The Cost of Moral Hazard and Limited Liability in the Principal-Agent Problem

Felipe Balmaceda[1], Santiago R. Balseiro[2],
Jose R. Correa[1], and Nicolas E. Stier-Moses[2]

[1] Departamento de Ingeniería Industrial, Universidad de Chile, Santiago, Chile
{fbalmace,jcorrea}@dii.uchile.cl
[2] Graduate School of Business, Columbia University, New York, USA
{sbalseiro13,stier}@gsb.columbia.edu

Abstract. In the classical principal-agent problem, a principal hires an agent to perform a task. The principal cares about the task's output but has no control over it. The agent can perform the task at different effort intensities, and that choice affects the task's output. To provide an incentive to the agent to work hard and since his effort intensity cannot be observed, the principal ties the agent's compensation to the task's output. If both the principal and the agent are risk-neutral and no further constraints are imposed, it is well-known that the outcome of the game maximizes social welfare. In this paper we quantify the potential social-welfare loss due to the existence of limited liability, which takes the form of a minimum wage constraint. To do so we rely on the worst-case welfare loss—commonly referred to as the Price of Anarchy—which quantifies the (in)efficiency of a system when its players act selfishly (i.e., they play a Nash equilibrium) versus choosing a socially-optimal solution. Our main result establishes that under the monotone likelihood-ratio property and limited liability constraints, the worst-case welfare loss in the principal-agent model is exactly equal to the number of efforts available.

1 Introduction

In this paper we analyze the classical principal-agent problem as put forward by Grossman and Hart [4]. The problem entails the following contracting situation: a principal hires an agent to perform a task. The principal cares about the task's output but cannot control it directly. Instead, the output is influenced by the agent's choice of effort intensity. The principal would like to induce the agent to choose the (in his view) optimal effort intensity but since the agent incurs a cost when making effort, the principal has to compensate the agent. Because the principal cannot observe the effort intensity chosen by the agent—this is the prevailing assumption in this type of models and leads to *moral hazard*—the principal can only tie the agent's compensation to the task's output, used as a proxy of effort. This compensation scheme entails a loss since the task's output is a random variable whose distribution depends on the effort chosen by the agent. Hence, the output is not completely determined by the agent's effort intensity. If the two were perfectly correlated, the principal could infer the effort by observing the outcome.

This class of principal-agent problems has been the workhorse to understand many interesting economic phenomena such as, to name a few, the theory of insurance under

A. Saberi (Ed.): WINE 2010, LNCS 6484, pp. 63–74, 2010.

moral hazard [12], the theory of managerial firms [1, 8], optimal sharecropping contracts between landowners and tenants [13], the efficiency wage theory [11], financial contracting [6], and job design and multi-tasking [5].

When both the principal and the agent are risk-neutral, the provision of a limited liability clause that restricts the exposure of the agent gives rise to an agency problem. If the principal wants to provide an incentive to the agent to work hard, he has to compensate the agent better when the realization of the task's output suggests that the effort intensity chosen by the agent was high. This imposes a gap between the marginal cost of the effort intensity experienced by the principal and the social marginal cost. Thus, the equilibrium contract will not maximize social welfare, meaning that a first-best outcome cannot be attained; instead, the constrained contract will be second-best.

In order to quantify the maximum social-welfare loss due to the existence of moral hazard and limited liability in a principal-agent setting, we rely on the concept of worst-case welfare loss, which quantifies the efficiency of a system when its players act selfishly (i.e., they play a Nash equilibrium) versus choosing a socially-optimal solution. The idea of using worst-case analysis to study non-cooperative games was introduced by Koutsoupias and Papadimitriou [7], and it is commonly referred to as the *Price of Anarchy* [9]. In our setting, the worst-case welfare loss is defined as the largest possible ratio between the social welfare of a socially-optimal solution—the sum of the principal's and agent's payoffs when the first-best effort intensity is chosen—and that of the sub-game perfect equilibrium. The worst ratio is with respect to the parameters that define an instance of the problem.

