Robust Contracts with Exploration

Chang Liu*

First version: November 30, 2022
This version: April 7, 2023

Abstract

We study a two-period moral hazard problem; there are two agents, with identical action sets that are unknown to the principal. The principal contracts with each agent sequentially, and seeks to maximize the worst-case discounted sum of payoffs, where the worst case is over the possible action sets. The principal observes the action chosen by the first agent, and then offers a new contract to the second agent based on this knowledge, thus having the opportunity to explore in the first period. We define a suitable rule of updating and characterize the principal's optimal payoff guarantee. Following nonlinear first-period contracts, optimal second-period contracts may also be nonlinear in some cases. Nonetheless, we find that linear contracts are optimal in both periods.

Keywords: Moral hazard, robustness, exploration, linear contracts, maxmin

*Department of Economics, Harvard University; chang_liu@fas.harvard.edu. I am deeply grateful to my advisors Shengwu Li, Tomasz Strzalecki, Eric Maskin and Benjamin Golub for their guidance and support throughout the project, and over the years. I would also like to thank Sejal Aggarwal, Ophir Averbuch, Benjamin Brooks, Shani Cohen, Juan Dodyk, Olivier Gossner, Jerry Green, Yingni Guo, Bård Harstad, Oliver Hart, Zoë Hitzig, Michihiro Kandori, Jacob Leshno, John Macke, Stephen Morris, Giorgio Saponaro, Brit Sharoni, Cassidy Shubatt, Kathryn Spier, Haoqi Tong, Alexander Wolitzky, and especially Yannai Gonczarowski for valuable comments and discussion. In addition, I thank the audience at UNSW Sydney for their insightful feedback.

1 Introduction

Moral hazard models, in which a principal designs a contract to incentivize an agent, have been extensively studied and widely applied. In many canonical moral hazard models, however, optimal contracts require precise knowledge of the environment: the set of all possible actions together with the (stochastic) mappings from actions to outcomes. This aspect raises practical concerns, because in reality the principal's knowledge is certainly not entirely correct. How should the principal design contracts that have robust guarantees even if some details are incorrect? The emerging area of robust contract design follows the *Wilson Doctrine* (Wilson, 1987), which advocates for realistic mechanisms that are detail free.

The pioneer work of Carroll (2015) assumes that the principal knows only some of the actions available to the agent, and evaluates contracts based on their worst-case performance, over the unknown actions the agent might take. The results show that, very generally, the optimal contract is linear, which provides new foundations for the common use of linear contracts in practice.

One suspicion, however, about the linear results in Carroll (2015) is how much they hinge on the principal's inability to explore the unknown, an opportunity that arises naturally in a model with multiple interactions.[1] It is not even clear how to model (non-Bayesian) exploration in the robust paradigm. Specifically, if the principal can observe the action chosen by an agent, then she can learn about initially unknown actions the agent may subsequently take again. Furthermore, she may also exclude some actions of the agent based on the rationality of the agent's choice. In such environments, how should the principal design contracts to best utilize exploration opportunities? Are linear contracts still robustly optimal with exploration?

A suitable class of applications of robust models in contract design involves the principal hiring or consulting agents who are more specialized than herself, which explains why the principal may not know all the actions available to the agents, and may not even have a prior belief about the unknown ones. Consider an individual who hires gig workers from online platforms. Long-term contracts are typically not enforceable, yet she has the opportunity to interact with a pool of workers. The workers have received similar professional training, and therefore her knowledge about the capability of the pool from past experience is valuable for improving future interactions. Within this example, the main theoretical question of this paper is twofold: First, how should the individual specify contracts to best respond to new knowledge gained from exploration? Second, in anticipation of such opportunities, what contracts are optimal for acquiring new knowledge?

[1] One related but different criticism of the robust mechanism design literature is that most models are static in construction but assumes commitment. We discuss this issue in the literature section. See also Libgober and Mu (2023) for a corresponding perspective in the area of informationally robust mechanism design.

In the baseline model of this paper (Section 2), we study a two-period moral hazard problem. There are two agents, with the identical action sets that are unknown to the principal. The principal contracts with each agent sequentially to provide incentives. She observes the action chosen by the first agent, and then offers a new contract to the second agent based on this knowledge, thus having the opportunity to explore in the first period. The principal and agents are all risk neutral, and payments are constrained by limited liability.

The baseline model assumes that the principal knows only one available action of the agents,[2] but other unknown actions may also exist, and the principal does not even have a well-defined prior belief about these unknown actions. Faced with this nonquantifiable uncertainty, the principal seeks to maximize her worst-case discounted sum of payoffs, where the worst case is over the possible action sets. In the first period, the principal believes that the true action set could be any set containing the known action. After the principal offers a contract to the first agent and observes his response, we specify the rule of updating as follows: she believes the action set could be any set that (i) contains the observed action in addition to the initially known action, and (ii) does not contain any action strictly better than the observed action under the first-period contract. We refer to such actions sets as *compatible* (Definition 1). Requirement (i) indicates that the principal learns the existence of the chosen action, and requirement (ii) captures the additional inference she can draw from the rationality of the first agent.

The main result of this paper is that linear contracts are robustly optimal not just in static settings, but also in dynamic environments with exploration. In order to obtain this conclusion, we characterize the principal's optimal payoff guarantee in both periods, under the above rule of updating *compatibility*. Following nonlinear first-period contracts, optimal second-period contracts may also be nonlinear in some cases. Nonetheless, it is robustly optimal to use linear first-period contracts, and therefore optimal second-period contracts are also linear on the equilibrium path.

We begin with the analysis of the second period of the dynamic relationship (Section 3), where the principal has offered some first-period contract and observed the response of the first agent. This is not a direct adaptation of the single-period problem in Carroll (2015), precisely because the principal draws additional inferences from the rationality of the first agent, which excludes certain actions. We fully characterize the principal's optimal second-period payoff guarantee, and identify the contract that attains it in various cases. The analysis reveals four ways the principal may respond to the knowledge gained from observing the chosen action (Lemma 1), and in particular shows that if the first-period contract is nonlinear, then the optimal

[2] Later in Subsection 5.1, we study how the results of the baseline model extend to the cases where the principal knows a general set of available actions.

second-period payoff guarantee may be achieved by a nonlinear second period contract. Specifically, the principal's optimal guarantee is achieved by offering the best among four contracts: (i) the first-period contract again, (ii) a modified version of the first-period contract with compensation for the second agent, and (iii) & (iv) two linear contracts that correspond to the optimal static contracts in Carroll (2015). As long as the first-period contract is nonlinear, and the observed action is such that one of the first two contracts is optimal, then the optimal guarantee is achieved by nonlinear contracts.

Going back to the first period the dynamic relationship (Section 4), where the principal chooses a first-period contract to maximize her overall payoff guarantee, we establish the optimality of a linear first-period contract (Theorem 1). The proof of the conclusion boils down to two steps. The first step shows that any nonlinear first-period contract can be improved into another linear contract, thereby (weakly) increasing the payoff's overall payoff guarantee (Lemma 2). To illustrate this, for any possible action taken by the first agent under the linear contract, we construct an alternative first-period action under the nonlinear contract, such that the principal's payoff guarantee is lower under the nonlinear one. In fact, our construction is stronger than necessary, so that the principal's payoff under the nonlinear contract is lower than under the linear one in both periods. The second step further shows that the maximum of the principal's first-period problem exists within the class of linear first-period contracts (Lemma 3). We set up a program to characterize the overall payoff guarantee of an arbitrary linear first-period contract, and prove that the guarantee is continuous in the first-period share. Therefore, it achieves a maximum, which is also the optimal overall guarantee among all linear first-period contracts. Combining these two steps, we show that, even with the opportunity to use first-period contracts as a means of exploration, no other more complicated form of contracts provides a better payoff guarantee to the principal than linear ones.

Finally, we consider several extensions of the baseline model and discuss further results (Section 5). First, we analyze the cases where the principal knows a set of actions the agents can take (Subsection 5.1). Just as in the baseline model, we characterize the principal's optimal second-period payoff guarantee in closed form, and identify the contract that attains it in various cases (Lemma 1′). In addition, as long as the set of known actions satisfies a condition called *lower bound on marginal cost* (Definition 3), linear contracts still outperform nonlinear ones (Theorem 1′). We then investigate an alternative environment to the baseline model (Subsection 5.2), where the agents' action set may potentially expand between periods. In this situation, the principal adopts a different rule of updating (instead of *compatibility*), and we show that linear contracts remain to be optimal in both periods (Theorem 2). Furthermore, we examine the structure of the optimal linear first-period contract in our dynamic model (Subsection 5.3) and compare it with the optimal static contract identified by Carroll (2015).

Related Literature Foundations for linear incentive contracts have received extensive research attention. The seminal work of Holmström and Milgrom (1987) considers a dynamic framework where output is produced gradually over time, the agent is aware of his own progress, and the principal pays the agent at the end. Although the principal is allowed to use the entire history of output to determine the payment, the optimal contract depends only on the number of realizations of each output level, and is linear in these counts. In a continuous time version of their problem where the agent controls the drift of a multidimensional Brownian motion, the optimal contract can be expressed as a linear function that depends only on the endpoint.[3] However, the stationary structure of their model is critical for this linearity result,[4] because linear contracts provide the agent with constant incentives to move forward independent of her past performance. In our model, the principal offers multiple contracts during the process, and exploration makes the principal's problem inherently non-stationary. Therefore, our paper considers a different form of foundation for linear contracts. Furthermore, Diamond (1998) and Barron, Georgiadis, and Swinkels (2020) provide arguments for linear contracts using static Bayesian frameworks.

More recently, pioneered by Carroll (2015), this issue has been investigated by a wave of research using robust models of contract design, which demands contract performance to be robust to limited knowledge of the environment. Carroll (2019) provides a comprehensive review of this approach, as well as an overview of the evolving field of robust mechanism design that adopts many other notions of robustness. Most work in robust contract design, however, analyzes static or one-shot models, which precludes the opportunity for designers to better understand parts of the environment they do not know. While starting with nonquantifiable uncertainty, designers may still be able to gradually gain a better understanding of the environment in which they repeatedly engage through exploration. Our dynamic model provides the principal with the opportunity to explore the unknown, in order to understand how the principal should design contracts that are robustly optimal given this exploration opportunity.

As stated by Carroll (2019), "another challenge is that trying to write dynamic models with non-Bayesian decision makers leads to well-known problems of dynamic inconsistency, except in special cases (e.g., Epstein and Schneider (2003)). This may be one reason why there has been relatively little work to date on robust mechanism design in dynamic settings." Knowing the difficulty, we carefully specify the principal's "beliefs" in the second period of our two-period model to follow a recursive structure analogous to Epstein and Schneider (2003), in order to avoid dynamic inconsistency issues.[5]

[3] Following Holmström and Milgrom (1987), Sung (1995) further shows that the optimal contract can still be linear when the agent controls the variance; Hellwig and Schmidt (2002) provide discrete time approximations of the continuous time model.

[4] For example, Schättler and Sung (1993) show that a time-dependent technology makes the optimal contract nonlinear.

[5] We require a specific rule of updating, *compatibility* (Definition 1), to restrict the set of actions the principal considers possible in the second period. This is an analogue of *consistency* in solution concepts like perfect Bayesian equilibrium.

This paper is relevant to the recent research that examines robust contracting in different organizational environments. Specifically, Dai and Toikka (2022) analyze moral hazard in teams, Marku, Ocampo, and Tondji (Forthcoming) study a common agency model, and Carroll and Bolte (Forthcoming) investigate a model with double moral hazard. Walton and Carroll (2022) provide a general framework that goes beyond simple bilateral relationships and allows for rich internal organizational structures. Our model analyzes a simple contracting environment, and aims to capture the main issue in terms of exploration. In particular, due to exploration, the analysis of any period in our dynamic model cannot be directly derived using the conclusions in Walton and Carroll (2022).[6]

From a broader perspective, Marku, Ocampo, and Tondji (Forthcoming) and Carroll and Bolte (Forthcoming) are in a similar spirit to our work on how the designers' robust objectives interact with their policy choices. In Marku, Ocampo, and Tondji (Forthcoming), several principals compete to contract with a common agent. In Carroll and Bolte (Forthcoming), the principal faces the choice of supplying input in the process of contracting with an agent. However, the maxmin objective in both studies is applied only once, whereas in our model it needs to be used in each of the two periods. In the area of informationally robust mechanism design, Libgober and Mu (2023) study durable good monopoly without commitment, and introduce the notion of *dynamically-consistent* worst-case information structure.

A number of other recent papers considering static models of robust contracts are related to our work, because the principal is aware of some additional characteristics of the unknown actions in addition to the concern that they may exist. As with Kambhampati (2022), who studies performance evaluation of agents, although we do not place any restrictions on the possible action sets of an individual agent, we assume that the two agents have identical action sets. However, our assumption is for a different reason, in order to make the principal's observations of chosen actions valuable. In addition, Antić (2021) assumes a lower bound on the productivity of all unknown actions of the principal. Furthermore, in Dütting, Roughgarden, and Talgam-Cohen (2020), the principal only knows the first moment of the distribution over output induced by each possible action, but not the full distribution.

The rest of the paper is organized as follows. Section 2 lays out the baseline model. The core analysis of the paper is in Sections 3 and 4. Section 3 analyzes the second period of the dynamic relationship and shows that, following nonlinear first-period contracts, optimal second-period contracts may also be nonlinear in some cases. Then, Section 4 studies the principal's first-period problem of maximizing the overall payoff

[6] We articulate the specific differences between our dynamic model and the general static framework in Walton and Carroll (2022) in Section 3.

guarantee, and establishes the optimality of linear first-period contracts. Section 5 analyzes several extensions to our baseline model and discusses further results, and Section 6 concludes. Appendix A contains the proofs of all results in the main text.

2 Model

2.1 Notation

We denote by $\Delta(X)$ the set of (Borel) probability measures on a set $X \subseteq \mathbb{R}$, equipped with the weak topology. For $x \in X$, we write δ_x for the degenerate distribution that puts probability one on x.

2.2 Setup

The baseline model is a two-period moral hazard problem, consisting of a principal (she) and two agents (he). The principal contracts with each agent sequentially to provide incentives, and the reservation payoff of the agents is zero. All parties are assumed to be risk neutral. The principal's discount factor is $\beta \in (0, \infty)$.

In each period ($t = 1, 2$), agent t takes a costly action that results in a stochastic output. The realized output y belongs to a set Y of possible output values. Assume Y is a compact subset of $\mathbb{R}$, either finite or infinite, and normalize the lowest possible output to zero: $\min(Y) = 0$.

An *action* of the agents, a, is a modeled as a pair $a = (F, c) \in \Delta(Y) \times \mathbb{R}^+$, with the interpretation that if an agent chooses action a, he incurs cost c, and output is drawn $y \sim F$. We equip $\Delta(Y) \times \mathbb{R}^+$ with the natural product topology.

A *techonology* is a (nonempty and) compact set of possible actions. The two agents have the same technology $A \subseteq \Delta(Y) \times \mathbb{R}^+$, which only they know but the principal does not. The principal knows only one action $a_0 = (F_0, c_0)$ available to the agents.[7] To ensure that the principal may, potentially, benefit from contracting with the agents, assume that $\mathbb{E}_{F_0}[y] - c_0 > 0$.

To capture the idea of exploration, assume that the principal observes the action chosen by agent 1, and then offers a new contract to agent 2 based on this knowledge. The chosen action itself, however, is not contractible.[8] Payments to the agents can only depend on the realized output, y. Assume that the agents

[7]Note that it is necessary for the principal to know at least one action to guarantee herself a strictly positive expected payoff, because otherwise it is always possible that the agents are not able to produce anything.

[8]It is a strong assumption that the chosen action becomes observable to the principal, especially since F represents a distribution. One interpretation is that each period summarizes (the "average" state of) a horizon for which the contract needs to remain fixed, while the agent is repeatedly taking action. During this process, the principal can keep observing him and figure out what action must be taken, in particular what F and c are. However, knowing that the action exists is still not the same as being able to write it into a contract. The action itself may be too complex to be accurately described in contract terms, or its inclusion into the contract may be directly prohibited by law.

have limited liability, so the payment to them can never be strictly negative. A *contract* is a continuous[9] function $w : Y \to \mathbb{R}^+$ such that $w(0) = 0$.

One foundation for $w(0) = 0$ is the standard *free disposal condition*, plus a *lowest support condition* on the agents' possible actions. We say a technology A satisfies the *lowest support condition* if, for all $(F, c) \in A$, the lowest output 0 is in the support of F. Under these two conditions, the principal will only offer contracts with $w(y) \geq w(0)$ for all y, because otherwise the agent may discard output to receive more payments. Given limited liability, it is then without loss of generality to focus on contracts with $w(0) = 0$, since a constant shift does not affect the agent's incentives, but only increases the principal's payoff.[10] Although we do not explicitly impose these two conditions, but only view them as a possible explanation for $w(0) = 0$, all results and derivations of the baseline model extend to the framework in which these two conditions hold. We maintain the assumption that $w(0) = 0$ throughout the analysis.

The timing within each period t is summarized as follows:

1. The principal offers a contract w_t.

2. Agent t chooses $a_t = (F_t, c_t) \in A$, or quits the relationship (zero payoff for both parties).

3. Output $y_t \sim F_t$ is realized.

4. Payoffs $y_t - w_t(y_t)$ to the principal and $w_t(y_t) - c_t$ to agent t.

The principal's objective is to maximize her worst-case expected discounted sum of payoffs over all possible technologies. In the first period, the principal only knows the action a_0, and believes that the true technology A could be any technology such that $A \ni a_0$. After the principal offers contract w_1 and observes the action a_1 chosen by agent 1, a rule of updating needs to be specified to determine the technologies that the principal considers possible. We say those possible technologies *compatible* with (w_1, a_1), formally defined as follows.

Definition 1 (Compatible)**.** *Given w_1 and $a_1 = (F_1, c_1)$, a technology A is* compatible *with (w_1, a_1) if*

1. $A \supseteq \{a_0, a_1\}$.

2. $\mathbb{E}_F[w_1(y)] - c \leq \mathbb{E}_{F_1}[w_1(y)] - c_1$ *for all* $(F, c) \in A$.

[9] The continuity assumption is made only to ensure the existence of best responses of the agents. This assumption becomes vacuous if Y is a finite set, and can also be weakened to upper semicontinuity with additional verifications. See also Carroll (2015, footnote 1), Walton and Carroll (2022, footnote 3), Carroll and Bolte (Forthcoming, footnote 1).

[10] If $w(0) > 0$, let $\tilde{w}(y) = w(y) - w(0) \geq 0$ be another valid contract. The agent's chosen action does not change if the principal instead offers $\tilde{w}$, but this increases the principal's payoff by $w(0)$.

Roughly speaking, a technology A is compatible with (w_1, a_1) if it contains a_1 (in addition to a_0), and does not contain any action strictly better than a_1 under w_1. The first requirement in Definition 1 indicates that the principal learns that action a_1 exists (in addition to the initially known a_0), and believes that agent 2 may also take this action again. The second requirement in Definition 1 captures the additional inference she can draw from agent 1's rationality: the true technology A cannot contain any action (F, c) that leads to a strictly higher payoff for agent 1, i.e., it is impossible that $\mathbb{E}_F [w_1 (y)] - c > \mathbb{E}_{F_1} [w_1 (y)] - c_1$.

2.3 Preliminary Analysis

It is relatively straightforward to describe the behavior of the agents. In each period t, given contract w and technology A, agent t chooses an action $(F, c) \in A$ to maximize his expected utility, so the best response correspondence is given by

$$BR (w|A) \equiv \underset{(F,c)\in A}{\arg\max} \{\mathbb{E}_F [w (y)] - c\} .$$

The principal's single-period expected payoff under technology A is denoted by

$$V (w|A) \equiv \max_{(F,c)\in BR(w|A)} \mathbb{E}_F [y - w (y)] ,$$

where we assume ties are broken in the principal's favor if the agent is indifferent among several actions.[11]

The principal's dynamic problem is solved via backward induction. In the second period, since the principal believes that A could be any technology compatible with (w_1, a_1), her problem is to choose a second-period contract w_2 to maximize her worst-case payoff:

$$V_2 (w_2|w_1, a_1) \equiv \inf_{A \text{ compatible with } (w_1,a_1)} V (w_2|A) .$$

Note that this is not a direct adaptation of the single-period problem in Carroll (2015) (where A could be any technology containing $\{a_0, a_1\}$), precisely because of her additional inference from agent 1's rationality in the definition of compatibility, which rules out the possibility that certain actions exist in A. In Section 3, we characterize the principal's *optimal second-period payoff guarantee*, $V_2^* (w_1, a_1)$, showing that this distinction matters. The maximum always exists, as we identify the contract that attains it; however, it may

[11]This tie-breaking assumption ensures the existence of optimal contracts, and minimizes the departure from standard models. Other tie-breaking rules will lead to essentially the same results, but may introduce technical complications. For example, the principal's optimal payoff guarantee may be approached, but not achieved, by linear contracts. See also Carroll (2015, Section D), Dai and Toikka (2022, footnote 4), Carroll and Bolte (Forthcoming).

be achieved by a nonlinear w_2 if the corresponding w_1 is nonlinear.

Going back to the first period, if the principal offers first-period contract w_1 and the true technology A is such that agent 1 chooses action $a_1 = (F_1, c_1)$, her *interim payoff guarantee*, defined as her payoff in the first period plus the discounted optimal second-period payoff guarantee, is given by

$$U(w_1|a_1) \equiv \mathbb{E}_{F_1}[y - w_1(y)] + \beta \cdot V_2^*(w_1, a_1).$$

Since she believes that the true technology A could be any technology such that $A \ni a_0$, her *overall payoff guarantee*, defined as the worst-case interim payoff guarantee over all possible technologies A, is given by

$$U(w_1) \equiv \inf_{A \ni a_0} \left\{ \max_{a_1 \in BR(w_1|A)} U(w_1|a_1) \right\},$$

where again we assume ties are broken in her favor.

The principal's first-period problem is to choose a first-period contract w_1 to maximize her overall payoff guarantee $U(w_1)$. In Section 4, we show the maximum exists and is achieved by a linear contract. That is, linear first-period contracts are optimal in terms of utilizing the exploration opportunity, making them even more robust.

3 Second Period Analysis

We begin our analysis with the second period of the dynamic relationship, where the principal has offered some first-period contract w_1 and observed agent 1's selected action a_1. We fully characterize the principal's optimal second-period payoff guarantee, $V_2^*(w_1, a_1)$, and identify the contract that attains it in various cases. The analysis reveals four ways the principal may respond to the knowledge gained from observing a_1, and in particular shows that if w_1 is nonlinear, then the optimal second-period payoff guarantee may be achieved by a nonlinear w_2.

The main result for the second period analysis is Lemma 1, which shows that $V_2^*(w_1, a_1)$ is achieved by offering the best among four contracts: (i) the first-period contract w_1 again, (ii) a modified w_1 with compensation for agent 2, and (iii) & (iv) two linear contracts that correspond to the optimal static contracts in Carroll (2015). As long as the first-period contract w_1 is nonlinear, and the observed action a_1 is such that one of the first two contracts is optimal, then $V_2^*(w_1, a_1)$ is achieved by nonlinear contracts.

Lemma 1 reveals that the analysis in this section is not a direct adaptation of the single-period problem in Carroll (2015), since optimal contracts may not be linear. This difference is precisely due to the second requirement of compatibility, where the principal draws additional inferences from the rationality of agent 1, excluding certain actions. Note that our analysis is also not covered by the recent work of Walton and Carroll (2022), which establishes a general static framework that allows for rich organizational structures, and identifies two properties of the counterparty's possible responses which jointly imply that a linear contract solves the principal's single-period maxmin problem. Specifically, their *Richness* property requires that the set of possible responses to a given contract be sufficiently and unboundedly broad. The Richness property, while natural in static analysis, is violated in our model exactly because of the principal's exploration and inference in the first period, since the true technology cannot contain any action that is strictly better for agent 1 than the observed action under the first-period contract.[12]

To demonstrate the intuition behind Lemma 1, in Subsection 3.1 we use a concrete example to illustrate how an earlier opportunity to explore may change the principal's payoff guarantee later, thereby affecting the subsequent optimal contracts. Then in Subsection 3.2, we formalize the observations from the example in order to obtain the main result.

3.1 An Example

Assume, as in the baseline model, that the principal contracts with each of the two agents sequentially, and the reservation payoff of the agents is zero. The principal does not know exactly what the action set is, but knows that the agents have some action a_0 available that can produce 2000 units of output in expectation, at a cost of 500 units. Additionally, she observes the action chosen by the first agent, and then offers a new contract to the second agent based on this knowledge.

Suppose that the principal offers a linear contract to the first agent in which they split the output equally. Then, even without any information about other available actions, the principal can be sure that the agent will get a payoff at least $2000/2 - 500 = 500$ units based on his rationality, and therefore she gets at least 500 units for herself because they equally split the output.[13] Note that if the principal has no knowledge about any action other than a_0, her worst-case payoff under this contract with the first agent is exactly 500 units. This is because, there may be another action a_0' that can produce $1000 + \varepsilon$ units of output without

[12]The other property in Walton and Carroll (2022), *Responsiveness*, indicates that the counterparty's behavior is responsive to the incentive provided by expected payment, and allows comparison of the principal's payoff guarantees from two different contracts. The Responsiveness property is satisfied in our model. As a converse result, Walton and Carroll (2022) also show that Responsiveness is necessary for linearity under a strengthened version of Richness. This result is in parallel with our analysis, since it is Richness that is not satisfied in our model.

[13]This argument was developed by Carroll (2015). In particular, given these numbers, splitting the output equally is the optimal static contract. To see this, suppose that the principal offers to pay the agent s share of output and keep $(1 - s)$ of output. The principal can be sure that the agent will get a payoff at least $2000s - 500$ units based on his rationality, and therefore she gets at least $(1 - s) \cdot (2000s - 500) / s = (1 - s) (2000 - 500/s)$ units for herself. This expression is maximized at $s = 1/2$, resulting in 500 units as the optimal guarantee.

incurring any cost. He strictly prefers a_0' to a_0 because he obtains a payoff of $500 + \varepsilon/2 - 0 > 500$ units. Moreover, a_0' leaves the principal with a payoff only slightly above 500 units.

However, if the first agent actually chooses the action a_0, then the principal knows that the action a_0' cannot exist, as well as any other action that might result in the agent's payoff exceeding 500 units. So, if she again offers the second agent the equal split contract, she knows that the second agent will choose the action a_0 because both agents have the same action sets. This guarantees herself a payoff of $2000/2 = 1000$ units, which is strictly higher than the worst case before she contracts with the first agent and gains knowledge. This does not occur in a static setting without exploration.

