification conditions.

We assume, as suggested by Gilboa and Schmeidler (1989), that both the principal and agent are risk-and-ambiguity averse with maxmin utilities. In particular, given a contract S, the agent chooses (e, v) to solve the following problem:

$$\sup_{e\in\mathcal{U}} \inf_{v\in\mathcal{D}} E^{e,v}\left[-\exp\left\{-\gamma_A\left(S-\int_0^1 c(e_t,t,Y)dt\right)\right\}\right]$$
$$\text{s.t.}\quad dY_t = f(e_t,\mu_t,\nu_t,t,Y)dt+\sigma(\nu_t,t,Y)dB_t^{e,v}.$$

It is well known that the case of a risk-neutral agent is the limiting case of this problem with $\gamma_A \downarrow 0$. Given this maxmin utility and effort process e, the agent chooses the worst prior $P^{e,v^*(e)}$, where $v^*(e) = (\mu^*(e), \nu^*(e))$. Under each prior $P^{e,v}$, the agent behaves as if the QV, $d\langle Y_t\rangle$, were equal to $\sigma^2(\nu, t, Y)dt$ with probability one, even though he knows that $\sigma^2(\nu, t, Y)$ bears no relation to the true $\sigma^2(\boldsymbol{\nu}, t, Y)$. In this paper, we call $\sigma^2(\nu, t, Y)$ the *ex ante* perceived QV (density) under prior $P^{\cdot,v}$, and $\sigma^2(\boldsymbol{\nu}, t, Y)$ the *ex post* realized QV.

As is the case with the perceived $\sigma^2(\nu, t, Y)$, the perceived $\mu(t, Y)$ is unlikely to be equal to the realized $\boldsymbol{\mu}(t, Y)$ which in turn conveys zero information about $\boldsymbol{\mu}(t+dt, Y)$. That is, brand-new ambiguity uncertainties on both mean and volatility keep arising continuously over time. Thus, neither the future mean nor the future volatility can be predicted through Bayesian learning. This is in agreement with the typical and implicit assumption in the ambiguity literature that ambiguity uncertainties are "generated by hard to interpret, ambiguity signals" continuously over time.[9] See Chen and Epstein (2002), and Epstein and Schneider (2010) for explanation of ambiguity aversion without learning.

On the other hand, the principal chooses S from the admissible class Ψ to maximize

$$\inf_{v\in\mathcal{D}} E^{e,v}\left[-\exp\left\{-\gamma_P\left(Y_1-S\right)\right\}\right],$$

[9] In other words, at each time t, the outcome and its volatility ambiguity uncertainties up to time t, i.e., $\{(s, Y_s, \sigma(\boldsymbol{\nu}, s, Y)); s \in [0, t]\}$, are completely resolved, but the resolution provides zero information about future ambiguity uncertainties, i.e., $\{(\boldsymbol{\mu}_s, \boldsymbol{\nu}_s); s \in (t, 1]\}$. One may imagine that, at each time t, time-t 'ambiguity urn' determining $(\boldsymbol{\mu}_t, \boldsymbol{\nu}_t)$ is completely revealed and replaced with a brand-new 'urn' for time $t + dt$. Consequently, neither (μ_t, ν_t) nor $(\boldsymbol{\mu}_{t+dt}, \boldsymbol{\nu}_{t+dt})$ has anything to do with $(\boldsymbol{\mu}_t, \boldsymbol{\nu}_t)$, and the agent simply chooses the worst pair of (μ_t, ν_t) based on information up to time t, $\{(s, Y_s); s \in [0, t]\}$, regardless of $\{(\boldsymbol{\mu}_s, \boldsymbol{\nu}_s) : s \in [0, t]\}$. Note that this consequence is directly from the definition of the ambiguity in the literature, and has nothing to do with the singular change of measures.

subject to appropriate constraints. The precise statement of the principal's problem shall be given, as we proceed. We assume that the admissible class Ψ of contracts consists of $\hat{\mathcal{F}}_1^{\mathcal{P}}$-measurable sharing schemes S's as follows:

$$\Psi := \left\{ S \;\middle|\; \begin{array}{l} S \text{ is } \hat{\mathcal{F}}_1^{\mathcal{P}}\text{-measurable, and under all } P^{e,v} \in \mathcal{P}, \\ -\exp\left\{-\gamma_A \left(S - \int_0^1 c(e_t, t, Y)dt\right)\right\} \in \mathbb{L}_{\mathcal{P}}^1 \end{array} \right\}, \tag{2}$$

where $\mathbb{L}_{\mathcal{P}}^1$ is the $L_{\mathcal{P}}^1$-closure of bounded uniformly continuous functions on Ω. See Nutz (2012b, Section 4) for the definition of $\mathbb{L}_{\mathcal{P}}^1$.

Before proceeding to the analysis of our contracting problem, let us define the agent's and principal's Hamiltonians H^A and H^P constrained by $D_t(Y)$, respectively as follows: for $(e, (\mu, \nu), p_1, p_2, t, Y) \in U \times D_t(Y) \times \mathcal{R} \times \mathcal{R} \times [0,1] \times \Omega$,

$$H^A(e, \mu, \nu; p_1, p_2, t, Y) := -c(e, t, Y) + \varphi_A(e; \mu, \nu; p_1, p_2, t, Y), \tag{3}$$

and for $(e, \beta, \theta, (\mu^P, \nu^P), p, t, Y) \in U \times \mathcal{R}^2 \times D_t(Y) \times \mathcal{R} \times [0,1] \times \Omega$,

$$\begin{aligned} &H^{0P}(e, \beta, \theta; \mu^P, \nu^P; p, t, Y) \\ &\quad := -c(e, t, Y) + \varphi_P(e, \beta, \theta; \mu^P, \nu^P; p, t, Y) + \min_{(\mu^A, \nu^A) \in D_t(Y)} \varphi_A(e; \mu^A, \nu^A; \beta, \theta, t, Y). \end{aligned} \tag{4}$$

Also let

$$\varphi_A(e; \mu, \nu; \beta, \theta, t, Y) := \beta f(e, \mu, \nu, t, Y) - \left(\frac{\gamma_A}{2}\beta^2 - \theta\right) \sigma^2(\nu, t, Y) \tag{5}$$

$$\begin{aligned} \varphi_P(e, \beta, \theta; \mu, \nu; p, t, Y) &:= (1 - \beta + p)\, f(e, \mu, \nu, t, Y) \\ &\quad - \left[\theta + \frac{\gamma_P}{2}(1 - \beta + p)^2\right] \sigma^2(\nu, t, Y), \end{aligned} \tag{6}$$

$$H^P(e, \theta; \mu^P, \nu^P; p, t, Y) := \left[\begin{array}{c} H^{0P}(e, \beta, \theta; \mu^P, \nu^P; p, t, Y) \\ \text{s.t.} \quad \beta \equiv \frac{c_e(e,t,Y)}{f_e(e,\mu,\nu,t,Y)} \end{array} \right]. \tag{7}$$

Assumption 1 *Existence of saddle points.*[10]

i. *For each $(p_1, p_2, t, Y) \in \mathcal{R} \times \mathcal{R} \times [0,1] \times \Omega$, there exists a saddle point $(e_t^*, \mu_t^*, \nu_t^*)$ such that for all $(e_t, (\mu_t, \nu_t)) \in U \times D_t(Y)$, $H^A(e_t, \mu_t^*, \nu_t^*, .) \le H^A(e_t^*, \mu_t^*, \nu_t^*, .) \le H^A(e_t^*, \mu_t, \nu_t, .)$.*

[10] See Sion (1958) and Rockafellar (1970) for minimax theorems.

ii. *For each* $(p, t, Y) \in \mathcal{R} \times [0, 1] \times \Omega$, *there exists a saddle point* $(e^*_t, \beta^*_t, \theta^*_t, \mu^*_t, \nu^*_t)$ *such that for all* $(e_t, \beta_t, \theta_t, (\mu_t, \nu_t)) \in U \times \mathcal{R}^2 \times D_t(Y)$, $H^P(e_t, \theta_t, \mu^*_t, \nu^*_t, .) \leq H^P(e^*_t, \theta^*_t, \mu^*_t, \nu^*_t, .) \leq H^P(e^*_t, \theta^*_t, \mu_t, \nu_t, .)$.

iii. *For each* $(p, t, Y) \in \mathcal{R} \times [0, 1] \times \Omega$, *there exists a saddle point* $(e^*_t, \beta^*_t, \theta^*_t, \mu^*_t, \nu^*_t)$ *such that for all* $(e_t, \beta_t, \theta_t, (\mu_t, \nu_t)) \in U \times \mathcal{R}^2 \times D_t(Y)$, $H^{0P}(e_t, \beta_t, \theta_t, \mu^*_t, \nu^*_t, .) \leq H^{0P}(e^*_t, \beta^*_t, \theta^*_t, \mu^*_t, \nu^*_t, .) \leq H^{0P}(e^*_t, \beta^*_t, \theta^*_t, \mu_t, \nu_t, .)$.

It shall be shown later that H^A and H^P, respectively, are related to the agent's and principal's second-best certainty-equivalent (CEQ) wealth levels; and H^{0P} to that of the principal's first best. The CEQ is defined in (A.11). Given the above formulation, a general method to deal with the principal's and agent's problems is developed in the Appendix, and utilized in the next sections.

3 Representation of Admissible Contracts

We start the analysis of our contracting problem, by narrowing down the original admissible class Ψ to a manageable subclass $\bar{\Psi}$ without affecting the principal's utility, as seen frequently in the standard continuous-time contracting literature, where $\bar{\Psi}$ shall be defined shortly. Let $\mathcal{W}^{S,e,v} (\in \mathcal{R})$ be the CEQ wealth level of the expected utility the agent achieves with an arbitrary choice of (e, v) given $S \in \Psi$, where $v = (\mu, \nu)$. Then,

$$- \exp\left(-\gamma \mathcal{W}^{S,e,v}\right) = E^{e,v} \left[- \exp\left\{ -\gamma_A \left(S - \int_0^1 c(e_s, \mu_s \nu_s, s, Y) ds \right) \right\} \right].$$

A comment is in order. For control problems with singular measures, an important concept is the quasi-sure (q.s.) measurability, which is defined by Denis and Martini (2006) and STZ (2011a): a property is said to hold $\mathcal{P}$-q.s., if it holds for all $P \in \mathcal{P}$, a.s.. Since $\mathcal{P}$ is an enlarged version of $\mathcal{P}^o$ with absolutely continuous measures, each of which is dominated by a member in $\mathcal{P}^o$, if a stochastic property holds $\mathcal{P}^o$-q.s., then so does it $\mathcal{P}$-q.s..

Theorem 1 *Given a contract $S \in \Psi$, suppose that the agent arbitrarily chooses admissible $(e, (\mu, \nu)) \in \mathcal{U} \times \mathcal{D}$. Then, there exists a unique $\mathcal{P}^o$-q.s. square integrable processes, (β_t, θ_t) such that S can be represented in the following form: $\mathcal{P}^o$-q.s.,*[11]

$$S = \mathcal{W}^{S,e,v} + \int_0^1 \left\{ c(e_t, t, Y) - \beta_t f(e_t, \mu_t, \nu_t, t, Y) + \left[\frac{\gamma_A}{2} \beta_t^2 - \theta_t \right] \sigma^2(\nu_t, t, Y) \right\} dt$$
$$+ \int_0^1 \theta_t d\langle Y_t \rangle + \int_0^1 \beta_t dY_t. \tag{8}$$

If θ_t were set to zero for all t, the representation (8) of the salary scheme S would be identical to that of Schättler and Sung (1993). Recall that $d\langle Y_t \rangle (\equiv \sigma^2(\boldsymbol{\nu}_t, t, Y)dt)$ is the realized QV, whereas $\sigma^2(\nu_t, t, Y)dt$ is the agent's ex-ante perceived QV under his worst prior. In the absence of volatility ambiguity, $\sigma^2(\boldsymbol{\nu}_t, t, Y) = \sigma^2(\nu_t, t, Y)$, a.s., and thus the dependence of S on the realized QV would be unnecessary. The salary scheme (8) implies that $\int_0^1 \theta_t(\sigma^2(\boldsymbol{\nu}_t, t, Y) - \sigma^2(\nu_t, t, Y))dt$ is an extra amount the agent demands in addition to the standard salary which he would be paid without ambiguity uncertainty. Of course this extra amount can be positive or negative depending on the realization of $\sigma(\boldsymbol{\nu}_t, t, Y)$.

In both the first- and second-best contracting cases, Theorem 1 suggests that, without loss of generality, the principal can only consider the following subclass $\bar{\Psi}$ of contracts which satisfy the agent's reservation utility:

$$\bar{\Psi} := \left\{ S \in \Psi \,|\, S \text{ in the form (8) with } \mathcal{W}^{S,e,v} = \mathcal{W}_0 \right\}. \tag{9}$$

Then, the principal chooses a contract S from $\bar{\Psi}$ by choosing a set of processes $(e_t, \mu_t, \nu_t, \beta_t, \theta_t)$, subject to appropriate constraints.

4 First-Best Contracting

Recall that given contract S, the agent chooses his effort $e \in \mathcal{U}$ and prior $(\mu^A, \nu^A) \in \mathcal{D}$ for the ambiguity uncertainty. In the first best, the principal instructs/forces the agent to

[11] A similar class is also considered in Cvitanić, Possamaï and Touzi (2014) who examine volatility control issues in their principal-agent problems without ambiguity uncertainties.

choose a specific effort process. Such a forcing contract is possible, perhaps because the agent's effort choice can be verifiably observed/perfectly monitored. However, the agent's choice of a prior may not be forced, as it is a matter of private perception. Thus, the principal's problem can be stated as fol1ows.

Problem 1 *(First-best contracting.) Choose a contract S by solving the following problem.*

$$\sup_{\substack{S \in \Psi, e \in \mathcal{U} \\ (\mu^A, \nu^A) \in \mathcal{D}}} \inf_{(\mu^P, \nu^P) \in \mathcal{D}} E^{e, v_P} \left[-\exp\left\{ -\gamma_P \left(Y_1 - S \right) \right\} \right]$$

$$s.t. \quad (i) \quad dY_t = f(e_t, \mu_t^P, \nu_t^P, t, Y)dt + \sigma(\nu_t^P, t, Y) dB_t^{e, v_P},$$

$$(ii) \quad (\mu^A, \nu^A) \in \arg \inf_{(\hat{\mu}, \hat{\nu}) \in \mathcal{D}} E^{e, \hat{v}} \left[-\exp\left\{ -\gamma_A \left(S - \int_0^1 c(e, t, Y) dt \right) \right\} \right]$$

$$s.t. \quad dY_t = f(e_t, \hat{\mu}, \hat{\nu}, t, Y)dt + \sigma(\hat{\nu}, t, Y) dB_t^{e, \hat{v}},$$

$$(iii) \quad E^{e, v_A} \left[-\exp\left\{ -\gamma_A \left(S - \int_0^1 c(e_t, t, Y) dt \right) \right\} \right] \geq -\exp\left(-\gamma_A \mathcal{W}_0 \right),$$

where $v_A = (\mu^A, \nu^A)$ and $v_P = (\mu^P, \nu^P)$. Constraint (ii) allows the agent to choose his worst pair of ambiguity parameter processes. Constraint (iii) is for his participation. As usual, we assume that if one contracting party is indifferent among many choices at optimum, then he/she chooses the one that is most favored by the other party.

Theorem 2 *(First best.) Assume that γ_A, $\gamma_P \geq 0$, but γ_A and γ_P are not simultaneously zero. Also assume that $\min_{\mu,\nu} \varphi_A(e; \mu, \nu; .)$ exists for each e, and that Assumption 1-iii holds.*[12] *Suppose that the optimizers for the principal and agent, $(e_t, \theta_t, \mu_t^P, \nu_t^P)$ and (μ_t^A, ν_t^A) lie in the interiors of their respective domains, for all $t \in [0,1]$. Then, in the first best, the worst priors of the principal and agent are symmetrized such that $(\mu_t^A, \nu_t^A) = (\mu_t^P, \nu_t^P) = (\mu_t^c, \nu_t^c) \in D_t(Y)$. Moreover, there exists a unique $\mathcal{P}^o$-q.s. square integrable process Z_t^{0P}*

[12] If $\gamma_A = \gamma_P = 0$, then one can show that there are infinitely many optimal contracts with their sensitivities β_t's to the outcome being any numbers between 0 and 1, for $t \in [0,1]$ a.e., and that the worst priors of the two parties are still equalized.

such that

$$(\mu_t^c, \nu_t^c) \in \min_{(\bar{\mu},\bar{\nu})\in D_t(Y)} f(e_t, \bar{\mu}, \bar{\nu}, t, Y) - \frac{1}{2}\frac{\gamma_A\gamma_P}{\gamma_A + \gamma_P}(1 + Z_t^{0P})\sigma^2(\bar{\nu}, t, Y), \tag{10}$$

and that the first-best optimal contract S is given as follows: $\mathcal{P}^o$-q.s.,

$$S = \mathcal{W}_0 + \int_0^1 \Big(c(e_t, t, Y) - \beta_t f(e_t, \mu_t^c, \nu_t^c, t, Y) + \frac{\gamma_A}{2}\beta_t^2\sigma^2(\nu_t^c, t, Y)\Big)\, dt + \int_0^1 \beta_t dY_t, \tag{11}$$

where $\beta_t = \frac{\gamma_P}{\gamma_A+\gamma_P}\left(1 + Z_t^{0P}\right)$, *and* $1 + Z_t^{0P} = \frac{c_e(e_t,t,Y)}{f_e(e_t,\mu_t^c,\nu_t^c,t,Y)}$.

The structure of the optimal contract (11) indicates that β_t and $1-\beta_t$ are, respectively, the agent's and principal's instantaneous nominal shares of the outcome Y. In addition to her nominal share $1-\beta_t$, the principal perceives an imaginary extra share, Z_t^{0P}, which is in fact the sensitivity of the CEQ of her expected remaining utility to the outcome process. If $Z_t^{0P} \neq 0$, then the principal effectively perceives the quantity, $1 - \beta_t + Z_t^{0P}$, as her instantaneous share of the outcome.[13] If $Z_t^{0P} > (<)0$ at optimum, the last statement of the theorem, $c_e/f_e = 1 + Z_t^{0P}$, implies that the marginal cost of effort c_e is greater (less) than the marginal product f_e. This deviation from the rule of "marginal product of labor" occurs, because current effort can affect not only current production, but future productivity. A similar process like Z_t^{0P} also appears in the second best in the next section, where we discuss in more detail.

Theorem 2 further indicates that the first-best contract, as usual, requires an outcome(-ambiguity uncertainty) sharing rule $\int_0^1 \beta_t dY_t$, but it does not require an extra sharing rule for volatility-ambiguity uncertainty. The reason is that the outcome-sharing rule alone symmetrizes overall exposures of the two parties to ambiguity uncertainties in terms of both the outcome and its volatility, with their risk aversion taken into account.

To see this, recall that the first-best contract is to optimally allocate the outcome uncertainty between the two parties. The outcome uncertainty is optimally shared when

[13]For the agent, his nominal and effective shares turn out to be optimally the same. See footnote 15.

marginal dollar-risk premia perceived by the two parties on their respective payoffs are equalized, with risk aversion taken into account. Note that the principal's and agent's effective exposures to the uncertainties, are $\gamma_A \beta_t dY_t$ and $\gamma_P(1-\beta_t+Z_t^{0P})dY_t$, respectively, which are equalized if β_t is given as in Theorem 2. With the already equalized exposures, the two parties demand the same dollar-risk premia, which implies their worst priors are symmetrized without necessity of an extra volatility-sharing rule.