In the principal-agent setting, Babaioff, Feldman, and Nisan [2, 3] introduced a combinatorial agency problem with multiple agents performing two-effort-two-outcome tasks. The authors studied the combinatorial structure of dependencies between agents' actions, and analyzed the worst-case welfare loss for a number of different classes of action dependencies. Our model, instead, deals with a single agent and its complexity lies in handling more sophisticated tasks, rather than the interaction between agents. The goal of this article is to evaluate the worst-case welfare loss with respect to the outcome vector, the vector of agent's costs of effort, and the probability distribution of outcomes for each level of effort. The main result, shown in Theorem 1, establishes that under the monotone likelihood-ratio property and when the principal and an agent protected by limited liability are risk-neutral, the worst-case welfare loss is exactly equal to the number of efforts available. In other words, for any instance of the problem the worst-case welfare loss cannot exceed the number of efforts available and there are instances where that loss is achieved.

Our result suggests that the worst equilibrium that may arise in the finite principal-agent problem with limited liability for the agent depends on the complexity of the delegated task, as measured by the number of available efforts. When the delegated task requires the choice between two different effort intensities (e.g., shirk or work) the worst-case welfare loss is 2, while when the delegated task demands the choice of one effort intensity among E possibilities, the worst-case welfare loss is E. Thus, the worst-case welfare loss increases with the complexity of the delegated task. Our result suggests that the principal-agent paradigm that studies the consequences of moral hazard for the efficiency of contracting and organizational design is sound. The potential

consequence of not dealing with a moral-hazard problem may have a non-negligible impact in the welfare of the system. For another interpretation, our results also quantify the impact of limited-liability in the utility of the principal, which is a way of measuring the inefficiency introduced by protecting the agent from carrying all the burden of the risk in the task's output.

Because the complexity of a principal-agent relationship is usually related to the number of tasks or projects rather than to the number of efforts or actions, we also study the worst-case welfare loss in an extension where there are multiple tasks. Here, the agent has to choose between working and shirking in each of several independent tasks. Surprisingly, we find that the worst-case welfare loss again equals 2, the number of efforts in each task, independently of how many tasks the agent has to work on. This confirms that, in terms of the potential welfare loss, the complexity of an agency relationship is better captured by the number of actions or efforts available rather than the number of tasks. Furthermore, it suggests that the incentive problem created by moral hazard is a natural source of economies of scope; that is, it is better to have one agent working in several different tasks than several agents working in one task each.

Most of our results arise from a characterization of the optimal wages that we provide. Working with the geometry of both the primal and the dual linear programs, we uncover the structure of the 'important' efforts, which we call *relevant*, and use them to bound the welfare of the solution to the principal-agent model with that arising when the agent chooses the socially-optimal effort.

The rest of the paper is organized as follows. In Sect. 2, we introduce the model with its main assumptions. Section 3 presents the main technical results. We start with the study of the two-effort-two-outcome case for an illustration of our techniques, continue with the general case, and present an example that shows that the lower bound is attained. We conclude with extensions in several directions in Sect. 4. For the missing proofs and details on the extensions, we refer the reader to the full version of the paper.

2 The Principal-Agent Model

In this section we describe the basic principal-agent model with $E \geq 2$ effort levels and $S \geq 2$ outcomes [4]. (Later on, in Sect. 4, we relax some of the assumptions presented below.) The agent chooses an effort $e \in \mathcal{E} \triangleq \{1, \ldots, E\}$, incurring a personal nonnegative cost of c_e. Efforts are sorted in increasing order with respect to costs; that is, $c_e \leq c_f$ if and only if $e \leq f$. Thus, a higher effort demands more work from the agent. The task's outcome depends on a random state of nature $s \in \mathcal{S} \triangleq \{1, \ldots, S\}$ whose distribution in turn depends on the effort level chosen by the agent. Each state has an associated nonnegative dollar amount that represents the principal's revenue. We denote the vector of outcomes indexed by state by $y = \{y^1, \ldots, y^S\}$. Without loss of generality, the outcomes are sorted in increasing order: $y^s \leq y^t$ if and only if $s \leq t$; hence, the principal's revenues are higher under states with a larger index. Finally, we let π_e^s be the common-knowledge probability of state $s \in \mathcal{S}$ when the agent chooses effort $e \in \mathcal{E}$. The probability mass function of the outcome under effort e is given by $\pi_e = \left\{\pi_e^1, \ldots, \pi_e^S\right\}$.