Now suppose the first agent chooses an action a_1 different from a_0, which can produce 4000 units of output in expectation, at a cost of 250 units. Then the principal knows that she can guarantee herself a higher payoff of $4000/2 = 2000$ units by offering the second agent the equal split contract. However, there are other contracts that would be more beneficial to her. If she offers to keep $3/4$ of output and pays the second agent $1/4$ of output, she can be sure that the agent will get a payoff at least $4000/4 - 250 = 750$ units, and therefore she gets at least $750 \times 3 = 2250 > 2000$ units for herself. The discovery of the low-cost, high-output action a_1 is very valuable to the principal, as she can offer a lower share to incentivize the second agent. We can further show that, following the equal split contract and the first agent's choice of a_1, this new contract is optimal for the principal.

This is not the only form in which subsequent optimal contracts may change. The principal might also increase the share offered to the second agent after observing a low-cost, low-output action. Suppose the first agent chooses another action a_1' that can produce 1200 units of output in expectation, at a cost of 90 units ($1200/2 - 90 = 510 > 500$). Then the principal's payoff guarantee from offering the second agent the equal split contract is $1200/2 = 600$ units, only slightly higher than the worst-case payoff when she only knows that a_0 exists (500 units). However, by increasing a little bit of the share offered to the second agent, to 55%, she can guarantee that her payoff is at least 810 units, a lot closer to the benchmark where agent 1 chooses the initially known action a_0.[14]

In Subsection 3.2 below, we characterize the principal's optimal payoff guarantee following different observed actions, and identify subsequent contracts that achieve it.

[14]The specific calculations leading to the 55% share and the payoff guarantee of 810 units are formalized in part 2 of Lemma 1 in Subsection 3.2. By contrast, in a static setting where the principal only knows the two actions $\{a_0, a_1'\}$ without any information about other actions, the optimal contract is to offer the agent a share of $\sqrt{90/1200} = 27.4\%$, resulting in a payoff guarantee of 632.7 units, strictly smaller than 810 units.

3.2 Optimal Second-Period Contracts

The previous example illustrates the possible variations of the principal's payoff guarantee and optimal contracts in the presence of exploration. Especially, we only discuss linear contracts in the example for the sake of convenience, and it is not a priori clear whether linear contracts are the best way to respond to the new knowledge. In addition, due to the foreseeable changes in the payoff guarantee later, optimal exploration contracts are not necessarily linear.

Suppose that in the first period the principal offers contract w_1 and observes agent 1's action $a_1 = (F_1, c_1)$. She learns that the true technology A is compatible with (w_1, a_1); that is, it contains a_0 and a_1, and does not contain any action strictly better than a_1 for agent 1 under w_1.

In the second period, if she offers the same contract $w_2 = w_1$, then she knows that agent 2 will choose a_1 again because the two agents have the same technology. This exactly repeats her first-period payoff $\mathbb{E}_{F_1}[y - w_1(y)]$ in the second period. Part 1 of Lemma 1 below shows that, in some cases, doing so is already optimal for the principal,[15] which means that an optimal second-period contract may be nonlinear following nonlinear first-period contracts.

Offering the same contract again is only one response of the principal to the knowledge gained by observing a_1, and there are plenty of other possible responses. For example, if the initially known action a_0 may lead to a higher payoff for the principal (i.e., $\mathbb{E}_{F_0}[y - w_1(y)] > \mathbb{E}_{F_1}[y - w_1(y)]$), then it might be tempting for the principal to try to obtain the payoff $\mathbb{E}_{F_0}[y - w_1(y)]$ instead. However, achieving this payoff requires the principal to use w_1 to induce action a_0, and this would violate agent 2's incentive constraint. Indeed, in the first period, the chosen action a_1 provides agent 1 with a (weakly) higher payoff compared to the known action a_0, and this relationship gets transferred to the second period because both agents have the same technology. This gives rise to the following notion of the *incentive gap*.

Definition 2 (Incentive gap)**.** *Given w_1 and $a_1 = (F_1, c_1)$, the* incentive gap*, $g(w_1, a_1)$, denotes the difference in agent 1's payoff between choosing a_1 and a_0. Formally,*

$$g(w_1, a_1) \equiv \left(\mathbb{E}_{F_1}[w_1(y)] - c_1\right) - \left(\mathbb{E}_{F_0}[w_1(y)] - c_0\right).$$

If the principal wants to induce action a_0 using a contract "similar to" w_1, then agent 2 needs to be compensated for not choosing a_1, and the amount of compensation increases with the incentive gap $g(w_1, a_1)$. Part 2 of Lemma 1 shows that the incentive gap sometimes becomes a real cost. Specifically, if

[15] In the numerical example, this corresponds to the case where agent 1 chooses the known action a_0 in response to the equal split contract.

$\mathbb{E}_{F_0}[y - w_1(y)] > g(w_1, a_1)$, then the principal can offer to agent 2 a modified version of w_1 with compensation in order to guarantee that her payoff in the second period is at least $\left(\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g(w_1, a_1)}\right)^2$.[16] Moreover, the proof of Lemma 1 shows that this is the optimal payoff guarantee using a modified version of w_1. Note that if the incentive gap is small, this value becomes close to $\mathbb{E}_{F_0}[y - w_1(y)]$, and may be better for the principal than simply offering $w_2 = w_1$ again.

After observing a_1, the principal learns that the true technology A must contain $\{a_0, a_1\}$. If this were all the knowledge the principal could acquire from this observation, then her second-period problem would be a single-period problem in Carroll (2015). The optimal second-period contract would be linear, and could be obtained by (i) first identifying the element in $\{a_0, a_1\}$ that attains $\max\left\{\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}, \sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right\}$, denoted by $a^* = (F^*, c^*)$, and then (ii) choosing $s_2 = \sqrt{c^*/\mathbb{E}_{F^*}[y]}$ to be the share and offering the linear contract $w_2(y) = s_2 y$.

However, as in the second requirement of compatibility (Definition 1), the principal can infer more from observing a_1, that the true technology A does not contain any action strictly better for agent 1. Note that this inference rules out the possibility that certain actions exist in A. Therefore, the payoff guarantee of the optimal contract found according to the procedure in Carroll (2015) can only get higher. That is, by offering the better one of the two linear contracts, $w_2(y) = s_2 y$ with $s_2 = \sqrt{c_0/\mathbb{E}_{F_0}[y]}$ or $s_2 = \sqrt{c_1/\mathbb{E}_{F_1}[y]}$, the principal can guarantee that her payoff in the second period is at least $\left(\max\left\{\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}, \sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right\}\right)^2$.[17] Parts 3 and 4 of Lemma 1 show that, when this payoff guarantee is larger than the previous two cases (w_1 again, or a modified w_1 with compensation), it is optimal for the principal to offer the better of the two linear contracts, and doing so exactly attains this payoff guarantee.

We are now ready to present the main result of this section, Lemma 1, which establishes the optimality of the aforementioned contracts. The principal's optimal second-period payoff guarantee is achieved by offering the best among the four contracts described above: w_1 again, modified w_1 with compensation, and the two linear contracts.

---

[16] In the numerical example, this corresponds to the case where agent 1 chooses the low-cost, low-output action a_1' in response to equal split contract. The incentive gap $g\left(w_1, a_1'\right) = (1200/2 - 90) - (2000/2 - 500) = 10$, and $\mathbb{E}_{F_0}[y - w_1(y)] = 2000/2 = 1000$. The principal's optimal second-period payoff guarantee is

$$\left(\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g(w_1, a_1)}\right)^2 = \left(\sqrt{1000} - \sqrt{10}\right)^2 = 810.$$

Part 2 of Lemma 1 shows that this guarantee is attained by offering agent 2 a contract with a share of

$$50\% + \sqrt{10/1000} \times 50\% = 55\%.$$

[17] In the numerical example, this corresponds to the case where agent 1 chooses the low-cost, high-output action a_1 in response to the equal split contract. We have $\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1} > \sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}$, and the optimal share is $\sqrt{c_1/\mathbb{E}_{F_1}[y]} = \sqrt{250/4000} = 1/4$.

Lemma 1. *Suppose the principal offers first-period contract w_1, and agent 1 chooses $a_1 = (F_1, c_1)$ in response. The principal's optimal second-period payoff guarantee is $V_2^*(w_1, a_1) = \Phi(w_1, a_1)^2$, where*

$$\Phi(w_1, a_1) \equiv \max\left\{ \sqrt{\mathbb{E}_{F_1}[y - w_1(y)]},\ \sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g(w_1, a_1)},\ \sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0},\ \sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1} \right\}$$

$$\text{(with } \sqrt{x} = -\infty \text{ for } x < 0 \text{ by convention).} \quad (1)$$

Specifically,

1. *If $\sqrt{\mathbb{E}_{F_1}[y - w_1(y)]}$ attains the maximum in equation (1), then the principal's optimal second-period payoff guarantee is achieved by $w_2 = w_1$.*
2. *If $\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g(w_1, a_1)}$ attains the maximum in equation (1), then the principal's optimal second-period payoff guarantee is achieved by*

$$w_2(y) = w_1(y) + m \cdot (y - w_1(y)) \quad \text{with} \quad m = \sqrt{\frac{g(w_1, a_1)}{\mathbb{E}_{F_0}[y - w_1(y)]}} \in [0, 1]. \quad (2)$$

3. *If $\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}$ attains the maximum in equation (1), then the principal's optimal second-period payoff guarantee is achieved by $w_2(y) = s_2 y$ with $s_2 = \sqrt{c_0 / \mathbb{E}_{F_0}[y]}$.*
4. *If $\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}$ attains the maximum in equation (1), then the principal's optimal second-period payoff guarantee is achieved by $w_2(y) = s_2 y$ with $s_2 = \sqrt{c_1 / \mathbb{E}_{F_1}[y]}$.*

Proof. All proofs of the results in the main text are in Appendix A. □

The statements in Lemma 1 are about the optimality within the entire set of contracts, including those that involve arbitrary kinks or curvatures.

The proof of Lemma 1 mainly consists of two parts. The first part is to prove that, when each element in the quadruple defined by equation (1) attains the maximum, the principal's payoff guarantee in the second period from offering the corresponding contract is exactly as claimed in the statement of Lemma 1. This requires providing lower bounds on the principal's second-period payoffs, and constructing worst-case technologies to show that the bounds are tight. The second part is to show that, under arbitrary second-period contracts, the principal's payoff guarantee is not strictly higher than $\Phi(w_1, a_1)^2$. This requires constructing worst-case technologies to show that the payoff guarantee is lower than (the square of) at least one of element in the quadruple.

Lemma 1 indicates that, as long as the first-period contract w_1 is nonlinear, and the observed action a_1 is such that one of the first two elements in the quadruple defined by equation (1) attains the maximum, then the principal's optimal second-period guarantee $V_2^*(w_1, a_1)$ is achieved by nonlinear contracts. On the other hand, for linear first-period contracts w_1, the four contracts mentioned in the statement of Lemma 1 are all linear. This shows that optimal way for the principal to respond to the knowledge gained is closely related to the specific approach she chooses to explore in the first period.

Moreover, the exact characterization of $V_2^*(w_1, a_1)$ is very useful for the analysis of the principal's first period in the next section. First, this provides a tool to compare the overall payoff guarantee between different first-period contracts. The comparison is crucial in the proof of Lemma 2, which shows that any nonlinear first-period contract can be improved by a linear one. Second, despite having a rather complicated form, the expression (1) is the maximum of four continuous functions (in the appropriate sense of continuity). The continuity is key to the proof of Lemma 3, which shows that within the class of linear first-period contracts, there exists an optimal one for the principal.

4 First Period Analysis

In the previous section, we have focused on principal's problem in the second period and fully characterized her optimal second-period payoff guarantee. This section analyzes the principal's first-period problem in the dynamic relationship, that is, choosing a first-period contract w_1 to maximize her overall payoff guarantee $U(w_1)$.

We first state the main result of this section, Theorem 1, which establishes the optimality of a linear first-period contract.

Theorem 1. *There exists a linear first-period contract w_1 that maximizes the principal's overall payoff guarantee $U(w_1)$.*

The principal's optimal overall payoff guarantee is achieved through a linear first-period contract, together with an optimally chosen linear second-period contract. Even with the opportunity to use the first-period contract as a means of exploration, no other more complicated form of contracts provides the principal with a better payoff guarantee than linear ones.

The proof of Theorem 1 boils down to two steps. The first step, Lemma 2, shows that any nonlinear first-period contract is outperformed by some linear one. The second step, Lemma 3, further shows that the maximum of the principal's first-period problem exists within the class of linear first-period contracts.

4.1 Proof Step 1: Improving Nonlinear Contracts

We start from any arbitrary first-period contract w_1, and construct another linear contract $\hat{w}_1$ that provides the principal with a weakly higher overall payoff guarantee. Thus, any nonlinear contract can be improved by a linear one.

For any first-period contract w_1, let $\hat{w}_1$ denote the following linear contract:

$$\hat{w}_1(y) = s_1 y \quad \text{with} \quad s_1 = \frac{\mathbb{E}_{F_0}[w_1(y)]}{\mathbb{E}_{F_0}[y]} \geq 0. \tag{3}$$

The procedure of constructing the linear $\hat{w}_1$ is depicted in Figure 1. The solid curve represents first-period

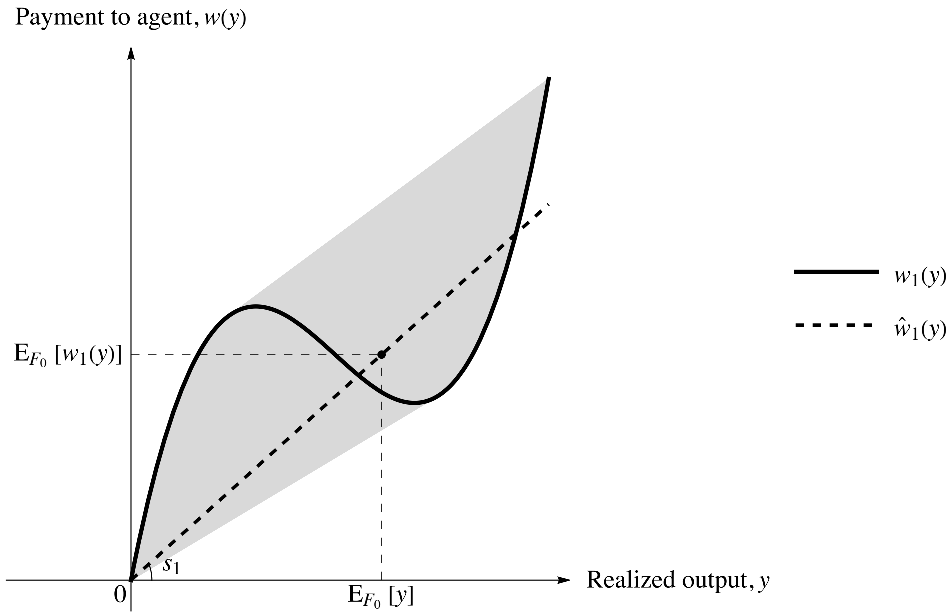

Figure 1: The linear contract $\hat{w}_1$ constructed from w_1.

contract w_1, which may be nonlinear and non-monotonic. Consider the point $(\mathbb{E}_{F_0}[y], \mathbb{E}_{F_0}[w_1(y)])$, whose coordinates are the expected output and the expected payment to agent 1 if he takes the known action $a_0 = (F_0, c_0)$. This point must lie within the convex hull of the curve w_1, represented by the shaded area in the figure. The constructed linear contract $\hat{w}_1$ is exactly the dashed line connecting the origin and this point, with a corresponding slope denoted by s_1.

Note that the linear contract $\hat{w}_1$ is chosen such that if agent 1 takes the action a_0, his payoff will be exactly equal under $\hat{w}_1$ as under w_1:

$$\mathbb{E}_{F_0}[\hat{w}_1(y)] - c_0 = s_1 \mathbb{E}_{F_0}[y] - c_0 = \mathbb{E}_{F_0}[w_1(y)] - c_0.$$

We will show that the principal's overall payoff guarantee is at least as high under $\hat{w}_1$ as it is under w_1; that is, $U(\hat{w}_1) \geq U(w_1)$.[18]

Lemma 2. *Let w_1 be any first-period contract. The linear contract $\hat{w}_1$ defined by equation (3) satisfies $U(\hat{w}_1) \geq U(w_1)$.*

Suppose the principal offers the linear first-period contract $\hat{w}_1$, and agent 1 chooses action a_1 from the true technology A. We need to show that the principal's interim payoff guarantee, $U(\hat{w}_1|a_1)$, is at least $U(w_1)$. If there exists another action a_1', which may be taken by agent 1 under w_1 and some other technology, such that

$$U(\hat{w}_1|a_1) \geq U\left(w_1\middle|a_1'\right) \tag{4}$$

holds, then $U(\hat{w}_1|a_1) \geq U\left(w_1\middle|a_1'\right) \geq U(w_1)$, and thus the desired conclusion is established. The proof of Lemma 2 explicitly constructs such an alternative action a_1' for each possible a_1.

Specifically, the principal's interim payoff guarantee consists of two parts, her payoff in the first period, plus the discounted optimal second-period payoff guarantee. The exact characterization of the second part in Lemma 1 is crucial for the construction of a_1', enabling the desired inequality (4) to hold period by period: under $(\hat{w}_1|a_1)$, the principal's payoff in the first period and her guarantee in the second period are both higher than under $\left(w_1\middle|a_1'\right)$.

By establishing Lemma 2, we have shown that any nonlinear first-period contract can be improved by a linear one. To finalize the proof of Theorem 1, it suffices to show that, within the class of linear contracts, the maximum of $U(w_1)$ exists. In the next subsection, we will set up a program that characterizes the principal's overall payoff guarantee of an arbitrary linear first-period contract, and prove the existence of maximum through its continuity in the first-period share.

4.2 Proof Step 2: Payoff Guarantee of a Linear Contract

To conclude the proof of Theorem 1, we need to establish the following Lemma 3.

Lemma 3. *Within the class of linear first-period contracts, there exists an optimal one for the principal.*

The proof of Lemma 3 requires to characterize the overall payoff guarantee of an arbitrary linear first-period contract, which is the focus of this subsection.

[18]Unlike the main text of Carroll (2015), which uses linear relations between the principal's and agent's payoffs to characterize the payoff guarantee of any contract, this is an adaptation of the alternative approach suggested by Lucas Maestri in Carroll (2015, Appendix C) to the two-period model.

Assume the principal offers a linear first-period contract $w_1(y) = s_1 y$ with $s_1 \in [0,1]$. If agent 1's payoff from taking the known action a_0 is strictly negative, i.e., $\mathbb{E}_{F_0}[w_1(y)] - c_0 = s_1\mathbb{E}_{F_0}[y] - c_0 < 0$, then the principal cannot guarantee any positive payoff in the first period, since it is possible that the agent can produce zero output at zero cost (i.e., $(\delta_0, 0) \in A$), and the agent would strictly prefer this action to a_0. Moreover, according to Lemma 1, the principal's optimal second-period payoff guarantee is $V_2^*(w_1, (\delta_0, 0)) = \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)^2$, the optimal static payoff guarantee in Carroll (2015), because the principal's discovery is of no use to her. This is already strictly worse than offering the alternative contract $s_1' y$ with $s_1' = \sqrt{c_0/\mathbb{E}_{F_0}[y]}$ instead, because doing so guarantees a strictly positive payoff $\left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)^2$ in the first period, and the payoff guarantee in the second period can only get better.

Therefore, when searching for optimal linear contracts, we may concentrate on those with $s_1 \geq c_0/\mathbb{E}_{F_0}[y]$ (i.e., agent 1 obtains nonnegative payoff from choosing a_0). For any such linear first-period contract, suppose that agent 1 chooses $a_1 = (F_1, c_1)$ in response. As is shown in Lemma 1, the principal's optimal second-period payoff guarantee is $\Phi(w_1, a_1)^2$, with Φ defined by equation (1). Thus, her interim payoff guarantee is

$$U(w_1|a_1) = \mathbb{E}_{F_1}[y - w_1(y)] + \beta \cdot \Phi(w_1, a_1)^2 = (1 - s_1)\mathbb{E}_{F_1}[y] + \beta \cdot \Phi(w_1, a_1)^2 .$$

The worst-case overall payoff guarantee minimizes the above expression over all a_1 that agent 1 may choose under some technology. Note that agent 1 prefers action a_1 over the known action a_0 if and only if the incentive gap is nonnegative, i.e., $g(w_1, a_1) \geq 0$, which is equivalent to

$$(\mathbb{E}_{F_1}[w_1(y)] - c_1) - (\mathbb{E}_{F_0}[w_1(y)] - c_0) = (s_1\mathbb{E}_{F_1}[y] - c_1) - (s_1\mathbb{E}_{F_0}[y] - c_0) \geq 0.$$

Hence, the following program yields a lower bound on the principal's overall payoff guarantee

$$\begin{aligned} \inf_{F_1, c_1} \quad & (1 - s_1)\mathbb{E}_{F_1}[y] + \beta \cdot \Phi(w_1, (F_1, c_1))^2 \\ \text{s.t.} \quad & (s_1\mathbb{E}_{F_1}[y] - c_1) - (s_1\mathbb{E}_{F_0}[y] - c_0) \geq 0, \end{aligned} \tag{5}$$

because the principal's interim payoff guarantee can never be strictly lower than the infimum given by program (5).

Conversely, if $s_1 \geq c_0/\mathbb{E}_{F_0}[y]$, then for any feasible $a_1 = (F_1, c_1)$ in program (5), agent 1 would take action a_1 in response to w_1 whenever the technology A is compatible with (w_1, a_1). The worst case over all such technologies leaves the principal with exactly her interim payoff guarantee, $U(w_1|a_1) =$

$(1 - s_1)\,\mathbb{E}_{F_1}[y] + \beta \cdot \Phi(w_1, a_1)^2$. Thus, if a solution to program (5) exists (i.e., if infimum may be replaced by minimum), then the principal's payoff guarantee cannot be strictly higher than its minimum value.

The above analysis shows that, for $s_1 \geq c_0/\mathbb{E}_{F_0}[y]$, the worst-case overall payoff guarantee of any linear first-period contract $w_1(y) = s_1 y$ is exactly characterized by program (5). In the proof of Lemma 3 in Appendix A.2, we formally show the existence of minimum in this program, and its continuity in the first-period share s_1. We first reformulate program (5) as an equivalent maximization problem with continuous objective function and compact feasible region, and then invoke Berge's maximum theorem to prove the required existence and continuity. Since the overall payoff guarantee of a linear first-period contract $w_1(y) = s_1 y$ is continuous in the first-period share s_1, it achieves a maximum. This maximum is also the optimal guarantee over all linear contracts.

Specifically, under a linear first-period contract w_1, the expression of $V_2^*(w_1, a_1) = \Phi(w_1, a_1)^2$ given by equation (1) gets simplified, thus showing that both the objective and the constraint of program (5) depend on the choice variables (F_1, c_1) only through the value of $(\mathbb{E}_{F_1}[y], c_1)$. Moreover, Φ is the maximum of four continuous functions, so it is itself continuous. To complete the proof, we only need to show that the value of $(\mathbb{E}_{F_1}[y], c_1)$ can be restricted to a compact region without affecting the infimum value of program (5), and that region changes in a continuous[19] manner when the first period share s_1 changes.

Combining Lemmas 2 and 3, we prove the main result of this section, Theorem 1, which establishes the optimality of a linear first-period contract. Although Lemma 1 in the previous section shows that, following nonlinear first-period contracts, optimal second-period contracts may also be nonlinear in some cases, here we demonstrate that the principal's optimal overall payoff guarantee is achieved by a linear first-period contract (along with an optimally chosen linear second-period contract). In our model, the principal has the opportunity to explore in the first period, and linear first-period contracts are optimal in terms of utilizing the exploration opportunity, making them even more robust.

5 Further Results

So far we have shown that, in the baseline model, linear contracts are optimal for the principal in both periods, even with the opportunity to use the first-period contract as a means of exploration. This section presents several natural extensions of the baseline model and further results.

In Subsection 5.1, we analyze the cases where the principal knows a *set of actions* available to the agents. The first main result is Lemma 1′, which characterizes the principal's optimal second-period payoff

[19] In the language of correspondences, both upper and lower hemicontinuous.

guarantee $V_2^*(w_1, a_1)$ in closed form and identifies the contract that attains it in various cases, analogous to Lemma 1 in the baseline model. Furthermore, we identify a sufficient condition on the set of known actions, *lower bound on marginal cost* (Definition 3), which ensures that linear contracts still outperform nonlinear ones. This leads to the second main result, Theorem 1′, which generalizes the optimality of linear contracts to richer environments.

Then in Subsection 5.2, we investigate situations where the agents' technology may advance between periods. In this case, the principal adopts a different rule of updating (instead of *compatibility*). As we show in Theorem 2, linear contracts remain optimal in both periods.

Furthermore, in Subsection 5.3 we examine the structure of the optimal linear first-period contract in our dynamic model, and compare it with the optimal static contract identified by Carroll (2015).

5.1 General Set of Known Actions

Suppose that the principal knows not only one action a_0 available to the agents, but a general compact set A_0 of available actions. To ensure that the principal may benefit from contracting with the agents, assume that there exists $(F, c) \in A_0$ such that $\mathbb{E}_F[y] - c > 0$.

In the first period, the principal believes that the true technology A could be any technology such that $A \supseteq A_0$. After the principal offers contract w_1 and observes the action a_1 chosen by agent 1, we adapt the rule of updating, *compatibility*, as follows:

Definition 1′ (Compatible). *Given w_1 and $a_1 = (F_1, c_1)$, a technology A is* compatible *with (w_1, a_1) if*

1. $A \supseteq A_0 \cup \{a_1\}$.
2. $\mathbb{E}_F[w_1(y)] - c \leq \mathbb{E}_{F_1}[w_1(y)] - c_1$ *for all* $(F, c) \in A$.

Throughout the analysis below, we denote the output distribution and cost associated with a generic action a by, respectively, F_a and c_a.

5.1.1 Second Period Analysis

As in the baseline model, we first consider the second period of the dynamic relationship, where the principal has offered some first-period contract w_1 and observed agent 1's chosen action $a_1 = (F_1, c_1)$. She learns that the true technology A is compatible with (w_1, a_1): it contains A_0 and a_1, and does not contain any action strictly better than a_1 for agent 1 under w_1.