Consequently, the structure of the first-best contract under ambiguity uncertainties remains the same as that of the standard first-best contract under risk uncertainties, except that commonly perceived parameter-pair process (μ_t^c, ν_t^c) has to be distinguished from true parameter-pair process $(\boldsymbol{\mu}_t, \boldsymbol{\nu}_t)$. In the second best, however, the joint ambiguity alters the well-known structure of the standard second-best contract, as is to be seen in the next section.

5 Second-Best Contracting

In the second best, the agent privately chooses not only his worst prior $P^{e,v}$ but effort process e. Taking into account his responses, the principal chooses a contract S as follows.

Problem 2 *(Second-best contracting.) Choose a contract S by solving the following problem.*

$$\sup_{\substack{S \in \Psi,\, e \in \mathcal{U} \\ (\mu^A, \nu^A) \in \mathcal{D}}} \inf_{(\mu^P,\nu^P)\in\mathcal{D}} E^{e,v_P}\left[-\exp\left\{-\gamma_P\left(Y_1 - S\right)\right\}\right]$$

$$s.t. \quad (i) \quad dY_t = f(e, \mu^P, \nu^P, t, Y)dt + \sigma(\nu^P, t, Y)dB_t^{e,v_P},$$

$$(ii) \quad (e, \mu^A, \nu^A) \in \arg\sup_{\hat{e}\in\mathcal{U}} \inf_{(\hat{\mu},\hat{\nu})\in\mathcal{D}} E^{\hat{e},\hat{v}}\left[-\exp\left\{-\gamma_A\left(S - \int_0^1 c(\hat{u}, t, Y)dt\right)\right\}\right]$$

$$s.t. \quad dY_t = f(\hat{e}, \hat{\mu}, \hat{\nu}, t, Y)dt + \sigma(\hat{\nu}, t, Y)dB_t^{\hat{e},\hat{v}},$$

$$(iii) \quad E^{e,v_A}\left[-\exp\left\{-\gamma_A\left(S - \int_0^1 c(e, t, Y)dt\right)\right\}\right] \geq -\exp\left(-\gamma_A \mathcal{W}_0\right).$$

The second constraint is for the incentive compatibility to allow him to privately choose both effort and a prior. The third is for the agent's participation. We make the usual assumption that given many indifferent choices, each party chooses one in favor of the other party. By Theorem 1, we also replace Ψ with $\bar{\Psi}$, without loss of generality.

5.1 Agent's Problem

Given a contract $S \in \bar{\Psi}$ in the form of (8) with (e^*, μ^*, ν^*), the agent may decide to choose (e, μ, ν), instead of (e^*, μ^*, ν^*). This possibility raises the well-known implementability issue. We say $S[e^*, \mu^*, \nu^*]$ is implementable if, given the same S, the agent chooses the same (e^*, μ^*, ν^*).[14] Suppose that the agent is given a salary function $S[e^*, \mu^*, \nu^*]$. Then, his problem at each time t is to optimize his expected remaining utility for unrealized future compensations as follows:

$$\sup_{e \in \mathcal{U}_t^1} \inf_{(\mu,\nu) \in \mathcal{D}_t^1} E_t^{e,v} \left[-\exp\left\{ -\gamma_A \left(\int_t^1 \left\{ -c(e_s, s, Y) - H^A(e_s^*, \mu_s^*, \nu_s^*; \beta_s, \theta_t, s, Y) \right\} ds \right.\right.\right.$$
$$\left.\left.\left. + \int_t^1 \theta_t d\langle Y_t \rangle + \int_t^1 \beta_s dY_s \right) \right\} \right],$$
$$\text{s.t.} \quad dY_s = f(e_s, \mu_s, \nu_s, s, Y)ds + \sigma(\nu_s, s, Y)dB_s^{e,v},$$

where $\mathcal{U}_t^1$ and $\mathcal{D}_t^1$ are, respectively, $\mathcal{U}$ and $\mathcal{D}$ restricted to period $[t, 1]$, and H^A is the agent's Hamiltonian (3), i.e.,

$$H^A = -c(e, t, Y) + \beta_t f(e, \mu_t, \nu_t, t, Y) + \left(\theta_t - \frac{\gamma_A}{2} \beta_t \right) \sigma^2(\nu, t, Y).$$

Theorem 3 *Suppose that Assumptions 1-i and -ii hold. Given a contract $S \in \bar{\Psi}$ with admissible $(e_t^*, (\mu_t^*, \nu_t^*); (\beta_t, \theta_t)) \in U \times D_t(Y) \times \mathcal{R}^2$, for all $t \in [0, 1]$, the agent chooses $(e_t^*, \mu_t^*, \nu_t^*)$ if and only if*

$$(e_t^*, \mu_t^*, \nu_t^*) \in \arg\max_{\hat{e}} \min_{\hat{\mu}, \hat{\nu}} H^A(\hat{e}, \hat{\mu}, \hat{\nu}; \beta_t, \theta_t, t, Y). \tag{12}$$

[14] See Schättler and Sung (1993, Definition 4.1) for the definition of implementable contracts.

That is, the contract $S \in \bar{\Psi}$ *with admissible* $(e_t^*, (\mu_t^*, \nu_t^*); (\beta_t, \theta_t)) \in U \times D_t(Y) \times \mathcal{R}^2$ *for all* $t \in [0, 1]$ *is implementable if and only if* $(e_t^*, (\mu_t^*, \nu_t^*))$ *is a saddle point of* H^A *given* (β_t, θ_t)*. Moreover, given an implementable salary function in* $\bar{\Psi}$*, the agent value function/optimal expected remaining utility* $\mathcal{V}_t$ *is constant over time: in particular,* $\mathcal{V}_t = -\exp(-\gamma_A \mathcal{W}_0)$*.*

Theorem 3 simplifies the agent's incentive compatibility condition to the problem of choosing an effort level e to maximize $-c+\beta f$, and a prior (μ, ν) to minimize $\varphi^A(e; \mu, \nu; \beta, \theta, t, Y)$. The maximization part is familiar from the existing literature, and it immediately implies the following.

Corollary 1 *If the optimal* e_t *lies in the interior* U*, then* $\beta_t = \frac{c_e(e,t,Y)}{f_e(e,\mu,\nu,t,Y)}$*.*

Theorem 3 also states that the agent's optimal expected remaining utility at each time t for unrealized future compensations is constant over time. Intuitively, the constant remaining utility is related to the well-known no-wealth effect with exponential utility. See Schättler and Sung (1997). This result turns out to be useful for intuitive understanding of the agent's dynamic decisions in general under the optimal contract. See footnote 15.

5.2 Principal's Problem

The principal is only concerned with implementable contracts in $\bar{\Psi}$, because of the following reason.

Proposition 1 *Suppose that Assumption 1-i holds. The principal does not gain by considering non-implementable contracts in* $\bar{\Psi}$*.*

Intuitively, for a nonimplementable contract in $\bar{\Psi}$, the principal can always find, at a lower cost to herself, an implementable contract which induces the agent to make same decisions on (e, μ^A, ν^A). Hence, Theorems 1 and 3, and Proposition 1 allow the principal to replace, without loss of generality, both the original class Ψ subject to the two constraints (ii) and

(iii) in Problem 2 with the subclass $\bar{\Psi}$ subject to (12). This simplification leads to the following result.

Theorem 4 *Suppose that Assumptions 1-i and -ii hold, that c is convex and f is concave in e, and that the optimal e_t lies in the interior U, for all $t \in [0,1]$. Then, there exists a unique $\mathcal{P}^o$-q.s. square integrable process Z_t^P such that the principal's optimal decision $(e_t, \mu_t^A, \nu_t^A, \mu_t^P, \nu_t^P, \theta_t)$ solves the following problem for all $t \in [0,1]$: $\mathcal{P}^o$-q.s.,*

$$\max_{\bar{e}_t, \bar{\theta}_t} \min_{(\bar{\mu}_t^P, \bar{\nu}_t^P) \in D_t(Y)} - c(\bar{e}_t, t, Y) + \varphi_P(\bar{e}_t, \bar{\theta}_t, \bar{\mu}_t^P, \bar{\nu}_t^P; Z_t^P, t, Y) + \varphi_A(\bar{\mu}_t^A, \bar{\nu}_t^A; \bar{e}_t, \bar{\beta}_t, \bar{\theta}_t, t, Y) \tag{13}$$

$$s.t. \quad \bar{\beta}_t = \frac{c_e(\bar{e}_t, t, Y)}{f_e(\bar{e}_t, \bar{\mu}_t^A, \bar{\nu}_t^A, t, Y)}, \tag{14}$$

$$(\bar{\mu}_t^A, \bar{\nu}_t^A) \in \arg \min_{(\hat{\mu}, \hat{\nu}) \in D_t(Y)} \varphi_A(\bar{e}_t, \hat{\mu}, \hat{\nu}; \bar{\theta}_t, \bar{\beta}_t, t, Y), \tag{15}$$

where (μ_t^P, ν_t^P) and (μ_t^A, ν_t^A) are, respectively, the principal's and agent's worst ambiguity parameter pairs. Moreover, the contract $S(\in \bar{\Psi})$ with $\{(e_t, \mu_t^A, \nu_t^A, \beta_t, \theta_t)\}$ is optimal.

The principal's Hamiltonian (13) represents her instantaneous CEQ gain, where the process Z_t^P results from the martingale representation theorem applied to her expected utility optimization. For ease of interpretation, let us rewrite the Hamiltonian in full as follows:

$$- c(\bar{e}_t, t, Y) + \left(1 - \bar{\beta}_t + Z_t^P\right) f(\bar{e}_t, \bar{\mu}_t^P, \bar{\nu}_t^P, t, Y) - \left[\frac{\gamma_P}{2}\left(1 - \bar{\beta}_t + Z_t^P\right)^2 + \bar{\theta}_t\right] \sigma^2(\bar{\nu}_t^P, t, Y) + \bar{\beta}_t f(\bar{e}_t, \bar{\mu}_t^A, \bar{\nu}_t^A, t, Y) - \left(\frac{\gamma_A}{2}\bar{\beta}_t^2 - \bar{\theta}_t\right) \sigma^2(\bar{\nu}_t^A, t, Y).$$

As noted in the first-best case, the process Z_t^P is the sensitivity of her CEQ to instantaneous changes in the outcome process, and thus the principal recognize Z_t^P as an imaginary extra share of the outcome. (See equation (A.16).) If $Z_t^P \geq 0$, then the marginal outcome is positively correlated with her marginal CEQ. (See Proposition 2 for an example where $Z_t^P > (<, =)\, 0$ if the asset/outcome exhibits increasing (decreasing, constant) returns to scale.) Then, the principal dynamically perceives, in effect, as if her instantaneous outcome share were $(1 - \beta_t + Z_t^P)$, rather than her nominal share $(1 - \beta_t)$, whereas the agent's effectively perceived share is the same as his nominal share, β_t.[15] Hence we call $1 - \beta_t + Z_t^P$ the principal's 'effective share' of the (instantaneous) outcome.

Next, we examine the general structure of the optimal volatility-sharing rule. The principal's choice of θ_t affects both her own and the agent's exposures to volatility uncertainty, as the structures of φ_P and φ_A suggest. Dollar-risk premia/burdens against overall uncertainty exposures at each instant are $\left[\frac{\gamma_P}{2}(1 - \beta_t + Z_t^P)^2 + \theta_t\right] \sigma^2(\nu_t^P, t, Y)$ to the principal and $\left[\frac{\gamma_A}{2}\beta_t^2 - \theta_t\right] \sigma^2(\nu_t^A, t, Y)$ to the agent. For period $(t, t + dt]$, her promise to transfer to him a realized dollar amount of $\theta_t d\langle Y_t \rangle$ at time $t + dt$ reduces his volatility-uncertainty exposure at time t. The expected time-t value of this amount is $\theta_t \sigma^2(\nu_t^P, t, Y)dt$ to the principal, and $\theta_t \sigma^2(\nu_t^A, t, Y)dt$ to the agent.

If θ_t is set to $\frac{\gamma_A}{2}\beta_t^2$, then the agent is induced to behave as if he were risk neutral, ignoring volatility uncertainties, and the principal behaves as if she alone had to take all volatility ambiguity uncertainties. If θ_t is set to $-\frac{\gamma_P}{2}(1 - \beta_t + Z_t^P)^2$, then it is now the principal that behaves as if she were risk neutral, and the agent that behaves as if he had to take them all alone. Neither case can be optimal to the principal, and she would like to strike balance between his and her ambiguity exposures by choosing θ_t. As a result, the optimal contract comprises both outcome- and volatility-sharing rules with corresponding sensitivities (β_t, θ_t) to the realized outcome and volatility, respectively.

In the following theorem, we show that the volatility-ambiguity uncertainty is optimally shared, when θ_t is set in such a way that the agent is induced to choose the same worst prior as that of the principal.

Theorem 5 *Suppose that assumptions for Theorem 4 hold, and that the optimizer*

[15] One may wonder if given S, the agent also effectively perceives his outcome share as $\beta_t + Z_t^A$ in general. This is true in general. However, if S is given in the form of (8), Theorem 3 tells us that the agent value function/expected remaining utility is constant over time, and thus $Z_t^A = 0$. That is, his nominal and effective shares of the outcome coincide with each other, at β_t. Roughly, the reason is that the agent's current incentive β_t rewards not only long- but short-term effects of his current effort on the outcome.

$(e_t, \mu_t^A, \nu_t^A, \mu_t^P, \nu_t^P, \theta_t)$ *lies in the interiors of their respective domains, for all* $t \in [0,1]$. *Then, the worst priors of the two contracting parties are symmetrized such that* $(\mu_t^A, \nu_t^A) = (\mu_t^P, \nu_t^P) = (\mu_t^c, \nu_t^c)$. *Under the symmetrized prior, there exists a unique* $\mathcal{P}^o$*-q.s. square integrable process* Z_t^P *such that the optimal outcome- and volatility-sharing sensitivities* (β_t, θ_t) *and the common prior* (μ_t^c, ν_t^c) *can be expressed as follows.*[16]

$$\beta_t = \frac{f_e + \gamma_P \beta_e (\sigma_t^c)^2}{f_e + (\gamma_P + \gamma_A)\beta_e(\sigma_t^c)^2}(1 + Z_t^P), \tag{16}$$

$$\theta_t = \frac{1}{2(1+Z_t^P)}\beta_t(1-\beta_t+Z_t^P)\left(\gamma_A\beta_t - \gamma_P(1-\beta_t+Z_t^P)\right) \geq 0, \tag{17}$$

where $\beta_e = \frac{\partial}{\partial e}\left(\frac{c_e(e_t,t,Y)}{f_e(e_t,\mu_t^c,\nu_t^c,t,Y)}\right)$, *and*

$$(\mu_t^c, \nu_t^c) \in \arg\min_{(\hat{\mu},\hat{\nu})\in D_t(Y)} (1+Z_t^P) f(e_t, \hat{\mu}_t, \hat{\nu}_t, t, Y_t) - \frac{1}{2}\left(\gamma_A \beta_t^2 + \gamma_P(1-\beta_t+Z_t^P)^2\right)\sigma^2(\hat{\nu}, t, Y_t). \tag{18}$$

Moreover, $0 < \beta_t \leq 1 + Z_t^P$.

Remark: If $\gamma_A = 0$, then $\beta_t = 1 + Z_t^P$ and $\theta_t = 0$; and if $\gamma_A > 0$, then $0 < \frac{\gamma_P}{\gamma_A+\gamma_P}(1+Z_t^P) < \beta_t < 1 + Z_t^P$ and $\theta_t > 0$.

The outcome-sharing sensitivity β_t in the form (16) distinguishes itself from that of the Holmstrom-Milgrom (1987) stationary case, in the following two aspects: (i) σ_t^c is endogenously determined through (18), and (ii) the multiplier $(1 + Z_t^P)$ is required for dynamic consideration. If marginal changes of the principal's future expected utility and those of the outcome are positively (negatively) correlated such that $Z_t^P > (<)0$, then the optimal outcome-sharing rule provides the agent with a higher-powered (lower-powered) work incentives than the one predicted by the stationary case where $Z_t^P = 0$. For example, see Proposition 2.

[16] If there were no mean ambiguity, then one can show that a volatility sharing rule is optimal as long as θ_t lies in the interval $[-\frac{\gamma_P}{2}(1-\beta_t+Z_t^P)^2, \frac{\gamma_A}{2}\beta^2]$. That is, θ_t can optimally be set at zero, and thus the presence of a volatility sharing rule is not necessary for optimal contracting. This observation is consistent with the statement in footnote 5, "Disagreement on the worst prior does not arise in contracting (even without a volatility sharing rule), if either the mean alone or the volatility alone of the outcome is ambiguous."

It is striking that the optimal volatility sharing rule symmetrizes the worst priors of the two contracting parties. The intuition for this result is as follows. Given an outcome-sharing rule, the equally ambiguity-averse principal and agent could perceive the most pessimistic prior over mean-volatility ambiguity uncertainties differently from each other, because of differences in payoff structure and risk aversion. Differential priors for the volatility would then result in differential risk premia even on identically uncertain payoffs. If the principal perceives a lower (higher) risk premium under her worst prior than the agent does under his worst prior, she would like to improve the contract by lowering (raising) his risk burden until the risk premia individually perceived by themselves are equalized. She achieves the equalization by choosing θ_t as in (17), under which their minimands φ_P and φ_A differ only by a scale factor, and thus their worst priors are equalized, and so are their risk premia.[17]

Consequently, the optimal volatility-sharing rule in general stipulates that the two contracting parties share the total risk premium or the sum of their individually perceived risk premia, in proportion to their effective outcome shares $(\beta_t, 1-\beta_t+Z_t^P)$. Recall that perceived marginal expected-dollar payoffs from an instantaneous change in the outcome are $\beta_t f dt$ to the agent, and $(1-\beta_t+Z_t^P) f dt$ to the principal. Also, recall that risk premia perceived by the principal and agent are $RP^P := \left[\frac{\gamma_P}{2}(1-\beta_t+Z_t^P)^2+\theta_t\right]\sigma^2(\nu_t^P, t, Y)$ and $RP^A := \left[\frac{\gamma_A}{2}\beta_t^2-\theta_t\right]\sigma^2(\nu_t^A, t, Y)$, respectively. The optimal volatility-sharing rule occurs as the principal's and agent's perceived risk premia per their expected dollar payoffs are equalized, i.e., $RP^P/((1-\beta_t+Z_t^P)f) = RP^A/(\beta_t f)$.

Mains results from the principal problem, thus far, can be summarized as follows: (i) the optimal contract consists of two sharing rules, one for the usual outcome sharing and the other for volatility sharing; (ii) the worst priors are symmetrized, and the volatility-sharing rule equalizes perceived risk premia per expected dollar payoff across the two contracting

[17] Given that the main concern in optimal contracting is to align interests of the two contracting parties, one can conjecture that the symmetrization result is generalizable to other types of ambiguity and utility functions. This issue remains to be of interest for future research.

parties under the common worst prior; (iii) the principal dynamically perceives, in effect, her outcome share as $(1-\beta_t+Z_t^P)$, rather than $(1-\beta_t)$; and (iv) the first-best volatility-sharing sensitivity is zero, whereas that for the second best is strictly positive, for $\gamma_A>0$ and $\gamma_P\geq 0$.

Even with the unspecified process Z_t^P, the above analysis has produced most of economic insights into effects of joint-ambiguity uncertainties on our principal-agent problem. In the next section, we provide an example where one can explicitly solve for Z_t^P with the well-known dynamics programming method.