Again, if the principal offers the same contract $w_2 = w_1$, agent 2 will choose a_1 since the two agents have the same technology. This exactly repeats the first-period payoff $\mathbb{E}_{F_1}[y - w_1(y)]$ in the second period. Moreover, if some initially known action $(F_0, c_0) \in A_0$ leads to a higher payoff for the principal, i.e., $\mathbb{E}_{F_0}[y - w_1(y)] > \mathbb{E}_{F_1}[y - w_1(y)]$, it might be tempting for the principal to try to obtain the payoff $\mathbb{E}_{F_0}[y - w_1(y)]$ instead. However, we have already seen that achieving this payoff would violate agent 2's incentive constraint, and agent 2 needs to be compensated for not choosing a_1. The amount of compensation increases with the *incentive gap*, which may now vary for different actions.

Definition 2′ (Incentive gap)**.** *Given w_1 and $a_1 = (F_1, c_1)$, the* incentive gap *with respect to an action a, $g(a|w_1, a_1)$, denotes the difference in agent* 1*'s payoff between choosing a_1 and a. Formally,*

$$g(a|w_1, a_1) \equiv (\mathbb{E}_{F_1}[w_1(y)] - c_1) - (\mathbb{E}_{F_a}[w_1(y)] - c_a).$$

Analogous to Lemma 1, part 1 of Lemma 1′ shows that if $\mathbb{E}_{F_0}[y - w_1(y)] > g(a_0|w_1, a_1)$, the principal can offer a modified version of w_1 with compensation in order to guarantee that her payoff in the second period is at least $\left(\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g(a_0|w_1, a_1)}\right)^2$. Let

$$\Phi_1(w_1, a_1) \equiv \max_{a \in A_0 \cup \{a_1\}} \left\{ \sqrt{\mathbb{E}_{F_a}[y - w_1(y)]} - \sqrt{g(a|w_1, a_1)} \right\}, \tag{6}$$

where we treat $w_2 = w_1$ as a special case of a modified version of w_1 (with no modification).[20] The proof of Lemma 1′ further shows that $\Phi_1(w_1, a_1)^2$ is the principal's optimal guarantee using a modified version of w_1.

Note that the optimal static contract in Carroll (2015) is still available to the principal:

1. Maximize $\sqrt{\mathbb{E}_{F_a}[y]} - \sqrt{c_a}$ over $a \in A_0 \cup \{a_1\}$, with solution $a^* = (F^*, c^*)$.
2. Set $s_2 = \sqrt{c^* / \mathbb{E}_{F^*}[y]}$ as the share, and offer linear contract $w_2(y) = s_2 y$.

Let

$$\Phi_2(a_1) \equiv \max_{a \in A_0 \cup \{a_1\}} \left\{ \sqrt{\mathbb{E}_{F_a}[y]} - \sqrt{c_a} \right\}. \tag{7}$$

By offering the optimal static contract following the procedure in Carroll (2015), the principal can guarantee that her payoff in the second period is at least $\Phi_2(a_1)^2$. Moreover, part 2 of Lemma 1′ shows that when

[20] By definition, $g(a_1|w_1, a_1) = 0$. Moreover, it follows from agent 1's rationality that $g(a_0|w_1, a_1) \geq 0$ for all $a_0 \in A_0$.

$\Phi_2(a_1) > \Phi_1(w_1, a_1)$, it is optimal for the principal to offer this optimal static contract in the second period, and doing so exactly attains payoff guarantee $\Phi_2(a_1)^2$.

We are now ready to present the main result of this subsection, Lemma 1′, which characterizes the principal's optimal second-period payoff guarantee $V_2^*(w_1, a_1)$, and establishes the optimality of the aforementioned contracts. It is optimal for the principal to offer either a modified version of w_1 with compensation, or a linear contract.

Lemma 1′. *Suppose the principal offers first-period contract w_1, and agent 1 chooses a_1 in response. The principal's optimal second-period payoff guarantee is*

$$V_2^*(w_1, a_1) = \left(\max\left\{\Phi_1(w_1, a_1), \Phi_2(a_1)\right\}\right)^2. \tag{8}$$

Specifically,

1. *If $\Phi_1(w_1, a_1) \geq \Phi_2(a_1)$ and $a^* \in A_0 \cup \{a_1\}$ attains the maximum in equation (6), then the principal's optimal second-period payoff guarantee is achieved by a modified version of w_1:*

$$w_2(y) = w_1(y) + m \cdot (y - w_1(y)) \quad \text{with} \quad m = \sqrt{\frac{g(a^*|w_1, a_1)}{\mathbb{E}_{F_{a^*}}[y - w_1(y)]}} \in [0, 1]. \tag{9}$$

2. *If $\Phi_1(w_1, a_1) < \Phi_2(a_1)$ and $a^* \in A_0 \cup \{a_1\}$ attains the maximum in equation (7), then the principal's optimal second-period payoff guarantee is achieved by a linear contract:*

$$w_2(y) = s_2 y \quad \text{with} \quad s_2 = \sqrt{\frac{c_{a^*}}{\mathbb{E}_{F_{a^*}}[y]}}. \tag{10}$$

5.1.2 First Period Analysis

So far, we have focused on principal's problem in the second period and fully characterized her optimal second-period payoff guarantee. Now we analyze the principal's first-period problem of choosing a first-period contract w_1 to maximize her overall payoff guarantee $U(w_1)$.

The following condition, *lower bound on marginal cost*, is sufficient to ensure that the principal's optimal overall payoff guarantee is achieved by a linear first-period contract.

Definition 3 (Lower bound on marginal cost)**.** *The known technology A_0 satisfies* lower bound on marginal cost *if, for any pair of actions $(F, c), (F', c') \in A_0$ with $0 < \mathbb{E}_F[y] < \mathbb{E}_{F'}[y]$, it holds that*

$$c' - c \geq \mathbb{E}_{F'}[y] - \mathbb{E}_F[y].$$

This condition provides linkage between different actions in the known technology A_0. Moreover, it contains the economic meaning that, between known actions, the change in costs cannot be too small compared with the change in expected output. Thus, this condition sets a lower bound on the marginal cost of the known technology in discrete form.

The main result of the first period analysis is Theorem 1′.

Theorem 1′. *Suppose the known technology A_0 satisfies lower bound on marginal cost. There exists a linear first-period contract w_1 that maximizes the principal's overall payoff guarantee $U(w_1)$.*

Analogous to Theorem 1, the proof of Theorem 1′ takes two steps: (1) Lemma 2′ improves any nonlinear first-period contract into a linear one; (2) Lemma 3′ shows that the maximum of the principal's first-period problem exists within the class of linear first-period contracts. We remark that the additional condition, lower bound on marginal cost, comes into play only in the first step of the proof (i.e., Lemma 2′).

We start from any arbitrary first-period contract w_1, and construct another linear contract $\hat{w}_1$ that provides the principal with a weakly higher overall payoff guarantee. Let $a_0 = (F_0, c_0)$ be the action agent 1 will choose if the true technology $A = A_0$. The procedure of constructing the linear $\hat{w}_1$ is exactly the same as in the proof of Lemma 2, given by equation (3). When the known technology satisfies lower bound on marginal cost, Lemma 2′ below shows that the principal's overall payoff guarantee is at least as high under $\hat{w}_1$ as it is under w_1.

Lemma 2′. *Suppose the known technology A_0 satisfies lower bound on marginal cost. Let w_1 be any first-period contract, and let $(F_0, c_0) \in A_0$ be agent 1's best response when the true technology A is just A_0. The linear contract $\hat{w}_1$ defined by equation (3) satisfies $U(\hat{w}_1) \geq U(w_1)$.*

Similar to the proof of Lemma 2, for any action that may be taken by agent 1 under $\hat{w}_1$ and some technology $A \supseteq A_0$, the proof of Lemma 2′ explicitly constructs an alternative action a_1' that may be taken by agent 1 under w_1 and some other technology. The difference between this general case and the baseline model is that the principal's optimal second-period payoff guarantee V_2^* is given by a more general expression (8), and in particular may be attained by $a^* \in A_0 \setminus \{a_0\}$. The condition *lower bound on marginal cost* disciplines the relationship between a_0 and a^*, which makes the proof method of Lemma 2 generalizable. In subsequent research, we hope to examine whether this (or any such) restriction is necessary, in the sense that there exists a counterexample when it is violated.

By establishing Lemma 2′, we have shown that any nonlinear first-period contract can be improved by a linear one. To finalize the proof of Theorem 1′, it suffices to show that, within the class of linear contracts, the maximum of $U(w_1)$ exists.

Lemma 3′. *Within the class of linear first-period contracts, there exists an optimal one for the principal.*

The proof of Lemma 3′ requires to characterize the overall payoff guarantee of an arbitrary linear first-period contract. Assume the principal offers a linear first-period contract $w_1(y) = s_1 y$ with $s_1 \in [0, 1]$, and agent 1 chooses $a_1 = (F_1, c_1)$ in response. As is shown in Lemma 1′, the principal's optimal second-period payoff guarantee $V_2^* (w_1, a_1) = (\max \{\Phi_1 (w_1, a_1), \Phi_2 (a_1)\})^2$, with Φ_1 defined by equation (6) and Φ_2 defined by equation (7). Thus, her interim payoff guarantee is

$$U (w_1|a_1) = \mathbb{E}_{F_1} [y - w_1(y)] + \beta \cdot V_2^* (w_1, a_1) = (1 - s_1) \mathbb{E}_{F_1}[y] + \beta \cdot V_2^* (w_1, a_1).$$

The worst-case overall payoff guarantee minimizes the above expression over all a_1 that agent 1 may choose under some technology. Note that agent 1 prefers action a_1 over all known actions $a \in A_0$ if and only if the incentive gap with respect to each $a \in A_0$ is nonnegative, i.e., $g (a|w_1, a_1) \geq 0$, which is equivalent to

$$(\mathbb{E}_{F_1} [w_1(y)] - c_1) - (\mathbb{E}_{F_a} [w_1(y)] - c_a) = (s_1\mathbb{E}_{F_1}[y] - c_1) - (s_1\mathbb{E}_{F_a}[y] - c_a) \geq 0, \quad \forall a \in A_0.$$

Moreover, agent 1 obtains at least his reservation payoff of zero, which can also be viewed as his payoff from the null action $(\delta_0, 0)$.

Hence, the following program yields a lower bound on the principal's overall payoff guarantee

$$\begin{aligned} \inf_{F_1, c_1} \quad & (1 - s_1) \mathbb{E}_{F_1} [y] + \beta \cdot V_2^* (w_1, (F_1, c_1)) \\ \text{s.t.} \quad & (s_1\mathbb{E}_{F_1}[y] - c_1) - (s_1\mathbb{E}_{F_a}[y] - c_a) \geq 0, \quad \forall a \in A_0 \cup \{(\delta_0, 0)\}, \end{aligned} \tag{11}$$

because the principal's interim payoff guarantee can never be strictly lower than the infimum given by program (11).

Conversely, for any feasible $a_1 = (F_1, c_1)$ in program (11), agent 1 would take action a_1 in response to w_1 whenever the technology A is compatible with (w_1, a_1). The worst case over all such technologies leaves the principal with exactly her interim payoff guarantee, $U (w_1|a_1) = (1 - s_1) \mathbb{E}_{F_1} [y] + \beta \cdot V_2^* (w_1, a_1)$. Thus, if a solution to program (11) exists, then the principal's payoff guarantee cannot be strictly higher than its minimum value.

Therefore, the worst-case overall payoff guarantee of any linear first-period contract $w_1 (y) = s_1 y$ is exactly characterized by program (11). In the proof of Lemma 3′ in Appendix A.3.2, we formally show the existence of minimum in this program, and its continuity in the first-period share s_1 using Berge's maximum theorem. Since the overall payoff guarantee is continuous in the first-period share s_1, it achieves a maximum.

This maximum is also the optimal guarantee over all linear contracts.

Combining Lemmas 2′ and 3′, we prove the main result of this section, Theorem 1′, which establishes the optimality of a linear first-period contract.

5.2 Technological Advances

The baseline model assumes that the two agents have the same technology. Now we analyze the case where the technology may advance between periods. Specifically, agent t has technology $A_t \subseteq \Delta(Y) \times \mathbb{R}^+$, with $A_1 \subseteq A_2$. The principal knows a general compact set A_0 of available actions, and there exists $(F, c) \in A_0$ such that $\mathbb{E}_F[y] - c > 0$.

The principal still maximizes her worst-case expected discounted sum of payoffs over all possible technologies. In the first period, she and believes that agent 1's technology A_1 could be any technology such that $A_1 \supseteq A_0$. Taking into account possible technological advances after the first period, the principal's rule of updating is defined as follows:

$$\text{After the principal offers contract } w_1 \text{ and observes the action } a_1 \text{ chosen by agent 1, she believes that agent 2's technology } A_2 \text{ could be any technology such that } A_2 \supseteq A_0 \cup \{a_1\}. \tag{12}$$

That is, as in the definition of compatibility, the principal learns that action a_1 exists in A_1 (in addition to the initially known set A_0), and believes that agent 2 may also take this action again (since $A_1 \subseteq A_2$). However, unlike in the definition of compatibility, the principal is not able to make the additional inference from agent 1's rationality that excludes certain actions in A_2, due to the possibility of technological advances.

Given the update rule (12), the principal's second-period problem becomes a single-period problem in Carroll (2015). The optimal second-period contract is linear, and the resulting optimal second-period payoff guarantee $\hat{V}_2^*(a_1) = \Phi_2(a_1)^2$ with Φ_2 defined by equation (7). Note that compared to the baseline model, the principal acquires less knowledge from the observation of a_1 under technological advances. As an implication, her optimal second-period payoff guarantee takes a simpler form that does not depend directly on the first-period contract w_1.

In the first period, the principal's interim payoff guarantee if agent 1 chooses $a_1 = (F_1, c_1)$ is

$$\hat{U}(w_1|a_1) \equiv \mathbb{E}_{F_1}[y - w_1(y)] + \beta \cdot \hat{V}_2^*(a_1),$$

and her overall payoff guarantee from w_1 is

$$\hat{U}(w_1) \equiv \inf_{A_1 \supseteq A_0} \left\{ \max_{a_1 \in BR(w_1|A_1)} \hat{U}(w_1|a_1) \right\}.$$

The main result of this subsection is Theorem 2.

Theorem 2. *Under technological advances, there exists a linear first-period contract w_1 that maximizes the principal's overall payoff guarantee $\hat{U}(w_1)$.*

Similar to Theorems 1 and 1′, the proof of Theorem 2 takes two steps: (1) improve any nonlinear first-period contract to a linear one; (2) prove that the maximum of the principal's first-period problem exists within the class of linear first-period contracts. Moreover, due to the simpler form of the principal's second-period payoff guarantee, the additional condition in the previous subsection (lower bound on marginal cost) is not needed for the proof of Theorem 2.

5.3 Optimal First-period Contract

Now we examine the structure of the optimal linear first-period contract in our dynamic model, and compare it with the optimal static contract identified by Carroll (2015). This requires an exact calculation of the overall payoff guarantee from an arbitrary linear first-period contract, which is complicated when the principal knows a general set A_0 of available actions.[21] For this reason, we focus on the case where the principal knows only one action $a_0 = (F_0, c_0)$ available.

In the previous subsection, we demonstrate that the principal's second-period payoff guarantee takes a simpler form under possible technological advances. It turns out that the principal's overall payoff guarantee is also easier to characterize in this situation. In the proof of Theorem 2 in Appendix A.3.3, we set up a program (A.24) that characterizes the principal's overall payoff guarantee from any linear first-period contract, which is an analogue to program (5) in the baseline model.

We explicitly solve the program (A.24) for any first-period share s_1, and the resulting overall payoff guarantee $\hat{U}$ is depicted in Figure 2. From this calculation, we can show that the optimal first-period share s_1^* exists and is unique. Moreover, in Figure 2, the optimal first-period share is larger than $s_0 \equiv \sqrt{c_0/\mathbb{E}_{F_0}[y]}$, the optimal static share.

[21] In particular, in response to a linear first-period contract $w(y) = s_1 y$, the optimal payoff that agent 1 can obtain from *known* actions, $\max_{a \in A_0} \{s_1 \mathbb{E}_{F_a}[y] - c_a\}$, changes with respect to s_1 in an intractable way. This payoff is a key component of the constraint in the program that characterizes the principal's overall payoff guarantee (e.g., program (11)).

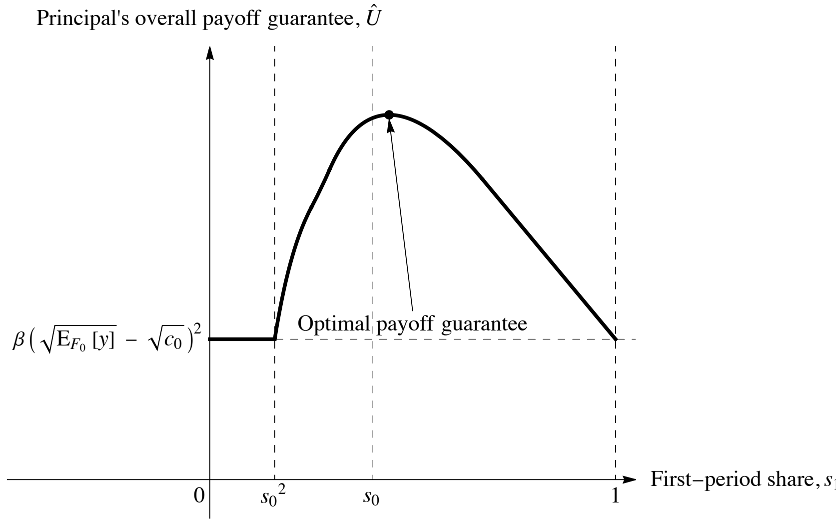

Figure 2: Overall payoff guarantee under technological advances ($s_0 = 0.4, \beta = 0.8$).

Proposition 1 formally establishes this observation and exactly characterizes the optimal first-period share. It reveals an *exploration effect* where the optimal first-period share offered to the agent is always larger than the optimal static share s_0. Moreover, the exploration effect gets larger as the principal becomes more patient (β increases) , provided that $\beta < 1$. When $\beta > 1$, it starts to decrease, and vanishes as $\beta \to \infty$.

Proposition 1. *Suppose the principal knows only one available action $a_0 = (F_0, c_0)$, and let $s_0 \equiv \sqrt{c_0/\mathbb{E}_{F_0}[y]}$ denote the optimal static share. Under technological advances, the optimal first-period share s_1^* is unique, and satisfies the following properties:*

1. *For all $\beta \in (0, \infty)$, the optimal first-period share is larger than the optimal static share, i.e., $s_1^* > s_0$.*
2. *In both limiting cases $\beta \to 0$ and $\beta \to \infty$, s_1^* approaches s_0.*
3. *s_1^* is strictly increasing in β if $\beta < 1$, and is strictly decreasing if $\beta > 1$.*

Proof of Proposition 1. Available upon request. □

The pattern identified by Proposition 1 is shown in Figure 3.

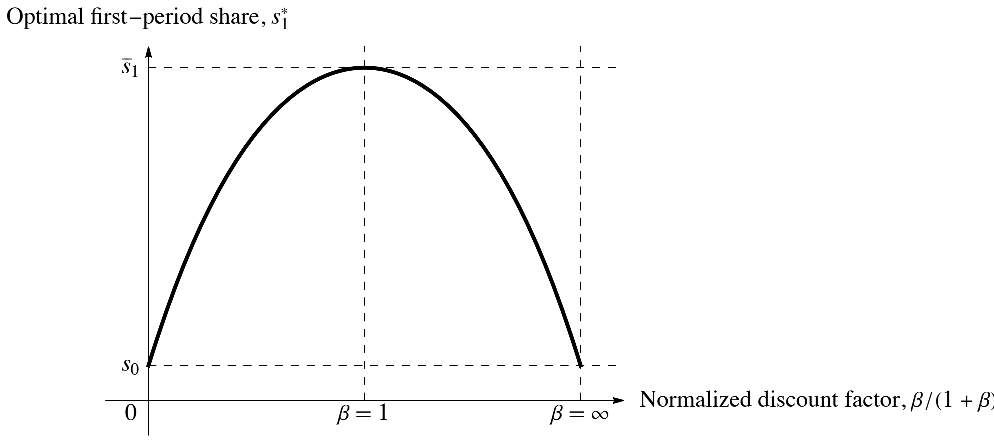

Figure 3: The optimal first-period share s_1^* under technological advances ($s_0 = 0.4$).

It is straightforward to understand the result that the dynamic model converges to the static model as the discount factor β approaches 0. To get intuition behind the opposite case, that is, when β approaches infinity, the optimal first-period share s_1^* approaches the optimal static share s_0 again, note that unlike in standard models where patience automatically leads to the option value of exploration, here the principal is concerned with the worst-case discovery. In the limiting case $\beta \to \infty$ where only the second period matters, there is no incentive for her to raise the first-period share s_1 from s_0, precisely because the worst-case technology always leaves the principal without any valuable discovery. The principal is essentially indifferent among any first-period contract in this limiting case, making the opportunity to explore in the first period completely useless to her.

In the baseline model without technological advances, the principal adopts a more complex rule of updating (i.e., compatibility). Under all possible parameters choices, we aim to compute the exact solution to program (5), which characterizes the overall payoff guarantee of any linear first-period contract $w_1(y) = s_1 y$. Current results show that, for a range of parameter values (specifically, β not too large), the resulting worst-case payoff guarantee U is a bell-shaped curve as depicted in Figure 4. From this figure, the optimal first-period share seems to be unique, and smaller than the optimal static share s_0.

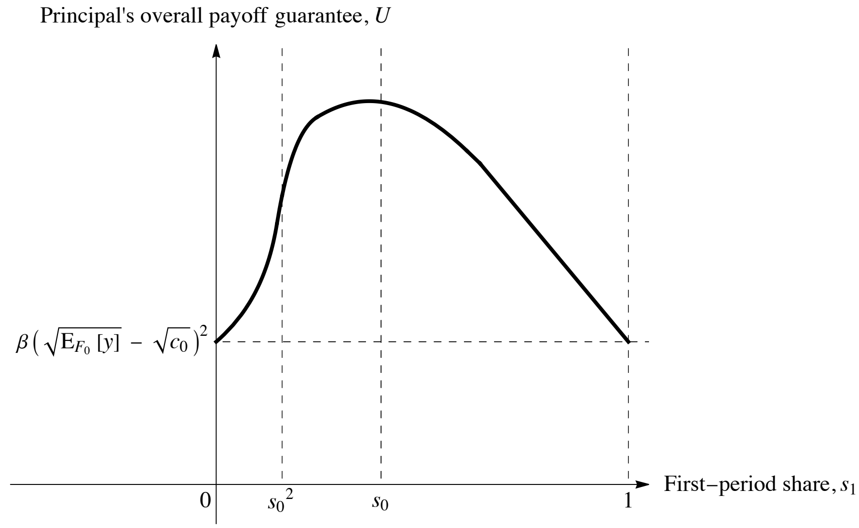

Figure 4: Overall payoff guarantee in the baseline model ($s_0 = 0.4, \beta = 0.8$).

We hope to finish the subsequent calculations to formally establish this observation, in order to better understand the exploration effect in the baseline model. In particular, we wonder whether the optimal first-period share s_1^* approaches the optimal static share s_0 again as the discount factor β approaches infinity.

6 Conclusion

In this paper, we study a two-period moral hazard problem, where the principal does not know the set of actions available to the agents and demands contracts to be robust to this uncertainty; she has the opportunity to explore in the first period and observes the chosen action, and then offers a new contract to the second agent with the same action set based on this knowledge. We define a suitable rule of updating and characterize the principal's optimal payoff guarantee, thereby identifying how the principal should respond to knowledge and design new contracts. The results show that linear contracts are robustly optimal not just in static settings, but also in dynamic environments with exploration.

We consider a contribution of this paper to propose one possible way to extend robust models in mechanism design to allow for multiple interactions and exploration. Despite the nonquantifiable uncertainty, designers can gradually gain a better understanding of the environment in which they repeatedly engage, using *compatibility* as a rule of updating. We hope the generality of this approach in other models will be further explored in future work.

References

Antić, Nemanja, (2021). "Contracting with Unknown Technologies." Working paper.

Barron, Daniel, George Georgiadis, and Jeroen M. Swinkels, (2020). "Optimal Contracts with a Risk-taking Agent." *Theoretical Economics* 15 (2):715–761.

Carroll, Gabriel, (2015). "Robustness and Linear Contracts." *American Economic Review* 105 (2):536–63.

———, (2019). "Robustness in Mechanism Design and Contracting." *Annual Review of Economics* 11 (1):139–166.

Carroll, Gabriel and Lukas Bolte, (Forthcoming). "Robust Contracting under Double Moral Hazard." *Theoretical Economics* .

Dai, Tianjiao and Juuso Toikka, (2022). "Robust Incentives for Teams." *Econometrica* 90 (4):1583–1613.

Diamond, Peter, (1998). "Managerial Incentives: On the Near Linearity of Optimal Compensation." *Journal of Political Economy* 106 (5):931–957.

Dütting, Paul, Tim Roughgarden, and Inbal Talgam-Cohen, (2020). "Simple versus Optimal Contracts." Working paper. Available at arXiv: https://arxiv.org/abs/1808.03713.

Epstein, Larry G. and Martin Schneider, (2003). "Recursive Multiple-Priors." *Journal of Economic Theory* 113 (1):1–31.

Hellwig, Martin F. and Klaus M. Schmidt, (2002). "Discrete-Time Approximations of the Holmström-Milgrom Brownian-Motion Model of Intertemporal Incentive Provision." *Econometrica* :2225–2264.

Holmström, Bengt and Paul Milgrom, (1987). "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* :303–328.

Kambhampati, Ashwin, (2022). "Robust Performance Evaluation of Independent and Identical Agents." Working paper.

Libgober, Jonathan and Xiaosheng Mu, (2023). "Coasian Dynamics under Informational Robustness." Working paper. Available at arXiv: https://arxiv.org/abs/2202.04616.

Marku, Keler, Sergio Ocampo, and Jean-Baptiste Tondji, (Forthcoming). "Robust Contracts in Common Agency." *The RAND Journal of Economics* .

Schättler, Heinz and Jaeyoung Sung, (1993). “The First-Order Approach to the Continuous-Time Principal-Agent Problem With Exponential Utility.” *Journal of Economic Theory* 61 (2):331–371.

Sung, Jaeyoung, (1995). “Linearity With Project Selection and Controllable Diffusion Rate in Continuous-Time Principal-Agent Problems.” *The RAND Journal of Economics* :720–743.

Walton, Daniel and Gabriel Carroll, (2022). “A General Framework for Robust Contracting Models.” *Econometrica* 90 (5):2129–2159.

Wilson, Robert, (1987). “Game-Theoretic Analyses of Trading Processes.” In *Advances in Economic Theory: Fifth World Congress*, edited by Truman Fassett Bewley. Cambridge University Press, 33–70.

Appendix A Proofs of Results in the Main Text

A.1 Proofs for Section 3

If the principal offers $w_2 = w_1$, agent 2 will choose a_1 again. This just repeats her first-period payoff $\mathbb{E}_{F_1}[y - w_1(y)]$ in the second period.

To prove Lemma 1, we start by establishing three lemmas, Lemmas A.1, A.2, A.3, to prove that the principal's payoff guarantee in the second period from offering the remaining three contracts, (i) $w_2(y) = w_1(y) + m \cdot (y - w_1(y))$ with m defined by equation (2), (ii) $w_2(y) = s_2 y$ with $s_2 = \sqrt{c_0/\mathbb{E}_{F_0}[y]}$, and (iii) $w_2(y) = s_2 y$ with $s_2 = \sqrt{c_1/\mathbb{E}_{F_1}[y]}$, is exactly as claimed in the statement of Lemma 1.

Lemma A.1. *If $\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g(w_1, a_1)}$ attains the maximum in equation (1), and the principal offers $w_2(y) = w_1(y) + m \cdot (y - w_1(y))$ with m defined by equation (2), then her payoff guarantee in the second period is exactly $\left(\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g(w_1, a_1)}\right)^2$.*

Proof of Lemma A.1. Let $g_0 \equiv g(w_1, a_1) = (\mathbb{E}_{F_1}[w_1(y)] - c_1) - (\mathbb{E}_{F_0}[w_1(y)] - c_0)$.

If $\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g_0}$ attains the maximum in equation (1), then it holds that $\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g_0} \geq \sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0} > 0$, which implies that $m \in [0, 1]$.

Suppose the principal offers $w_2(y) = w_1(y) + m \cdot (y - w_1(y))$ with m defined by equation (2). We first show that this guarantees her at least $\left(\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g_0}\right)^2$.

Let (F_2, c_2) be the action chosen by agent 2. By agent 1's rationality, we have

$$\mathbb{E}_{F_1}[w_1(y)] - c_1 \geq \mathbb{E}_{F_2}[w_1(y)] - c_2.$$

By agent 2's rationality, we have

$$\mathbb{E}_{F_2}[w_2(y)] - c_2 \geq \mathbb{E}_{F_0}[w_2(y)] - c_0.$$

Summing up the two inequalities, we obtain

$$\begin{aligned} m \cdot \mathbb{E}_{F_2}[y - w_1(y)] = \mathbb{E}_{F_2}[w_2(y) - w_1(y)] &\geq (\mathbb{E}_{F_0}[w_2(y)] - c_0) - (\mathbb{E}_{F_1}[w_1(y)] - c_1) \\ &= m \cdot \mathbb{E}_{F_0}[y - w_1(y)] - g_0, \end{aligned}$$

implying that

$$\mathbb{E}_{F_2}[y - w_1(y)] \geq \mathbb{E}_{F_0}[y - w_1(y)] - g_0/m.$$

Therefore, the principal's payoff in the second period is

$$\begin{aligned} \mathbb{E}_{F_2}[y - w_2(y)] &= \mathbb{E}_{F_2}[y - w_1(y)] - m \cdot \mathbb{E}_{F_2}[y - w_1(y)] = (1 - m)\,\mathbb{E}_{F_2}[y - w_1(y)] \\ &\geq (1 - m)\left(\mathbb{E}_{F_0}[y - w_1(y)] - g_0/m\right) = \left(\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g_0}\right)^2, \end{aligned}$$

as desired.

Next we show that her payoff guarantee from $w_2(y) = w_1(y) + m \cdot (y - w_1(y))$ cannot be strictly higher than $\left(\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g_0}\right)^2$, since this is exactly her payoff when the technology is $A = \{a_0, a_1, (F', c')\}$, with $F' = (1 - m) F_0 + m \cdot \delta_0$ and $c' = c_0 - (m \cdot \mathbb{E}_{F_0}[w_1(y)] + g_0)$.

The proof takes three steps.

Step 1 $\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g_0} \geq \sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}$ implies $c_0 \geq m \cdot \mathbb{E}_{F_0}[w_1(y)] + g_0$, so c' is indeed nonnegative.

It suffices to show

$$
\begin{aligned}
& \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} + \sqrt{g_0} \right)^2 \geq m \cdot \mathbb{E}_{F_0}\left[w_1(y)\right] + g_0 \\
\Leftrightarrow \quad & \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \right)^2 \geq m \cdot \mathbb{E}_{F_0}\left[w_1(y)\right] - 2\sqrt{g_0} \cdot \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \right) \\
\Leftrightarrow \quad & \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \right)^2 \geq m \cdot \left(\mathbb{E}_{F_0}\left[w_1(y)\right] - 2\sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \cdot \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \right) \right).
\end{aligned} \tag{A.1}
$$

Note that

$$
\begin{aligned}
& \mathbb{E}_{F_0}\left[w_1(y)\right] - 2\sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \cdot \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \right) \\
= & \mathbb{E}_{F_0}\left[w_1(y)\right] - 2\sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \cdot \frac{\mathbb{E}_{F_0}\left[w_1(y)\right]}{\sqrt{\mathbb{E}_{F_0}[y]} + \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]}} \\
= & \frac{\mathbb{E}_{F_0}\left[w_1(y)\right]}{\sqrt{\mathbb{E}_{F_0}[y]} + \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]}} \cdot \left(\sqrt{\mathbb{E}_{F_0}[y]} + \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} - 2\sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \right) \\
= & \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \right) \cdot \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \right) = \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \right)^2.
\end{aligned}
$$

Therefore, inequality (A.1) is equivalent to

$$
\left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \right)^2 \geq m \cdot \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \right)^2,
$$

which is implied by the assumption that $\sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} \geq \sqrt{g_0}$ (or equivalently, $m \leq 1$).

Step 2 $A = \{a_0, a_1, (F', c')\}$ is compatible with (w_1, a_1). That is, agent 1 chooses a_1 in response to w_1.

Agent 1's payoff from (F', c') is

$$
\begin{aligned}
\mathbb{E}_{F'}\left[w_1(y)\right] - c' &= (1-m)\,\mathbb{E}_{F_0}\left[w_1(y)\right] - c_0 + \left(m \cdot \mathbb{E}_{F_0}\left[w_1(y)\right] + g_0\right) \\
&= \left(\mathbb{E}_{F_0}\left[w_1(y)\right] - c_0\right) + g_0 = \mathbb{E}_{F_1}\left[w_1(y)\right] - c_1,
\end{aligned}
$$

so he would choose $a_1 = (F_1, c_1)$ in response to w_1.

Note that agent 1 is actually indifferent between (F_1, c_1) and (F', c'), and we will show below that agent 2 is indifferent between (F_0, c_0) and (F', c'). Technically to ensure that agent 1 chooses (F_1, c_1) and agent 2 chooses (F', c') we can set $F' = (1 - m + \varepsilon) F_0 + (m - \varepsilon)\delta_0$ and $c' = c_0 - (m \cdot \mathbb{E}_{F_0}\left[w_1(y)\right] + g_0) + \varepsilon \cdot \mathbb{E}_{F_0}\left[w_1(y) + (m/2) \cdot (y - w_1(y))\right]$ then let $\varepsilon \downarrow 0$. Many of the following cases of potential indifference shall be treated similarly, and we omit them for brevity.

Step 3 If $A = \{a_0, a_1, (F', c')\}$, then agent 2 chooses (F', c') in response to w_2, leading to a payoff of $\left(\sqrt{\mathbb{E}_{F_0}\left[y - w_1(y)\right]} - \sqrt{g_0} \right)^2$ for the principal.

Agent 2's payoff from (F', c') is

$$
\begin{aligned}
\mathbb{E}_{F'}\left[w_2(y)\right] - c' &= (1-m)\,\mathbb{E}_{F_0}\left[w_1(y) + m \cdot (y - w_1(y))\right] - c_0 + \left(m \cdot \mathbb{E}_{F_0}\left[w_1(y)\right] + g_0\right) \\
&= \mathbb{E}_{F_0}\left[w_1(y)\right] + m \cdot \mathbb{E}_{F_0}\left[y - w_1(y)\right] - m^2 \cdot \mathbb{E}_{F_0}\left[y - w_1(y)\right] - c_0 + g_0 \\
&= \mathbb{E}_{F_0}\left[w_2(y)\right] - g_0 - c_0 + g_0 = \mathbb{E}_{F_0}\left[w_2(y)\right] - c_0,
\end{aligned}
$$

and his payoff from $a_1 = (F_1, c_1)$ is

$$\begin{aligned}
\mathbb{E}_{F_1}[w_2(y)] - c_1 &= \mathbb{E}_{F_1}[w_1(y) + m \cdot (y - w_1(y))] - c_1 \\
&= m \cdot \mathbb{E}_{F_1}[y - w_1(y)] + (\mathbb{E}_{F_0}[w_1(y)] - c_0) + g_0 \\
&\leq m \cdot \left(\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g_0}\right)^2 + (\mathbb{E}_{F_0}[w_1(y)] - c_0) + g_0 \\
&\leq m \cdot \sqrt{\mathbb{E}_{F_0}[y - w_1(y)]}\left(\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g_0}\right) + (\mathbb{E}_{F_0}[w_1(y)] - c_0) + g_0 \\
&= m \cdot \mathbb{E}_{F_0}[y - w_1(y)] - g_0 + (\mathbb{E}_{F_0}[w_1(y)] - c_0) + g_0 = \mathbb{E}_{F_0}[w_2(y)] - c_0,
\end{aligned}$$

so he would choose (F', c') in response to w_2.

This leaves the principal with payoff of

$$\begin{aligned}
\mathbb{E}_{F'}[y - w_2(y)] &= \mathbb{E}_{F'}[y - w_1(y)] - m \cdot \mathbb{E}_{F'}[y - w_1(y)] = (1 - m)\,\mathbb{E}_{F'}[y - w_1(y)] \\
&= (1 - m)^2\,\mathbb{E}_{F_0}[y - w_1(y)] = \left(\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g_0}\right)^2,
\end{aligned}$$

as desired.

This completes the proof. □

Lemma A.2. *If $\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}$ attains the maximum in equation (1), and the principal offers the linear contract $w_2(y) = s_2 y$ with $s_2 = \sqrt{c_0/\mathbb{E}_{F_0}[y]}$, then her payoff guarantee in the second period is exactly $\left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)^2$.*

Proof of Lemma A.2. Suppose that $\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}$ attains the maximum in equation (1), and the principal offers the linear contract $w_2(y) = s_2 y$ with $s_2 = \sqrt{c_0/\mathbb{E}_{F_0}[y]}$. We first show that this guarantees her at least $\left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)^2$.[22]

Let (F_2, c_2) be the action chosen by agent 2. By agent 2's rationality, we have

$$\mathbb{E}_{F_2}[w_2(y)] - c_2 \geq \mathbb{E}_{F_0}[w_2(y)] - c_0,$$

which further implies that

$$s_2\mathbb{E}_{F_2}[y] = \mathbb{E}_{F_2}[w_2(y)] \geq \mathbb{E}_{F_2}[w_2(y)] - c_2 \geq \mathbb{E}_{F_0}[w_2(y)] - c_0 = s_2\mathbb{E}_{F_0}[y] - c_0,$$

and hence

$$\mathbb{E}_{F_2}[y] \geq \mathbb{E}_{F_0}[y] - c_0/s_2.$$

Therefore, the principal's payoff in the second period is

$$\mathbb{E}_{F_2}[y - w_2(y)] = \mathbb{E}_{F_2}[(1 - s_2)\,y] \geq (1 - s_2)\,(\mathbb{E}_{F_0}[y] - c_0/s_2) = \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)^2,$$

as desired.

Next we show that her payoff guarantee from this linear contract cannot be strictly higher, since $\left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)^2$ is exactly her payoff when the technology is $A = \{a_0, a_1, (F', 0)\}$, with $F' = \lambda F_0 + (1-\lambda)\delta_0$ where $\lambda = 1 - \sqrt{c_0/\mathbb{E}_{F_0}[y]} \in [0, 1]$.

The proof takes two steps. Let $g_0 \equiv g(w_1, a_1) = (\mathbb{E}_{F_1}[w_1(y)] - c_1) - (\mathbb{E}_{F_0}[w_1(y)] - c_0)$.

[22] This is exactly the static payoff guarantee in Carroll (2015).

Step 1 $A = \{a_0, a_1, (F', 0)\}$ is compatible with (w_1, a_1). That is, agent 1 chooses a_1 in response to w_1.

Agent 1's payoff from $(F', 0)$ is $\mathbb{E}_{F'}[w_1(y)] = \lambda \mathbb{E}_{F_0}[w_1(y)] = \left(1 - \sqrt{c_0/\mathbb{E}_{F_0}[y]}\right)\mathbb{E}_{F_0}[w_1(y)]$, and we have

$$\left(1 - \sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}}\right)\mathbb{E}_{F_0}[w_1(y)] \le \mathbb{E}_{F_1}[w_1(y)] - c_1 \quad \Leftrightarrow \quad \left(1 - \sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}}\right)\mathbb{E}_{F_0}[w_1(y)] \le (\mathbb{E}_{F_0}[w_1(y)] - c_0) + g_0$$
$$\Leftrightarrow \quad \sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}}\mathbb{E}_{F_0}[w_1(y)] - c_0 + g_0 \ge 0.$$

From

$$\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0} \ge \sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g_0},$$

we obtain

$$\mathbb{E}_{F_0}[w_1(y)] \ge \mathbb{E}_{F_0}[y] - \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0} + \sqrt{g_0}\right)^2,$$

and thus

$$\begin{aligned}\sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}}\mathbb{E}_{F_0}[w_1(y)] - c_0 + g_0 &\ge \sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}} \cdot \left(\mathbb{E}_{F_0}[y] - \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0} + \sqrt{g_0}\right)^2\right) - c_0 + g_0 \\ &= \left(1 - \sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}}\right)\left(\sqrt{c_0} - \sqrt{g_0}\right)^2 \ge 0,\end{aligned}$$

as desired. So we indeed have $\mathbb{E}_{F'}[w_1(y)] \le \mathbb{E}_{F_1}[w_1(y)] - c_1$, implying that agent 1 would choose $a_1 = (F_1, c_1)$ in response to w_1.

Step 2 If $A = \{a_0, a_1, (F', 0)\}$, then agent 2 chooses $(F', 0)$ in response to w_2, leading to a payoff of $\left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)^2$ for the principal.

Agent 2's payoff from $(F', 0)$ is

$$\begin{aligned}\mathbb{E}_{F'}[w_2(y)] = \lambda \mathbb{E}_{F_0}[s_2 y] &= \left(1 - \sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}}\right) \cdot \sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}} \cdot \mathbb{E}_{F_0}[y] \\ &= \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)\sqrt{c_0} = \sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}} \cdot \mathbb{E}_{F_0}[y] - c_0 \\ &= s_2 \mathbb{E}_{F_0}[y] - c_0 = \mathbb{E}_{F_0}[w_2(y)] - c_0.\end{aligned}$$

His payoff from $a_1 = (F_1, c_1)$ is $\mathbb{E}_{F_1}[w_2(y)] - c_1 = \sqrt{c_0/\mathbb{E}_{F_0}[y]} \cdot \mathbb{E}_{F_1}[y] - c_1$, and we have

$$\sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}} \cdot \mathbb{E}_{F_1}[y] - c_1 \le \mathbb{E}_{F_0}[w_2(y)] - c_0 \quad \Leftrightarrow \quad \sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}} \cdot \mathbb{E}_{F_1}[y] - c_1 \le \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)\sqrt{c_0}.$$

From $\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0} \ge \sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}$, we obtain $\mathbb{E}_{F_1}[y] \le \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0} + \sqrt{c_1}\right)^2$, and thus

$$\begin{aligned}&\left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)\sqrt{c_0} - \left(\sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}} \cdot \mathbb{E}_{F_1}[y] - c_1\right) \\ \ge &\left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)\sqrt{c_0} - \left(\sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}} \cdot \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0} + \sqrt{c_1}\right)^2 - c_1\right) \\ = &\left(1 - \sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}}\right)\left(\sqrt{c_0} - \sqrt{c_1}\right)^2 \ge 0,\end{aligned}$$

as desired. So we indeed have $\mathbb{E}_{F_1}[w_2(y)] - c_1 \le \mathbb{E}_{F_0}[w_2(y)] - c_0 = \mathbb{E}_{F'}[w_2(y)]$, implying that agent 2 would choose $(F', 0)$ in response to w_2.

This leaves the principal with payoff of

$$\begin{aligned}\mathbb{E}_{F'}[y - w_2(y)] = \lambda \mathbb{E}_{F_0}[(1-s_2)y] &= \left(1 - \sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}}\right)\left(1 - \sqrt{\frac{c_0}{\mathbb{E}_{F_0}[y]}}\right) \cdot \mathbb{E}_{F_0}[y] \\ &= \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)^2,\end{aligned}$$

as desired.

This completes the proof. □

Lemma A.3. *If* $\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}$ *attains the maximum in equation* (1)*, and the principal offers the linear contract* $w_2(y) = s_2 y$ *with* $s_2 = \sqrt{c_1/\mathbb{E}_{F_1}[y]}$*, then her payoff guarantee in the second period is exactly* $\left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)^2$.

Proof of Lemma A.3. If $\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}$ attains the maximum in equation (1), then it holds that $\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1} \ge \sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0} > 0$, which implies that $c_1/\mathbb{E}_{F_1}[y] \in [0,1]$.

Suppose the principal offers the linear contract $w_2(y) = s_2 y$ with $s_2 = \sqrt{c_1/\mathbb{E}_{F_1}[y]}$. We first show that this guarantees her at least $\left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)^2$.[23]

Let (F_2, c_2) be the action chosen by agent 2. By agent 2's rationality, we have

$$\mathbb{E}_{F_2}[w_2(y)] - c_2 \ge \mathbb{E}_{F_1}[w_2(y)] - c_1,$$

which further implies that

$$s_2 \mathbb{E}_{F_2}[y] = \mathbb{E}_{F_2}[w_2(y)] \ge \mathbb{E}_{F_2}[w_2(y)] - c_2 \ge \mathbb{E}_{F_1}[w_2(y)] - c_1 = s_2 \mathbb{E}_{F_1}[y] - c_1,$$

and hence

$$\mathbb{E}_{F_2}[y] \ge \mathbb{E}_{F_1}[y] - c_1/s_2.$$

Therefore, the principal's payoff in the second period is

$$\mathbb{E}_{F_2}[y - w_2(y)] = \mathbb{E}_{F_2}[(1-s_2)y] \ge (1-s_2)(\mathbb{E}_{F_1}[y] - c_1/s_2) = \left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)^2,$$

as desired.

Next we show that her payoff guarantee from this linear contract cannot be strictly higher, since this is exactly her payoff when the technology is $A = \{a_0, a_1, (F', 0)\}$, with $F' = \lambda F_1 + (1-\lambda)\delta_0$ where $\lambda = 1 - \sqrt{c_1/\mathbb{E}_{F_1}[y]} \in [0,1]$.

The proof takes two steps.

Step 1 $A = \{a_0, a_1, (F', 0)\}$ is compatible with (w_1, a_1). That is, agent 1 chooses a_1 in response to w_1.

Agent 1's payoff from $(F', 0)$ is $\mathbb{E}_{F'}[w_1(y)] = \lambda \mathbb{E}_{F_1}[w_1(y)] = \left(1 - \sqrt{c_1/\mathbb{E}_{F_1}[y]}\right)\mathbb{E}_{F_1}[w_1(y)]$, and we

[23] This is exactly the static payoff guarantee in Carroll (2015).

have

$$\left(1-\sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}}\right)\mathbb{E}_{F_1}[w_1(y)] \le \mathbb{E}_{F_1}[w_1(y)] - c_1 \quad\Leftrightarrow\quad \left(1-\sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}}\right)\mathbb{E}_{F_1}[w_1(y)] \le \mathbb{E}_{F_1}[w_1(y)] - c_1$$
$$\Leftrightarrow\quad \sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}}\mathbb{E}_{F_1}[w_1(y)] - c_1 \ge 0.$$

From $\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1} \ge \sqrt{\mathbb{E}_{F_1}[y - w_1(y)]}$, we obtain $\mathbb{E}_{F_1}[w_1(y)] \ge \mathbb{E}_{F_1}[y] - \left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)^2$, and thus

$$\sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}}\mathbb{E}_{F_1}[w_1(y)] - c_1 \ge \sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}} \cdot \left(\mathbb{E}_{F_1}[y] - \left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)^2\right) - c_1 = \left(1-\sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}}\right)c_1 \ge 0,$$

as desired. So we indeed have $\mathbb{E}_{F'}[w_1(y)] \le \mathbb{E}_{F_1}[w_1(y)] - c_1$, implying that agent 1 would choose (F_1, c_1) in response to w_1.

Step 2 If $A = \{a_0, a_1, (F', 0)\}$, then agent 2 chooses $(F', 0)$ in response to w_2, leading to a payoff of $\left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)^2$ for the principal.

Agent 2's payoff from $(F', 0)$ is

$$\begin{aligned}
\mathbb{E}_{F'}[w_2(y)] = \lambda \mathbb{E}_{F_1}[s_2 y] &= \left(1-\sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}}\right)\cdot \sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}} \cdot \mathbb{E}_{F_1}[y] \\
&= \left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)\sqrt{c_1} = \sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}} \cdot \mathbb{E}_{F_1}[y] - c_1 \\
&= s_2 \mathbb{E}_{F_1}[y] - c_1 = \mathbb{E}_{F_1}[w_2(y)] - c_1.
\end{aligned}$$

His payoff from (F_0, c_0) is $\mathbb{E}_{F_0}[w_2(y)] - c_0 = \sqrt{c_1/\mathbb{E}_{F_1}[y]} \cdot \mathbb{E}_{F_0}[y] - c_0$, and we have

$$\sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}} \cdot \mathbb{E}_{F_0}[y] - c_0 \le \mathbb{E}_{F_1}[w_2(y)] - c_1 \quad\Leftrightarrow\quad \sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}} \cdot \mathbb{E}_{F_0}[y] - c_0 \le \left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)\sqrt{c_1}.$$

From $\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1} \ge \sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}$, we obtain $\mathbb{E}_{F_0}[y] \le \left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1} + \sqrt{c_0}\right)^2$, and thus

$$\begin{aligned}
&\left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)\sqrt{c_1} - \left(\sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}} \cdot \mathbb{E}_{F_0}[y] - c_0\right) \\
&\ge \left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)\sqrt{c_1} - \left(\sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}} \cdot \left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1} + \sqrt{c_0}\right)^2 - c_0\right) \\
&= \left(1-\sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}}\right)\left(\sqrt{c_1} - \sqrt{c_0}\right)^2 \ge 0,
\end{aligned}$$

as desired. So we indeed have $\mathbb{E}_{F_0}[w_2(y)] - c_0 \le \mathbb{E}_{F_1}[w_2(y)] - c_1 = \mathbb{E}_{F'}[w_2(y)]$, implying that agent 2 would choose $(F', 0)$ in response to w_2.

This leaves the principal with payoff of

$$\begin{aligned}
\mathbb{E}_{F'}[y - w_2(y)] = \lambda \mathbb{E}_{F_0}[(1-s_2)y] &= \left(1-\sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}}\right)\left(1-\sqrt{\frac{c_1}{\mathbb{E}_{F_1}[y]}}\right)\cdot \mathbb{E}_{F_1}[y] \\
&= \left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)^2,
\end{aligned}$$

as desired.

This completes the proof. □

We are now ready to prove Lemma 1.

Proof of Lemma 1. If the principal offers $w_2 = w_1$, this guarantees her payoff in the first-period, which is equal to $\mathbb{E}_{F_1}[y - w_1(y)]$. Note that her payoff guarantee from $w_2 = w_1$ cannot be strictly higher, since this is exactly her payoff when the technology is $A = \{a_0, a_1\}$, which is compatible with (w_1, a_1).

Together with Lemmas A.1, A.2 and A.3, we have shown that by offering the best among the four contracts: (i) $w_2 = w_1$, (ii) $w_2(y) = w_1(y) + m \cdot (y - w_1(y))$ with m defined by equation (2), (iii) $w_2(y) = s_2 y$ with $s_2 = \sqrt{c_0/\mathbb{E}_{F_0}[y]}$, and (iv) $w_2(y) = s_2 y$ with $s_2 = \sqrt{c_1/\mathbb{E}_{F_1}[y]}$, the principal's payoff guarantee in the second period is exactly given by $\Phi(w_1, a_1)^2$, where Φ is defined by equation (1). The principal's optimal second-period payoff guarantee, $V_2^*(w_1, a_1)$, is thus at least $\Phi(w_1, a_1)^2$.

Now consider an arbitrary second-period contract w_2. It suffices to show that the principal's payoff guarantee is not strictly higher than $\Phi(w_1, a_1)^2$ under w_2.

Consider the following two cases.

Case 1. $\mathbb{E}_{F_1}[w_2(y)] - c_1 \geq \mathbb{E}_{F_0}[w_2(y)] - c_0$.

1. If $\mathbb{E}_{F_1}[w_2(y)] \geq \mathbb{E}_{F_1}[w_1(y)]$, consider the second-period contract w_2 when the technology is $A = \{a_0, a_1\}$, which is compatible with (w_1, a_1). Agent 2 would prefer to take action $a_1 = (F_1, c_1)$. This leaves the principal with a payoff of

$$\mathbb{E}_{F_1}[y - w_2(y)] \leq \mathbb{E}_{F_1}[y - w_1(y)] \leq \Phi(w_1, a_1)^2,$$

as desired.

2. If $\mathbb{E}_{F_1}[w_2(y)] < c_1$, consider the second-period contract w_2 when $A = \{a_0, a_1, (\delta_0, 0)\}$, which is compatible with (w_1, a_1). Agent 2's payoff from $(\delta_0, 0)$ is

$$w_2(0) \geq 0 > \mathbb{E}_{F_1}[w_2(y)] - c_1,$$

so he would prefer to take action $(\delta_0, 0)$. This leaves the principal with a payoff of

$$-w_2(0) \leq 0 \leq \Phi(w_1, a_1)^2,$$

as desired.

3. If $c_1 \leq \mathbb{E}_{F_1}[w_2(y)] < \mathbb{E}_{F_1}[w_1(y)]$, let $\lambda = 1 - c_1/\mathbb{E}_{F_1}[w_2(y)] \in [0, 1]$ and let F' be the mixture $\lambda F_1 + (1 - \lambda)\delta_0$. Consider the technology $A = \{a_0, a_1, (F', 0)\}$.

We proceed with two steps.