6 A Linear-Quadratic Case

We consider a special case of the linear-quadratic Markovian moral hazard model examined in Schättler and Sung (1997), and extend the case by introducing joint ambiguity to it.[18] Assume that $c(e)=(\kappa/2)e^2$ for $e,\kappa>0$, and that the gross profit of the firm is driven by the following dynamics:

$$dY_t=(\eta Y_t+e_t+\mu_t)dt+\nu_t dB_t^{u,v},\quad Y_0\in\mathcal{R},$$

where $\eta\in\mathcal{R}$, and $\nu>0$. If $\eta>(<)\,0$ exhibiting increasing (decreasing) returns to scale, then the greater the outcome, the higher (lower) the production efficiency.

Let $0<\underline{\mu}<\overline{\mu}$ and $0<\underline{\nu}<\overline{\nu}$ and $\alpha>0$. Also assume that the mean-volatility joint ambiguity parameter set D is given by

$$D=\left\{(\mu,\nu)\in[\underline{\mu},\overline{\mu}]\times[\underline{\nu},\overline{\nu}]\;\middle|\;\pi=\mu-\frac{\alpha}{2}(\nu-\nu^0)^2\geq 0\right\}.$$

This constraint set is a slightly modified version of 'the quadratic ambiguity' introduced by Epstein and Schneider (2010) and Epstein and Ji (2013): the smaller the coefficient α,

[18] In order to compute Z_t^P for non-Markovian cases, one may consider the dynamic programming method by Cvitanić, Possamaï and Touzi (2015). The process Z_t^P can also be viewed as a component of the solution to a second-order backward stochastic differential equation (2BSDE): see Nutz (2012b) and STZ (2012).

the greater the degree of mean-volatility joint ambiguity. Thus, one may regard $1/\alpha$ as a measure of the degree of joint ambiguity uncertainties.

By Theorems 1 and 3, and Corollary 1, it is without loss of generality that the principal only considers contracts S in the following form:

$$S = \mathcal{W}_0 + \int_0^1 \left\{ \frac{\kappa}{2} e_t^2 - \kappa e_t(\eta Y_t + e_t) - \min_{(\hat{\mu},\hat{\nu})\in D_t} \left[\kappa e_t \hat{\mu}_t^A - \left(\frac{\gamma_A}{2}\kappa^2 e_t^2 - \theta_t \right) (\hat{\nu}_t^A)^2 \right] \right\} dt$$
$$+ \int_0^1 \theta_t d\langle Y_t \rangle + \int_0^1 \kappa e_t dY_t.$$

Let $(\mu_t^A, \nu_t^A) \in \arg\min_{(\hat{\mu},\hat{\nu})\in D_t} \kappa e_t \hat{\mu}_t^A - \left(\frac{\gamma_A}{2}\kappa^2 e_t^2 - \theta_t\right)(\hat{\nu}_t^A)^2$. Then, the principal's utility is

$$\sup_{e,\mu^A,\nu^A} \inf_{\mu^P,\nu^P} \quad \hat{E}^{e,\mu^P,\nu^P} \left[-\exp\left\{ -\gamma^P \left(\begin{array}{l} Y_1 - S + \int_0^1 \frac{1}{2\gamma_P} \frac{(\eta Y_t + e_t + \mu_t^P)^2}{(\nu_t^P)^2} dt \\ - \int_0^1 \frac{1}{\gamma_P} \frac{\eta Y_t + e_t + \mu_t^P}{(\nu_t^P)^2} dY_t \end{array} \right) \right\} \right]$$
$$\text{s.t.} \quad dY_t = \nu_t^P dW_t, \quad Y_0 \in \mathcal{R},$$

where W_t is a $\mathcal{P}^o$-universal standard BM. (See the Appendix.) Define the principal's expected remaining utility at time t as follows.

$$J(e,\theta,\mu^P,\nu^P;t,Y_t)$$
$$= \hat{E}_t^{e,\mu^P,\nu^P} \left[-\exp\left\{ -\gamma_P \left(\begin{array}{l} Y_0 - \mathcal{W}_0 \\ + \int_t^1 \left\{ \begin{array}{l} \frac{1}{2\gamma_P} \frac{(\eta Y_s + e_s + \mu_s^P)^2}{(\nu_s^P)^2} - \frac{\kappa}{2} e_s^2 + \kappa e_s(\eta Y_s + e_s) \\ + \kappa e_s \mu_s^A - \left(\frac{\gamma_A}{2}\kappa^2 e_s^2 - \theta_s\right)(\nu_s^A)^2 - \theta_s(\nu_s^P)^2 \end{array} \right\} ds \\ + \int_t^1 \left\{ 1 - \kappa e_s - \frac{1}{\gamma_P} \frac{\eta Y_s + e_s + \mu_s^P}{(\nu_s^P)^2} \right\} \nu_s^P dW_s \end{array} \right) \right\} \right].$$

The principal optimal choice of $(e,\theta;\mu^P,\nu^P)$ becomes a dynamic programming problem under a Markovian environment. Let

$$\mathcal{V}_t := \operatorname{ess\,sup}_{e,\theta} \inf_{\mu^P,\nu^P} J(e,\theta,\mu^P,\nu^P;t,Y_t)$$
$$\text{s.t.} \quad (\mu_t^A, \nu_t^A) \in \arg\min_{(\hat{\mu},\hat{\nu})\in D_t} \kappa e_t \hat{\mu}_t^A - \left(\frac{\gamma_A}{2}\kappa^2 e_t^2 - \theta_t \right) (\hat{\nu}_t^A)^2.$$

Then, one can utilize the Hamilton-Jacobi-Bellman-Isaac (HJBI) equation, $\mathcal{P}^o$-quasi surely, in order to find $\mathcal{V}_t$, which is the well-known HJB equation with the Isaac condition for a saddle point.[19]

Define

$$\Upsilon(\nu_t) \equiv \left\{1 + (1 + R_t)\,\gamma_P \kappa \nu_t^2\right\} \frac{\gamma_A \exp(\eta(1-t))}{\alpha R_t^2} + \frac{\nu^0}{\nu_t} - 1, \tag{19}$$

where $R_t := 1 + (\gamma_A + \gamma_P)\kappa\nu_t^2$. The function Υ arises as the agent faces a tradeoff between the mean and volatility of the ambiguity uncertainties, while choosing his worst prior.

Proposition 2 *Suppose that $\nu^0 > \underline{\nu} > 0$, and that $\Upsilon(\overline{\nu}) > 0$, $\Upsilon(\underline{\nu}) < 0$: i.e., $\underline{\nu}$ and $\overline{\nu}$ are sufficiently small and large, respectively, enough to ensure an interior optimum in ν. Then, the common worst prior (μ_t^c, ν_t^c) at optimum is a unique solution to the following equations,*

$$\mu_t^c = \frac{\alpha}{2}(\nu_t^c - \nu^0)^2; \tag{20}$$

$$\Upsilon(\nu_t^c) = 0. \tag{21}$$

Moreover, the value function $\mathcal{V}_t$ is given by $\mathcal{V}(t, Y_t) = -e^{-\gamma_P(\zeta(t)Y_t + \rho(t))}$, $\mathcal{P}^o$-q.s., where

$$\zeta(t) = e^{\eta(1-t)} - 1 = Z_t^P, \tag{22}$$

$$\rho(t) = Y_0 - \mathcal{W}_0 + \int_t^1 \left[e^{\eta(1-s)}\mu_s^c + e^{2\eta(1-s)} \frac{1 + \kappa\gamma_P(\nu_s^c)^2\left(1 - \kappa\gamma_A(\nu_s^c)^2\right)}{2\kappa\left(1 + \kappa(\gamma_P + \gamma_A)(\nu_s^c)^2\right)} \right] ds. \tag{23}$$

The sharing sensitivities (β_t, θ_t) of the optimal contract are:

$$\beta_t = \frac{1 + \gamma_P\kappa(\nu_t^c)^2}{1 + (\gamma_A + \gamma_P)\kappa(\nu_t^c)^2} e^{\eta(1-t)}, \tag{24}$$

$$\theta_t = \frac{1}{2e^{\eta(1-t)}}\beta_t \left(e^{\eta(1-t)} - \beta_t\right)\left[\gamma_A\beta_t - \left(e^{\eta(1-t)} - \beta_t\right)\gamma_P\right]. \tag{25}$$

Remark: Equations (20) and (21) imply that the common worst prior depends on time t, but not on state Y_t.

[19] Note that $\mathcal{V}_t$ is called 'the lower value function' in the literature on two-person zero-sum games. However, because of our saddle point assumption, its HJB becomes the same as the HJBI which assumes the lower and upper value functions to be equalized. See Pham and Zhang (2014) for the justification of the HJBI equation with controlled volatilities. Our HJBI equation is of the same form as that of the HJB in Schättler and Sung (1993,1995), except that, roughly speaking, volatilities are treated as controllable in this paper, but not controllable in that paper. Mastrolia and Possamaï (2015) also use a similar HJBI to ours with $\eta = 0$ and $\mu_t = 0$.

Note that if $\eta = 0$, then we have a joint-ambiguity version of the Holmstrom-Milgrom stationary case: (20), (21), (24) and (25) imply that the common prior (μ_t^c, ν_t^c) and (β_t, θ_t) are constant over time, and thus the optimal contract is linear in Y_1 and $\langle Y_1 \rangle$. Intuitively and not surprisingly, the linearity of the optimal contract is due to the stationarity of the principal's problem resulting from the time-state independence of $(c_t, f_t, \sigma_t, D_t(Y))$.[20] Because of this stationarity, the volatility of the principal's CEQ process turns out to be zero, i.e., $Z_t^P = 0$ over time, which is consistent with Corollary A.1. See Chen and Sung (2018) for more economic implications.

Finally, straightforward computations produce the following comparative statics.

Corollary 2 *(Comparative statics.)*

i. *Both the commonly perceived mean and volatility increase with the returns-to-scale parameter η, the degree of ambiguity $1/\alpha$, and the agent's ability $1/\kappa$. That is, $\frac{\partial \mu_t^c}{\partial \eta}, \frac{\partial \nu_t^c}{\partial \eta} > 0$, and $\frac{\partial \mu_t^c}{\partial \alpha}, \frac{\partial \nu_t^c}{\partial \alpha}, \frac{\partial \mu_t^c}{\partial \kappa}, \frac{\partial \nu_t^c}{\partial \kappa} < 0$.*

ii. *The outcome-sharing sensitivity increases with the agent's ability, but decreases with the degree of ambiguity: i.e., $\frac{\partial \beta_t}{\partial \kappa} < 0$, and $\frac{\partial \beta_t}{\partial \alpha} > 0$. Moreover, if the principal is risk neutral, i.e., $\gamma_P = 0$, then the sensitivity also increases with the returns-to-scale parameter: i.e., $\frac{\partial \beta_t}{\partial \eta} > 0$.*

Note that both η and $1/\kappa$ are production-efficiency parameters. Intuitively, the agent's high ability in production motivates the principal to provide more incentives to the agent with a high outcome-sharing sensitivity: $\frac{\partial \beta_t}{\partial \kappa} < 0$. The same conclusion holds with respect to η, if the principal is risk neutral: i.e., $\frac{\partial \beta_t}{\partial \eta} > 0$. If the principal is highly risk averse, this inequality can be reversed, because her high risk-risk aversion can cause the common volatility perception to increase with η too fast.

[20] See Schättler and Sung (1997) or Hellwig and Schmidt (2002) for the well-known intuition together with a rigorous discrete-time justification of Holmstrom-Milgrom's linearity result.

Also, intuitively, an increase in the degree of ambiguity, $1/\alpha$, induces both the principal and agent to perceive more pessimistically about ambiguity uncertainties. As a result, both parties commonly perceive a higher volatility, i.e., $\frac{\partial \nu_t^c}{\partial \alpha} < 0$. Thus, we also have: the higher the degree of ambiguity, the higher the perceived mean, i.e., $\frac{\partial \mu_t^c}{\partial \alpha} < 0$. Given the quadratic ambiguity constraint D, a higher volatility perception leads him to perceive a higher mean along the boundary of the constraint. Moreover, as the agent become more pessimistic, his perceived volatility increases, negatively affecting incentives for the agent to work. Thus we have $\frac{\partial \beta_t}{\partial \alpha} > 0$.

For $\frac{\partial \nu_t^c}{\partial \eta} > 0$ and $\frac{\partial \nu_t^c}{\partial \kappa} < 0$, it is somewhat striking that given ambiguity uncertainties, a marginal increase in production efficiency induces the two contracting parties to perceive a higher volatility. The reason is that a marginal increase in efficiency marginally increases the outcome-sharing sensitivity and thus the ambiguity uncertainty exposures to both parties, leading to greater volatility perceptions.

7 Conclusion

We have examined effects of ambiguity uncertainties about the mean and volatility of the outcome on optimal contracting. Unlike the existing literature, we distinguish between ex-post realized and ex-ante perceived volatilities.

In order to investigate our contracting problem under ambiguity, we have extended the well-known martingale method by allowing volatility controls. Developed for exponential utility, this method can be easily modified for other utilities such as additively separable utilities. See, for instance, Schättler and Sung (1993, footnote 11). Although it can be challenging to obtain general closed-form solutions, we believe that the CARA case in this paper can serve as a benchmark in conjecturing economics of solutions, if any, to other general utility cases. See, for instance, Baker and Hall (2004, p775) who use the CARA case to infer economics of managerial compensations for managers with constant-relative-risk-aversion (CRRA) utilities.

Applying the method to the contracting problem, we have shown that the second-best contract in the presence of the joint ambiguity uncertainties contains two sharing rules, one for outcome sharing and the other for volatility sharing, and that the realized compensation is positively associated with the realized volatility. The positive association can be consistent with frequent executive compensation practices of granting stock options, as stock option values are sensitive to changes in the realized volatility of the underlying asset.

However, the first-best contract does not require a volatility-sharing arrangement. Recall that the first-best contract is about uncertainty-sharing, whereas the second-best contract is about not only uncertainty-sharing but incentives. In the first best, under the outcome-sharing rule alone, the outcome ambiguity-uncertainties are already optimally shared between the two parties. No separate side contract can further improve the outcome-uncertainty sharing. On the other hand, in the second best, the outcome-sharing rule should take into consideration effort incentives, as well as uncertainty sharing. For effort incentives, the outcome sharing sensitivity has to be increased beyond that of the first best. The increase can however upset the balance in uncertainty exposures of the two parties, inducing differential risk perceptions about ambiguity uncertainties. One can easily see that differential perceptions, if any, can cause inefficiency in contracting like information asymmetry problems (without learning) can. We have argued that the inefficiency can be mitigated by the volatility-sharing side contract, as it induces the symmetrization of their perceptions.

A Appendix

A.1 Admissible Probability Measures

Recall that the outcome process Y is the coordinate process. We consider a family of probability measures, under which the process evolves like Itô processes with different drift and diffusion rates. It is well known that two probability measures generating different volatilities/diffusion rates cannot be absolutely continuous with respect to each other. Thus, needs for singular measure changes arise.

We first construct a class $\mathcal{P}^o$ of (fully or partially) singular Wiener measures. For this, we consider sub-sample spaces Ω^σ's of Ω, each of which consists of all sample paths in Ω with equal quadratic variations over all time $t \in [0,1]$, and for each σ, we define a Wiener measure on Ω using Ω^σ as the support of the measure. Then, we expand $\mathcal{P}^o$ to $\mathcal{P}$ by adding probability measures which are absolutely continuous with respect to each measure in $\mathcal{P}^o$. All probability measures including singular measures are defined on the universal filtration $\{\hat{\mathcal{F}}_t\}$ which is described in Section 2. We shall precisely define Ω^σ, $\mathcal{P}^o$ and $\mathcal{P}$ as we proceed.

Let us first recall the QV process,

$$\langle Y_t \rangle (\omega) = \lim_{\Delta t \to 0} \sum_{k=1}^{n} \left| Y_{t_{k+1}}(\omega) - Y_{t_k}(\omega) \right|^2,$$

where $0 < t_1 < t_2 < ... < t_n = t$ and Δt is the mesh size of the partition. If $d\langle Y_t \rangle = \sigma_t^2 dt$, then σ_t is the volatility/diffusion rate of Y_t.

For the general method to be presented in this Appendix, we redefine U and D as subsets of $\mathcal{R}^n$ and $\mathcal{R}^m$, respectively, and $\mathcal{U}$ and $\mathcal{D}$ as the classes of $\hat{\mathcal{F}}^{\mathcal{P}}$-progressively measurable processes such that $u(\in \mathcal{U}) : [0,1] \times \Omega \to U(\subset \mathcal{R}^n)$, and $v(\in \mathcal{D}) : [0,1] \times \Omega \to D(\subset \mathcal{R}^m)$.

Assumption A.1 *The class* $\mathbf{\Sigma}$ *consists of diffusion coefficients,* $\sigma : U \times D \times [0,1] \times \Omega \longrightarrow \mathcal{R}_+$*, which are uniformly bounded,* $\hat{\mathcal{F}}^{\mathcal{P}}$*-progressively measurable, and strictly positive real-valued functionals, satisfying the uniform Lipschitz condition: there exists a constant* K

such that for all $t \in [0,1]$, $(u,v) \in \mathcal{U} \times \mathcal{D}$, *and* $Y, \bar{Y} \in \Omega$,

$$|\sigma(u(t,Y), v(t,Y), t, Y) - \sigma(u(t,\bar{Y}), v(t,\bar{Y}), t, \bar{Y})| \leq K \sup_{0 \leq s \leq t} |Y_s - \bar{Y}_s|.$$

Let Ω^σ be a subspace of the original sample space Ω such that, for each admissible $\sigma(\in \mathbf{\Sigma})$ and each $t \in (0,1]$,

$$\Omega_t^\sigma := \left\{ \omega \in \Omega_t \mid \langle \omega_\tau \rangle = \langle \omega_t \rangle + \int_t^\tau \sigma_s^2 ds, \, \forall \tau \in (t,1] \right\}. \tag{A.1}$$

Then, Ω_t^σ consists of all sample paths, each of which has a QV density of σ_s^2 for all $s \in (t,1]$, conditional on ω_t.[21] Note that Ω_t^σ and $\Omega_t^{\sigma'}$ are disjoint subsets of Ω_t, if $\sigma_s \neq \sigma'_s$, for all $s \in (t,1]$, and Ω_t is a super set of the uncountable union of Ω_t^σ's, for $\sigma \in \mathbf{\Sigma}$. We write Ω_0^σ as Ω^σ for brevity.

In what follows, we let $\sigma(t,Y) \equiv \sigma(u(t,Y), v(t,Y), t, Y)$, again for brevity. Then, $\sigma^2(t,Y)$ is the QV density of the coordinate process $Y_t(\omega)(\equiv \omega_t)$, i.e., $\langle Y_t \rangle = \int_0^t \sigma^2(s,Y)ds$. For $\sigma \in \mathbf{\Sigma}$, we define process W_t^σ as follows:

$$W_t^\sigma \equiv \int_0^t \frac{1}{\sigma(s,Y)} dY_s. \tag{A.2}$$

Note that the QV density of $W_t^\sigma(\omega)$ is always equal to one for all $\omega \in \Omega^\sigma$. Then one can construct a Wiener measure P^σ under which W_t^σ becomes a standard BM, as follows.