Step 1 A is compatible with (w_1, a_1). That is, agent 1 chooses a_1 in response to w_1.

Agent 1's payoff from $(F', 0)$ is

$$\mathbb{E}_{F'}[w_1(y)] = \lambda \mathbb{E}_{F_1}[w_1(y)] = \mathbb{E}_{F_1}[w_1(y)] - \frac{\mathbb{E}_{F_1}[w_1(y)]}{\mathbb{E}_{F_1}[w_2(y)]} c_1 < \mathbb{E}_{F_1}[w_1(y)] - c_1,$$

so he would prefer to take action $a_1 = (F_1, c_1)$ when $A = \{a_0, a_1, (F', 0)\}$.

Step 2 Agent 2 chooses $(F', 0)$ in response to w_2, resulting in the principal's payoff no more than $\left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)^2$.

Agent 2's payoff from $(F', 0)$ is

$$\begin{aligned}\mathbb{E}_{F'}[w_2(y)] &= \lambda \mathbb{E}_{F_1}[w_2(y)] + (1-\lambda) w_2(0) \\ &\geq \lambda \mathbb{E}_{F_1}[w_2(y)] = \mathbb{E}_{F_1}[w_2(y)] - c_1,\end{aligned}$$

which is also larger than $\mathbb{E}_{F_0}[w_2(y)] - c_0$ by assumption. So he would prefer to take action $(F', 0)$.

This leaves the principal with a payoff of

$$\begin{aligned}\mathbb{E}_{F'}[y - w_2(y)] &= \lambda \mathbb{E}_{F_1}[y - w_2(y)] + (1-\lambda)(0 - w_2(0)) \\ &\leq \lambda \mathbb{E}_{F_1}[y - w_2(y)] = \left(1 - \frac{c_1}{\mathbb{E}_{F_1}[w_2(y)]}\right)(\mathbb{E}_{F_1}[y] - \mathbb{E}_{F_1}[w_2(y)]) \\ &\leq \left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)^2, \end{aligned} \tag{A.2}$$

which is no more than $\Phi(w_1, a_1)^2$, as desired. The last inequality (A.2),

$$\begin{aligned} &\left(1 - \frac{c_1}{\mathbb{E}_{F_1}[w_2(y)]}\right)(\mathbb{E}_{F_1}[y] - \mathbb{E}_{F_1}[w_2(y)]) \leq \left(\sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right)^2 \\ \Leftrightarrow \quad &\left(\sqrt{\mathbb{E}_{F_1}[w_2(y)]} - \sqrt{\frac{c_1 \mathbb{E}_{F_1}[y]}{\mathbb{E}_{F_1}[w_2(y)]}}\right)^2 \geq 0, \end{aligned}$$

which always holds.

Case 2. $\mathbb{E}_{F_1}[w_2(y)] - c_1 < \mathbb{E}_{F_0}[w_2(y)] - c_0$.

1. If $\mathbb{E}_{F_0}[w_2(y)] < c_0$, consider the second-period contract w_2 when $A = \{a_0, a_1, (\delta_0, 0)\}$, which is compatible with (w_1, a_1). Agent 2's payoff from $(\delta_0, 0)$ is

$$w_2(0) \geq 0 > \mathbb{E}_{F_0}[w_2(y)] - c_0,$$

so he would prefer to take action $(\delta_0, 0)$. This leaves the principal with a payoff of

$$-w_2(0) \leq 0 \leq \Phi(w_1, a_1)^2,$$

as desired.

2. If $\mathbb{E}_{F_0}[w_2(y)] \geq c_0$, and it holds that

$$\begin{aligned} \text{either} \quad &\text{(i)} \quad \mathbb{E}_{F_0}[w_1(y)] \leq \mathbb{E}_{F_1}[w_1(y)] - c_1, \\ \text{or} \quad &\text{(ii)} \quad \mathbb{E}_{F_0}[w_2(y)] < \frac{\mathbb{E}_{F_0}[w_1(y)]}{\mathbb{E}_{F_0}[w_1(y)] - (\mathbb{E}_{F_1}[w_1(y)] - c_1)} c_0, \end{aligned} \tag{A.3}$$

let $\lambda = 1 - c_0 / \mathbb{E}_{F_0}[w_2(y)] \in [0, 1]$ and let F' be the mixture $\lambda F_0 + (1-\lambda)\delta_0$. Consider the technology $A = \{a_0, a_1, (F', 0)\}$.

We proceed with two steps.

Step 1 A is compatible with (w_1, a_1). That is, agent 1 chooses a_1 in response to w_1.

Agent 1's payoff from $(F', 0)$ is

$$\begin{aligned}\mathbb{E}_{F'}[w_1(y)] = \lambda \mathbb{E}_{F_0}[w_1(y)] &= \mathbb{E}_{F_0}[w_1(y)] - \frac{\mathbb{E}_{F_0}[w_1(y)]}{\mathbb{E}_{F_0}[w_2(y)]} c_0 \\ &< \mathbb{E}_{F_1}[w_1(y)] - c_1. \end{aligned} \tag{A.4}$$

Note that inequality (A.4) holds exactly due to the assumptions in (A.3). So agent 1 would prefer to take action $a_1 = (F_1, c_1)$ when $A = \{a_0, a_1, (F', 0)\}$.

Step 2 Agent 2 chooses $(F', 0)$ in response to w_2, resulting in the principal's payoff no more than $\left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)^2$.

Agent 2's payoff from $(F', 0)$ is

$$\begin{aligned}\mathbb{E}_{F'}[w_2(y)] &= \lambda \mathbb{E}_{F_0}[w_2(y)] + (1-\lambda) w_2(0) \\ &\geq \lambda \mathbb{E}_{F_0}[w_2(y)] = \mathbb{E}_{F_0}[w_2(y)] - c_0, \end{aligned}$$

which is also larger than $\mathbb{E}_{F_1}[w_2(y)] - c_1$ by assumption. So he would prefer to take action $(F', 0)$ when $A = \{a_0, a_1, (F', 0)\}$.

This leaves the principal with a payoff of

$$\begin{aligned}\mathbb{E}_{F'}[y - w_2(y)] &= \lambda \mathbb{E}_{F_0}[y - w_2(y)] + (1-\lambda)(0 - w_2(0)) \\ &\leq \lambda \mathbb{E}_{F_0}[y - w_2(y)] = \left(1 - \frac{c_0}{\mathbb{E}_{F_0}[w_2(y)]}\right)(\mathbb{E}_{F_0}[y] - \mathbb{E}_{F_0}[w_2(y)]) \\ &\leq \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)^2, \end{aligned} \tag{A.5}$$

which is no more than $\Phi(w_1, a_1)^2$, as desired. The last inequality (A.15) holds for the same reason as (A.2).

3. If both inequalities in (A.3) are reversed, i.e.,

$$\mathbb{E}_{F_0}[w_1(y)] > \mathbb{E}_{F_1}[w_1(y)] - c_1 \quad \text{and} \quad \mathbb{E}_{F_0}[w_2(y)] \geq \frac{\mathbb{E}_{F_0}[w_1(y)]}{\mathbb{E}_{F_0}[w_1(y)] - (\mathbb{E}_{F_1}[w_1(y)] - c_1)} c_0,$$

let

$$\begin{aligned}\lambda &= \frac{(\mathbb{E}_{F_0}[w_2(y)] - c_0) - (\mathbb{E}_{F_1}[w_1(y)] - c_1)}{\mathbb{E}_{F_0}[w_2(y)] - \mathbb{E}_{F_0}[w_1(y)]}, \\ c' &= \frac{\mathbb{E}_{F_0}[w_1(y)](\mathbb{E}_{F_0}[w_2(y)] - c_0) - \mathbb{E}_{F_0}[w_2(y)](\mathbb{E}_{F_1}[w_1(y)] - c_1)}{\mathbb{E}_{F_0}[w_2(y)] - \mathbb{E}_{F_0}[w_1(y)]}, \end{aligned}$$

and let F' be the mixture $\lambda F_0 + (1-\lambda)\delta_0$. Consider the technology $A = \{a_0, a_1, (F', c')\}$.

We proceed with three steps.

Step 1 $\lambda \in [0, 1]$ and $c' \geq 0$, so (F', c') is a valid action.

Note that

$$\mathbb{E}_{F_0}\left[w_2(y)\right] \geq \frac{\mathbb{E}_{F_0}\left[w_1(y)\right]}{\mathbb{E}_{F_0}\left[w_1(y)\right]-\left(\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1\right)} c_0 \geq \frac{\mathbb{E}_{F_0}\left[w_1(y)\right]}{\mathbb{E}_{F_0}\left[w_1(y)\right]-\left(\mathbb{E}_{F_0}\left[w_1(y)\right]-c_0\right)} c_0=\mathbb{E}_{F_0}\left[w_1(y)\right],$$

so the denominator of λ and c' is positive.

Moreover,

$$\begin{aligned}
\mathbb{E}_{F_0}\left[w_2(y)\right]-c_0 & \geq \frac{\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1}{\mathbb{E}_{F_0}\left[w_1(y)\right]-\left(\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1\right)} c_0 \\
& \geq \frac{\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1}{\mathbb{E}_{F_0}\left[w_1(y)\right]-\left(\mathbb{E}_{F_0}\left[w_1(y)\right]-c_0\right)} c_0=\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1,
\end{aligned}$$

so the numerator of λ is positive.

The numerator of c' is positive because

$$\begin{aligned}
& \mathbb{E}_{F_0}\left[w_1(y)\right]\left(\mathbb{E}_{F_0}\left[w_2(y)\right]-c_0\right) \geq \mathbb{E}_{F_0}\left[w_2(y)\right]\left(\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1\right) \\
\Leftrightarrow \quad & \mathbb{E}_{F_0}\left[w_2(y)\right] \geq \frac{\mathbb{E}_{F_0}\left[w_1(y)\right]}{\mathbb{E}_{F_0}\left[w_1(y)\right]-\left(\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1\right)} c_0 .
\end{aligned}$$

Finally,

$$\begin{aligned}
& \left(\mathbb{E}_{F_0}\left[w_2(y)\right]-c_0\right)-\left(\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1\right) \leq \mathbb{E}_{F_0}\left[w_2(y)\right]-\mathbb{E}_{F_0}\left[w_1(y)\right] \\
\Leftrightarrow \quad & \mathbb{E}_{F_0}\left[w_1(y)\right]-c_0 \leq \mathbb{E}_{F_1}\left[w_1(y)\right]-c_1,
\end{aligned}$$

so λ is indeed smaller than 1.

Step 2 A is compatible with (w_1, a_1). That is, agent 1 chooses a_1 in response to w_1.

Agent 1's payoff from (F', c') is

$$\mathbb{E}_{F'}\left[w_1(y)\right]-c'=\lambda \mathbb{E}_{F_0}\left[w_1(y)\right]-c'=\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1,$$

so he would prefer to take action $a_1=(F_1, c_1)$ when $A=\{a_0, a_1,(F', c')\}$.

Step 3 Agent 2 chooses (F', c') in response to w_2, resulting in the principal's payoff no more than $\left(\sqrt{\mathbb{E}_{F_0}\left[y-w_1(y)\right]}-\sqrt{g\left(w_1, a_1\right)}\right)^2$.

Agent 2's payoff from (F', c') is

$$\begin{aligned}
\mathbb{E}_{F'}\left[w_2(y)\right]-c' & =\lambda \mathbb{E}_{F_0}\left[w_2(y)\right]+(1-\lambda) w_2(0)-c' \\
& \geq \lambda \mathbb{E}_{F_0}\left[w_2(y)\right]-c'=\mathbb{E}_{F_0}\left[w_2(y)\right]-c_0,
\end{aligned}$$

which is also larger than $\mathbb{E}_{F_1}\left[w_2(y)\right]-c_1$ by assumption. So he would prefer to take action (F', c') when $A=\{a_0, a_1,(F', c')\}$.

This leaves the principal with a payoff of

$$\begin{aligned}
\mathbb{E}_{F'}\left[y-w_2(y)\right] &= \lambda \mathbb{E}_{F_0}\left[y-w_2(y)\right]+(1-\lambda)(0-w_2(0)) \\
&\leq \lambda \mathbb{E}_{F_0}\left[y-w_2(y)\right]=\frac{\left(\mathbb{E}_{F_0}\left[w_2(y)\right]-c_0\right)-\left(\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1\right)}{\mathbb{E}_{F_0}\left[w_2(y)\right]-\mathbb{E}_{F_0}\left[w_1(y)\right]}\left(\mathbb{E}_{F_0}[y]-\mathbb{E}_{F_0}\left[w_2(y)\right]\right) \\
&\leq\left(\sqrt{\mathbb{E}_{F_0}\left[y-w_1(y)\right]}-\sqrt{g\left(w_1, a_1\right)}\right)^2, \qquad (A.6)
\end{aligned}$$

which is no more than $\Phi\left(w_1, a_1\right)^2$, as desired. The last inequality (A.6),

$$\begin{aligned}
&\frac{\left(\mathbb{E}_{F_0}\left[w_2(y)\right]-c_0\right)-\left(\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1\right)}{\mathbb{E}_{F_0}\left[w_2(y)\right]-\mathbb{E}_{F_0}\left[w_1(y)\right]}\left(\mathbb{E}_{F_0}[y]-\mathbb{E}_{F_0}\left[w_2(y)\right]\right) \leq\left(\sqrt{\mathbb{E}_{F_0}\left[y-w_1(y)\right]}-\sqrt{g\left(w_1, a_1\right)}\right)^2 \\
\Leftrightarrow \quad &\frac{\left(\mathbb{E}_{F_0}\left[w_2(y)\right]-\mathbb{E}_{F_0}\left[w_1(y)\right]-\sqrt{\mathbb{E}_{F_0}\left[y-w_1(y)\right]} \cdot \sqrt{g\left(w_1, a_1\right)}\right)^2}{\mathbb{E}_{F_0}\left[w_2(y)\right]-\mathbb{E}_{F_0}\left[w_1(y)\right]} \geq 0,
\end{aligned}$$

which always holds. (Recall that $g\left(w_1, a_1\right)=\left(\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1\right)-\left(\mathbb{E}_{F_0}\left[w_1(y)\right]-c_0\right) \geq 0$.)

Summing up the above cases, we prove that the principal's payoff guarantee is not strictly higher than $\Phi\left(w_1, a_1\right)^2$ under any second-period contract w_2.

This completes the proof. □

A.2 Proofs for Section 4

Proof of Lemma 2. Consider an arbitrary action $a_1=\left(F_1, c_1\right)$ agent 1 would take under contract $\hat{w}_1$. We need to show that the principal's interim payoff guarantee, $U\left(\hat{w}_1 \mid a_1\right)$, is at least $U\left(w_1\right)$. The incentive gap is

$$g\left(\hat{w}_1, a_1\right)=\left(\mathbb{E}_{F_1}\left[\hat{w}_1(y)\right]-c_1\right)-\left(\mathbb{E}_{F_0}\left[\hat{w}_1(y)\right]-c_0\right) \geq 0,$$

and Lemma 1 shows that the principal's optimal second-period payoff guarantee is $V_2^*\left(\hat{w}_1, a_1\right)=\Phi\left(\hat{w}_1, a_1\right)^2$, where

$$\Phi\left(\hat{w}_1, a_1\right)=\max \left\{\sqrt{\mathbb{E}_{F_1}\left[y-\hat{w}_1(y)\right]}, \ \sqrt{\mathbb{E}_{F_0}\left[y-\hat{w}_1(y)\right]}-\sqrt{g\left(\hat{w}_1, a_1\right)}, \sqrt{\mathbb{E}_{F_0}[y]}-\sqrt{c_0}, \sqrt{\mathbb{E}_{F_1}[y]}-\sqrt{c_1}\right\},$$
$$\text{(with } \sqrt{x}=-\infty \text{ for } x<0 \text{ by convention).} \qquad (A.7)$$

The principal's interim payoff guarantee is

$$U\left(\hat{w}_1 \mid a_1\right)=\mathbb{E}_{F_1}\left[y-\hat{w}_1(y)\right]+\beta \cdot V_2^*\left(\hat{w}_1, a_1\right).$$

It suffices to construct another action a_1', which may be taken by agent 1 under w_1 and some other technology, such that $U\left(w_1 \middle| a_1'\right) \leq U\left(\hat{w}_1 \mid a_1\right)$. Note that an action may be taken by agent 1 if and only if the incentive gap is nonnegative, i.e., $g\left(w_1, a_1'\right) \geq 0$.

Case 1. $\mathbb{E}_{F_1}[y] \geq \mathbb{E}_{F_0}[y]$.

Consider $a_1'=a_0$. The corresponding incentive gap is $g\left(w_1, a_0\right)=0$. When agent 1 takes action a_0 in response, the principal's resulting payoff in the first period is

$$\mathbb{E}_{F_0}\left[y-w_1(y)\right]=\left(1-s_1\right) \mathbb{E}_{F_0}[y] \leq\left(1-s_1\right) \mathbb{E}_{F_1}[y]=\mathbb{E}_{F_1}\left[y-\hat{w}_1(y)\right],$$

so her payoff in the first period under $(w_1|a_0)$ is weakly lower than under $(\hat{w}_1|a_1)$.

Moreover, it follows from Lemma 1 that the principal's optimal second-period payoff guarantee is $V_2^*(w_1, a_0) = \Phi(w_1, a_0)^2$, where

$$\Phi(w_1, a_0) = \max\left\{\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]},\ \sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right\}.$$

Note that we have shown $\mathbb{E}_{F_0}[y - w_1(y)] \leq \mathbb{E}_{F_1}[y - \hat{w}_1(y)]$, so $\Phi(w_1, a_0)$ is also weakly smaller than $\Phi(\hat{w}_1, a_1)$ (given by equation (A.7)). This implies that $V_2^*(w_1, a_0) \leq V_2^*(\hat{w}_1, a_1)$.

Therefore, the principal's interim payoff guarantee is

$$\begin{aligned} U(w_1|a_0) &= \mathbb{E}_{F_0}[y - w_1(y)] + \beta \cdot V_2^*(w_1, a_0) \\ &\leq \mathbb{E}_{F_1}[y - \hat{w}_1(y)] + \beta \cdot V_2^*(\hat{w}_1, a_1) = U(\hat{w}_1|a_1), \end{aligned}$$

as desired.

Case 2. $\mathbb{E}_{F_1}[y] < \mathbb{E}_{F_0}[y]$.

Let $\lambda = \mathbb{E}_{F_1}[y]/\mathbb{E}_{F_0}[y] \in [0, 1]$ and let F_1' be the mixture $\lambda F_0 + (1 - \lambda)\delta_0$. Note that $\mathbb{E}_{F_1'}[y] = \mathbb{E}_{F_1}[y]$. Consider $a_1' = \left(F_1', c_1\right)$. The corresponding incentive gap is

$$g\left(w_1, a_1'\right) = \left(\mathbb{E}_{F_1'}[w_1(y)] - c_1\right) - \left(\mathbb{E}_{F_0}[w_1(y)] - c_0\right).$$

Note that

$$\mathbb{E}_{F_1'}[w_1(y)] - c_1 = \lambda\mathbb{E}_{F_0}[w_1(y)] - c_1 = \lambda s_1\mathbb{E}_{F_0}[y] - c_1 = s_1\mathbb{E}_{F_1}[y] - c_1 = \mathbb{E}_{F_1}[\hat{w}_1(y)] - c_1,$$

and

$$\mathbb{E}_{F_0}[w_1(y)] - c_0 = s_1\mathbb{E}_{F_0}[y] - c_0 = \mathbb{E}_{F_0}[\hat{w}_1(y)] - c_0.$$

Thus,

$$\begin{aligned} g\left(w_1, a_1'\right) &= \left(\mathbb{E}_{F_1'}[w_1(y)] - c_1\right) - \left(\mathbb{E}_{F_0}[w_1(y)] - c_0\right) \\ &= \left(\mathbb{E}_{F_1}[\hat{w}_1(y)] - c_1\right) - \left(\mathbb{E}_{F_0}[\hat{w}_1(y)] - c_0\right) \\ &= g(\hat{w}_1, a_1) \geq 0. \end{aligned}$$

When agent 1 takes action a_1' in response, the principal's resulting payoff in the first period is

$$\mathbb{E}_{F_1'}[y - w_1(y)] = \lambda\mathbb{E}_{F_0}[y - w_1(y)] = \lambda(1 - s_1)\mathbb{E}_{F_0}[y] = (1 - s_1)\mathbb{E}_{F_1}[y] = \mathbb{E}_{F_1}[y - \hat{w}_1(y)],$$

so her payoff in the first period under $\left(w_1\middle|a_1'\right)$ and under $(\hat{w}_1|a_1)$ are exactly equal.

Moreover, the quadruple in equation (1) with respect to $\left(w_1, a_1'\right)$,

$$\left\{\sqrt{\mathbb{E}_{F_1'}[y - w_1(y)]},\ \sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g\left(w_1, a_1'\right)},\ \sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0},\ \sqrt{\mathbb{E}_{F_1'}[y]} - \sqrt{c_1}\right\},$$

takes the same value as the quadruple in equation (1) with respect to $(\hat{w}_1, a_1)$,

$$\left\{\sqrt{\mathbb{E}_{F_1}[y - \hat{w}_1(y)]},\ \sqrt{\mathbb{E}_{F_0}[y - \hat{w}_1(y)]} - \sqrt{g(\hat{w}_1, a_1)},\ \sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0},\ \sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right\}.$$

It follows from Lemma 1 that the principal's optimal second-period payoff guarantee also takes the same value: $V_2^*\left(w_1, a_1'\right) = V_2^*\left(\hat{w}_1, a_1\right)$.

Therefore, the principal's interim payoff guarantee is

$$\begin{aligned} U\left(w_1 \middle| a_1'\right) &= \mathbb{E}_{F_1'}\left[y - w_1(y)\right] + \beta \cdot V_2^*\left(w_1, a_1'\right) \\ &= \mathbb{E}_{F_1}\left[y - \hat{w}_1(y)\right] + \beta \cdot V_2^*\left(\hat{w}_1, a_1\right) = U\left(\hat{w}_1 | a_1\right), \end{aligned}$$

as desired.

This completes the proof. □

Proof of Lemma 3. Suppose $s_1 \geq c_0/\mathbb{E}_{F_0}[y]$. We first reformulate program (5) as an equivalent maximization problem with continuous objective function and compact feasible region. Slightly abusing notation, we use $U(s_1)$ instead of $U(w_1)$ to denote the infimum value of program (5).

Plug $w_1(y) = s_1 y$ into equation (1) and let $s_0 \equiv \sqrt{c_0/\mathbb{E}_{F_0}[y]}$. We may rewrite $\Phi(w_1, a_1)$ as

$$\Phi\left(w_1, a_1\right) = \max\left\{\sqrt{\left(1 - s_1\right)\mathbb{E}_{F_1}[y]},\ \sqrt{\left(1 - s_1\right)\mathbb{E}_{F_0}[y]} - \sqrt{g\left(w_1, a_1\right)}, \left(1 - s_0\right)\sqrt{\mathbb{E}_{F_0}[y]},\ \sqrt{\mathbb{E}_{F_1}[y]} - \sqrt{c_1}\right\}.$$

Similarly,

$$g\left(w_1, a_1\right) = \left(s_1 \mathbb{E}_{F_1}[y] - c_1\right) - \left(s_1 - s_0^2\right)\mathbb{E}_{F_0}[y] \geq 0.$$

Note that both the objective and the constraints of program (5) depend on the choice variables (F_1, c_1) only through the value of $\left(\mathbb{E}_{F_1}[y], c_1\right)$. Rewrite $\mathbb{E}_{F_1}[y] = x\mathbb{E}_{F_0}[y]$, $c_1 = z\mathbb{E}_{F_0}[y]$, and let $g\left(w_1, a_1\right) = h\mathbb{E}_{F_0}[y]$ with $x, z, h \geq 0$. Plugging into the original program (5) and cancelling out $\mathbb{E}_{F_0}[y]$ from both sides of the constraints, we obtain an equivalent program

$$\begin{aligned} U\left(s_1\right) = \inf_{x,z,h} \quad & \left(\left(1 - s_1\right)x + \beta \cdot \hat{\Phi}\left(x, z, h; s_1\right)^2\right)\mathbb{E}_{F_0}[y] \\ \text{s.t.} \quad & h = s_1 x - z - \left(s_1 - s_0^2\right) \geq 0, \quad x, z \geq 0, \end{aligned} \tag{A.8}$$

where

$$\hat{\Phi}\left(x, z, h; s_1\right) \equiv \max\left\{\sqrt{\left(1 - s_1\right)x},\ \sqrt{1 - s_1} - \sqrt{h}, 1 - s_0,\ \sqrt{x} - \sqrt{z}\right\}. \tag{A.9}$$

Note that $(x, z, h) = \left(1, s_0^2, 0\right)$ is feasible in program (A.8) and leads to objective value

$$\left(\left(1 - s_1\right) + \beta \cdot \max\left\{\sqrt{1 - s_1}, 1 - s_0\right\}^2\right)\mathbb{E}_{F_0}[y].$$

If $x \geq 1 + \beta$, then

$$\begin{aligned} \left(1 - s_1\right)x + \beta \cdot \hat{\Phi}\left(x, z, h; s_1\right)^2 &\geq \left(1 - s_1\right)(1 + \beta) + \beta\left(1 - s_0\right)^2 \\ &= \left(1 - s_1\right) + \beta\left(1 - s_1\right) + \beta\left(1 - s_0\right)^2 \\ &\geq \left(1 - s_1\right) + \beta \cdot \max\left\{\sqrt{1 - s_1}, 1 - s_0\right\}^2. \end{aligned}$$

Therefore, restricting $x \in [0, 1 + \beta]$ will not change the infimum of program (A.8). Moreover,

$$\max\{z, h\} \leq z + h = s_1 x - \left(s_1 - s_0^2\right) \leq s_1 x \leq x,$$

so restricting $(x, z, h) \in [0, 1+\beta]^3$ will not change the infimum of program (A.8).

Consider the following program

$$\begin{aligned} \Psi^*(s_1) \equiv \sup_{x,z,h} \quad & \Psi(x, z, h; s_1) \equiv -\left((1-s_1)x + \beta \cdot \hat{\Phi}(x, z, h; s_1)^2\right) \\ \text{s.t.} \quad & (x, z, h) \in \Gamma(s_1), \end{aligned} \tag{A.10}$$

where $\hat{\Phi}$ is defined by equation (A.9), and Γ is defined as follows:

$$\Gamma(s_1) \equiv \left\{(x, z, h) \in [0, 1+\beta]^3 : h = s_1 x - z - \left(s_1 - s_0^2\right)\right\}.$$

By definition, $\Psi : [0, 1+\beta]^3 \times \left[s_0^2, 1\right] \to \mathbb{R}$ is a continuous function, and $\Gamma : \left[s_0^2, 1\right] \rightrightarrows [0, 1+\beta]^3$ is a compact-valued and nonempty-valued correspondence. Moreover, the infimum of program (A.8), $U(s_1)$, is given by $(-\Psi^*(s_1)) \cdot \mathbb{E}_{F_0}[y]$.

Note that for each s_1, $\Gamma(s_1)$ defines a plane intersecting a cube, and that the plane shifts linearly in s_1. Thus, Γ is both upper and lower hemicontinuous. It then follows from Berge's maximum theorem that Ψ^* is continuous, and

$$\Gamma^*(s_1) \equiv \{(x, z, h) \in \Gamma(s_1) : \Psi(x, z, h; s_1) = \Psi^*(s_1)\}$$

is upper hemicontinuous with nonempty and compact values. As a consequence, a solution to program (A.10) exists for all s_1, and the supremum can be replaced by maximum.

It follows that the infimum in program (A.8) and therefore the original program (5) can both be replaced by minimum, and the resulting minimum value $U(s_1) = (-\Psi^*(s_1)) \cdot \mathbb{E}_{F_0}[y]$ is continuous in s_1. Hence, $U(s_1)$ achieves a maximum over $\left[s_0^2, 1\right]$. This maximum is also the optimal guarantee over all linear contracts. □

Proof of Theorem 1. According to Lemma 3, among all linear first-period contracts, there exists an optimal one, call it w_1^*. If w_1 is any other (nonlinear) first-period contract that outperforms w_1^*, then by Lemma 2, there is a linear contract that in turn does at least as well as w_1. But this contradicts the fact that w_1^* is an optimal linear contract. Therefore, w_1^* is optimal among all first-period contracts. □

A.3 Proofs for Section 5

A.3.1 Proofs for Subsection 5.1.1

To prove Lemma 1′, we start by establishing two lemmas, Lemmas A.4 and A.5, to show that the principal's payoff guarantees in the second period from offering the two contracts, (i) $w_2(y) = w_1(y) + m \cdot (y - w_1(y))$ with m defined by equation (9), and (ii) $w_2(y) = s_2 y$ with s_2 defined by equation (10), are exactly as claimed in the statement of Lemma 1′.

Lemma A.4. *If $\Phi_1(w_1, a_1) \geq \Phi_2(a_1)$ and $a^* = (F^*, c^*) \in A_0 \cup \{a_1\}$ attains the maximum in equation (6), and the principal offers $w_2(y) = w_1(y) + m \cdot (y - w_1(y))$ with m defined by equation (9), then her payoff guarantee in the second period is exactly*

$$\Phi_1(w_1, a_1)^2 = \left(\sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g(a^*|w_1, a_1)}\right)^2.$$

Proof of Lemma A.4. Let $g^* \equiv g(a^*|w_1, a_1) = (\mathbb{E}_{F_1}[w_1(y)] - c_1) - (\mathbb{E}_{F^*}[w_1(y)] - c^*) \geq 0$. We have

$\Phi_1(w_1, a_1) = \sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g^*}$. From $\Phi_1(w_1, a_1) \geq \Phi_2(a_1) > 0$, it holds that

$$m = \sqrt{\frac{g^*}{\mathbb{E}_{F^*}[y - w_1(y)]}} \in [0, 1].$$

Suppose the principal offers $w_2(y) = w_1(y) + m \cdot (y - w_1(y))$ with m defined by equation (9). We first show that this guarantees her at least $\left(\sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g^*}\right)^2$.

Let (F_2, c_2) be the action chosen by agent 2. By agent 1's rationality, we have

$$\mathbb{E}_{F_1}[w_1(y)] - c_1 \geq \mathbb{E}_{F_2}[w_1(y)] - c_2.$$

By agent 2's rationality, we have

$$\mathbb{E}_{F_2}[w_2(y)] - c_2 \geq \mathbb{E}_{F^*}[w_2(y)] - c^*.$$

Summing up the two inequalities, we obtain

$$\begin{aligned} m \cdot \mathbb{E}_{F_2}[y - w_1(y)] = \mathbb{E}_{F_2}[w_2(y) - w_1(y)] &\geq (\mathbb{E}_{F^*}[w_2(y)] - c^*) - (\mathbb{E}_{F_1}[w_1(y)] - c_1) \\ &= m \cdot \mathbb{E}_{F^*}[y - w_1(y)] - g^*, \end{aligned}$$

implying that

$$\mathbb{E}_{F_2}[y - w_1(y)] \geq \mathbb{E}_{F^*}[y - w_1(y)] - g^*/m.$$

Therefore, the principal's payoff in the second period is

$$\begin{aligned} \mathbb{E}_{F_2}[y - w_2(y)] = \mathbb{E}_{F_2}[y - w_1(y)] - m \cdot \mathbb{E}_{F_2}[y - w_1(y)] &= (1 - m)\,\mathbb{E}_{F_2}[y - w_1(y)] \\ \geq (1 - m)(\mathbb{E}_{F^*}[y - w_1(y)] - g^*/m) &= \left(\sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g^*}\right)^2, \end{aligned}$$

as desired.

Next we show that her payoff guarantee from $w_2(y) = w_1(y) + m \cdot (y - w_1(y))$ cannot be strictly higher than $\left(\sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g^*}\right)^2$, since this is exactly her payoff when the technology is $A = A_0 \cup \{a_1, (F', c')\}$, with $F' = (1 - m) F^* + m \cdot \delta_0$ and $c' = c^* - (m \cdot \mathbb{E}_{F^*}[w_1(y)] + g^*)$.

The proof takes three steps.

Step 1 $c^* \geq m \cdot \mathbb{E}_{F^*}[w_1(y)] + g^*$, so c' is indeed nonnegative.

From $\Phi_1(w_1, a_1) \geq \Phi_2(a_1)$, we obtain

$$\sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g^*} = \Phi_1(w_1, a_1) \geq \Phi_2(a_1) \geq \sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*},$$

which implies that

$$c^* \geq \left(\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} + \sqrt{g^*}\right)^2$$

It suffices to show

$$
\begin{aligned}
& \left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}+\sqrt{g^*}\right)^2 \geq m\cdot\mathbb{E}_{F^*}\left[w_1(y)\right]+g^* \\
\Leftrightarrow \quad & \left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\right)^2 \geq m\cdot\mathbb{E}_{F^*}\left[w_1(y)\right]-2\sqrt{g^*}\cdot\left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\right) \\
\Leftrightarrow \quad & \left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\right)^2 \geq m\cdot\left(\mathbb{E}_{F^*}\left[w_1(y)\right]-2\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\cdot\left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\right)\right).
\end{aligned}
\tag{A.11}
$$

Note that

$$
\begin{aligned}
& \mathbb{E}_{F^*}\left[w_1(y)\right]-2\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\cdot\left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\right) \\
= & \mathbb{E}_{F^*}\left[w_1(y)\right]-2\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\cdot\frac{\mathbb{E}_{F^*}\left[w_1(y)\right]}{\sqrt{\mathbb{E}_{F^*}[y]}+\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}} \\
= & \frac{\mathbb{E}_{F^*}\left[w_1(y)\right]}{\sqrt{\mathbb{E}_{F^*}[y]}+\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}}\cdot\left(\sqrt{\mathbb{E}_{F^*}[y]}+\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}-2\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\right) \\
= & \left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\right)\cdot\left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\right)=\left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\right)^2.
\end{aligned}
$$

Therefore, inequality (A.11) is equivalent to

$$
\left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\right)^2 \geq m\cdot\left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}\right)^2,
$$

which is implied by the assumption that $\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]} \geq \sqrt{g^*}$ (or equivalently, $m \leq 1$).

Step 2 $A=A_0\cup\{a_1,(F',c')\}$ is compatible with (w_1,a_1). That is, agent 1 chooses a_1 in response to w_1.

Agent 1's payoff from (F',c') is

$$
\begin{aligned}
\mathbb{E}_{F'}\left[w_1(y)\right]-c' &= (1-m)\,\mathbb{E}_{F^*}\left[w_1(y)\right]-c^*+\left(m\cdot\mathbb{E}_{F^*}\left[w_1(y)\right]+g^*\right) \\
&= \left(\mathbb{E}_{F^*}\left[w_1(y)\right]-c^*\right)+g^*=\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1,
\end{aligned}
$$

so he would choose $a_1=(F_1,c_1)$ in response to w_1.

Note that agent 1 is actually indifferent between (F_1,c_1) and (F',c'), and we will show below that agent 2 is indifferent between (F^*,c^*) and (F',c'). Technically to ensure that agent 1 chooses (F_1,c_1) and agent 2 chooses (F',c') we can set $F'=(1-m+\varepsilon)\,F^*+(m-\varepsilon)\,\delta_0$ and $c'=c^*-(m\cdot\mathbb{E}_{F^*}\left[w_1(y)\right]+g^*)+\varepsilon\cdot\mathbb{E}_{F^*}\left[w_1(y)+(m/2)\cdot(y-w_1(y))\right]$, and then let $\varepsilon\downarrow 0$. Many of the following cases of potential indifference shall be treated similarly, and we omit them for brevity.

Step 3 If $A=A_0\cup\{a_1,(F',c')\}$, then agent 2 chooses (F',c') in response to w_2, leading to a payoff of $\left(\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}-\sqrt{g^*}\right)^2$ for the principal.

Agent 2's payoff from (F',c') is

$$
\begin{aligned}
\mathbb{E}_{F'}\left[w_2(y)\right]-c' &= (1-m)\,\mathbb{E}_{F^*}\left[w_1(y)+m\cdot(y-w_1(y))\right]-c^*+\left(m\cdot\mathbb{E}_{F^*}\left[w_1(y)\right]+g^*\right) \\
&= \mathbb{E}_{F^*}\left[w_1(y)\right]+m\cdot\mathbb{E}_{F^*}\left[y-w_1(y)\right]-m^2\cdot\mathbb{E}_{F^*}\left[y-w_1(y)\right]-c^*+g^* \\
&= \mathbb{E}_{F^*}\left[w_2(y)\right]-g^*-c^*+g^*=\mathbb{E}_{F^*}\left[w_2(y)\right]-c^*.
\end{aligned}
$$

For any action $a_0 = (F_0, c_0) \in A_0 \cup \{a_1\}$, let $g_0 \equiv g(a_0|w_1, a_1) = (\mathbb{E}_{F_1}[w_1(y)] - c_1) - (\mathbb{E}_{F_0}[w_1(y)] - c_0) \geq 0$. Agent 2's payoff from a_0 is

$$\mathbb{E}_{F_0}[w_2(y)] - c_0 = \mathbb{E}_{F_0}[w_1(y) + m \cdot (y - w_1(y))] - c_0 = m \cdot \mathbb{E}_{F_0}[y - w_1(y)] + (\mathbb{E}_{F_0}[w_1(y)] - c_0).$$

Note that

$$\begin{aligned} &\sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g^*} = \Phi_1(w_1, a_1) \geq \sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g_0} \\ \Rightarrow \quad & \mathbb{E}_{F_0}[y - w_1(y)] \leq \left(\sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g^*} + \sqrt{g_0}\right)^2. \end{aligned}$$

Moreover,

$$\mathbb{E}_{F_0}[w_1(y)] - c_0 = (\mathbb{E}_{F_1}[w_1(y)] - c_1) - g_0 = (\mathbb{E}_{F^*}[w_1(y)] - c^*) + g^* - g_0.$$

Thus, agent 2's payoff from a_0,

$$\begin{aligned} \mathbb{E}_{F_0}[w_2(y)] - c_0 &= m \cdot \mathbb{E}_{F_0}[y - w_1(y)] + (\mathbb{E}_{F_0}[w_1(y)] - c_0) \\ &\leq m \cdot \left(\sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g^*} + \sqrt{g_0}\right)^2 + (\mathbb{E}_{F^*}[w_1(y)] - c^*) + g^* - g_0 \\ &\leq m \cdot \mathbb{E}_{F^*}[y - w_1(y)] + (\mathbb{E}_{F^*}[w_1(y)] - c^*) \qquad \text{(A.12)} \\ &= \mathbb{E}_{F^*}[w_2(y)] - c^* = \mathbb{E}_{F'}[w_2(y)] - c', \end{aligned}$$

so he would choose (F', c') in response to w_2. Recall $m = \sqrt{g^*/\mathbb{E}_{F^*}[y - w_1(y)]}$, so the last inequality (A.12) is equivalent to

$$\begin{aligned} & m \cdot \left(\sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g^*} + \sqrt{g_0}\right)^2 + g^* - g_0 \leq m \cdot \mathbb{E}_{F^*}[y - w_1(y)] \\ \Leftrightarrow \quad & \left(1 - \sqrt{\frac{g^*}{\mathbb{E}_{F^*}[y - w_1(y)]}}\right)\left(\sqrt{g_0} - \sqrt{g^*}\right) \geq 0, \end{aligned}$$

which always holds.

This leaves the principal with a payoff of

$$\begin{aligned} \mathbb{E}_{F'}[y - w_2(y)] &= \mathbb{E}_{F'}[y - w_1(y)] - m \cdot \mathbb{E}_{F'}[y - w_1(y)] = (1 - m)\mathbb{E}_{F'}[y - w_1(y)] \\ &= (1 - m)^2 \mathbb{E}_{F^*}[y - w_1(y)] = \left(\sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g^*}\right)^2, \end{aligned}$$

as desired.

This completes the proof. $\square$

Lemma A.5. *If $\Phi_1(w_1, a_1) < \Phi_2(a_1)$ and $(F^*, c^*) \in A_0 \cup \{a_1\}$ attains the maximum in equation (7), and the principal offers the linear contract $w_2(y) = s_2 y$ with s_2 defined by equation (10), then her payoff guarantee in the second period is exactly*

$$\Phi_2(a_1)^2 = \left(\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*}\right)^2.$$

Proof of Lemma A.5. Suppose the principal offers the linear contract $w_2(y) = s_2 y$ with s_2 defined by equation (10). We first show that this guarantees her at least $\left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{c^*}\right)^2$.[24]

Let (F_2, c_2) be the action chosen by agent 2. By agent 2's rationality, we have

$$\mathbb{E}_{F_2}[w_2(y)] - c_2 \geq \mathbb{E}_{F^*}[w_2(y)] - c^*,$$

which further implies that

$$s_2\mathbb{E}_{F_2}[y] = \mathbb{E}_{F_2}[w_2(y)] \geq \mathbb{E}_{F_2}[w_2(y)] - c_2 \geq \mathbb{E}_{F^*}[w_2(y)] - c^* = s_2\mathbb{E}_{F^*}[y] - c^*,$$

and hence

$$\mathbb{E}_{F_2}[y] \geq \mathbb{E}_{F^*}[y] - c^*/s_2.$$

Therefore, the principal's payoff in the second period is

$$\mathbb{E}_{F_2}[y - w_2(y)] = \mathbb{E}_{F_2}[(1-s_2)y] \geq (1-s_2)(\mathbb{E}_{F^*}[y] - c^*/s_2) = \left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{c^*}\right)^2,$$

as desired.

Next we show that her payoff guarantee from this linear contract cannot be strictly higher, since $\left(\sqrt{\mathbb{E}_{F^*}[y]}-\sqrt{c^*}\right)^2$ is exactly her payoff when the technology is $A = A_0 \cup \{a_1, (F', 0)\}$, with $F' = \lambda F^* + (1-\lambda)\delta_0$ where $\lambda = 1 - \sqrt{c^*/\mathbb{E}_{F^*}[y]} \in [0,1]$.

The proof takes two steps. Let $g^* \equiv g(a^*|w_1, a_1) = (\mathbb{E}_{F_1}[w_1(y)] - c_1) - (\mathbb{E}_{F^*}[w_1(y)] - c^*) \geq 0$.

Step 1 $A = A_0 \cup \{a_1, (F', 0)\}$ is compatible with (w_1, a_1). That is, agent 1 chooses a_1 in response to w_1.

Agent 1's payoff from $(F', 0)$ is $\mathbb{E}_{F'}[w_1(y)] = \lambda\mathbb{E}_{F^*}[w_1(y)] = \left(1-\sqrt{c^*/\mathbb{E}_{F^*}[y]}\right)\mathbb{E}_{F^*}[w_1(y)]$, and we have

$$\begin{aligned}\left(1-\sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}}\right)\mathbb{E}_{F^*}[w_1(y)] \leq \mathbb{E}_{F_1}[w_1(y)] - c_1 \quad &\Leftrightarrow \quad \left(1-\sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}}\right)\mathbb{E}_{F^*}[w_1(y)] \leq (\mathbb{E}_{F^*}[w_1(y)] - c^*) + g^* \\ &\Leftrightarrow \quad \sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}}\mathbb{E}_{F^*}[w_1(y)] - c^* + g^* \geq 0.\end{aligned}$$

From

$$\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*} = \Phi_2(a_1) > \Phi_1(w_1, a_1) \geq \sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g^*},$$

we obtain

$$\mathbb{E}_{F^*}[w_1(y)] \geq \mathbb{E}_{F^*}[y] - \left(\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*} + \sqrt{g^*}\right)^2,$$

and thus

$$\begin{aligned}\sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}}\mathbb{E}_{F^*}[w_1(y)] - c^* + g^* &\geq \sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}} \cdot \left(\mathbb{E}_{F^*}[y] - \left(\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*} + \sqrt{g^*}\right)^2\right) - c^* + g^* \\ &= \left(1-\sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}}\right)\left(\sqrt{c^*} - \sqrt{g^*}\right)^2 \geq 0,\end{aligned}$$

as desired. So we indeed have $\mathbb{E}_{F'}[w_1(y)] \leq \mathbb{E}_{F_1}[w_1(y)] - c_1$, implying that agent 1 would choose $a_1 = (F_1, c_1)$ in response to w_1.

[24] This is exactly the static payoff guarantee in Carroll (2015).

Step 2 If $A = A_0 \cup \{a_1, (F', 0)\}$, then agent 2 chooses $(F', 0)$ in response to w_2, leading to a payoff of $\left(\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*}\right)^2$ for the principal.

Agent 2's payoff from $(F', 0)$ is

$$\begin{aligned}\mathbb{E}_{F'}[w_2(y)] = \lambda \mathbb{E}_{F^*}[s_2 y] &= \left(1 - \sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}}\right) \cdot \sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}} \cdot \mathbb{E}_{F^*}[y] \\ &= \left(\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*}\right)\sqrt{c^*} = \sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}} \cdot \mathbb{E}_{F^*}[y] - c^* \\ &= s_2 \mathbb{E}_{F^*}[y] - c^* = \mathbb{E}_{F^*}[w_2(y)] - c^*.\end{aligned}$$

For any action $a_0 = (F_0, c_0) \in A_0 \cup \{a_1\}$, agent 2's payoff from a_0 is

$$\mathbb{E}_{F_0}[w_2(y)] - c_0 = \sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}} \cdot \mathbb{E}_{F_0}[y] - c_0,$$

and we have

$$\sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}} \cdot \mathbb{E}_{F_0}[y] - c_0 \leq \mathbb{E}_{F^*}[w_2(y)] - c^* \quad \Leftrightarrow \quad \sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}} \cdot \mathbb{E}_{F_0}[y] - c_0 \leq \left(\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*}\right)\sqrt{c^*}.$$

From $\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*} = \Phi_2(a_1) \geq \sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}$, we obtain $\mathbb{E}_{F_0}[y] \leq \left(\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*} + \sqrt{c_0}\right)^2$, and thus

$$\begin{aligned}&\left(\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*}\right)\sqrt{c^*} - \left(\sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}} \cdot \mathbb{E}_{F_0}[y] - c_0\right) \\ \geq &\left(\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*}\right)\sqrt{c^*} - \left(\sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}} \cdot \left(\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*} + \sqrt{c_0}\right)^2 - c_0\right) \\ = &\left(1 - \sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}}\right)\left(\sqrt{c^*} - \sqrt{c_0}\right)^2 \geq 0,\end{aligned}$$

as desired. So we indeed have $\mathbb{E}_{F_0}[w_2(y)] - c_0 \leq \mathbb{E}_{F^*}[w_2(y)] - c^* = \mathbb{E}_{F'}[w_2(y)]$, implying that agent 2 would choose $(F', 0)$ in response to w_2.

This leaves the principal with a payoff of

$$\begin{aligned}\mathbb{E}_{F'}[y - w_2(y)] = \lambda \mathbb{E}_{F^*}[(1 - s_2)\, y] &= \left(1 - \sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}}\right)\left(1 - \sqrt{\frac{c^*}{\mathbb{E}_{F^*}[y]}}\right) \cdot \mathbb{E}_{F^*}[y] \\ &= \left(\sqrt{\mathbb{E}_{F^*}[y]} - \sqrt{c^*}\right)^2,\end{aligned}$$

as desired.

This completes the proof. □

We are now ready to prove Lemma 1′.

Proof of Lemma 1′. Combining Lemmas A.4 and A.5, we have shown that by offering the best of the two contracts: (i) $w_2(y) = w_1(y) + m \cdot (y - w_1(y))$ with m defined by equation (9), and (ii) $w_2(y) = s_2 y$ with s_2 defined by equation (10), the principal's payoff guarantee in the second period is exactly given by $(\max\{\Phi_1(w_1, a_1), \Phi_2(a_1)\})^2$. The principal's optimal second-period payoff guarantee, $V_2^*(w_1, a_1)$, is thus at least $(\max\{\Phi_1(w_1, a_1), \Phi_2(a_1)\})^2$.

Now consider an arbitrary second-period contract w_2. It suffices to show that the principal's payoff guarantee is not strictly higher than $(\max\{\Phi_1(w_1, a_1), \Phi_2(a_1)\})^2$ under w_2.

Let $a_0 = (F_0, c_0)$ be the action agent 2 will choose if the true technology is exactly $A_0 \cup \{a_1\}$. Consider the following three cases.

Case 1. $\mathbb{E}_{F_0}[w_2(y)] < c_0$.

Consider the second-period contract w_2 when $A = A_0 \cup \{a_1, (\delta_0, 0)\}$, which is compatible with (w_1, a_1). Agent 2's payoff from $(\delta_0, 0)$ is

$$w_2(0) \geq 0 > \mathbb{E}_{F_0}[w_2(y)] - c_0,$$

so he would prefer to take action $(\delta_0, 0)$. This leaves the principal with a payoff of

$$-w_2(0) \leq 0 \leq \Phi_2(a_1)^2,$$

as desired.

Case 2. $\mathbb{E}_{F_0}[w_2(y)] \geq c_0$, and it holds that

$$\begin{aligned} &\text{either} \quad \text{(i)} \quad \mathbb{E}_{F_0}[w_1(y)] \leq \mathbb{E}_{F_1}[w_1(y)] - c_1, \\ &\text{or} \quad \text{(ii)} \quad \mathbb{E}_{F_0}[w_2(y)] < \frac{\mathbb{E}_{F_0}[w_1(y)]}{\mathbb{E}_{F_0}[w_1(y)] - (\mathbb{E}_{F_1}[w_1(y)] - c_1)} c_0. \end{aligned} \tag{A.13}$$

Let $\lambda = 1 - c_0/\mathbb{E}_{F_0}[w_2(y)] \in [0, 1]$ and let F' be the mixture $\lambda F_0 + (1-\lambda)\delta_0$. Consider the technology $A = A_0 \cup \{a_1, (F', 0)\}$. We proceed with two steps.

Step 1 A is compatible with (w_1, a_1). That is, agent 1 chooses a_1 in response to w_1.

Agent 1's payoff from $(F', 0)$ is

$$\begin{aligned} \mathbb{E}_{F'}[w_1(y)] = \lambda \mathbb{E}_{F_0}[w_1(y)] &= \mathbb{E}_{F_0}[w_1(y)] - \frac{\mathbb{E}_{F_0}[w_1(y)]}{\mathbb{E}_{F_0}[w_2(y)]} c_0 \\ &< \mathbb{E}_{F_1}[w_1(y)] - c_1. \end{aligned} \tag{A.14}$$

Note that inequality (A.14) holds exactly due to the assumptions in (A.13). So agent 1 would prefer to take action $a_1 = (F_1, c_1)$ when $A = A_0 \cup \{a_1, (F', 0)\}$.

Step 2 Agent 2 chooses $(F', 0)$ in response to w_2, resulting in the principal's payoff no more than $\Phi_2(a_1)^2$.

Agent 2's payoff from $(F', 0)$ is

$$\begin{aligned} \mathbb{E}_{F'}[w_2(y)] &= \lambda \mathbb{E}_{F_0}[w_2(y)] + (1-\lambda) w_2(0) \\ &\geq \lambda \mathbb{E}_{F_0}[w_2(y)] = \mathbb{E}_{F_0}[w_2(y)] - c_0. \end{aligned}$$

So he would prefer to take action $(F', 0)$ when $A = A_0 \cup \{a_1, (F', 0)\}$.

This leaves the principal with a payoff of

$$\begin{aligned} \mathbb{E}_{F'}[y - w_2(y)] &= \lambda \mathbb{E}_{F_0}[y - w_2(y)] + (1-\lambda)(0 - w_2(0)) \\ &\leq \lambda \mathbb{E}_{F_0}[y - w_2(y)] = \left(1 - \frac{c_0}{\mathbb{E}_{F_0}[w_2(y)]}\right)(\mathbb{E}_{F_0}[y] - \mathbb{E}_{F_0}[w_2(y)]) \\ &\leq \left(\sqrt{\mathbb{E}_{F_0}[y]} - \sqrt{c_0}\right)^2, \end{aligned} \tag{A.15}$$

which is no more than $\Phi_2(a_1)^2$, as desired. The last inequality (A.15),

$$\begin{aligned}
&\left(1-\frac{c_0}{\mathbb{E}_{F_0}[w_2(y)]}\right)\left(\mathbb{E}_{F_0}[y]-\mathbb{E}_{F_0}[w_2(y)]\right)\le\left(\sqrt{\mathbb{E}_{F_0}[y]}-\sqrt{c_0}\right)^2\\
\Leftrightarrow\quad&\left(\sqrt{\mathbb{E}_{F_0}[w_2(y)]}-\sqrt{\frac{c_0\mathbb{E}_{F_0}[y]}{\mathbb{E}_{F_0}[w_2(y)]}}\right)^2\ge 0,
\end{aligned}$$

which always holds.