Lemma A.1 *For each* $\sigma \in \mathbf{\Sigma}$, *there exists a probability measure* P^σ *under which* W_t^σ *is a standard BM, and* Y_t *is a martingale with its QV density being* σ_t^2, P^σ*-a.s..*

Proof: Let P^1 be the classical Wiener measure on Ω such that $P^1(\Omega^1) = 1$ and that the coordinate process Y_t becomes a standard BM under $(P^1, \hat{\mathcal{F}}, \Omega)$. Choose $\sigma \in \mathbf{\Sigma}$. Then, there exists P^σ on Ω such that $P^\sigma(\Omega^\sigma) = 1$, and $\int_0^t \frac{1}{\sigma_s} dY_s$ is a standard BM under $(P^\sigma, \hat{\mathcal{F}}, \Omega)$. The reason is that for each $\omega^1 \in \Omega^1$ one can find $\omega^\sigma \in \Omega^\sigma$ such that $\omega_t^\sigma \equiv \int_0^t \sigma_s d\omega_s^1$ for

[21] This view is also consistent with STZ (2011a, Section 8).

all t. The converse is true as well. That is, given $\{\omega_s : 0 \leq s \leq t\}$, Ω_t^1 and Ω_t^σ are in one-to-one correspondence for each $t \in [0,1]$. In fact, $\Omega_t^1 \equiv \{\omega \in \Omega \,|\, \omega_\tau = \int_t^\tau \frac{1}{\sigma_s} d\omega_s^\sigma, \forall \tau \in (t,1], \omega^\sigma \in \Omega_t^\sigma\}$, and $\Omega_t^\sigma \equiv \{\omega^\sigma \in \Omega \,|\, \omega_\tau^\sigma = \int_t^\tau \sigma_s d\omega_s^1, \forall \tau \in (t,1], \forall \omega^1 \in \Omega_t^1\}$. For all $A \in \hat{\mathcal{F}}$, set $P^\sigma(A^\sigma) = P^1(A)$, where $A^\sigma = \left\{\omega^\sigma \,|\, \omega_t^\sigma = \int_0^t \sigma_s d\omega_s, \forall t \in [0,1], \omega \in A\right\}$. Then, P^σ defined on Ω with $P^\sigma(\Omega^\sigma) = 1$ becomes the Wiener measure under which the process $W_t^\sigma (= \int_0^t \frac{1}{\sigma_s} dY_s)$ is a standard BM. Thus, the law of $(\{Y_t\}, P^1, \Omega)$ is identical to that of $(\{W_t^\sigma\}, P^\sigma, \Omega)$, except for their null sets. □

Remark If $\sigma_s \neq \sigma_s'$ for all $s \in (t,1]$, then conditional probability measures P_t^σ and $P_t^{\sigma'}$ are mutually singular, because Ω_t^σ and $\Omega_t^{\sigma'}$ are disjoint subsets of the original conditional state space Ω_t, and $P_t^\sigma(\Omega_t^{\sigma'}) = P_t^{\sigma'}(\Omega_t^\sigma) = 0$. For example, if $\sigma_s = 1$ and $\sigma_s' = 2$, for all $s \in [0,1]$, then P_t^σ and $P_t^{\sigma'}$ are mutually singular for all $t \in [0,1]$. □

Thus, Ω^σ is the support of the Wiener measure P^σ, under which W_t^σ is a standard BM on Ω. Note that W_t^σ is defined on the whole Ω, not just on Ω^σ. The same is true for all other stochastic processes and random variables in this paper. Let $\mathcal{P}^o$ be a class of singular Wiener measures such that $\mathcal{P}^o := \{P^\sigma; \sigma \in \mathbf{\Sigma}\}$. Then $\{W^\sigma, P^\sigma \in \mathcal{P}^o\}$ is a family of standard BM's.

In fact, one can construct a $\mathcal{P}^o$-universal standard BM. Let $\boldsymbol{\sigma}_t^2(\omega)dt \equiv d\langle Y_t(\omega)\rangle, \forall \omega \in \Omega$, i.e., $\boldsymbol{\sigma}_t^2(\omega)$ is the QV density of ω. Also let, for $\boldsymbol{\sigma}_s > 0$, $s, t \in [0,1]$,

$$W_t := \int_0^t \frac{1}{\boldsymbol{\sigma}_s(\omega)} dY_s(\omega).$$

Then W_t is a $\mathcal{P}^o$-standard BM, i.e., $W_t = W_t^\sigma$ a.s., under each $P^\sigma \in \mathcal{P}^o$. In particular, W_t shares identical sample paths with W_t^σ on Ω^σ, i.e., $W_t(\omega) = W_t^\sigma(\omega)$, and $\boldsymbol{\sigma}_t(\omega) = \sigma(t, Y)$ on Ω^σ. Hence, W_t is a $\mathcal{P}^o$-universal standard BM, or a $\mathcal{P}^o$-*aggregator* of the family $\{W^\sigma, P^\sigma \in \mathcal{P}^o\}$.

Next, we expand $\mathcal{P}^o$ to $\mathcal{P}$ by adding a collection of probability measures that are absolutely continuous with respect to each $P^\sigma \in \mathcal{P}^o$. For this, we introduce another class Φ of $\hat{\mathcal{F}}^{\mathcal{P}^o}$-progressively measurable functionals.

Assumption A.2 *The class Φ is the collection of functionals, $f : U \times D \times [0,1] \times \Omega \to \mathcal{R}$, with the following properties: for each $(u,v) \in \mathcal{U} \times \mathcal{D}$ and $\sigma(u,v,t,Y) \in \mathbf{\Sigma}$, $f(u,v,t,Y)$ is $\hat{\mathcal{F}}^{\mathcal{P}^o}$-progressively measurable, $\int_0^t \frac{f^2(u,v,t,Y)}{\sigma^2(u,v,t,Y)}ds < \infty$ path by path for all $t \in [0,1]$, and there exists a constant K such that*

$$\frac{f(u,v,t,Y)}{\sigma(u,v,t,Y)} \leq K\left(1 + \max_{0\leq s\leq t}|Y_s|\right), \quad \forall(u,v,t,Y).$$

Let

$$\vartheta_t^{\sigma,f} := \exp\left(\int_0^t \frac{f(u,v,s,Y)}{\sigma(u,v,s,Y)}dW_s - \frac{1}{2}\int_0^t \frac{f^2(u,v,s,Y)}{\sigma^2(u,v,s,Y)}ds\right). \tag{A.3}$$

Then, this is a P^σ-exponential martingale for $\sigma \in \mathbf{\Sigma}$. We define a new probability measure $P^{\sigma,f}$ by $dP^{\sigma,f} = \vartheta_1^{\sigma,f}dP^\sigma$ on $\hat{\mathcal{F}}_1$. That is, $P^{\sigma,f}$ is absolutely continuous with respect to P^σ. If $f \equiv 0$, then $P^{\sigma,f} = P^\sigma \in \mathcal{P}^o$. In general, by the Girsanov theorem,

$$B_t^{\sigma,f} := W_t - \int_0^t \frac{f_s}{\sigma_s}ds$$

is a standard BM under $P^{\sigma,f}$ for each $(\sigma, f) \in \mathbf{\Sigma} \times \Phi$. Since $dY_t(\omega) = \boldsymbol{\sigma}_t(\omega)dW_t$ under $\mathcal{P}^o$, and $\boldsymbol{\sigma}_t(\omega) = \sigma_t$ under $P^\sigma \in \mathcal{P}^o$, we have, under each P^σ,

$$dY_t = \boldsymbol{\sigma}_t(\omega)dW_t = \sigma_t dW_t \tag{A.4}$$

Then, Assumption (A.1) implies that there exists a unique strong solution to this stochastic differential equation (A.4).[22] (See Elliott (1982, Theorem 14.6)). Moreover, there is a unique $\mathcal{P}^o$-quasi surely aggregated strong solution. See STZ (2011a) and Nutz (2012), for the aggregation result.

[22]The classical Peng's G-Brownian Motion can be interpreted as the coordinate process Y of Ω under a subfamily of $\mathcal{P}^o$.

Using (A.4), we expand $\mathcal{P}^o$ to the family $\mathcal{P}$ of admissible probability measures on Ω as follows:

$$\mathcal{P} := \left\{ P^{\sigma,f} \,|\, dP^{\sigma,f} = \vartheta_1^{\sigma,f} dP^{\sigma}, f \in \Phi,\, P^{\sigma} \in \mathcal{P}^o \right\}. \tag{A.5}$$

For brevity, we let

$$P^{\sigma,f} \equiv P^{\sigma(u,v,.),f(u,v,.)} \equiv P^{u,v}.$$

There can be multiple (u,v)'s corresponding to $P^{\sigma,f}$ such that $\sigma(u,v,.) = \sigma(u',v',.)$ and $f(u,v,.) = f(u',v',.)$. Then, we simply have $P^{u,v} = P^{u',v'}$, without affecting our results. Under $P^{u,v} \in \mathcal{P}$, we rewrite the dynamics of process Y in (A.4) as follows:

$$dY_t = f(u,v,t,Y)dt + \sigma(u,v,t,Y)dB_t^{u,v}. \tag{A.6}$$

The general method in this Appendix is based on the outcome dynamics in (A.6). For our agency problem in the text, however, we consider a special case under the following simplifying assumptions: $n = 1$, $m = 2$, $u = e$, $v = (\mu, \nu)$ and σ is independent of (e, μ). Then we can write the dynamics of the outcome as in (1), and $P^{u,v} = P^{e,v} = P^{e,(\mu,\nu)}$.

A.2 The mean-volatility control problem

In this section, we extend the martingale method of Schättler and Sung (1993), by allowing the agent to control both the mean and volatility of the outcome process. We derive a Hamiltonian for our general mean-volatility control problem in weak formulation.

Given the family $\mathcal{P}$ of singular and absolutely continuous probability measures as defined in (A.5), we work on a general mean-volatility control problem of which our agency problem is a special case. We assume the classes $\mathcal{U}$ and $\mathcal{D}$ of admissible control and ambiguity parameter processes $\{(u,v)\}$ are constrained by the time-state dependent compact and convex sets $U_t(Y)(\subset U)$ and $D_t(Y)(\subset D)$ with nonempty interiors for each t, such that $u_t \in U_t(Y)$ and $v_t \in D_t(Y)$. Moreover, we let $\mathcal{U}_t^1$ and $\mathcal{D}_t^1$, respectively, be the sets of admissible u's and v's restricted by $U_t(Y)$ and $D_t(Y)$ for time period $[t,1]$. Then, $\mathcal{U} \equiv \mathcal{U}_0^1$ and $\mathcal{D} \equiv \mathcal{D}_0^1$. We consider the following general problem:

$$\sup_{u\in\mathcal{U}} \inf_{v\in\mathcal{D}} \quad E^{u,v}\left[-\exp\left\{-\gamma\left(\xi(Y)+\int_0^1 g(.)ds+\int_0^1 q(.)d\langle Y_s\rangle+\int_0^1 h(.)dY_s\right)\right\}\right] \tag{A.7}$$
$$\text{s.t.} \quad dY_t = f(u,v,t,Y)dt+\sigma(u,v,t,Y)dB_t^{u,v},$$

where $B_t^{u,v}(= W_t - \int_0^t \frac{f_s}{\sigma_s}ds)$ is the standard BM under $P^{u,v}$. We assume $\gamma > 0$ to derive all results of this section, and interpret the limit as the risk-neutral agent case.[23] Let g, q, h, f, and σ be real-valued functionals such that

$$g,q,h,f : U \times D \times [0,1] \times \Omega \to \mathcal{R},$$
$$\sigma : U \times D \times [0,1] \times \Omega \to \mathcal{R}_+.$$

The general problem (A.7) covers our principal's and agent's problems as special cases. Let

$$\hat{\vartheta}_1 = \frac{dP^{u,v}}{d\hat{P}^{u,v}} = \exp\left(\int_0^1 \frac{f}{\sigma^2}dY_t - \frac{1}{2}\int_0^1 \frac{f^2}{\sigma_s^2}ds\right). \tag{A.8}$$

Then, the problem can be equivalently written as

$$\sup_{u\in\mathcal{U}} \inf_{v\in\mathcal{D}} \quad \hat{E}^{u,v}\left[-\exp\left\{-\gamma\xi(Y)-\int_0^1 \hat{G}_s ds-\int_0^1 \Gamma_s dY_s\right\}\right] \tag{A.9}$$
$$\text{s.t.} \quad dY_t = \sigma(u,v,t,Y)dW_t,$$

where

$$\hat{G}_t = \gamma g(.) + \gamma q(.)\sigma^2 + \frac{1}{2}\frac{f^2}{\sigma^2},$$
$$\Gamma_t = \gamma h(.) - \frac{f}{\sigma^2}.$$

Hence, the original general problem is transformed into a pure volatility control problem. By the construction of $\mathcal{P}^o$ and $\mathcal{P}$, $\hat{P}^{u,v} \in \mathcal{P}^o$ if and only if $P^{u,v} \in \mathcal{P}$ where $dP^{u,v} \equiv \hat{\vartheta}_1 d\hat{P}^{u,v}$.

[23] That is, all results in this paper with $\gamma = 0$ hold for the risk-neutral case, which can be shown by repeating basically the same reasonings as presented in this section.

Let

$$\phi_t(u,v,Y) := -\exp\left\{-\gamma\xi(Y) - \int_t^1 \hat{G}(u,v,s,Y)ds - \int_t^1 \Gamma(u,v,s,Y)dY_s\right\}, \tag{A.10}$$

We define the CEQ wealth $\mathcal{Q}$ and value functions $(V, \mathcal{V})$ as follows:

$$\mathcal{Q}_t^{u,v} := -\frac{1}{\gamma}\ln\left(-\hat{E}_t^{u,v}\left[\phi_t(u,v,Y)\right]\right), \tag{A.11}$$

$$V_t(u,v^*(u)) = \operatorname*{ess\,inf}_{v\in\mathcal{D}_t^1} -\exp\left(-\gamma\mathcal{Q}_t^{u,v}\right), \tag{A.12}$$

$$\mathcal{V}_t = \operatorname*{ess\,sup}_{u\in\mathcal{U}_t^1} V_t(u,v^*(u)). \tag{A.13}$$

Let H^o be the Hamiltonian given as follows: for $(u_t, v_t, p, t, Y) \in U_t(Y) \times D_t(Y) \times \mathcal{R} \times [0,1] \times \Omega$,

$$H^o(u_t,v_t,p,t,Y) := pK(u_t,v_t,t,Y) + G(u_t,v_t,t,Y) - \frac{\gamma}{2}\left(p\sigma(u_t,v_t,t,Y)\right)^2, \tag{A.14}$$

where

$$G := g + hf + \left[q - \frac{\gamma}{2}h^2\right]\sigma^2,$$
$$K := f - \gamma h\sigma^2.$$

Assumption A.3 *For each (p,t,Y), there exists a saddle point $(u_t^*, v_t^*) \in U_t(Y) \times D_t(Y)$ such that, for all $u_t \in U_t(Y)$ and $v_t \in D_t(Y)$,*

$$H^o(u_t,v_t^*,p,t,Y) \le H^o(u_t^*,v_t^*,p,t,Y) \le H^o(u_t^*,v_t,p,t,Y),$$

and that $(u^*, v^*) \in \mathcal{U} \times \mathcal{D}$.

We note that the existence of saddle-point value functions in general is a challenging mathematical issue, particularly for diffusion control problems. See for example Pham and Zhang (2014). In this paper, we just content ourselves with providing closed-form solutions to some specific cases as in Proposition 2, without looking into general conditions for the existence.

In the following lemma, we characterize the value function process $\mathcal{V}_t$ and thus the optimal CEQ process $\mathcal{Q}_t^*(\equiv \mathcal{Q}_t^{u^*,v^*})$.

Lemma A.2 *Let* $\phi_0(u,v,Y) \in \mathbb{L}^1_{\mathcal{P}}$*, and also let Assumptions A.1, through A.3 hold. Then, there exist a unique* $\mathcal{P}^o$*-q.s. square integrable process* Z^*_t *and a saddle point process* (u^*, v^*) *such that*

$$(u^*_t, v^*_t) \in \arg\max_{u \in U_t(Y)} \min_{v \in D_t(Y)} H^o(u, v, Z^*_t, t, Y), \tag{A.15}$$

$$d\mathcal{Q}^*_t = -H^o(u^*_t, v^*_t, Z^*_t, t, Y)dt + \frac{\gamma}{2}(Z^*_t)^2 \left(d\langle Y_t \rangle - (\sigma^*_t)^2 dt\right) + Z^*_t dY_t, \quad \mathcal{Q}^*_1 = \xi(Y), \tag{A.16}$$

and $\mathcal{V}_t = -\exp(-\gamma \mathcal{Q}^*_t) = ess\sup_{u \in \mathcal{U}^1_t} \inf_{v \in \mathcal{D}^1_t} [-\exp(-\gamma \mathcal{Q}^{u,v}_t)]$.

Remark: Under $\tilde{P}^{u^*,v^*}$, $d\mathcal{Q}^*_t = -H^o(u^*_t, v^*_t, Z^*_t, t, Y)dt + Z^*_t dY_t, \quad \mathcal{Q}^*_1 = \xi(Y)$.