Case 3. Both inequalities in (A.13) are reversed, i.e.,

$$\mathbb{E}_{F_0}[w_1(y)]>\mathbb{E}_{F_1}[w_1(y)]-c_1\quad\text{and}\quad\mathbb{E}_{F_0}[w_2(y)]\ge\frac{\mathbb{E}_{F_0}[w_1(y)]}{\mathbb{E}_{F_0}[w_1(y)]-(\mathbb{E}_{F_1}[w_1(y)]-c_1)}c_0.$$

Let

$$\begin{aligned}
\lambda&=\frac{(\mathbb{E}_{F_0}[w_2(y)]-c_0)-(\mathbb{E}_{F_1}[w_1(y)]-c_1)}{\mathbb{E}_{F_0}[w_2(y)]-\mathbb{E}_{F_0}[w_1(y)]},\\
c'&=\frac{\mathbb{E}_{F_0}[w_1(y)](\mathbb{E}_{F_0}[w_2(y)]-c_0)-\mathbb{E}_{F_0}[w_2(y)](\mathbb{E}_{F_1}[w_1(y)]-c_1)}{\mathbb{E}_{F_0}[w_2(y)]-\mathbb{E}_{F_0}[w_1(y)]},
\end{aligned}$$

and let F' be the mixture $\lambda F_0+(1-\lambda)\delta_0$. Consider the technology $A=A_0\cup\{a_1,(F',c')\}$. We proceed with three steps.

Step 1 $\lambda\in[0,1]$ and $c'\ge 0$, so (F',c') is a valid action.

Note that

$$\mathbb{E}_{F_0}[w_2(y)]\ge\frac{\mathbb{E}_{F_0}[w_1(y)]}{\mathbb{E}_{F_0}[w_1(y)]-(\mathbb{E}_{F_1}[w_1(y)]-c_1)}c_0\ge\frac{\mathbb{E}_{F_0}[w_1(y)]}{\mathbb{E}_{F_0}[w_1(y)]-(\mathbb{E}_{F_0}[w_1(y)]-c_0)}c_0=\mathbb{E}_{F_0}[w_1(y)],$$

so the denominator of λ and c' is positive.

Moreover,

$$\begin{aligned}
\mathbb{E}_{F_0}[w_2(y)]-c_0&\ge\frac{\mathbb{E}_{F_1}[w_1(y)]-c_1}{\mathbb{E}_{F_0}[w_1(y)]-(\mathbb{E}_{F_1}[w_1(y)]-c_1)}c_0\\
&\ge\frac{\mathbb{E}_{F_1}[w_1(y)]-c_1}{\mathbb{E}_{F_0}[w_1(y)]-(\mathbb{E}_{F_0}[w_1(y)]-c_0)}c_0=\mathbb{E}_{F_1}[w_1(y)]-c_1,
\end{aligned}$$

so the numerator of λ is positive.

The numerator of c' is positive because

$$\begin{aligned}
&\mathbb{E}_{F_0}[w_1(y)](\mathbb{E}_{F_0}[w_2(y)]-c_0)\ge\mathbb{E}_{F_0}[w_2(y)](\mathbb{E}_{F_1}[w_1(y)]-c_1)\\
\Leftrightarrow\quad&\mathbb{E}_{F_0}[w_2(y)]\ge\frac{\mathbb{E}_{F_0}[w_1(y)]}{\mathbb{E}_{F_0}[w_1(y)]-(\mathbb{E}_{F_1}[w_1(y)]-c_1)}c_0.
\end{aligned}$$

Finally,

$$\begin{aligned}
&(\mathbb{E}_{F_0}[w_2(y)]-c_0)-(\mathbb{E}_{F_1}[w_1(y)]-c_1)\le\mathbb{E}_{F_0}[w_2(y)]-\mathbb{E}_{F_0}[w_1(y)]\\
\Leftrightarrow\quad&\mathbb{E}_{F_0}[w_1(y)]-c_0\le\mathbb{E}_{F_1}[w_1(y)]-c_1,
\end{aligned}$$

so λ is indeed smaller than 1.

Step 2 A is compatible with (w_1, a_1). That is, agent 1 chooses a_1 in response to w_1.
Agent 1's payoff from (F', c') is

$$\mathbb{E}_{F'}[w_1(y)] - c' = \lambda \mathbb{E}_{F_0}[w_1(y)] - c' = \mathbb{E}_{F_1}[w_1(y)] - c_1,$$

so he would prefer to take action $a_1 = (F_1, c_1)$ when $A = A_0 \cup \{a_1, (F', c')\}$.

Step 3 Agent 2 chooses (F', c') in response to w_2, resulting in the principal's payoff no more than $\Phi_1(w_1, a_1)^2$.
Agent 2's payoff from (F', c') is

$$\begin{aligned}\mathbb{E}_{F'}[w_2(y)] - c' &= \lambda \mathbb{E}_{F_0}[w_2(y)] + (1-\lambda) w_2(0) - c' \\ &\geq \lambda \mathbb{E}_{F_0}[w_2(y)] - c' = \mathbb{E}_{F_0}[w_2(y)] - c_0.\end{aligned}$$

So he would prefer to take action (F', c') when $A = A_0 \cup \{a_1, (F', c')\}$.
This leaves the principal with a payoff of

$$\begin{aligned}\mathbb{E}_{F'}[y - w_2(y)] &= \lambda \mathbb{E}_{F_0}[y - w_2(y)] + (1-\lambda)(0 - w_2(0)) \\ &\leq \lambda \mathbb{E}_{F_0}[y - w_2(y)] = \frac{(\mathbb{E}_{F_0}[w_2(y)] - c_0) - (\mathbb{E}_{F_1}[w_1(y)] - c_1)}{\mathbb{E}_{F_0}[w_2(y)] - \mathbb{E}_{F_0}[w_1(y)]}(\mathbb{E}_{F_0}[y] - \mathbb{E}_{F_0}[w_2(y)]) \\ &\leq \left(\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g(a_0|w_1, a_1)}\right)^2,\end{aligned} \tag{A.16}$$

which is no more than $\Phi(w_1, a_1)^2$, as desired. The last inequality (A.16),

$$\begin{aligned}&\frac{(\mathbb{E}_{F_0}[w_2(y)] - c_0) - (\mathbb{E}_{F_1}[w_1(y)] - c_1)}{\mathbb{E}_{F_0}[w_2(y)] - \mathbb{E}_{F_0}[w_1(y)]}(\mathbb{E}_{F_0}[y] - \mathbb{E}_{F_0}[w_2(y)]) \leq \left(\sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} - \sqrt{g(a_0|w_1, a_1)}\right)^2 \\ \Leftrightarrow \quad &\frac{\left(\mathbb{E}_{F_0}[w_2(y)] - \mathbb{E}_{F_0}[w_1(y)] - \sqrt{\mathbb{E}_{F_0}[y - w_1(y)]} \cdot \sqrt{g(a_0|w_1, a_1)}\right)^2}{\mathbb{E}_{F_0}[w_2(y)] - \mathbb{E}_{F_0}[w_1(y)]} \geq 0,\end{aligned}$$

which always holds. (Recall that $g(a_0|w_1, a_1) = (\mathbb{E}_{F_1}[w_1(y)] - c_1) - (\mathbb{E}_{F_0}[w_1(y)] - c_0) \geq 0$.)

Summing up the above three cases, we prove that the principal's payoff guarantee is not strictly higher than $(\max\{\Phi_1(w_1, a_1), \Phi_2(a_1)\})^2$ under any second-period contract w_2.
This completes the proof. □

A.3.2 Proofs for Subsection 5.1.2

To prove Lemma 2′, we start by establishing the following Lemma A.6.

Lemma A.6. *Suppose the known technology A_0 satisfies lower bound on marginal cost. If $\Phi_1(w_1, a_1) \geq \Phi_2(a_1)$ and $a^* = (F^*, c^*) \in A_0$ attains the maximum in equation (6), then (i) $c^* \leq c_0$, (ii) $\mathbb{E}_{F^*}[y] \leq \mathbb{E}_{F_0}[y]$, and (iii) $\mathbb{E}_{F^*}[w_1(y)] \leq \mathbb{E}_{F^*}[\hat{w}_1(y)] = s_1 \mathbb{E}_{F^*}[y]$, where $\hat{w}_1$ is defined by equation (3).*

Proof of Lemma A.6. Let $g_0 \equiv g(a_0|w_1, a_1) = (\mathbb{E}_{F_1}[w_1(y)] - c_1) - (\mathbb{E}_{F_0}[w_1(y)] - c_0) \geq 0$, and $g^* \equiv g(a^*|w_1, a_1) = (\mathbb{E}_{F_1}[w_1(y)] - c_1) - (\mathbb{E}_{F^*}[w_1(y)] - c^*) \geq 0$. By assumption, we have

$$\mathbb{E}_{F_0}[w_1(y)] - c_0 \geq \mathbb{E}_{F^*}[w_1(y)] - c^* \quad \Rightarrow \quad g^* \geq g_0.$$

Note that

$$\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}-\sqrt{g^*}=\Phi_1(w_1,a_1)\geq\sqrt{\mathbb{E}_{F_0}\left[y-w_1(y)\right]}-\sqrt{g_0}. \tag{A.17}$$

We first argue that $c^* \leq c_0$ must hold, otherwise there will be a contradiction to the assumption that A_0 satisfies lower bound on marginal cost.

Suppose not, i.e., $c^* > c_0$. Consider the following two cases.

Case 1. $\sqrt{\mathbb{E}_{F_0}\left[y-w_1(y)\right]}\geq\sqrt{g_0}$.

From equation (A.17) we obtain

$$\begin{aligned}\frac{\left(\mathbb{E}_{F^*}\left[y-w_1(y)\right]\right)-\left(\mathbb{E}_{F_0}\left[y-w_1(y)\right]\right)}{\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}+\sqrt{\mathbb{E}_{F_0}\left[y-w_1(y)\right]}}&=\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}-\sqrt{\mathbb{E}_{F_0}\left[y-w_1(y)\right]}\\&\geq\sqrt{g^*}-\sqrt{g_0}=\frac{\left(\mathbb{E}_{F_0}\left[w_1(y)\right]-c_0\right)-\left(\mathbb{E}_{F^*}\left[w_1(y)\right]-c^*\right)}{\sqrt{g^*}+\sqrt{g_0}}.\end{aligned}$$

Since $\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}>\sqrt{g^*}$ and $\sqrt{\mathbb{E}_{F_0}\left[y-w_1(y)\right]}\geq\sqrt{g_0}$, the above expression implies that

$$\begin{aligned}&\left(\mathbb{E}_{F^*}\left[y-w_1(y)\right]\right)-\left(\mathbb{E}_{F_0}\left[y-w_1(y)\right]\right)>\left(\mathbb{E}_{F_0}\left[w_1(y)\right]-c_0\right)-\left(\mathbb{E}_{F^*}\left[w_1(y)\right]-c^*\right)\\\Rightarrow\quad&\mathbb{E}_{F^*}[y]-\mathbb{E}_{F_0}[y]>c^*-c_0>0,\end{aligned}$$

a contradiction to the assumption that A_0 satisfies lower bound on marginal cost!

Case 2. $\sqrt{\mathbb{E}_{F_0}\left[y-w_1(y)\right]}<\sqrt{g_0}$.

We have

$$\mathbb{E}_{F_0}\left[y-w_1(y)\right]<g_0=\left(\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1\right)-\left(\mathbb{E}_{F_0}\left[w_1(y)\right]-c_0\right)\quad\Rightarrow\quad\mathbb{E}_{F_0}[y]-c_0<\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1.$$

Similarly, from $\Phi_1(w_1,a_1)\geq\Phi_2(a_1)>0$, we have $\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}-\sqrt{g^*}>0$, and thus

$$\mathbb{E}_{F^*}\left[y-w_1(y)\right]>g^*=\left(\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1\right)-\left(\mathbb{E}_{F^*}\left[w_1(y)\right]-c^*\right)\quad\Rightarrow\quad\mathbb{E}_{F^*}[y]-c^*>\mathbb{E}_{F_1}\left[w_1(y)\right]-c_1.$$

It follows that

$$\mathbb{E}_{F^*}[y]-c^*>\mathbb{E}_{F_0}[y]-c_0\quad\Rightarrow\quad\mathbb{E}_{F^*}[y]-\mathbb{E}_{F_0}[y]>c^*-c_0>0,$$

another contradiction to the assumption that A_0 satisfies lower bound on marginal cost!

Summing up the above two cases, we show that $c^*\leq c_0$. It follows from lower bound on marginal cost that $\mathbb{E}_{F^*}[y]\leq\mathbb{E}_{F_0}[y]$.

Moreover, $\mathbb{E}_{F_0}\left[w_1(y)\right]-c_0\geq\mathbb{E}_{F^*}\left[w_1(y)\right]-c^*$ implies that

$$\mathbb{E}_{F_0}\left[w_1(y)\right]-\mathbb{E}_{F^*}\left[w_1(y)\right]\geq c_0-c^*\geq0\quad\Rightarrow\quad\mathbb{E}_{F_0}\left[w_1(y)\right]\geq\mathbb{E}_{F^*}\left[w_1(y)\right].$$

Equation (A.17) implies that

$$\sqrt{\mathbb{E}_{F^*}\left[y-w_1(y)\right]}-\sqrt{\mathbb{E}_{F_0}\left[y-w_1(y)\right]}\geq\sqrt{g^*}-\sqrt{g_0}\geq0\quad\Rightarrow\quad\mathbb{E}_{F^*}\left[y-w_1(y)\right]\geq\mathbb{E}_{F_0}\left[y-w_1(y)\right].$$

Combining the above two inequalities, we have

$$\begin{aligned}
&\frac{\mathbb{E}_{F^*}[y - w_1(y)]}{\mathbb{E}_{F^*}[w_1(y)]} \geq \frac{\mathbb{E}_{F_0}[y - w_1(y)]}{\mathbb{E}_{F_0}[w_1(y)]} \\
\Rightarrow \quad &\frac{\mathbb{E}_{F^*}[y]}{\mathbb{E}_{F^*}[w_1(y)]} \geq \frac{\mathbb{E}_{F_0}[y]}{\mathbb{E}_{F_0}[w_1(y)]} = \frac{1}{s_1} \\
\Rightarrow \quad &\mathbb{E}_{F^*}[w_1(y)] \leq s_1 \mathbb{E}_{F^*}[y],
\end{aligned} \tag{A.18}$$

as desired. The equality in (A.18) follows from the definition in (3). □

Proof of Lemma 2′. Consider an arbitrary action $a_1 = (F_1, c_1)$ agent 1 would take under contract $\hat{w}_1$. We need to show that the principal's interim payoff guarantee, $U(\hat{w}_1|a_1)$, is at least $U(w_1)$. Lemma 1′ shows that the principal's optimal second-period payoff guarantee is

$$V_2^*(\hat{w}_1, a_1) = (\max\{\Phi_1(w_1, a_1), \Phi_2(a_1)\})^2,$$

where

$$\begin{aligned}
\Phi_1(\hat{w}_1, a_1) &= \max_{a \in A_0 \cup \{a_1\}} \left\{ \sqrt{\mathbb{E}_{F_a}[y - \hat{w}_1(y)]} - \sqrt{g(a|\hat{w}_1, a_1)} \right\}, \\
\Phi_2(a_1) &= \max_{a \in A_0 \cup \{a_1\}} \left\{ \sqrt{\mathbb{E}_{F_a}[y]} - \sqrt{c_a} \right\},
\end{aligned}$$

and her interim payoff guarantee is

$$U(\hat{w}_1|a_1) = \mathbb{E}_{F_1}[y - \hat{w}_1(y)] + \beta \cdot V_2^*(\hat{w}_1, a_1).$$

It suffices to construct another action a_1', which may be taken by agent 1 under w_1 and some other technology, such that $U(w_1|a_1') \leq U(\hat{w}_1|a_1)$. By assumption, a_0 is agent 1's best response if $A = A_0$, so an action a_1' may be taken by agent 1 under w_1 if and only if the incentive gap with respect to a_0 is nonnegative, i.e., $g(a_0|w_1, a_1') \geq 0$. Consider the following two cases.

Case 1. $\mathbb{E}_{F_1}[y] \geq \mathbb{E}_{F_0}[y]$.

Let $a_1' = a_0$. When agent 1 takes action a_0 in response to w_1, the principal's resulting payoff in the first period is

$$\mathbb{E}_{F_0}[y - w_1(y)] = (1 - s_1)\mathbb{E}_{F_0}[y] \leq (1 - s_1)\mathbb{E}_{F_1}[y] = \mathbb{E}_{F_1}[y - \hat{w}_1(y)],$$

so her payoff in the first period under $(w_1|a_0)$ is weakly lower than under $(\hat{w}_1|a_1)$.

Moreover, it follows from Lemma 1′ that the principal's optimal second-period payoff guarantee is

$$V_2^*(w_1, a_0) = (\max\{\Phi_1(w_1, a_0), \Phi_2(a_0)\})^2.$$

We now show that $V_2^*(w_1, a_0) \leq V_2^*(\hat{w}_1, a_1)$, which is equivalent to

$$\max\{\Phi_1(w_1, a_0), \Phi_2(a_0)\} \leq \max\{\Phi_1(w_1, a_1), \Phi_2(a_1)\}.$$

Note that

$$\Phi_1(w_1, a_0) = \max_{a \in A_0} \left\{ \sqrt{\mathbb{E}_{F_a}[y - w_1(y)]} - \sqrt{g(a|w_1, a_0)} \right\},$$
$$\Phi_2(a_0) = \max_{a \in A_0} \left\{ \sqrt{\mathbb{E}_{F_a}[y]} - \sqrt{c_a} \right\}.$$

By definition we have $0 < \Phi_2(a_0) \leq \Phi_2(a_1)$. Thus, it suffices to show that whenever $\Phi_1(w_1, a_0) > \Phi_2(a_0)$, it holds that $\Phi_1(w_1, a_0) \leq \Phi_1(\hat{w}_1, a_1)$.

Let $a^* = (F^*, c^*) \in A_0$ attains the maximum in $\Phi_1(w_1, a_0)$. It follows from Lemma A.6 that $\mathbb{E}_{F^*}[y] \leq \mathbb{E}_{F_0}[y] \leq \mathbb{E}_{F_1}[y]$ and $\mathbb{E}_{F^*}[w_1(y)] \leq s_1 \mathbb{E}_{F^*}[y]$.

We claim that

$$\Phi_1(w_1, a_0) = \sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g(a^*|w_1, a_0)} \leq \sqrt{\mathbb{E}_{F_1}[y - \hat{w}_1(y)]} \leq \Phi_1(\hat{w}_1, a_1).$$

must hold. Suppose not, then

$$\sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g(a^*|w_1, a_0)} > \sqrt{\mathbb{E}_{F_1}[y - \hat{w}_1(y)]},$$

which implies that

$$\sqrt{(1 - s_1)\mathbb{E}_{F^*}[y]} \geq \sqrt{\mathbb{E}_{F^*}[y - w_1(y)]} - \sqrt{g(a^*|w_1, a_0)} > \sqrt{\mathbb{E}_{F_1}[y - \hat{w}_1(y)]} = \sqrt{(1 - s_1)\mathbb{E}_{F_1}[y]},$$

a contradiction to $\mathbb{E}_{F^*}[y] \leq \mathbb{E}_{F_1}[y]$!

Therefore, whenever $\Phi_1(w_1, a_0) > \Phi_2(a_0)$, it holds that $\Phi_1(w_1, a_0) \leq \Phi_1(\hat{w}_1, a_1)$, which implies $V_2^*(w_1, a_0) \leq V_2^*(\hat{w}_1, a_1)$. The principal's interim payoff guarantee is

$$\begin{aligned} U(w_1|a_0) &= \mathbb{E}_{F_0}[y - w_1(y)] + \beta \cdot V_2^*(w_1, a_0) \\ &\leq \mathbb{E}_{F_1}[y - \hat{w}_1(y)] + \beta \cdot V_2^*(\hat{w}_1, a_1) = U(\hat{w}_1|a_1), \end{aligned}$$

as desired.

Case 2. $\mathbb{E}_{F_1}[y] < \mathbb{E}_{F_0}[y]$.

Let $\lambda = \mathbb{E}_{F_1}[y]/\mathbb{E}_{F_0}[y] \in [0, 1]$ and let F_1' be the mixture $\lambda F_0 + (1 - \lambda)\delta_0$. Note that $\mathbb{E}_{F_1'}[y] = \mathbb{E}_{F_1}[y]$. Consider $a_1' = (F_1', c_1)$. For any action a, the corresponding incentive gap with respect to a is

$$g(a|w_1, a_1') = \left(\mathbb{E}_{F_1'}[w_1(y)] - c_1\right) - \left(\mathbb{E}_{F_a}[w_1(y)] - c_a\right).$$

Note that

$$\mathbb{E}_{F_1'}[w_1(y)] - c_1 = \lambda \mathbb{E}_{F_0}[w_1(y)] - c_1 = \lambda s_1 \mathbb{E}_{F_0}[y] - c_1 = s_1 \mathbb{E}_{F_1}[y] - c_1 = \mathbb{E}_{F_1}[\hat{w}_1(y)] - c_1,$$

and

$$\mathbb{E}_{F_0}[w_1(y)] - c_0 = s_1 \mathbb{E}_{F_0}[y] - c_0 = \mathbb{E}_{F_0}[\hat{w}_1(y)] - c_0.$$

Thus,

$$\begin{aligned} g\left(a_0\middle|w_1, a_1'\right) &= \left(\mathbb{E}_{F_1'}\left[w_1(y)\right] - c_1\right) - \left(\mathbb{E}_{F_0}\left[w_1(y)\right] - c_0\right) \\ &= \left(\mathbb{E}_{F_1}\left[\hat{w}_1(y)\right] - c_1\right) - \left(\mathbb{E}_{F_0}\left[\hat{w}_1(y)\right] - c_0\right) \\ &= g\left(a_0|\hat{w}_1, a_1\right) \geq 0, \end{aligned}$$

implying that a_1' may be chosen by agent 1 in response to w_1 under some technology.

When agent 1 chooses action a_1' in response, the principal's resulting payoff in the first period is

$$\mathbb{E}_{F_1'}\left[y - w_1(y)\right] = \lambda\mathbb{E}_{F_0}\left[y - w_1(y)\right] = \lambda\left(1 - s_1\right)\mathbb{E}_{F_0}\left[y\right] = \left(1 - s_1\right)\mathbb{E}_{F_1}\left[y\right] = \mathbb{E}_{F_1}\left[y - \hat{w}_1(y)\right],$$

so her payoff in the first period under $\left(w_1\middle|a_1'\right)$ and under $\left(\hat{w}_1|a_1\right)$ are exactly equal.

Moreover, it follows from Lemma 1′ that the principal's optimal second-period payoff guarantee under $\left(w_1\middle|a_1'\right)$ is

$$V_2^*\left(w_1, a_1'\right) = \left(\max\left\{\Phi_1\left(w_1, a_1'\right), \Phi_2\left(a_1'\right)\right\}\right)^2.$$

We now show that $V_2^*\left(w_1, a_1'\right) \leq V_2^*\left(\hat{w}_1, a_1'\right)$, which is equivalent to

$$\max\left\{\Phi_1\left(w_1, a_1'\right), \Phi_2\left(a_1'\right)\right\} \leq \max\left\{\Phi_1\left(w_1, a_1\right), \Phi_2\left(a_1\right)\right\}.$$

Note that

$$\begin{aligned} \Phi_1\left(w_1, a_1'\right) &= \max_{a \in A_0 \cup \left\{a_1'\right\}}\left\{\sqrt{\mathbb{E}_{F_a}\left[y - w_1(y)\right]} - \sqrt{g\left(a\middle|w_1, a_1'\right)}\right\}, \\ \Phi_2\left(a_1'\right) &= \max_{a \in A_0 \cup \left\{a_1'\right\}}\left\{\sqrt{\mathbb{E}_{F_a}[y]} - \sqrt{c_a}\right\}. \end{aligned}$$

From $\mathbb{E}_{F_1'}[y] = \mathbb{E}_{F_1}[y]$, it follows that $\Phi_2\left(a_1'\right) = \Phi_2\left(a_1\right) > 0$. Thus, it suffices to show that whenever $\Phi_1\left(w_1, a_1'\right) > \Phi_2\left(a_1'\right)$, it holds that $\Phi_1\left(w_1, a_1'\right) \leq \Phi_1\left(\hat{w}_1, a_1\right)$.

Let $a^* = \left(F^*, c^*\right) \in A_0 \cup \left\{a_1'\right\}$ attains the maximum in $\Phi_1\left(w_1, a_1'\right)$.

1. If $a^* = a_1'$, then

$$\begin{aligned} \Phi_1\left(w_1, a_1'\right) &= \sqrt{\mathbb{E}_{F_1'}\left[y - w_1(y)\right]} - \sqrt{g\left(a_1'\middle|w_1, a_1'\right)} \\ &= \sqrt{\mathbb{E}_{F_1}\left[y - \hat{w}_1(y)\right]} - \sqrt{g\left(a_1|\hat{w}_1, a_1\right)} \leq \Phi_1\left(\hat{w}_1, a_1\right), \end{aligned}$$

   as desired.

2. If $a^* \in A_0$, then it follows from Lemma A.6 that $\mathbb{E}_{F^*}\left[w_1(y)\right] \leq \mathbb{E}_{F^*}\left[\hat{w}_1(y)\right]$.

   From $\Phi_1\left(w_1, a_1'\right) > \Phi_2\left(a_1'\right) > 0$, we have $\Phi_1\left(w_1, a_1'\right) = \sqrt{\mathbb{E}_{F^*}\left[y - w_1(y)\right]} - \sqrt{g\left(a^*\middle|w_1, a_1'\right)} > 0$, and thus

$$\begin{aligned} & \mathbb{E}_{F^*}\left[y - w_1(y)\right] > g\left(a^*\middle|w_1, a_1'\right) = \left(\mathbb{E}_{F_1'}\left[w_1(y)\right] - c_1\right) - \left(\mathbb{E}_{F^*}\left[w_1(y)\right] - c^*\right) \\ \Rightarrow\ & \mathbb{E}_{F^*}\left[y\right] - c^* > \mathbb{E}_{F_1'}\left[w_1(y)\right] - c_1 = \mathbb{E}_{F_1}\left[\hat{w}_1(y)\right] - c_1 \\ \Rightarrow\ & \mathbb{E}_{F^*}\left[y - \hat{w}_1(y)\right] > \left(\mathbb{E}_{F_1}\left[\hat{w}_1(y)\right] - c_1\right) - \left(\mathbb{E}_{F^*}\left[\hat{w}_1(y)\right] - c^*\right) = g\left(a^*\middle|\hat{w}_1, a_1\right). \end{aligned}$$

We claim that

$$\Phi_1\left(w_1, a_1'\right) = \sqrt{\mathbb{E}_{F^*}\left[y - w_1(y)\right]} - \sqrt{g\left(a^*\middle|w_1, a_1'\right)} \le \sqrt{\mathbb{E}_{F^*}\left[y - \hat{w}_1(y)\right]} - \sqrt{g\left(a^*|\hat{w}_1, a_1\right)} \le \Phi_1\left(\hat{w}_1, a_1\right).$$

must hold. Suppose not, then

$$\begin{aligned}
& \sqrt{\mathbb{E}_{F^*}\left[y - w_1(y)\right]} - \sqrt{g\left(a^*\middle|w_1, a_1'\right)} \le \sqrt{\mathbb{E}_{F^*}\left[y - \hat{w}_1(y)\right]} - \sqrt{g\left(a^*|\hat{w}_1, a_1\right)} \\
\Leftrightarrow\quad & \sqrt{\mathbb{E}_{F^*}\left[y - w_1(y)\right]} - \sqrt{\mathbb{E}_{F^*}\left[y - \hat{w}_1(y)\right]} \le \sqrt{g\left(a^*\middle|w_1, a_1'\right)} - \sqrt{g\left(a^*|\hat{w}_1, a_1\right)} \\
\Leftrightarrow\quad & \frac{\mathbb{E}_{F^*}\left[y - w_1(y)\right] - \mathbb{E}_{F^*}\left[y - \hat{w}_1(y)\right]}{\sqrt{\mathbb{E}_{F^*}\left[y - w_1(y)\right]} + \sqrt{\mathbb{E}_{F^*}\left[y - \hat{w}_1(y)\right]}} \le \frac{g\left(a^*\middle|w_1, a_1'\right) - g\left(a^*|\hat{w}_1, a_1\right)}{\sqrt{g\left(a^*\middle|w_1, a_1'\right)} + \sqrt{g\left(a^*|\hat{w}_1, a_1\right)}}.
\end{aligned} \tag{A.19}$$

Note that

$$\mathbb{E}_{F^*}\left[y - w_1(y)\right] - \mathbb{E}_{F^*}\left[y - \hat{w}_1(y)\right] = \mathbb{E}_{F^*}\left[\hat{w}_1(y)\right] - \mathbb{E}_{F^*}\left[w_1(y)\right] \ge 0,$$

and that

$$\begin{aligned}
& g\left(a^*\middle|w_1, a_1'\right) - g\left(a^*|\hat{w}_1, a_1\right) \\
= & \left(\left(\mathbb{E}_{F_1'}\left[w_1(y)\right] - c_1\right) - \left(\mathbb{E}_{F^*}\left[w_1(y)\right] - c^*\right)\right) - \left(\left(\mathbb{E}_{F_1}\left[\hat{w}_1(y)\right] - c_1\right) - \left(\mathbb{E}_{F^*}\left[\hat{w}_1(y)\right] - c^*\right)\right) \\
= & \mathbb{E}_{F^*}\left[\hat{w}_1(y)\right] - \mathbb{E}_{F^*}\left[w_1(y)\right] \ge 0.
\end{aligned}$$

Therefore, inequality (A.21) is equivalent to

$$\frac{\mathbb{E}_{F^*}\left[\hat{w}_1(y)\right] - \mathbb{E}_{F^*}\left[w_1(y)\right]}{\sqrt{\mathbb{E}_{F^*}\left[y - w_1(y)\right]} + \sqrt{\mathbb{E}_{F^*}\left[y - \hat{w}_1(y)\right]}} \le \frac{\mathbb{E}_{F^*}\left[\hat{w}_1(y)\right] - \mathbb{E}_{F^*}\left[w_1(y)\right]}{\sqrt{g\left(a^*\middle|w_1, a_1'\right)} + \sqrt{g\left(a^*|\hat{w}_1, a_1\right)}},$$

which is implied by $\mathbb{E}_{F^*}\left[y - w_1(y)\right] > g\left(a^*\middle|w_1, a_1'\right)$ and $\mathbb{E}_{F^*}\left[y - \hat{w}_1(y)\right] > g\left(a^*|\hat{w}_1, a_1\right)$.

Therefore, whenever $\Phi_1\left(w_1, a_1'\right) > \Phi_2\left(a_1'\right)$, it holds that $\Phi_1\left(w_1, a_1'\right) \le \Phi_1\left(\hat{w}_1, a_1'\right)$, which implies $V_2^*\left(w_1, a_1'\right) \le V_2^*\left(\hat{w}_1, a_1\right)$. The principal's interim payoff guarantee is

$$\begin{aligned}
U\left(w_1\middle|a_1'\right) &= \mathbb{E}_{F_1'}\left[y - w_1(y)\right] + \beta \cdot V_2^*\left(w_1, a_1'\right) \\
&\le \mathbb{E}_{F_1}\left[y - \hat{w}_1(y)\right] + \beta \cdot V_2^*\left(\hat{w}_1, a_1\right) = U\left(\hat{w}_1|a_1\right),
\end{aligned}$$

as desired.

This completes the proof. □

Proof of Lemma 3′. We first reformulate program (11) as an equivalent maximization problem with continuous objective function and compact feasible region. Slightly abusing notation, we use $U(s_1)$ instead of $U(w_1)$ to denote the infimum value of program (11).

Plug $w_1(y) = s_1 y$ into equation (6). We may rewrite $\Phi_1(w_1, a_1)$ as

$$\Phi_1(w_1, a_1) = \max_{a \in A_0 \cup \{a_1\}} \left\{ \sqrt{(1 - s_1)\,\mathbb{E}_{F_a}[y]} - \sqrt{g(a|w_1, a_1)} \right\}.$$

Similarly, for $a \in A_0 \cup \{a_1\}$,

$$g(a|w_1, a_1) = (s_1 \mathbb{E}_{F_1}[y] - c_1) - (s_1 \mathbb{E}_{F_a}[y] - c_a) \geq 0.$$

Note that both the objective and the constraints of program (11) depend on the choice variables (F_1, c_1) only through the value of $(\mathbb{E}_{F_1}[y], c_1)$. Rewrite $\mathbb{E}_{F_1}[y] = x$ and $c_1 = z$ with $x, z \geq 0$. Plugging into the original program (11), we obtain an equivalent program

$$\begin{aligned} U(s_1) = \inf_{x,z} \quad & (1 - s_1) x + \beta \cdot \max\left\{\hat{\Phi}_1(x, z; s_1), \hat{\Phi}_2(x, z)\right\}^2 \\ \text{s.t.} \quad & s_1 x - z \geq \max_{a \in A_0 \cup \{(\delta_0, 0)\}} \{s_1 \mathbb{E}_{F_a}[y] - c_a\}, \quad x, z \geq 0, \end{aligned} \tag{A.20}$$

where

$$\hat{\Phi}_1(x, z; s_1) \equiv \max\left\{\sqrt{(1 - s_1) x}, \max_{a \in A_0}\left\{\sqrt{(1 - s_1)\mathbb{E}_{F_a}[y]} - \sqrt{(s_1 x - z) - (s_1 \mathbb{E}_{F_a}[y] - c_a)}\right\}\right\}, \tag{A.21}$$

$$\hat{\Phi}_2(x, z) \equiv \max\left\{\sqrt{x} - \sqrt{z}, \max_{a \in A_0}\left\{\sqrt{\mathbb{E}_{F_a}[y]} - \sqrt{c_a}\right\}\right\}. \tag{A.22}$$

Let $\overline{x} \equiv \max_{a \in A_0} \mathbb{E}_{F_a}[y] > 0$, and $\overline{v} \equiv \max_{a \in A_0}\left\{\sqrt{\mathbb{E}_{F_a}[y]} - \sqrt{c_a}\right\} > 0$. Suppose

$$(F_0, c_0) \in \underset{a \in A_0 \cup \{(\delta_0, 0)\}}{\arg\max} \{s_1 \mathbb{E}_{F_a}[y] - c_a\}.$$

Note that $(x_0, z_0) = (\mathbb{E}_{F_0}[y], c_0)$ is feasible in program (A.20) and leads to objective value

$$(1 - s_1) x_0 + \beta \cdot \max\left\{\hat{\Phi}_1(x_0, z_0; s_1), \hat{\Phi}_2(x_0, z_0)\right\}^2 \leq (1 - s_1)\overline{x} + \beta \cdot \max\left\{\sqrt{(1 - s_1)\overline{x}}, \overline{v}\right\}^2.$$

If $x \geq (1 + \beta)\overline{x}$, then

$$\begin{aligned} (1 - s_1) x + \beta \cdot \max\left\{\hat{\Phi}_1(x, z; s_1), \hat{\Phi}_2(x, z)\right\}^2 &\geq (1 - s_1)(1 + \beta)\overline{x} + \beta \cdot \overline{v}^2 \\ &= (1 - s_1)\overline{x} + \beta(1 - s_1)\overline{x} + \beta \cdot \overline{v}^2 \\ &\geq (1 - s_1) + \beta \cdot \max\left\{\sqrt{(1 - s_1)\overline{x}}, \overline{v}\right\}^2. \end{aligned}$$

Therefore, restricting $x \in [0, (1 + \beta)\overline{x}]$ will not change the infimum of program (A.20). Moreover,

$$s_1 x - z \geq 0 \quad \Rightarrow \quad z \leq s_1 x \leq x,$$

so restricting $(x, z) \in [0, (1 + \beta)\overline{x}]^2$ will not change the infimum of program (A.20).

Consider the following program

$$\begin{aligned} \Psi^*(s_1) \equiv \sup_{x,z} \quad & \Psi(x, z; s_1) \equiv -\left((1 - s_1) x + \beta \cdot \max\left\{\hat{\Phi}_1(x, z; s_1), \hat{\Phi}_2(x, z)\right\}^2\right) \\ \text{s.t.} \quad & (x, z) \in \Gamma(s_1), \end{aligned} \tag{A.23}$$

where $\hat{\Phi}_1$ is defined by equation (A.21), $\hat{\Phi}_2$ is defined by equation (A.22), and Γ is defined as follows:

$$\Gamma(s_1) \equiv \left\{(x, z) \in [0, (1 + \beta)\overline{x}]^2 : s_1 x - z \geq \max_{a \in A_0 \cup \{(\delta_0, 0)\}} \{s_1 \mathbb{E}_{F_a}[y] - c_a\}\right\}.$$

By definition, $\Psi : [0,(1+\beta)\overline{x}]^2 \times [0,1] \to \mathbb{R}$ is a continuous function, and $\Gamma : [0,1] \rightrightarrows [0,(1+\beta)\overline{x}]^2$ is a compact-valued and nonempty-valued correspondence. Moreover, the infimum of program (A.20), $U(s_1)$, is given by $-\Psi^*(s_1)$.

Note that for each s_1, $\Gamma(s_1)$ defines a half plane intersecting a square, and that the half plane shifts linearly in s_1. Thus, Γ is both upper and lower hemicontinuous. It then follows from Berge's maximum theorem that Ψ^* is continuous, and

$$\Gamma^*(s_1) \equiv \{(x,z) \in \Gamma(s_1) : \Psi(x,z;s_1) = \Psi^*(s_1)\}$$

is upper hemicontinuous with nonempty and compact values. As a consequence, a solution to program (A.26) exists for all s_1, and the supremum can be replaced by maximum.

It follows that the infimum in program (A.20) and therefore the original program (11) can both be replaced by minimum, and the resulting minimum value $U(s_1) = -\Psi^*(s_1)$ is continuous in s_1. Hence, $U(s_1)$ achieves a maximum over $[0,1]$. This maximum is also the optimal guarantee over all linear contracts. □

Proof of Theorem 1′. According to Lemma 3′, among all linear first-period contracts, there exists an optimal one, call it w_1^*. If w_1 is any other (nonlinear) first-period contract that outperforms w_1^*, then by Lemma 2′, there is a linear contract that in turn does at least as well as w_1. But this contradicts the fact that w_1^* is an optimal linear contract. Therefore, w_1^* is optimal among all first-period contracts. □

A.3.3 Proofs for Subsection 5.2

To prove Theorem 2, we start by establishing two lemmas, Lemmas A.7 and A.8. Lemma A.7 shows that any nonlinear contract is outperformed by some linear one, and Lemma A.8 further shows that the maximum of the principal's first-period problem exists within the class of linear first-period contracts.

Lemma A.7. *Let w_1 be any first-period contract, and let $(F_0, c_0) \in A_0$ be agent 1's best response when his technology $A_1 = A_0$. Under technological advances, the linear contract $\hat{w}_1$ defined by equation (3) satisfies $\hat{U}(\hat{w}_1) \geq \hat{U}(w_1)$.*

Proof of Lemma A.7. Consider an arbitrary action $a_1 = (F_1, c_1)$ agent 1 would take under contract $\hat{w}_1$. We need to show that the principal's interim payoff guarantee, $\hat{U}(\hat{w}_1|a_1)$, is at least $\hat{U}(w_1)$. Note that

$$\hat{U}(\hat{w}_1|a_1) = \mathbb{E}_{F_1}[y - \hat{w}_1(y)] + \beta \cdot \hat{V}_2^*(a_1),$$

where $\hat{V}_2^*(a_1) = \Phi_2(a_1)^2$ with

$$\Phi_2(a_1) = \max_{a \in A_0 \cup \{a_1\}} \left\{ \sqrt{\mathbb{E}_{F_a}[y]} - \sqrt{c_a} \right\}.$$

It suffices to construct another action a_1', which may be taken by agent 1 under w_1 and some other technology, such that $\hat{U}\left(w_1 \middle| a_1'\right) \leq \hat{U}(\hat{w}_1|a_1)$. By assumption, a_0 is agent 1's best response if $A_1 = A_0$, so an action a_1' may be taken by agent 1 under w_1 if and only if the incentive gap with respect to a_0 is nonnegative, i.e., $g\left(a_0 \middle| w_1, a_1'\right) \geq 0$. Consider the following two cases.

Case 1. $\mathbb{E}_{F_1}[y] \geq \mathbb{E}_{F_0}[y]$.

Let $a_1' = a_0$. When agent 1 takes action a_0 in response to w_1, the principal's resulting payoff in the first period is

$$\mathbb{E}_{F_0}[y - w_1(y)] = (1 - s_1)\mathbb{E}_{F_0}[y] \leq (1 - s_1)\mathbb{E}_{F_1}[y] = \mathbb{E}_{F_1}[y - \hat{w}_1(y)],$$

so her payoff in the first period under $(w_1|a_0)$ is weakly lower than under $(\hat{w}_1|a_1)$.

Moreover, the principal's optimal second-period payoff guarantee is $\hat{V}_2^* (a_0) = \Phi_2 (a_0)^2$ with

$$\Phi_2 (a_0) = \max_{a \in A_0} \left\{ \sqrt{\mathbb{E}_{F_a}[y]} - \sqrt{c_a} \right\}.$$

By definition we have $0 < \Phi_2 (a_0) \leq \Phi_2 (a_1)$, which implies $\hat{V}_2^* (a_0) \leq \hat{V}_2^* (a_1)$. The principal's interim payoff guarantee is

$$\begin{aligned}\hat{U} (w_1 | a_0) &= \mathbb{E}_{F_0} [y - w_1(y)] + \beta \cdot \hat{V}_2^* (a_0) \\ &\leq \mathbb{E}_{F_1} [y - \hat{w}_1(y)] + \beta \cdot \hat{V}_2^* (a_1) = \hat{U} (\hat{w}_1 | a_1),\end{aligned}$$

as desired.

Case 2. $\mathbb{E}_{F_1} [y] < \mathbb{E}_{F_0} [y]$.

Let $\lambda = \mathbb{E}_{F_1}[y]/\mathbb{E}_{F_0}[y] \in [0, 1]$ and let F_1' be the mixture $\lambda F_0 + (1 - \lambda) \delta_0$. Note that $\mathbb{E}_{F_1'} [y] = \mathbb{E}_{F_1}[y]$. Consider $a_1' = \left(F_1', c_1\right)$. Note that

$$\mathbb{E}_{F_1'} [w_1 (y)] - c_1 = \lambda \mathbb{E}_{F_0} [w_1 (y)] - c_1 = \lambda s_1 \mathbb{E}_{F_0} [y] - c_1 = s_1 \mathbb{E}_{F_1} [y] - c_1 = \mathbb{E}_{F_1} [\hat{w}_1 (y)] - c_1,$$

and

$$\mathbb{E}_{F_0} [w_1 (y)] - c_0 = s_1 \mathbb{E}_{F_0} [y] - c_0 = \mathbb{E}_{F_0} [\hat{w}_1 (y)] - c_0.$$

Thus,

$$\begin{aligned} g\left(a_0 \middle| w_1, a_1'\right) &= \left(\mathbb{E}_{F_1'} [w_1 (y)] - c_1\right) - (\mathbb{E}_{F_0} [w_1 (y)] - c_0) \\ &= (\mathbb{E}_{F_1} [\hat{w}_1 (y)] - c_1) - (\mathbb{E}_{F_0} [\hat{w}_1 (y)] - c_0) \\ &= g (a_0 | \hat{w}_1, a_1) \geq 0, \end{aligned}$$

implying that a_1' may be chosen by agent 1 in response to w_1 under some technology.

When agent 1 chooses action a_1' in response, the principal's resulting payoff in the first period is

$$\mathbb{E}_{F_1'} [y - w_1(y)] = \lambda \mathbb{E}_{F_0} [y - w_1(y)] = \lambda (1 - s_1) \mathbb{E}_{F_0} [y] = (1 - s_1) \mathbb{E}_{F_1} [y] = \mathbb{E}_{F_1} [y - \hat{w}_1(y)],$$

so her payoff in the first period under $\left(w_1 \middle| a_1'\right)$ and under $(\hat{w}_1 | a_1)$ are exactly equal.

Moreover, the principal's optimal second-period payoff guarantee is $\hat{V}_2^* \left(w_1, a_1'\right) = \Phi_2 \left(a_1'\right)^2$ with

$$\Phi_2 \left(a_1'\right) = \max_{a \in A_0 \cup \{a_1'\}} \left\{ \sqrt{\mathbb{E}_{F_a}[y]} - \sqrt{c_a} \right\}.$$

From $\mathbb{E}_{F_1'} [y] = \mathbb{E}_{F_1}[y]$, it follows that $\Phi_2 \left(a_1'\right) = \Phi_2 (a_1) > 0$, which implies that $\hat{V}_2^* \left(a_1'\right) = \hat{V}_2^* (a_1)$. The principal's interim payoff guarantee is

$$\begin{aligned}\hat{U} \left(w_1 \middle| a_1'\right) &= \mathbb{E}_{F_1'} [y - w_1(y)] + \beta \cdot \hat{V}_2^* \left(a_1'\right) \\ &= \mathbb{E}_{F_1} [y - \hat{w}_1(y)] + \beta \cdot \hat{V}_2^* (a_1) = \hat{U} (\hat{w}_1 | a_1),\end{aligned}$$

as desired.

This completes the proof. □

Lemma A.8. *Under technological advances, within the class of linear first-period contracts, there exists an optimal one for the principal.*

Proof of Lemma A.8. Assume the principal offers a linear first-period contract $w_1(y) = s_1 y$ with $s_1 \in [0,1]$, and agent 1 chooses $a_1 = (F_1, c_1)$ in response. The principal's optimal second-period payoff guarantee $\hat{V}_2^*(a_1) = \Phi_2(a_1)^2$, with Φ_2 defined by equation (7). Thus, her interim payoff guarantee is

$$\hat{U}(w_1|a_1) = \mathbb{E}_{F_1}[y - w_1(y)] + \beta \cdot \hat{V}_2^*(a_1) = (1 - s_1)\mathbb{E}_{F_1}[y] + \beta \cdot \Phi_2(a_1)^2 .$$

The worst-case overall payoff guarantee minimizes the above expression over all a_1 that agent 1 may choose under some technology A_1. Note that agent 1 prefers action a_1 over all known actions $a \in A_0$ if and only if the incentive gap with respect to each $a \in A_0$ is nonnegative, i.e., $g(a|w_1, a_1) \geq 0$, which is equivalent to

$$(\mathbb{E}_{F_1}[w_1(y)] - c_1) - (\mathbb{E}_{F_a}[w_1(y)] - c_a) = (s_1\mathbb{E}_{F_1}[y] - c_1) - (s_1\mathbb{E}_{F_a}[y] - c_a) \geq 0, \quad \forall a \in A_0.$$

Moreover, agent 1 obtains at least his reservation payoff of zero, which can also be viewed as his payoff from the null action $(\delta_0, 0)$. Hence, the following program yields a lower bound on the principal's overall payoff guarantee

$$\begin{aligned} \inf_{F_1, c_1} \quad & (1 - s_1)\mathbb{E}_{F_1}[y] + \beta \cdot \Phi_2(F_1, c_1)^2 \\ \text{s.t.} \quad & (s_1\mathbb{E}_{F_1}[y] - c_1) - (s_1\mathbb{E}_{F_a}[y] - c_a) \geq 0, \quad \forall a \in A_0 \cup \{(\delta_0, 0)\}, \end{aligned} \tag{A.24}$$

because the principal's interim payoff guarantee can never be strictly lower than the infimum given by program (A.24).

Conversely, for any feasible $a_1 = (F_1, c_1)$ in program (A.24), agent 1 would take action a_1 in response to w_1 when his technology $A_1 = A_0 \cup \{a_1\}$. The worst case over all such technologies leaves the principal with exactly her interim payoff guarantee, $\hat{U}(w_1|a_1) = (1 - s_1)\mathbb{E}_{F_1}[y] + \beta \cdot \Phi_2(a_1)^2$. Thus, if a solution to program (A.24) exists, then the principal's payoff guarantee cannot be strictly higher than its minimum value.

Therefore, the worst-case overall payoff guarantee of any linear first-period contract $w_1(y) = s_1 y$ is exactly characterized by program (A.24).

Now we reformulate program (A.24) as an equivalent maximization problem with continuous objective function and compact feasible region. Slightly abusing notation, we use $\hat{U}(s_1)$ instead of $\hat{U}(w_1)$ to denote the infimum value of program (A.24). Note that both the objective and the constraints of program (A.24) depend on the choice variables (F_1, c_1) only through the value of $(\mathbb{E}_{F_1}[y], c_1)$. Rewrite $\mathbb{E}_{F_1}[y] = x$ and $c_1 = z$ with $x, z \geq 0$. Plugging into the original program (A.24), we obtain an equivalent program

$$\begin{aligned} \hat{U}(s_1) = \inf_{x,z} \quad & (1 - s_1)x + \beta \cdot \hat{\Phi}_2(x, z)^2 \\ \text{s.t.} \quad & s_1 x - z \geq \max_{a \in A_0 \cup \{(\delta_0, 0)\}} \{s_1\mathbb{E}_{F_a}[y] - c_a\}, \quad x, z \geq 0, \end{aligned} \tag{A.25}$$

where $\hat{\Phi}_2$ is defined by equation (A.22).

Let $\overline{x} \equiv \max_{a \in A_0} \mathbb{E}_{F_a}[y] > 0$, and $\overline{v} \equiv \max_{a \in A_0} \left\{ \sqrt{\mathbb{E}_{F_a}[y]} - \sqrt{c_a} \right\} > 0$. Suppose

$$(F_0, c_0) \in \operatorname*{arg\,max}_{a \in A_0 \cup \{(\delta_0, 0)\}} \{s_1\mathbb{E}_{F_a}[y] - c_a\}.$$

Note that $(x_0, z_0) = (\mathbb{E}_{F_0}[y], c_0)$ is feasible in program (A.25) and leads to objective value

$$(1 - s_1)\, x_0 + \beta \cdot \hat{\Phi}_2 (x_0, z_0)^2 \leq (1 - s_1)\, \overline{x} + \beta \cdot \overline{v}^2.$$

If $x \geq \overline{x}$, then

$$(1 - s_1)\, x + \beta \cdot \hat{\Phi}_2 (x, z)^2 \geq (1 - s_1)\, \overline{x} + \beta \cdot \overline{v}^2.$$

Therefore, restricting $x \in [0, \overline{x}]$ will not change the infimum of program (A.25). Moreover,

$$s_1 x - z \geq 0 \quad \Rightarrow \quad z \leq s_1 x \leq x,$$

so restricting $(x, z) \in [0, \overline{x}]^2$ will not change the infimum of program (A.25).

Consider the following program

$$\begin{aligned} \hat{\Psi}^* (s_1) \equiv \sup_{x,z} \quad & \hat{\Psi}(x, z; s_1) \equiv -\left((1 - s_1)\, x + \beta \cdot \hat{\Phi}_2 (x, z)^2\right) \\ \text{s.t.} \quad & (x, z) \in \hat{\Gamma}(s_1), \end{aligned} \tag{A.26}$$

where $\hat{\Phi}_2$ is defined by equation (A.22), and $\hat{\Gamma}$ is defined as follows:

$$\hat{\Gamma}(s_1) \equiv \left\{(x, z) \in [0, \overline{x}]^2 : s_1 x - z \geq \max_{a \in A_0 \cup \{(\delta_0, 0)\}} \{s_1 \mathbb{E}_{F_a}[y] - c_a\}\right\}.$$

By definition, $\hat{\Psi} : [0, \overline{x}]^2 \times [0, 1] \to \mathbb{R}$ is a continuous function, and $\Gamma : [0, 1] \rightrightarrows [0, \overline{x}]^2$ is a compact-valued and nonempty-valued correspondence. Moreover, the infimum of program (A.25), $\hat{U}(s_1)$, is given by $-\hat{\Psi}^*(s_1)$.

Note that for each s_1, $\hat{\Gamma}(s_1)$ defines a half plane intersecting a square, and that the half plane shifts linearly in s_1. Thus, $\hat{\Gamma}$ is both upper and lower hemicontinuous. It then follows from Berge's maximum theorem that $\hat{\Psi}^*$ is continuous, and

$$\hat{\Gamma}^* (s_1) \equiv \left\{(x, z) \in \hat{\Gamma}(s_1) : \hat{\Psi}(x, z; s_1) = \hat{\Psi}^*(s_1)\right\}$$

is upper hemicontinuous with nonempty and compact values. As a consequence, a solution to program (A.26) exists for all s_1, and the supremum can be replaced by maximum.

It follows that the infimum in program (A.25) and therefore the original program (A.24) can both be replaced by minimum, and the resulting minimum value $\hat{U}(s_1) = -\hat{\Psi}^*(s_1)$ is continuous in s_1. Hence, $\hat{U}(s_1)$ achieves a maximum over $[0, 1]$. This maximum is also the optimal guarantee over all linear contracts. □

Proof of Theorem 2. According to Lemma A.8, among all linear first-period contracts, there exists an optimal one, call it w_1^*. If w_1 is any other (nonlinear) first-period contract that outperforms w_1^*, then by Lemma A.7, there is a linear contract that in turn does at least as well as w_1. But this contradicts the fact that w_1^* is an optimal linear contract. Therefore, w_1^* is optimal among all first-period contracts. □