Proof: For admissible (u,v) and (u',v'), let

$$\varphi^{u,v}_t(u',v') := \hat{E}^{u',v'}\left[-\exp\left\{-\gamma\xi(Y) - \int_0^1 \hat{G}(u,v,s,Y)ds - \int_0^1 \Gamma(u,v,s,Y)dY_s\right\} \Big| \hat{\mathcal{F}}_t\right].$$

Note that $\varphi^{u,v}_t(u',v')$ is a martingale under $\hat{P}^{u'v'}$ and $\{\varphi^{u,v}_t(u'v') \,|\, (u',v') \in \mathcal{U} \times \mathcal{D}\}$ is a family of martingales under different measures. Then, there exist a unique $\mathcal{P}^o$-q.s. $\hat{\mathcal{F}}$-progressively measurable aggregating process $\varphi^{u,v}_t$ and a unique $\mathcal{P}^o$-q.s. square integrable process $\hat{Z}^{u,v}_t$ such that under each $\hat{P}^{u',v'}(\in \mathcal{P})$, $\varphi^{u,v}_t = \varphi^{u,v}_t(u',v')$, and

$$\varphi^{u,v}_t = -\exp\left\{-\gamma\xi(Y) - \int_0^1 \hat{G}(.)ds - \int_0^1 \Gamma(.)dY_s\right\} + \gamma \int_t^1 \varphi^{u,v}_s \hat{Z}^{u,v}_s dY_s \tag{A.17}$$

is a $(\hat{\mathcal{F}}, P^{u'v'})$-martingale, where the suppressed arguments are (u_s, v_s, s, Y). (See Nutz (2012a) or STZ (2011) for the aggregation.) That is, $\varphi^{u,v}_t$ is a $\mathcal{P}^o$-q.s. martingale with the following dynamics:

$$d\varphi^{u,v}_t = -\gamma \varphi^{u,v}_t \hat{Z}^{u,v}_t dY_t, \ \mathcal{P}^o\text{-q.s.} \tag{A.18}$$

On the other hand, let us define a process $\mathcal{Q}^{u,v}_t$, $\mathcal{P}^o$-q.s. through the following equation:

$$\varphi^{u,v}_t = -\exp\left\{-\gamma \mathcal{Q}^{u,v}_t - \int_0^t \hat{G}(.)ds - \int_0^t \Gamma(.)dY_s\right\}.$$

(Note that $\mathcal{Q}_t^{u,v}$ is well defined under all $P \in \mathcal{P}$, and that it is the CEQ wealth process under $P^{u,v}$.) Then, the Itô formula implies, $\mathcal{P}^o$-q.s.,

$$d\varphi_t^{u,v} = \varphi_t^{u,v} \left\{ -\gamma d\mathcal{Q}_t^{u,v} - \hat{G}(.)dt - \Gamma(.)dY_t \right\} + \frac{1}{2}\varphi_t^{u,v} d\left\langle \gamma\mathcal{Q}_t^{u,v} + \int_0^t \Gamma_s dY_s \right\rangle. \quad \text{(A.19)}$$

Equating (A.18) and (A.19), we have, $\mathcal{P}^o$-q.s.,

$$\gamma d\mathcal{Q}_t^{u,v} = -\hat{G}(.)dt + \frac{\gamma^2}{2}(\hat{Z}_t^{u,v})^2 d\langle Y_t\rangle + \left(\gamma\hat{Z}_t^{u,v} - \Gamma(.)\right) dY_t, \quad \mathcal{Q}_1^{u,v} = \xi(Y).$$

Let $\gamma Z_t^{u,v} := \gamma\hat{Z}_t^{u,v} - \Gamma(.)$. Then

$$\gamma d\mathcal{Q}_t^{u,v} = -\hat{G}(.)dt + \frac{1}{2}\left(\gamma Z_t^{u,v} + \Gamma(.)\right)^2 d\langle Y_t\rangle + \gamma Z_t^{u,v} dY_t, \quad \mathcal{Q}_1^{u,v} = \xi(Y). \quad \text{(A.20)}$$

Now, let $v^*(u)$ be the worst parameter process given u such that

$$v_t^*(u) \in \arg\operatorname{ess}\inf_{v\in\mathcal{D}} - \exp\left(-\gamma\mathcal{Q}_t(u,v)\right).$$

Consider an admissible (concatenated) pair process (u_s, v_s) with the following structure: for $t, \tau \in [0,1]$,

$$(u,v) = \begin{cases} (u_s, v_s^*(u)) & \text{for } \tau < s \le 1, \\ (u_s, v_s) & \text{for } t \le s \le \tau. \end{cases}$$

Let $\mathcal{Q}_\tau^*(u) \equiv \mathcal{Q}_\tau^{u,v^*(u)}$, and

$$J_\tau^{u,v} := -\exp\left\{ -\gamma\mathcal{Q}_\tau^*(u) - \int_t^\tau \hat{G}(u_s, v_s, .)ds - \int_t^\tau \Gamma(u_s, v_s, .)dY_s \right\}. \quad \text{(A.21)}$$

If $\tau = t$, then $J_t^{u,v} = -\exp\left\{-\gamma\mathcal{Q}_t^*(u)\right\} = V_t(u, v^*(u))$. For $t \le s \le \tau$, the dynamic programming principle implies that [24]

$$\begin{aligned} \hat{E}_s^{u,v}[J_\tau^{u,v}] &= \exp\left\{ -\int_t^s \hat{G}_w(.)dw - \int_t^s \Gamma_w dY_w \right\} \\ &\qquad \times \hat{E}_s^{u,v}\left[-\exp\left\{ -\gamma\mathcal{Q}_\tau^*(u) - \int_s^\tau \hat{G}_w(.)dw - \int_s^\tau \Gamma_w dY_w \right\} \right] \\ &\ge J_s^{u,v}. \end{aligned}$$

[24]For the dynamic programming principle with controlled diffusions, see, for example, Nutz (2012b, Theorem 5.2), STZ (2012, Proposition 5.14), Bouchard and Nutz (2012, Theorems 4.2 and 4.6), Nutz and Soner (2012, Theorem 4.10), and Pham and Zhang (2014, Theorem 4.6, Proposition 7.3). For that with controlled drifts, see, for instance, Rishel (1970), Davis (1973), Davis and Veraiya (1973, Theorem 3.1), and Davis (1979).

That is, $J_\tau^{u.v}$ is a $(\hat{\mathcal{F}}, \hat{P}^{u,v})$-submartingale on $[t,1]$. Then the Doob-Meyer decomposition implies that there exist a unique square integrable process $M_t^{u,v}$ and a unique increasing process $A_t^{u,v}$ under $\hat{P}^{u,v}$ such that

$$J_\tau^{u,v} = \int_t^\tau M_s^{u,v} dY_s + A_\tau^{u,v},$$

where $\hat{E}_t^{u,v}[\int_t^1 (M_s^{u,v}\sigma_s^{u,v})^2\, ds] < \infty$.

On the other hand, (A.21) implies, $\mathcal{P}^o$-q.s.,

$$dJ_\tau^{u,v} = J_\tau^{u,v}\left\{-\gamma d\mathcal{Q}_\tau^*(u) - \hat{G}(.)d\tau - \Gamma_\tau dY_\tau\right\} + \frac{\gamma}{2} J_\tau^{u,v} d\left\langle \gamma\mathcal{Q}_\tau^*(u) + \int_t^\tau \Gamma_s dY_s \right\rangle,$$

and (A.20) implies that under $P^{u,v^*(u)}$ a.s., the CEQ process with $(u, v^*(u))$ evolves as follows;

$$\gamma d\mathcal{Q}_t^*(u) = -\hat{G}_t^*(u)dt + \frac{1}{2}\left(\gamma Z_t^*(u) + \Gamma_t^*(u)\right)^2 (\sigma_t^*)^2 dt + \gamma Z_t^*(u) dY_t. \tag{A.22}$$

where $\hat{G}_t^*(u)$, $Z_t^*(u)$ and $\Gamma_t^*(u)$ are short for $\hat{G}(u_t, v_t^*(u), t, Y)$, $Z_t^{u,v^*(u)}$ and $\Gamma(u_t, v_t^*(u), t, Y)$, respectively. Recall that under $\hat{P}^{u,v^*(u)}$, $d\langle Y_t\rangle = (\sigma_t^*(u))^2 dt \equiv (\sigma_t^*(u, v^*(u), t, Y))^2 dt$. Also, recall that (u, v) is the concatenated pair satisfying $(u_s, v_s) \equiv (u_s, v_s^*(u))$ for $\tau < s \leq 1$. Then, with the superscript (u, v) on J_τ suppressed, under $\hat{P}^{u,v}$,

$$\begin{aligned} dJ_\tau &= J_\tau \left\{ \hat{G}_\tau^*(u) - \frac{1}{2}\left(\gamma Z_\tau^*(u) + \Gamma_\tau^*(u)\right)^2 (\sigma_\tau^*(u))^2 - \hat{G}(u, v, .) + \frac{1}{2}\left(\gamma Z_\tau^*(u) + \Gamma\right)^2 \sigma_\tau^2 \right\} d\tau \\ &\quad - J_\tau \left(\gamma Z_\tau^*(u) + \Gamma_\tau\right) dY_\tau \\ &= \gamma J_\tau \left\{ H^o(u_\tau, v_\tau^*(u), Z_\tau^*(u), \tau, Y) - H^o(u_\tau, v_\tau, Z_\tau^*(u), \tau, Y) \right\} d\tau - J_\tau \left(\gamma Z_\tau^*(u) + \Gamma_\tau\right) dY_\tau, \end{aligned}$$

where $H^o(u, v, Z, t, Y)$ is as defined in (A.14). However, since J_τ is a $(\hat{\mathcal{F}}, \tilde{P}^{u,v})$-submartingale, the uniqueness of the Doob-Meyer decomposition implies $-J_\tau\left(\gamma Z_\tau^*(u) + \Gamma_\tau\right) = M_\tau^{u,v}$ on $\Omega^{\sigma(u,v)}$, and $\hat{E}_t^{u,v}[\int_t^1 \left(J_s\left(\gamma Z_s^*(u) + \Gamma_\tau\right)\sigma_s\right)^2 ds] < \infty$. That is, the stochastic integral is a $(\hat{\mathcal{F}}, \hat{P}^{u,v})$-martingale, and for $t, \tau \in [0,1]$,

$$-\exp\left(-\gamma\mathcal{Q}_t^{u,v}\right) = -\exp\left(-\gamma\mathcal{Q}_t^*(u)\right) + \hat{E}_t^{u,v}\left[\int_t^\tau \gamma J_s \Delta H^o(u_s, v_s, Z_s^*(u), s, Y) ds\right], \tag{A.23}$$

where

$$\Delta H^o(u_s, v_s, Z_s^*(u), s, Y) := H^o(u_s, v_s^*(u), Z_s^*(u), s, Y) - H^o(u_s, v_s, Z_s^*(u), s, Y).$$

Note that $\mathcal{Q}_t^{u,v}$ in (A.23) is CEQ at time t when arbitrary admissible (u, v) during period $[t, \tau]$, and $(u, v^*(u))$ afterwards are to be chosen. Then, since $J_s < 0$, the application of the optimality principle to (A.23) yields the following inequality: given each $u \in \mathcal{U}$, for all $v \in \mathcal{D}$, and $s \in [0, 1]$,

$$H^o(u_s, v_s, Z_s^*(u), s, Y) \geq H^o(u_s, v_s^*(u), Z_s^*(u), s, Y). \tag{A.24}$$

Next, let u^* be such that

$$u_t^* \in \arg \operatorname*{ess\,sup}_u - \exp\left(-\gamma \mathcal{Q}_t^*(u)\right),$$

and consider an admissible pair (u, v) with the following structure: for $t, \tau \in [0, 1]$,

$$(u, v) = \begin{cases} (u_s^*, v_s^*) & \text{for } \tau < s \leq 1, \\ (u_s, v_s^*(u)) & \text{for } t \leq s \leq \tau, \end{cases}$$

where v^* denotes $v^*(u^*)$. Let $\mathcal{Q}_\tau^* = \mathcal{Q}_\tau^*(u^*)$, and

$$J_\tau^{u,v} := -\exp\left\{-\gamma \mathcal{Q}_\tau^* - \int_t^\tau \hat{G}(u_s, v_s^*(u), .)ds - \int_t^\tau \Gamma(u_s, v_s^*(u), .)dY_s\right\}.$$

Then by using a similar argument to the previous case, one can show that $J_\tau^{u,v}$ is a $(\hat{\mathcal{F}}, \hat{P}^{u,v})$-supermartingale on $[t, 1]$, and that u^* satisfies, for all $u \in \mathcal{U}$ and $s \in [0, 1]$,

$$H^o(u_s, v_s^*(u), Z_s^*, s, Y) \leq H^o(u_s^*, v_s^*, Z_s^*, s, Y), \tag{A.25}$$

where $Z_s^* \equiv Z_s^*(u^*)$. The inequalities (A.24) and (A.25) imply that the optima pair (u^*, v^*) at each $s \in [0, 1]$, satisfies the following inequalities: for all $(u, v) \in \mathcal{U} \times \mathcal{D}$, i.e., $\mathcal{P}^o$-q.s.,

$$H^o(u_s, v_s^*(u), Z_s^*, s, Y) \leq H^o(u_s^*, v_s^*, Z_s^*, s, Y) \leq H^o(u_s^*, v_s, Z_s^*, s, Y).$$

In other words, (u_s^*, v_s^*) is a saddle point of H^o, confirming (A.15). Then (A.22) directly implies (A.16). □

The following corollary turns out to be convenient to ravel out the principal's problem when the decision environment is stationary as in Holmstrom and Milgrom (1987).

Corollary A.1 *Suppose that $\xi(Y) = k$, where k is a constant, and that D_t, $G(.)$, f, and σ are all state-independent such that $\pi(v,t)$, $G(u,v,t)$, $f(u,v,t)$, and $\sigma(u,v,t)$. Also suppose that G is a saddle function with respect to maximizing in u and minimizing in v. Let $(\hat{u}_s, \hat{v}_s) \in \arg\max_{u\in U} \min_{v\in D} G(u,v,s)$ for $s \in [0,1]$. Then, $(\hat{u}_s, \hat{v}_s) = (u_s^*, v_s^*)$, $Z_t^* \equiv 0$, $t \in [0,1]$, and $H^o(u_t^*, v_t^*, Z^*, t, Y) = G(u_t^*, v_t^*, t,)$.*

Proof: Let

$$-e^{-\gamma Q^{u,v}} = \hat{E}_t^{u,v}\left[-\exp\left\{-\gamma\left(k + \int_t^1 \hat{G}(u_s, v_s, s)ds + \int_t^1 \Gamma(u_s, v_s, s)dY_s\right)\right\}\right].$$

Then, we have

$$\begin{aligned}-e^{-\gamma Q^{u,v}} &= \tilde{E}^{u,v}\left[-\exp\left\{-\gamma\left(k + \int_t^1 \left(\hat{G}(u_s, v_s, s) - \frac{\gamma}{2}\Gamma^2(u_s, v_s, s)\sigma^2(u_s, v_s, s)\right)ds\right)\right\}\right] \\ &= \tilde{E}^{u,v}\left[-\exp\left\{-\gamma\left(k + \int_t^1 G(u_s, v_s, s)ds\right)\right\}\right],\end{aligned}$$

where

$$\frac{d\hat{P}^{u,v}}{d\tilde{P}^{u,v}} = \exp\left(-\gamma\int_t^1 \Gamma(u_s, v_s, s)dY_s - \frac{\gamma^2}{2}\int_t^1 \Gamma^2(u_s, v_s, s)\sigma^2(u_s, v_s,)ds\right).$$

Since $(\hat{u}, \hat{v})$ is a pair of state-independent processes, we have

$$-e^{-\gamma Q^{\hat{u},\hat{v}}} = -\exp\left\{-\gamma\left(k + \int_t^1 G(\hat{u}_s, \hat{v}_s, s)ds\right)\right\},$$

which implies $Q^{\hat{u},\hat{v}}$ is deterministic. Moreover, since G is a saddle function, given each (t, Y), for all (u, v), a.s.,

$$\begin{aligned}-e^{-\gamma Q^{\hat{u},v}} &\geq \tilde{E}_t^{\hat{u},v}\left[-\exp\left\{-\gamma\left(k + \int_t^1 G(\hat{u}_s, \hat{v}_s, s)ds\right)\right\}\right] \\ &= -e^{-\gamma Q^{\hat{u},\hat{v}}} = \tilde{E}_t^{u,\hat{v}}\left[-\exp\left\{-\gamma\left(k + \int_t^1 G(\hat{u}_s, \hat{v}_s, s)ds\right)\right\}\right] \geq -e^{-\gamma Q^{u,\hat{v}}}.\end{aligned}$$

Thus for all (t, Y) and (u, v), $(\hat{u}, \hat{v})$ is optimal and $-e^{-\gamma Q^{\hat{u},\hat{v}}}$ is the value function, $\mathcal{V}_t$, i.e., $(\hat{u}, \hat{v}) = (u^*, v^*)$. Since $\mathcal{V}_t$ is deterministic, $Z_t^{u^*,v^*} \equiv 0$, $t \in [0,1]$, and by (A.15), $G = H^o$. □

A.3 Proof of Theorem 1

Given an admissible contract $S \in \Psi$, suppose that agent arbitrary chooses (e, μ, ν). Then, the agent's expected utility is

$$E^{u,v}\left[-\exp\left\{-\gamma_A\left(S(Y) - \int_0^1 c(e_t, t, Y)ds\right)\right\}\right]$$
$$\text{s.t.} \quad dY_t = f(e_t, \mu_t, \nu, t, Y)dt + \sigma(\nu_t, t, Y)dB_t^{e,v}.$$

This quantity can be equivalently rewritten as follows:

$$\hat{E}^{e,v}\left[-\exp\left\{-\gamma_A\left(S(Y) - \int_0^1 \left(c(e_t, t, Y) - \frac{1}{2\gamma_A}\frac{f^2}{\sigma^2}\right)dt - \frac{1}{\gamma_A}\int_0^1 \frac{f}{\sigma^2}dY_t\right)\right\}\right]$$
$$\text{s.t.} \quad dY_t = \sigma(\nu_t, t, Y)dW_t,$$

where W_t is a $\mathcal{P}^o$-universal standard BM, and $\frac{d\hat{P}^{e,v}}{dP^{e,v}} = \exp\left(\frac{1}{2}\int_0^1 \frac{f^2}{\sigma^2}dt - \int_0^1 \frac{f}{\sigma^2}dY_t\right)$. Let

$$\varphi_t^{e',v';e,v} := \hat{E}_t^{e',v'}\left[-\exp\left\{-\gamma_A\left(S(Y) - \int_0^1 \left(c(e_t, t, Y) - \frac{1}{2\gamma_A}\frac{f^2}{\sigma^2}\right)dt - \frac{1}{\gamma_A}\int_0^1 \frac{f}{\sigma^2}dY_t\right)\right\}\right].$$

Then $\varphi_t^{e,v}(e', v')$ is a $\hat{P}^{e',v'}$-martingale, and thus there exists a unique $\hat{Z}^{e',v';e,v}$. Moreover $\{\varphi_t^{e,v}(e', v') \,|\, (e', v') \in \mathcal{U}\times\mathcal{D}\}$ is a family of martingales under different probability measures. Then, by Nutz (2012) (or STZ (2011)), there exist unique $\mathcal{P}^o$-q.s. $\hat{\mathcal{F}}_t$-progressively measurable and aggregating, process $\varphi_t^{e,v}$ and square integrable process $\hat{Z}_t^{e,v}$ such that under $\hat{P}^{e',v'} \in \mathcal{P}$, $\varphi_t^{e,v} = \varphi_t^{e,v}(e', v')$, and $\hat{Z}_t^{e,v,} = \hat{Z}_t^{e,v:e',v'}$, a.s.. In particular, $\mathcal{P}^o$-q.s.,

$$\varphi_t^{e,v} = -\exp\left\{-\gamma_A\left(S(Y) - \int_0^1 \left(c(e_s, s, Y) - \frac{1}{2\gamma_A}\frac{f^2}{\sigma^2}\right)ds - \frac{1}{\gamma_A}\int_0^1 \frac{f}{\sigma^2}dY_s\right)\right\}$$
$$+ \gamma_A \int_t^1 \varphi_s^{e,v}\hat{Z}_s^{e,v}dY_s.$$

Hence, we have

$$d\varphi_t^{e,v} = -\gamma_A \varphi_t^{e,v}\hat{Z}_t^{e,v}dY_t, \tag{A.26}$$

which is the same as Eq.(A.18), the general case.

On the other hand, given $\varphi_t^{e,v}$, $\mathcal{P}^o$-q.s., we define a process $Q_t^{e,v}$, $\mathcal{P}^o$-q.s. through the following equation:

$$\varphi_t^{e,v} \equiv -\exp\left\{-\gamma_A\left(Q_t^{e,v} - \int_0^t \left(c(e,s,Y) - \frac{1}{2\gamma_A}\frac{f^2}{\sigma^2}\right)ds - \frac{1}{\gamma_A}\int_0^t \frac{f}{\sigma^2}dY_s\right)\right\}.$$

(Note that $Q_1^{e,v} \equiv S(Y)$, $\mathcal{P}^o$-q.s., and that, under $P^{e,v}$, $Q_t^{e,v}$ is the CEQ of the agent's remaining expected utility at t.) Then, by the Itô formula, $\mathcal{P}^o$-q.s.,

$$d\varphi_t^{e,v} = \varphi_t^{e,v}\left\{-\gamma_A dQ_t^{e,v} + \gamma_A\left(c(e,t,Y) - \frac{1}{2\gamma_A}\frac{f^2}{\sigma^2}\right)dt + \frac{f}{\sigma^2}dY_t\right\}$$
$$+\frac{1}{2}\varphi^{e,v}d\left\langle \gamma_A Q_t^{e,v} - \int_0^t \frac{f}{\sigma^2}dY_s\right\rangle.$$

This equation, together with Eq.(A.26) implies that the CEQ process dynamics are given by, $\mathcal{P}^o$-q.s.,

$$dQ_t^{e,v} = \left(c(e,t,Y) - \frac{1}{2\gamma_A}\frac{f^2}{\sigma^2}\right)dt + \left(\frac{1}{\gamma_A}\frac{f}{\sigma^2} + \hat{Z}_t^{e,v}\right)dY_t + \frac{\gamma_A}{2}\left(\hat{Z}_t^{e,v}\right)^2\langle dY_t\rangle.$$

Since $Q_1^{u,v} \equiv S(Y)$, $\mathcal{P}^o$-q.s.,

$$S(Y) = \mathcal{W}_0^{S,u,v} + \int_0^1 \left(c(e,t,Y) - \frac{1}{2\gamma_A}\frac{f^2}{\sigma^2}\right)dt + \int_0^1 \left(\frac{1}{\gamma_A}\frac{f}{\sigma^2} + \hat{Z}_t^{e,v}\right)dY_t$$
$$+\frac{\gamma_A}{2}\int_0^1 \left(\hat{Z}_t^{e,v}\right)^2 d\langle Y_t\rangle.$$

Then, by letting

$$\beta_t := \frac{1}{\gamma_A}\frac{f}{\sigma^2} + \hat{Z}_t^{e,v}, \quad \theta_t := \frac{\gamma_A}{2}\left(\hat{Z}_t^{e,v}\right)^2,$$

we have (8). □

A.4 Proof of Theorem 2

By Theorem 1, it is without loss of generality for the principal to consider only the subclass $\bar{\Psi}$ defined in (9) where each $S \in \bar{\Psi}$ is given in the following form: $\mathcal{P}^o$-q.s.,

$$S = \mathcal{W}_0 + \int_0^1 \left\{c(e_t,t,Y) - \beta_t f(e_t,\mu_t,\nu_t,t,Y) + \left[\frac{\gamma_A}{2}\beta_t^2 - \theta_t\right]\sigma^2(\nu_t,t,Y)\right\}dt$$
$$+\int_0^1 \theta_t d\langle Y_t\rangle + \int_0^1 \beta_t dY_t.$$

Recall that even in the first best, the agent is allowed to choose privately his worst prior. Thus, there can arise the implementability issue about his worst prior choice. (We discuss this issue in Section 5.1.) Theorem 3 implies that given effort process e_t, each implementable contract $S \in \bar{\Psi}$ has to satisfy the following condition for the agent's choice of the worst prior:

$$(\mu_t^A, \nu_t^A) \in \arg \min_{(\mu,\nu)\in D} \varphi_A(e_t; \mu, \nu; \beta_t, \theta_t; t, Y) = \beta_t f + \left(\theta_t - \frac{\gamma_A}{2}\beta_t^2\right)\sigma^2. \tag{A.27}$$

Thus, the principal's problem can be rewritten as

$$\sup_{e,\theta,\beta} \inf_{(\mu^P,\nu^P)\in\mathcal{D}} E^{e,v}\left[-\exp\left\{-\gamma_P\left(Y_1 - S\right)\right\}\right],$$

where $v = (\mu^P, \nu^P)$ and, $\mathcal{P}^o$-q.s.,

$$\begin{aligned} Y_1 - S = Y_0 - \mathcal{W}_0 + \int_0^1 \left(\theta_t(\sigma_t^A)^2 + \beta_t f_t^A - c(e_t, t, Y) - \frac{\gamma_A}{2}\beta_t^2(\sigma_t^A)^2\right) dt \\ - \int_0^1 \theta_t d\langle Y_t\rangle + \int_0^1 (1-\beta_t)\, dY_t, \end{aligned}$$

where $f_t^A = f(e_t, \mu_t^A, \nu_t^A, , t, Y)$, and $\sigma_t^A = \sigma(\nu_t^A, t, Y)$, with (μ^A, ν^A) being subject to (A.27). Then, by Lemma A.2, there exists a unique $\mathcal{P}^o$-q.s. square integrable process Z_t^{0P} such that the principal's optimal solution $(e_t, \theta_t, \beta_t, \mu_t^P, \nu_t^P, \mu_t^A, \nu_t^A)$ results from optimizing the following Hamiltonian:

$$\begin{aligned} \max_{\bar{e},\bar{\beta},\bar{\theta}} \min_{(\bar{\mu},\bar{\nu})\in D} H_t^P &= Z_t^{0P}(f - \gamma_P(1-\bar{\beta})\bar{\sigma}_t^2) + (1-\bar{\beta})f - c(\bar{e}, t, Y) \\ &\quad - \left[\bar{\theta} + \frac{\gamma_P}{2}\left(1-\bar{\beta}\right)^2\right]\bar{\sigma}_t^2 - \frac{\gamma_P}{2}(Z_t^{0P})^2\bar{\sigma}_t^2 + \min_{(\hat{\mu},\hat{\nu})\in D} \varphi_A(\hat{\mu}, \hat{\nu}; \bar{e}, \bar{\beta}, \bar{\theta}; t, Y). \end{aligned}$$

From this, note that given $(\bar{e}, \bar{\beta}, \bar{\theta})$, the principal's worst parameter pair (μ_t^P, ν_t^P) is determined independently of the agent's. Thus, her problem can be rewritten as

$$\begin{aligned} \max_{\bar{e},\bar{\beta},\bar{\theta}} \quad \hat{H}_t^P = -c(\bar{e}, t, Y) + \min_{(\hat{\mu}^P,\hat{\nu}^P)\in D} \varphi_P(\hat{\mu}^P, \hat{\nu}^P; \bar{e}, \bar{\beta}, \bar{\theta}; t, Y, Z_t^{0P}) \\ + \min_{(\hat{\mu}^A,\hat{\nu}^A)\in D} \varphi_A(\hat{\mu}^A, \hat{\nu}^A; \bar{e}, \bar{\beta}, \bar{\theta}; t, Y). \end{aligned}$$

where φ_P is defined in (6). Since both φ_A and φ_P are continuously differentiable and since $D_t(Y)$ satisfies the KKT constraint qualification conditions, the FOCs with respect to (μ, ν) become necessary conditions for the principal's and agent's minimization problems, and thus one can apply the Envelope Theorem to obtain FOCs with respect to (e, b, β) as follows.

$$(1 - \beta_t + Z_t^{0P}) \frac{\partial f_t^P}{\partial e} + \beta_t \frac{\partial f_t^A}{\partial e} - c_e = 0, \tag{A.28}$$

$$(f_t^A - f_t^P) - \gamma_A \beta_t (\sigma_t^A)^2 + \gamma_P (1 - \beta_t + Z_t^{0P})(\sigma_t^P)^2 = 0, \tag{A.29}$$

$$(\sigma_t^A)^2 - (\sigma_t^P)^2 = 0, \tag{A.30}$$

where $f_t^A = f(e_t, \mu_t^A, \nu_t^A, t, Y)$ and $f_t^P = f(e_t, \mu_t^P, \nu_t^P, t, Y)$. By FOC (A.30), $\sigma_t^A = \sigma_t^P = \sigma_t^c$ at optimum, where $\sigma_t^A = \sigma(\nu_t^A(\beta_t, \theta_t), t, Y) \equiv \sigma^A(\beta_t, \theta_t, t, Y)$ and $\sigma_t^P = \sigma(\nu_t^P(\beta_t, \theta_t), t, Y) \equiv \sigma^P(\beta_t, \theta_t, t, Y)$. Since σ is increasing in ν by assumption, we have $\nu_t^A(\beta_t, \theta_t) = \nu_t^P(\beta_t, \theta_t) = \nu_t^c$, i.e., volatility perceptions are symmetrized across the two contracting parties, under their own worst priors.

With volatility perceptions symmetrized at ν_t^c, the principal's decisions on (b, β) are affected by γ_A, γ_P, and potential perceptional difference $(f_t^A - f_t^P)$ between the two parties. In order to understand the difference, given ν_t^c, note that the principal and agent choose their perceptions (μ_t^P, ν_t^c) and (μ_t^A, ν_t^c), respectively as follows:

$$(\mu_t^P, \nu_t^c) \in \arg \min_{(\mu,\nu) \in D} \varphi_P(\mu, \nu; e_t, \beta_t, \theta_t, t, Y, Z_t^{0P}),$$

$$(\mu_t^A, \nu_t^c) \in \arg \min_{(\mu,\nu) \in D} \varphi_A(\mu, \nu; e_t, \beta_t, \theta_t, t, Y).$$

That is, given (β_t, θ_t) and the state variable Z_t^{0P},

$$\mu_t^P \in \arg \min_{\mu \in D(\nu_t^c)} (1 - \beta_t + Z_t^{0P}) f(e, \mu, \nu_t^c, t, Y),$$

$$\mu_t^A \in \arg \min_{\mu \in D(\nu_t^c)} \beta_t f(e, \mu, \nu_t^c, t, Y),$$

where $D(\nu_t^c) := \{\mu \in \mathcal{R} \,|\, \pi(\mu, \nu_t^c, t, Y) \geq 0\}$. Let $f_t^A = f(e_t, \mu_t^A, \nu_t^c, , t, Y)$ and $f_t^P = f(e_t, \mu_t^P, \nu_t^c, , t, Y)$. Then, since $f_\mu > 0$, one can see that if $1 - \beta_t + Z_t^{0P} > 0$, then $\mu_t^P =$

$\min_{D(\nu_t^c)} \mu(\nu_t^c)$, and if $1 - \beta_t + Z_t^{0P} < 0$, then $\mu_t^P = \max_{D(\nu_t^c)} \mu(\nu_t^c)$; and if $\beta_t > 0$, then $\mu_t^A = \min_{D(\nu_t^c)} \mu(\nu_t^c)$, and if $\beta_t < 0$, then $\mu_t^A = \max_{D(\nu_t^c)} \mu(\nu_t^c)$.

Since $\gamma_A, \gamma_P \geq 0$ and they are not simultaneously zero, we must have $0 \leq \beta_t \leq 1 + Z_t^{0P}$, because of the following reasons.

- If $\beta_t < 0$ and $1 + Z_t^{0P} < \beta_t$, condition (A.28) is contradicted.
- If $\beta_t < 0$ and $1 + Z_t^{0P} > \beta_t$, then $\mu_t^P = \min \mu(\nu_t^c)$. Thus $f_t^A - f_t^P \geq 0$ contradicting FOC (A.29), because γ_A and γ_P are not simultaneously zero.
- If $\beta_t > 0$ and $1 + Z_t^{0P} < \beta_t$, then $\mu_t^A = \min \mu(\nu_t^c)$. Thus $f_t^A - f_t^P \leq 0$ contradicting FOC (A.29), again because γ_A and γ_P are not simultaneously zero.
- If $0 < \beta_t < 1 + Z_t^{0P}$, then $\mu_t^A = \mu_t^P = \min \mu(\nu_c)$.
- If $\beta = 0$, then by (A.28), $(1 + Z_t^{0P}) f_e^P = c_e$ and thus $(1 + Z_t^{0P}) > 0$, and $\mu_t^P = \min \mu(\nu_c)$, which implies $f^A - f^P \geq 0$, contradicting FOC (A.29), if $\gamma_P > 0$. Thus, in this case, we must have $\gamma_P = 0$, $\gamma_A > 0$, and $f_t^A = f_t^P$, or $\mu_t^A = \mu_t^P = \min \mu(\nu_c)$.
- If $\beta = 1 + Z_t^{0P}$, then $\beta > 0$ by (A.28), and $\mu_t^A = \min \mu(\nu_c)$, which implies $f^A - f^P \leq 0$, contradicting FOC (A.29), if $\gamma_A > 0$. Thus, in this case, we must have $\gamma_A = 0$, $\gamma_P > 0$, and $f_t^A = f_t^P$, or $\mu_t^A = \mu_t^P = \min \mu(\nu_c)$.

Therefore, $0 \leq \beta \leq 1 + Z_t^{0P}$ and the worst priors are symmetrized with $\mu_t^A = \mu_t^P = \min \mu(\nu_c)$. Moreover, under the symmetrized prior, since $\gamma_A + \gamma_P > 0$ and $f_e > 0$, the FOCs imply

$$1 + Z_t^{0P} = \frac{c_e}{f_e} > 0, \quad \beta_t = \frac{\gamma_P}{(\gamma_A + \gamma_P)} \frac{c_e}{f_e} > 0. \tag{A.31}$$

These equations imply that if $\gamma_A, \gamma_P > 0$, then $1 - \beta_t + Z_t^{0P} > 0$ at optimum.

Let (μ_t^A, ν_t^A) and (μ_t^P, ν_t^P) be symmetrized at (μ_t^c, ν_t^c). That is,

$$(\mu_t^A, \nu_t^A) \in \arg \min_{(\mu,\nu)\in D} \varphi_A, \quad \text{and}$$

$$(\mu_t^P, \nu_t^P) \in \arg \min_{(\mu,\nu)\in D} \varphi_P.$$

Also let

$$B_{Pt} := \frac{\gamma_P}{2}(1 - \beta_t + Z_t^{0P})^2 + \theta_t, \tag{A.32}$$

$$B_{At} := \frac{\gamma_A}{2}\beta_t^2 - \theta_t. \tag{A.33}$$

Then, the FOCs for these two problems with respect to (μ, ν) are as follows: there are Largrange multipliers $\lambda_{Pt}, \lambda_{At} \geq 0$ such that

$$-(1 - \beta_t + Z_t^{0P}) f_\mu + \lambda_{Pt}\pi_\mu = 0, \tag{A.34}$$

$$-(1 - \beta_t + Z_t^{0P}) f_\nu + 2B_{Pt}\sigma_t \frac{\partial \sigma_t}{\partial \nu} + \lambda_{Pt}\pi_\nu = 0, \tag{A.35}$$

$$\lambda_{Pt}\pi = 0, \quad \pi \geq 0, \tag{A.36}$$

$$-\beta_t f_\mu + \lambda_{At}\pi_\mu = 0, \tag{A.37}$$

$$-\beta_t f_\nu + 2B_{At}\sigma_t \frac{\partial \sigma_t}{\partial \nu} + \lambda_{At}\pi_\nu = 0, \tag{A.38}$$

$$\lambda_{At}\pi = 0, \quad \pi \geq 0. \tag{A.39}$$

Above FOCs imply $\pi_\mu \neq 0$, at optimum. To see this, suppose $\pi_\mu = 0$. Then by (A.34), $\beta_t = 1 + Z_t^{0P}$. However, (A.37) implies $\beta_t = 0$, contradicting (A.31).

Since $\pi_\mu \neq 0$, FOCs (A.34) and (A.37) imply $(1 - \beta_t + Z_t^{0P})\lambda_{At} = \beta_t \lambda_{Pt}$. Together with this relationship, (A.35) and (A.38) imply

$$\beta_t B_{Pt}\sigma_t \frac{\partial \sigma_t}{\partial \nu} - (1 - \beta_t + Z_t^{0P}) B_{At}\sigma_t \frac{\partial \sigma_t}{\partial \nu} = 0.$$

Then, since $\sigma_t \frac{\partial \sigma_t}{\partial \nu} > 0$ by assumption, we have $\beta_t B_{Pt} = (1 - \beta_t + Z_t^{0P}) B_{At}$, which in turn implies

$$\theta_t = \frac{\beta(1 - \beta + Z_t^{0P})}{2(1 + Z_t^{0P})} \left(\gamma_A \beta_t - \gamma_P (1 - \beta_t + Z_t^{0P})\right) = 0.$$

The last equality holds because of (A.31). Given $\theta_t = 0$, and β_t satisfying (A.31), the common worst prior minimizes both φ^A and φ^P as it solves (10), and the optimal compensation scheme S is given as stated in the theorem. □

A.5 Proof of Theorem 3

Let us rewrite the agent's expected remaining utility as follows.

$$\operatorname*{ess\,sup}_{e\in\mathcal{U}_t^1} \inf_{(\mu,\nu)\in\mathcal{D}_t^1} E_t^{e,v}\left[-\exp\left\{-\gamma_A\left(\int_t^1 \left\{-c(e_s,s,Y) - H^A(e_s^*,\mu_s^*,\nu_s^*;\beta_s,\theta_s,s,Y)\right\}ds \right.\right.\right.$$
$$\left.\left.\left.+\int_t^1\left[\theta_s - \frac{\gamma_A}{2}\beta_s^2\right]\sigma^2 ds + \int_t^1 \beta_s f ds\right)\right\}\frac{d\tilde{P}_t^{e,v}}{dP_t^{e,v}}\right],$$

where

$$\frac{d\tilde{P}_t^{e,v}}{dP_t^{e,v}} = \exp\left\{-\gamma_A\int_t^1 \beta_s\left(dY_s - f ds\right) - \frac{\gamma_A^2}{2}\int_t^1 \beta_s^2\sigma^2(\nu_s,s,Y)ds\right\}.$$

Hence, we have

$$\operatorname*{ess\,sup}_{e\in\mathcal{U}_t^1} \inf_{(\mu,\nu)\in\mathcal{D}_t^1} \tilde{E}_t^{e,v}\left[-\exp\left\{-\gamma_A\left(\int_t^1 \left\{H^A(e_s,\mu_s,\nu_s;.) - H^A(e_s^*,\mu_s^*,\nu_s^*;.)\right\}ds\right)\right\}\right]. \tag{A.40}$$

The 'if' part. Suppose (e^*,μ^*,ν^*) is a saddle point of H^A: given (β_s,θ_s,s,Y), for all $(e_s,(\mu_s,\nu_s)) \in U \times D_s(Y)$,

$$H^A(e_s,\mu_s^*,\nu_s^*;.) \le H^A(e_s^*,\mu_s^*,\nu_s^*;.) \le H^A(e_s^*,\mu_s,\nu_s;.).$$

Let us define $\tilde{G}(e_s,\mu_s,\nu_s;e_s^*,\mu_s^*,\nu_s^*;.) \equiv H^A(e_s,\mu_s,\nu_s;.) - H^A(e_s^*,\mu_s^*,\nu_s^*;.)$. Then, (e_s^*,μ_s^*,ν_s^*) is also a saddle point of $\tilde{G}$ given (β_s,θ_s,s,Y), and the agent's remaining utility for all (s,Y) satisfies

$$\begin{aligned}&-\exp\left\{-\gamma_A\int_t^1 \tilde{G}(e_s,\mu_s^*,\nu_s^*;.)ds\right\}\\ &\le -\exp\left\{-\gamma_A\int_t^1 \tilde{G}(e_s^*,\mu_s^*,\nu_s^*;.)ds\right\} = -1\\ &\le -\exp\left\{-\gamma_A\int_t^1 \tilde{G}(e_s^*,\mu_s,\nu_s;.)ds\right\}.\end{aligned} \tag{A.41}$$

The first inequality of (A.41) implies

$$E_t^{e,v^*}\left[-\exp\left\{-\gamma_A\int_t^1\tilde{G}(e_s,\mu_s^*,\nu_s^*;.)ds\right\}\frac{d\tilde{P}_t^{e,v^*}}{dP_t^{e,v^*}}\right]$$

$$\leq E_t^{e,v^*}\left[-\exp\left\{-\gamma_A\int_t^1\tilde{G}(e_s^*,\mu_s^*,\nu_s^*;.)ds\right\}\frac{d\tilde{P}_t^{e,v^*}}{dP_t^{e,v^*}}\right]=-1$$

$$=E_t^{e^*,v^*}\left[-\exp\left\{-\gamma_A\int_t^1\tilde{G}(e_s^*,\mu_s^*,\nu_s^*;.)ds\right\}\frac{d\tilde{P}_t^{e^*,v^*}}{dP_t^{e^*,v^*}}\right].$$

For both the above inequality and equality, we have utilized the following facts:

$$\tilde{G}(e_t^*,\mu_t^*,\nu_t^*;.)=0,\quad\text{and}\quad E_t^{e,v}\left[\frac{d\tilde{P}_t^{e,v}}{dP_t^{e,v}}\right]=1,\quad\forall(t,e,v).$$

Similarly, by the second inequality of (A.41), we have

$$E_t^{e^*,v^*}\left[-\exp\left\{-\gamma_A\int_t^1\tilde{G}(e_s^*,\mu_s^*,\nu_s^*;.)ds\right\}\frac{d\tilde{P}_t^{e^*,v^*}}{dP_t^{e^*,v^*}}\right]=-1$$

$$=E_t^{e^*,v}\left[-\exp\left\{-\gamma_A\int_t^1\tilde{G}(e_s^*,\mu_s^*,\nu_s^*;.)ds\right\}\frac{d\tilde{P}_t^{e^*,v}}{dP_t^{e^*,v}}\right]$$

$$\leq E_t^{e^*,v}\left[-\exp\left\{-\gamma_A\int_t^1\tilde{G}(e_s^*,\mu_s,\nu_s;.)ds\right\}\frac{d\tilde{P}_t^{e^*,v}}{dP_t^{e^*,v}}\right].$$

Note that the first equality holds even when $P_t^{e^*,v^*}$ and $P_t^{e^*,v}$ are mutually singular, because $\tilde{G}\equiv 0$ at (e^*,μ^*,ν^*). Hence, by the principle of optimality, (e^*,μ^*,ν^*) is an optimal response by the agent, and his value function $\mathcal{V}_t(\equiv V_t^{e^*,v^*})$, or the optimal expected remaining utility, is constant over time at $-\exp\left(-\gamma_A\mathcal{W}_0\right)$.

The 'only if' part. Suppose that agent is given a contract $S\in\bar{\Psi}$ with (e^*,μ^*,ν^*), and the agent in fact, chooses (e^*,μ^*,ν^*). Then it must be true that (e^*,μ^*,ν^*) is a saddle point of H^A. To see this, we apply Lemma A.2. Let $u\equiv e$, $v\equiv(\mu,\nu)$, $\xi(Y)=\mathcal{W}_0$, $g_t=c(\bar{e}_t,t,Y)-\beta f(\bar{e}_t,\bar{\mu}_t,\bar{\nu}_t,t,Y)+\left(\frac{\gamma_A}{2}\beta_t^2-\theta_t\right)\bar{\sigma}_t^2(\bar{\nu}_t,t,Y)-c(\hat{e},t,Y)$, $q_t=\theta_t$, and $h=\beta$. Also let

$$G=\bar{c}-\beta\bar{f}+\left(\frac{\gamma_A}{2}\beta_t^2-\theta_t\right)\bar{\sigma}_t^2-c(\hat{e},t,Y)+\beta_t f(\hat{e},\hat{\mu}_t,\hat{\nu}_t,t,Y)+\left(q-\frac{\gamma_A}{2}\beta^2\right)\sigma^2(\hat{\nu}_t,t,Y),$$

$$K=f(\hat{e},\hat{\mu}_t,\hat{\nu}_t,t,Y)-\gamma_A\beta_t\sigma^2(\hat{\nu}_t,t,Y).$$

Since the contract S induces the agent to optimally choose $(e_t^*, \mu_t^*, \nu_t^*)$, the Lemma implies that there exists a $\mathcal{P}$-q.s. square-integrable process $Z_t^{u,v}$ such that

$$(e_t^*, \mu_t^*, \nu_t^*) \in \arg\max_{\hat{e}\in U} \min_{(\hat{\mu},\hat{\nu})\in D_t(Y)} \quad H^A(\hat{e}, \hat{\mu}, \hat{\nu}, Z_t^{u,v}, t, Y), \tag{A.42}$$

where for each $(Z_t^{u,v}, t, Y)$,

$$\begin{aligned} H^A(\hat{e}, \hat{\mu}, \hat{\nu}, Z_t, t, Y) &= Z_t^{u,v} K_t + G - \frac{\gamma_A}{2}(Z_t^{u,v})^2 \sigma^2(\hat{\nu}, t, Y) \\ &= (\beta_t + Z_t^{u,v})\, f(\hat{e}, \hat{\mu}_t, \hat{\nu}_t, t, Y) + \left(q_t - \frac{\gamma_A}{2}\,(\beta_t + Z_t^{u,v})^2\right) \sigma^2(\hat{\nu}, t, Y) \\ &\quad + c^* - \beta f^* + \left(\frac{\gamma_A}{2}\beta_t^2 - q_t\right)(\sigma_t^*)^2 - c(\hat{e}, t, Y). \end{aligned}$$

Lemma A.2 indicates that this condition is necessary and sufficient for the optimality of $(e_t^*, \mu_t^*, \nu_t^*)$. However, (A.40) implies the agent's utility is constant over time. That is $Z_t^{u,v} \equiv 0$. Then this Hamiltonian coincides with the one stated in the Proposition. □

A.6 Proof of Proposition 1

Consider a non-saddle point process $(e, (\mu, \nu)) \in \mathcal{U} \times \mathcal{D}$ and a contract $S(e, \mu, \nu; \beta, \theta) \in \bar{\Psi}$. Suppose that the agent accepts this contract, and optimally chooses (e^*, μ^*, ν^*) with his initial CEQ wealth equal to $\mathcal{W}_0^*$. Then, by Theorem 3, (e^*, μ^*, ν^*) is a saddle point process. Without loss of generality, we can assume that $\mathcal{W}_0^* \geq \mathcal{W}_0$, because otherwise, the agent rejects the contract. Then, by Theorem 1, there exists a pair of unique $\mathcal{P}^o$-q.s. square-integrable processes (β', θ') such that the contract $S(e, \mu, \nu; \beta, \theta)$ can be represented $\mathcal{P}^o$-q.s. in distribution by contract $S(e^*, \mu^*, \nu^*; \beta', \theta') - \mathcal{W}_0 + \mathcal{W}_0^*$, where $S(e^*, \mu^*, \nu^*; \beta', \theta') \in \bar{\Psi}$. That is, given $S(e^*, \mu^*, \nu^*; \beta', \theta') - \mathcal{W}_0 + \mathcal{W}_0^*$, the agent chooses the same saddle point process (e^*, μ^*, ν^*), with his initial CEQ wealth equal to $\mathcal{W}_0^*$. However, the no-wealth effect with exponential utility implies that the agent would also choose the same saddle point process $(e^*, \mu^*, \nu^*; \beta', \theta')$, with his initial CEQ wealth equal to $\mathcal{W}_0$, even if he were given $S(e^*, \mu^*, \nu^*; \beta', \theta')$. That is, $S(e^*, \mu^*, \nu^*; \beta', \theta')$ is implementable. On the other hand, the principal is indifferent between the two contracts, $S(e, \mu, \nu; \beta, \theta)$ and

$S(e^*,\mu^*,\nu^*;\beta',\theta')-\mathcal{W}_0+\mathcal{W}_0^*$, because they are identically distributed. Hence, she (weakly) prefers $S(e^*,\mu^*,\nu^*;\beta',\theta')$ to either of the two contracts, whenever $\mathcal{W}_0^* \geq \mathcal{W}_0$, and the statement follows. □

A.7 Proof of Theorem 4

The principal's problem is to choose (e,μ^A,ν^A), θ, and (μ^P,ν^P) in order to optimize her expected utility as follows:

$$\sup_e \inf_{v^P} E^{e,v^P}\left[-\exp\left\{-\gamma_P\left(Y_1 - S\right)\right\}\right]$$

$$\text{s.t.} \quad (i)\, dY_t = f(e_t,\mu_t^P,\nu_t^P,t,Y)dt + \sigma(\nu_t^P,t,Y)dB_t^{u,v^P},$$

$$(ii)\, (e_t,\mu_t^A,\nu_t^A) \in \arg\max_{\hat{e}} \min_{(\hat{\mu},\hat{\nu})\in D} \beta_t f(\hat{e},\hat{\mu},\hat{\nu},t,Y) - c(\hat{e},t,Y) + \left[\theta_t - \frac{\gamma_A}{2}\beta_t^2\right]\sigma^2(\hat{\nu},t,Y),$$

$$(iii)\, \beta_t = \frac{c_e(e_t,t,Y)}{f_e(e_t,\mu_t^A,\nu_t^A,t,Y)},$$

where $v^P = (\mu^P,\nu^P)$, and

$$\begin{aligned} Y_1 - S = -\mathcal{W}_0 &- \int_0^1 \left\{c(e_t,t,Y) - \beta_t f(e_t,\mu_t^A,\nu_t^A,t,Y) + \left[\frac{\gamma_A}{2}\beta_t^2 - \theta_t\right]\sigma^2(\nu_t^A,t,Y)\right\}dt \\ &- \int_0^1 \theta_t d\langle Y_t\rangle + \int_0^1 (1-\beta_t)dY_t \\ = -\mathcal{W}_0 &+ \int_0^1 H^A(e,\mu_t^A,\nu_t^A;\beta_t,\theta_t,t,Y)dt - \int_0^1 \theta_t d\langle Y_t\rangle + \int_0^1 (1-\beta_t)dY_t. \end{aligned}$$

The constraint (ii) is for incentive compatibility. Given that β_t satisfies (iii) where c is convex and f is concave in e, the agent's FOC with respect to $\hat{e}$ implies that given a contract with $\beta(e,.)$, the agent always chooses $\hat{e}=e$. Thus, given β as in (iii), the incentive compatibility for e is always satisfied, and therefore (ii) can be simplified as follows.

$$(\mu_t^A,\nu_t^A) \in \arg\min_{(\hat{\mu},\hat{\nu})\in D} \varphi_A(e,\hat{\mu},\hat{\nu};\theta_t,\beta_t,t,Y).$$

Then, by Lemma A.2, there exists a unique $\mathcal{P}^o$-q.s. square integrable process Z_t^P such that the principal optimizes the following Hamiltonian H_t^P:

$$H^P = Z_t^P\left(f - \gamma_P(1-\beta_t)(\sigma_t^P)^2\right) + G_t - \frac{\gamma_P}{2}(Z_t^P)^2\sigma^2.$$

This Hamiltonian can be rearranged as stated in (13). The constraints (14) and (15) are, respectively, for incentive compatibility for the agent's effort and worst prior choices. □

A.8 Proof of Theorem 5

Since both φ_A and φ_P are continuously differentiable and since D satisfies the KKT constraint qualification conditions, FOCs of the principal's and agent's minimization problems with respect to (μ, ν) become necessary conditions, and thus one can apply the Envelope Theorem to the principal's problem (13) for the following FOCs with respect to e and b:

$$- c_e(e_t, t, Y) + \frac{\partial \varphi_A}{\partial e} + \frac{\partial \varphi_P}{\partial e} = 0, \tag{A.43}$$

$$(\sigma_t^A)^2 - (\sigma_t^P)^2 = 0, \tag{A.44}$$

where $\sigma_t^k \equiv \sigma(\nu_t^k, t, Y)$, $k = A, P$. Note that given $\sigma_t^A = \sigma_t^P$, FOC (A.43), together with $\beta = c_e / f_e^A$, implies

$$\beta_e(f^A - f^P) + (f_e^P + \gamma_P \beta_e \sigma^2)(1 + Z_t^P) = \beta \left(f_e^P + (\gamma_P + \gamma_A)\beta_e \sigma^2 \right), \tag{A.45}$$

where the subscript e denotes the partial derivative of the corresponding function with respect to e, and f^k is short for $f(e_t, \mu_t^k, \nu_t^k, t, Y)$, $k = A, P$.

We claim that at optimum, the worst priors of the two contracting parties are symmetrized, and $0 < \beta \leq 1 + Z_t^P$, where the equality holds only when $\gamma_A = 0$. Since f is concave, $\beta_e > 0$. Since c_e and f_e are strictly positive, we have $\beta = c_e / f_e > 0$, trivially. FOC (A.44) implies $\nu_t^A = \nu_t^P = \nu_t^c$. Let (μ_t^P, ν_t^c) and (μ_t^A, ν_t^c) be the optimal perceptions chosen by the principal and agent, respectively. Then, for all $(\mu, \nu) \in D$,

$$\beta_t f(\mu_t^A, \nu_t^c, t, Y) - B_{At}(\sigma_t^c)^2 \leq \beta f(\mu, \nu, t, Y) - B_{At}\sigma^2 \quad \text{and}$$

$$(1 - \beta + Z_t^P) f(\mu_t^P, \nu_t^c, t, Y) - B_{Pt}(\sigma_t^c)^2 \leq (1 - \beta + Z_t^P) f(\mu, \nu, t, Y) - B_{Pt}\sigma^2.$$

where B_{Pt} and B_{At} are as defined in (A.32) and (A.33), respectively. At $\nu = \nu_t^c$, $\sigma = \sigma_t^c$, and for all μ,

$$\beta_t(f(\mu_t^A, \nu_t^c, t, Y) - f(\mu, \nu_t^c, t, Y)) \le 0, \tag{A.46}$$

$$(1 - \beta + Z_t^P)(f(\mu_t^P, \nu_t^c, t, Y) - f(\mu, \nu_t^c, t, Y)) \le 0. \tag{A.47}$$

If $\beta > 1 + Z_t^P$, then by the two inequalities, $f(\mu_t^A, \nu_t^c, t, Y) - f(\mu_t^P, \nu_t^c, t, Y) \le 0$. However, Eq.(A.45) implies

$$\beta_e(f^A - f^P) + (f_e^P + \gamma_P \beta_e \sigma^2)\beta > \beta \left(f_e^P + (\gamma_P + \gamma_A)\beta_e \sigma^2\right),$$

Since $\beta, \beta_e > 0$,

$$\beta_e(f^A - f^P) > \gamma_A \beta \beta_e (\sigma_t^c)^2 \ge 0.$$

Thus, $f^A - f^P > 0$, contradiction. That is, we must have $0 < \beta \le 1 + Z_t^P$.

Thus, if $\beta < 1 + Z_t^P$, then the two inequalities (A.46) and (A.47) imply that for all μ, $f(\mu_t^A, \nu_t^c, t, Y) \le f(\mu, \nu_t^c, t, Y)$, and $f(\mu_t^P, \nu_t^c, t, Y) \le f(\mu, \nu_t^c, t, Y)$. That is, $f(\mu_t^A, \nu_t^c, t, Y) = f(\mu_t^P, \nu_t^c, t, Y) = \min_\mu f(\mu, \nu_t^c, t, Y)$, i.e., $\mu_t^A = \mu_t^P$. Hence the worst priors are symmetrized.

If $\beta_t = 1 + Z_t^P$, then we always have $f(\mu_t^A, \nu_t^c, t, Y) = \min_\mu f(\mu, \nu_t^c, t, Y)$, and thus $f^A - f^P \le 0$. However, Eq.(A.45) implies

$$\beta_e(f^A - f^P) + (f_e^P + \gamma_P \beta_e \sigma^2)\beta = \beta \left(f_e^P + (\gamma_P + \gamma_A)\beta_e \sigma^2\right).$$

Thus, we have $\beta_e(f^A - f^P) = \beta \gamma_A \beta_e \sigma^2$. If $\gamma_A > 0$, $f^A - f^P > 0$, contradiction. Therefore, if $\beta = 1 + Z_t^P$, then $\gamma_A = 0$, but then by the same FOC, $f^A = f^P$. That is, the worst priors are symmetrized. That is, in all cases with $0 < \beta \le 1 + Z_t^P$, the worst priors are symmetrized. Then under the symmetrized worst prior, FOC (A.45) yields (16) which holds for both cases, $\gamma_A > 0$ and $\gamma_A = 0$.

Now, let (μ_t^A, ν_t^A) and (μ_t^P, ν_t^P) be symmetrized at (μ_t^c, ν_t^c). That is,

$$(\mu_t^c, \nu_t^c) \in \arg \min_{(\mu,\nu) \in D} \varphi_A, \quad \text{and}$$

$$(\mu_t^c, \nu_t^c) \in \arg \min_{(\mu,\nu) \in D} \varphi_P.$$

Then, using the same reasoning as in the last part of the proof of Theorem 2, one can show that θ_t is given by Eq.(17). $\square$

A.9 Proof of Proposition 2

Equipped with the Itô formula in the presence of singular measures, and with the saddle-point assumption (called the Issacs condition), one can derive the following Hamilton-Jacobi-Bellman-Issacs (HJBI) equation: $\mathcal{P}^o$-q.s.,

$$\begin{aligned} 0 \quad \equiv \quad & \frac{\partial V}{\partial t} + \sup_{e,\theta} \inf_{\mu^P,\nu^P} \left[-\gamma_P \frac{\partial V}{\partial y} \left\{ (1-\kappa e_t)(\nu_t^P)^2 - \frac{1}{\gamma_P}(\eta Y_t + e_t + \mu_t^P) \right\} + \frac{1}{2}\frac{\partial^2 V}{\partial y^2}(\nu_t^P)^2 \right. \\ & + \gamma_P V \left\{ \theta_t (\nu_t^P)^2 - \frac{1}{2\gamma_P} \frac{(\eta Y_t + e_t + \mu_t^P)^2}{(\nu_t^P)^2} + \frac{\kappa}{2} e_t^2 - \kappa e_t(\eta Y_t + e_t) - \kappa e_t \mu_t^A \right. \\ & \left. \left. + \left(\frac{\gamma_A}{2}\kappa^2 e_t^2 - \theta_t \right)(\nu_t^A)^2 + \frac{\gamma_P}{2}\left(1 - \kappa e_t - \frac{1}{\gamma_P}\frac{\eta Y_t + e_t + \mu_t^P}{(\nu_t^P)^2} \right)^2 (\nu_t^P)^2 \right\} \right], \end{aligned}$$

with $V(1,Y_1) = Y_0 - \mathcal{W}_0$. Also suppose that $E^{u,v}\left[\int_0^1 \left(\frac{\partial}{\partial y} V(t,Y) \right)^2 \nu_t^2 dt \right] < \infty$ for all $P^{u,v} \in \mathcal{P}^o$. Then, the well-known verification theorem still holds: that is, if there is a function V satisfying the HJBI with the square integrability condition, then V is the value function, i.e., $V \equiv \mathcal{V}$, and the optimand in the HJBI is the principal's Hamiltonain, H^P.

Let us consider a function $V(t,y) = -\exp\{-\gamma_P(\zeta(t)y + \rho(t))\}$ with $\zeta(1) = 0$ and $\rho(1) = Y_0 - \mathcal{W}_0$. Then, $V(1,y) = -\exp(-\gamma_P(Y_0 - \mathcal{W}_0))$,

$$\frac{\partial V}{\partial t} = -\gamma_P(y\dot{\zeta} + \dot{\rho})V, \quad \frac{\partial V}{\partial y} = -\gamma_P \zeta(t) V, \quad \text{and} \quad \frac{\partial^2 V}{\partial y^2} = \gamma_P^2 \zeta^2(t) V.$$

Set $\dot{\zeta} = -(1+\zeta(t))\eta$. Since $\zeta(1) = 0$, we have $\zeta(t) = e^{\eta(1-t)} - 1$. Suppose the function V satisfies the HJBI. Then, $V_t = \mathcal{V}_t$, and $\mathcal{Q}_t = \zeta(t)Y_t + \rho(t)$ which in turn implies $\zeta(t) = Z_t^P$ by (A.16). Moreover, from the HJBI,

$$\begin{aligned} 0 \quad \equiv \quad & \dot{\rho} + \sup_{e,\theta} \inf_{\mu^P,\nu^P} \left[-(\zeta(t) + 1 - \kappa e_t)^2 \frac{\gamma_P}{2}(\nu_t^P)^2 - \left(\frac{\gamma_A}{2}\kappa^2 e_t^2 - \theta_t \right)(\nu_t^A)^2 \right. \\ & \left. -\theta_t(\nu_t^P)^2 + \frac{\kappa}{2}e_t^2 + \kappa e_t \mu_t^A + (1 - \kappa e_t + \zeta(t))(e_t + \mu_t^P) \right]. \end{aligned}$$

For some $\lambda_t^A, \lambda_t^P \geq 0$, the FOCs are

$$\kappa(\mu_t^A - \mu_t^P) + \left(\kappa\gamma_P(\nu_t^P)^2 + 1\right)(1+\zeta(t))$$
$$= \kappa e_t \left(1 + \kappa(\gamma_P(\nu_t^P)^2 + \gamma_A(\nu_t^A)^2)\right), \tag{A.48}$$

$$(\nu_t^A)^2 - (\nu_t^P)^2 = 0, \tag{A.49}$$

$$(1 - \kappa e_t + \zeta(t)) - \lambda_t^P \pi_{\mu^P} = 0, \tag{A.50}$$

$$-\left(\zeta(t) + 1 - \kappa e_t\right)^2 \gamma_P \nu_t^P - 2\theta_t \nu_t^P - \lambda_t^P \pi_{\nu^P} = 0, \tag{A.51}$$

$$\kappa e_t - \lambda_t^A \pi_{\mu^A} = 0, \tag{A.52}$$

$$-2\left(\frac{\gamma_A}{2}\kappa^2 e_t^2 - \theta_t\right)\nu_t^A - \lambda_t^A \pi_{\nu^A} = 0, \tag{A.53}$$

$$\lambda_t^P \pi(\mu_t^P, \nu_t^P) = 0, \quad \pi(\mu_t^P, \nu_t^P) \geq 0, \tag{A.54}$$

$$\lambda_t^A \pi(\mu_t^A, \nu_t^A) = 0, \quad \pi(\mu_t^A, \nu_t^A) \geq 0. \tag{A.55}$$

By using the same procedure as in the proof of Theorem 5, one can show that $\mu_t^P = \mu_t^A = \mu_t^c$, $\nu_t^P = \nu_t^A = \nu_t^c$, and

$$\theta_t = \frac{1}{2(1+\zeta(t))}\beta_t(1 - \beta_t + \zeta(t))\left[\gamma_A\beta_t - (1 - \beta_t + \zeta(t))\,\gamma_P\right]. \tag{A.56}$$

In fact, this equation also follows from FOCs (A.49) to (A.53). Then from (A.48), we have (24). Moreover, since $\pi_\mu = 1$ and $\pi_\nu = -\alpha(\nu_t^c - \nu^0)$, FOCs (A.52) and (A.53) imply

$$\left(\gamma_A\beta_t^2 - 2\theta_t\right)\nu_t^c = \alpha\beta_t(\nu_t^c - \nu^0).$$

This equation, together with (A.56) and (24), implies that $\nu_t^c - \nu^0 \geq 0$, and $\Upsilon(\nu_t^c) = 0$. Note that Υ_ν

$$\Upsilon_\nu \quad = \quad -4\frac{\gamma_A\kappa\nu_t^c}{R_t^3}\frac{\gamma_A(1+\zeta(t))}{\alpha} - \frac{\nu^0}{(\nu_t^c)^2}.$$

Thus, $\Upsilon_\nu < 0$. $\Upsilon(0) = \infty$. Note that $\Upsilon(\infty) < 0$ if $\frac{\gamma_A\gamma_P}{\gamma_A+\gamma_P}\frac{e^{\eta(1-t)}}{\alpha} - 1 < 0$ for all t. Then there is a unique interior solution for ν_t^c for all t. Since the volatility ambiguity interval is $[\underline{\nu}, \overline{\nu}]$

where $\Upsilon(\overline{\nu}) > 0$ and $\Upsilon(\underline{\nu}) < 0$, there exists a unique optimal ν_t^c in the interior of $[\underline{\nu}, \overline{\nu}]$. Consequently, the common worst prior (μ^c, ν^c) can depend only on time, but not on state.

On the other hand, the HJBI yields

$$\begin{aligned}\dot{\rho} &= e^{\eta(1-t)}\left(e^{\eta(1-t)} - 2\beta_t\right)\frac{\gamma_P}{2}(\nu_t^c)^2 - e^{\eta(1-t)}\left(\frac{\beta_t}{\kappa} + \frac{\alpha}{2}(\nu_t^c - \nu^0)^2\right) + \frac{\beta_t^2}{2\kappa}\left[1 + (\gamma_A + \gamma_P)\kappa(\nu_t^c)^2\right] \\ &= -\left[e^{\eta(1-t)}\frac{\alpha}{2}(\nu_t^c - \nu^0)^2 + e^{2\eta(1-t)}\frac{1 + \kappa\gamma_P(\nu_t^c)^2 - \kappa^2\gamma_A\gamma_P(\nu_t^c)^4}{2\kappa\left(1 + \kappa(\gamma_P + \gamma_A)(\nu_t^c)^2\right)}\right].\end{aligned}$$

Thus, we obtain (23), because $\rho(1) = Y_0 - \mathcal{W}_0$.

Finally, we use the KKT conditions to check if the unique interior solution is a saddle point. Assume $\kappa + (\gamma_A - \gamma_P)\kappa^2(\nu_t^c)^2 > 0$. It is easy to check the concavity of the Hamiltonian in (e, θ), because $H_{ee}^P = -\kappa - \gamma_A\kappa^2(\nu_t^A)^2 + \gamma_P\kappa^2(\nu_t^P)^2 = -\kappa - (\gamma_A - \gamma_P)\kappa^2(\nu_t^c)^2 < 0$, and $H_{\theta\theta} = H_{e\theta} = 0$. Thus, H is concave in (e, θ). To check the convexity of the constrained Hamiltonian in (μ^P, ν^P), let $\mathcal{L} = H^P - \lambda^P(\mu_t^P - \frac{\alpha}{2}(\nu_t^P - \nu^0)^2)$: $\mathcal{L}_{\lambda\mu} = -1$ $\mathcal{L}_{\lambda\nu} = \alpha(\nu_t^P - \nu^0) > 0$, $\mathcal{L}_{\mu\nu} = 0$, $\mathcal{L}_{\mu\mu} = 0$, $\mathcal{L}_{\nu\nu} = -(\zeta(t) + 1 - \kappa e_t)^2\gamma_P - 2\theta + \lambda^P\alpha > 0$. For the last inequality, we have used (A.51). Thus, the negative of the determinant of the bordered Hessian is positive, confirming the convexity. □

A.10 Proof of Corollary 2:

Comparative statics: Note that $\mu^c = \frac{\alpha}{2}(\nu^c - \nu^0)^2$. This implies that $\frac{d\mu^c}{d\nu^c} > 0$. Recall that the optimal volatility satisfies $\Upsilon(\nu_t^c; \eta, \alpha, \kappa) = 0$. From this, we find comparative statics for (μ^c, ν^c) wrt (η, α, κ). Straightforward computation shows

$$\Upsilon_\nu = -\frac{4\kappa\gamma_A\nu^c}{R_t^3}\frac{\gamma_A e^{\eta(1-t)}}{\alpha} - \frac{\nu^0}{(\nu_t^c)^2} < 0,$$

$$\Upsilon_\eta = \left\{1 + (1 + R_t)\,\gamma_P\kappa(\nu_t^c)^2\right\}\frac{(1-t)\gamma_A e^{\eta(1-t)}}{\alpha R_t^2} > 0,$$

$$\Upsilon_\alpha = -\left\{1 + (1 + R_t)\,\gamma_P\kappa(\nu_t^c)^2\right\}\frac{\gamma_A e^{\eta(1-t)}}{\alpha^2 R_t^2} < 0,$$

$$\Upsilon_\kappa = -2\gamma_A(\nu^c)^2\frac{\gamma_A e^{\eta(1-t)}}{\alpha R_t^3} < 0.$$

From these inequalities, we immediately have the following comparative statics:

$$\frac{d\nu}{d\eta} = -\frac{\Upsilon_\eta}{\Upsilon_\nu} > 0, \quad \frac{d\nu}{d\alpha} = -\frac{\Upsilon_\alpha}{\Upsilon_\nu} < 0, \quad \frac{d\nu}{d\kappa} = -\frac{\Upsilon_\kappa}{\Upsilon_\nu} < 0.$$

Next, for comparative statics with respect to the outcome sensitivity β, note that

$$\beta_\nu = \frac{-2\kappa\gamma_A\nu_t^c}{R_t^2}e^{\eta(1-t)} < 0, \quad \beta_\kappa = \frac{-\gamma_A(\nu_t^c)^2}{R_t^2}e^{\eta(1-t)}, \quad \text{and} \quad \beta_\eta = \beta(1-t).$$

Hence, we have the following result: for $t < 1$,

$$\frac{d\beta_t}{d\eta} = \beta_\nu\frac{\partial\nu_t}{\partial\eta} + \beta_\eta = -\frac{1}{\Upsilon_\nu}\left[\beta_\nu\Upsilon_\eta - \beta_\eta\Upsilon_\nu\right] > 0, \text{if } \gamma_P = 0.$$

This inequality holds because

$$\beta_\nu\Upsilon_\eta - \beta_\eta\Upsilon_\nu = (1-t)\frac{e^{\eta(1-t)}}{\alpha R_t^4}\left[-2\kappa\gamma_A\nu_t^c\gamma_A e^{\eta(1-t)}\left\{\kappa(\gamma_A+\gamma_P)(\nu_t^c)^2\gamma_P\kappa(\nu_t^c)^2 - 1\right\}\right.$$
$$\left.+(1+\gamma_P\kappa(\nu_t^c)^2)\frac{\nu^0\alpha R_t^3}{(\nu_t^c)^2}\right] > 0, \quad \text{if } \gamma_P = 0.$$

On the other hand,

$$\frac{d\beta}{d\alpha} = \beta_\nu\frac{\partial\nu}{\partial\alpha} > 0,$$
$$\frac{d\beta}{d\kappa} = \beta_\nu\frac{\partial\nu}{\partial\kappa} + \beta_\kappa = \beta_\nu(-\frac{\Upsilon_\kappa}{\Upsilon_\nu}) + \beta_\kappa = -\frac{1}{\Upsilon_\nu}\left[\beta_\nu\Upsilon_\kappa - \beta_\kappa\Upsilon_\nu\right] < 0.$$

The last inequality holds, because

$$\beta_\nu\Upsilon_\kappa - \beta_\kappa\Upsilon_\nu = -\frac{\gamma_A(\nu^c)^2}{R_t^2}e^{\eta(1-t)}\frac{\nu^0}{(\nu_t^c)^2} < 0. \quad \square$$

References

1. Anscombe, F.J. and R.J. Aumann (1963): "A definition of subjective probability" *Annuals of Mathematical Statistics*, 34, 199-205.

2. Baker, G. P., and B. J. Hall (2004): "CEO Incentives and Firm Size," *Journal of Labor Economics*, 22 (4), 767-798.

3. Bazaraa, M.S., H.D. Sherali and C.M. Shetty (2006): *Nonlinear Programming: Theory and Algorithms*, 3rd ed., Wieley-interscience, Hoboken, New Jersey.

4. Bouchard, B. and M. Nutz (2102): "Weak Dynamic Programming for Generalized State Constraints," *SIAM J. Control and Optim*, 50(6), 3344-3373.

5. Campbell, J.Y., and G.B. Taksler (2003): "Equity Volatility and Corporate Bond Yields," *Journal of Finance*, 58(6), 2321-2349.

6. Chen, X., and J. Sung (2018): "Managerial Compensation and Outcome Volatility," Working paper.

7. Chen, Z., and L.G. Epstein (2002): "Ambiguity, Risk, and Asset Returns in Continuous Time," *Econometrica*, 70, 1403-1443.

8. Choquet, G. (1955): "Theory of Capacities," *Ann. Inst. Fourier*, 5, 131-295.

9. Coles, J.L., N.D. Daniel, and L. Naveen (2006): "Managerial Incentives and Risk-taking," *Journal of Financial Economics*, 76, 431-468.

10. Core and Guay (2002): "The Other Side of the Trade-Off: The Impact of Risk on Executive Compensation: a revised comment," Working Paper, University of Pennsylvania.

11. Cvitanić, J., D. Possamaï and N. Touzi (2016): " Moral Hazard in Dynamic Risk Management," *Management Science*, forthcoming.

12. Cvitanić, J., D. Possamaï and N. Touzi (2016): " Dynamic Programming Approach to Principal Agent Problems," *Finance and Stochastics*, forthcoming.

13. Davis, M. H. A. (1973): "On the Existence of Optimal Policies in Stochastic Control", *SIAM J. Control* 11, 587-594.

14. Davis, M. H. A. (1979): "Martingale Methods in Stochastic Control," in *Stochastic Differential Systems*, ed. by M. Kohlmann and W. Vogel, *Lecture Notes in Control and Information Science* 16, Springer Verlag, New York, 1979, 85 - 117.

15. Davis, M. H. A., and P. Varaiya (1973): Dynamic Programming Conditions for Partially Observable Stochastic Systems, *SIAM J. Control* 11, 226 - 261.

16. Denis, L., and C. Martini (2006): A Theoretical Framework for the Pricing of Contingent Claims in the Presence of Model Uncertainty, *Annals of Applied Probability* 16 (2), 827-852.

17. Elliott, R. (1982): *Stochastic Calculus and Applications*, Springer Verlag, New York.

18. Ellsberg, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 643-669.

19. Epstein, L.G. and S. Ji (2013): "Ambiguous Volatility and Asset Pricing in Continuous Time," *Review of Financial Studies*, 26 (7): 1740-1786.

20. Epstein, L.G. and M. Schneider (2008): "Ambiguity, Information Quality and Asset Pricing," *Journal of Finance* 63, 197-228.

21. Epstein, L.G. and M. Schneider (2010): "Ambiguity, and Asset Markets," *Annual Review of Financial Economics* 2, 315-46.

22. Frydman, C. and D. Jenter (2010): "CEO compensation," NBER Working Paper No. 16585.

23. Frydman, C. and R.E. Saks (2010): "Executive Compensation: A New View from a Long-term Perspective, 1936-2005," *Review of Financial Studies* 23, 2099-2138.

24. Gilboa, I.(1987): "Expected utility with purely subjective non-additive probabilities *Journal of Mathematical Economics*, 16, p. 65-88.

25. Gilboa, I., and D. Schmeidler (1989): "Maximin Expected Utility with Non-unique Priors," *Journal of Mathematical Economics*, 18, 141-153.

26. Guay (1999): "The Sensitivity of CEO wealth to Equity Risk: and Analysis of the Magnitude and Determinants," *Journal of Financial Economics*, 53, 43-71.

27. Hellwig, M. and K.M. Schmidt (2002): " Discrete-time Approximation of Holmström-Milgrom Brownian-motion Model of Intertemporal Incentive Provision," *Econometrica*, 70, 2225-264.

28. Holmstrom, B. and P. Milgrom (1987): Aggregation and Linearity in the Provision of Intertemporal Incentives, *Econometrica* 55, 303-328.

29. Klibanoff, P., M. Marinacci, and S. Mukerji (2005): A Smooth Model of Decision Making Under Ambiguity, *Econometrica*, 73, 18491892.

30. Mastrolia, T. and D. Possamaï (2015): "Moral Hazard under Ambiguity," Working Paper, Universitê Paris Dauphine.

31. Miao, J. and A. Rivera (2015): "Robust Contracts in Continuous Time," Working Paper, Boston University

32. Murphy, K. J. (2012): "Executive Compensation: Where We Are, and How We Got There," forthcoming in *Handbook of the Economics and Finance*, ed. by G. Constantinides, M. Harris, and R. Stulz, Elsevier Science North Holland.

33. Nutz, M. (2012a): "Pathwise Construction of Stochastic Integrals," *Electronic Communications in Probability*, 17(24), 1-7.

34. Nutz, M. (2012b): "A Quasi-sure Approach to the Control of Non-Markovian Stochastic Differential Equations," *Electronic Journal of Probability*, 17(23), 1-23.

35. Nutz, M, and H.M. Soner (2012): "Superhedging and Dynamic Risk Measures under Volatility Uncertainty," *SIAM J. Control Optim.*, 50(4), 2065-2089.

36. Ou-Yang, H. (2003): "Optimal contracts in a continuous-time delegated portfolio management problem," *Review of Financial Studies*, 16:173208.

37. Peng, S. (2006): "G-Expectation, G-Brownian Motion and Related Stochastic Calculus of Itô type," in *The Abel Symposium 2005, Abel Symposia 2*, ed. by Benth et. al., Springer-Verlag, 541-567.

38. Pham, T. and J. Zhang: "Two Person Zero-Sum Game in Weak Formulation and Path Dependent Bellman-Isaacs Equation," *SIAM J. Control Optim.*, 52(4), 2090-2121.

39. Rockafellar, R.T. (1970): *Convex Analysis*, Vol. 28 of Princeton Math. Series, Princeton Univ. Press.

40. Savage, L.J. (1954): *The Foundation of Statistics*, John Wiley and Sons, New York. Revised and enlarged edition, Dover, New York, 1972.

41. Sannikov, Y. (2008): "A Continuous-Time Version of the Principal-Agent Problem," *Review of Economic Studies*, 75, 957-984.

42. Schättler, H., and J. Sung, "The First-Order Approach to Continuous-Time Principal-Agent Problem with Exponential Utility," *Journal of Economic Theory*, 61, 1993, 331-371.

43. Schättler, H., and J. Sung, "On Optimal Sharing Rules in Discrete- and Continuous-Time Principal-Agent Problems with Exponential Utility," *Journal of Economic Dynamics and Control*, Vol 21, 1997, 551 - 574.

44. Schmeidler, D. (1989): Subjective probability and expected utility without additivity *Econometrica*, 57, p. 517587.

45. Sion, M. (1958), “On General Minimax Theorems,” *Pacific Journal of Mathematics*, Vol 8, 1958, 171-176.

46. Soner, M., N. Touzi, and J. Zhang (2011a): “Quasi-sure Stochastic Analysis though Aggregation,” *Electronic Journal of Probability*, 16, 1844-1879.

47. Soner, M., N. Touzi, and J. Zhang (2011b): “Martingale Representation Theorem under G-expectation,” *Stochastic Process and Their Applications*, 121, 265-287.

48. Soner, M., N. Touzi, and J. Zhang (2012): “Wellposedness of Second Order Backward SDEs,” *Probability and Related Fields*, 153, 149-190.

49. Soner, M., N. Touzi, and J. Zhang (2013): “Dual Formulation of Second Order Target Problems,” *Annals of Applied Probability* 23, 308-347.

50. Sung, J. (1995): “Linearity with Project Selection and Controllable Diffusion Rate in Continuous-Time Principal-Agent Problems,” *RAND Journal of Economics*, Vol 24, 720-743.

51. Sung, J. and X. Wan (2015): “A General Equilibrium Model of a Multi-Firm Moral-Hazard Economy with Financial Markets,” *Mathematical Finance*, Vol 25, 2015, 827-868.

52. Szydlowski, M. (2012): “Ambiguity in Dynamic Contracts,” Working Paper, University of Minnesota.

53. Takayama, A. (1985): *Mathematical Economics*, 2nd ed., Cambridge University Press, Cambridge, U.K.

54. Weinschenk, P. (2010): “Moral Hazard and Ambiguity,” Working Paper, Bonn Graduate School of Economics.