The principal can contract wages to the agent that depend on the outcome y but cannot observe the agent's chosen effort e. Indeed, the principal offers a take-it-or-leave-it

contract to the agent that specifies a state-dependent wage schedule $w = \{w^1, \ldots, w^S\}$. The agent decides whether to accept or reject the offer, and if accepted, then he chooses an effort level before learning the realized state. The rational agent should accept the contract if the *individual rationality* (IR) and *limited liability* (LL) constraints are satisfied. The former specifies that the contract must yield an expected utility to the agent greater than or equal to that of choosing the *outside option*. The latter specifies that the wage must be nonnegative in every state occurring with positive probability. After accepting a contract specifying a wage schedule w, the risk-neutral agent has to choose an effort $e \in \mathcal{E}$. He does so by maximizing the expected payoff, which is given by $\pi_e w - c_e$, the difference between the expected wage and the cost incurred in the effort chosen.

Putting it all together, the principal's problem consists on choosing a wage schedule w and an effort intensity e for the agent that solve the following problem:

$$u^P \triangleq \max_{e \in \mathcal{E}, w} \pi_e (y - w) \tag{1}$$

$$\text{s.t. } \pi_e w - c_e \geq 0 \qquad \text{(IR)} \tag{2}$$

$$e \in \arg\max_{f \in \mathcal{E}} \{\pi_f w - c_f\} \qquad \text{(IC)} \tag{3}$$

$$w \geq 0\,. \qquad \text{(LL)} \tag{4}$$

The objective measures the difference between the principal's expected revenue and payment, hence computing his expected profit. Constraints (IR) and (LL) were described earlier. The *incentive compatibility* (IC) constraints guarantee that the agent will choose the principal's desired effort since he does not find it profitable to deviate from e.

Equivalently, one can formulate the principal's problem as $u^P = \max_{e \in \mathcal{E}} \{\pi_e y - z_e\}$ $= \max_{e \in \mathcal{E}} \{u_e^P\}$. Here, we have defined z_e to be the minimum expected payment incurred by the principal so the agent accepts the contract and picks effort e. In addition, we denote by $u_e^P \triangleq \pi_e y - z_e$ the principal's maximum expected utility when effort e is implemented, and by $\mathcal{E}^P$ the set of optimal efforts for the principal, $\mathcal{E}^P \triangleq \arg\max_{e \in \mathcal{E}} \{u_e^P\}$. Exploiting that the set of efforts is finite, we can write the IC constraint (3) explicitly to obtain the *minimum payment linear program* corresponding to effort e, which we denote by MPLP(e):

$$z_e = \min_{w \in \mathbb{R}^S} \pi_e w \tag{5}$$

$$\text{s.t. } \pi_e w - c_e \geq 0 \tag{6}$$

$$\pi_e w - c_e \geq \pi_f w - c_f \quad \forall f \in \mathcal{E} \setminus e \tag{7}$$

$$w \geq 0\,. \tag{8}$$

Notice that this problem is independent of the output y.

We say that the principal *implements* effort $e \in \mathcal{E}$ when the wage schedule w is consistent with the agent choosing effort e. For a fixed effort e, (2), (3), and (4) characterize the polyhedron of feasible wages that implement e. The principal will choose a wage belonging to that set that achieves z_e by minimizing the expected payment $\pi_e w$. We

are only interested in efforts that are attainable under some wage schedule, which we refer to as *feasible efforts*. An effort is feasible if the polyhedron corresponding to it is nonempty.

2.1 The Monotone Likelihood-Ratio Property

We make the assumption that the probability distributions π_e satisfy the well-known *monotone likelihood-ratio property* (MLRP). That is, $\{\pi_e\}_{e\in\mathcal{E}}$ verifies $\pi_e^s/\pi_f^s \geq \pi_e^t/\pi_f^t$ for all states $s < t$ and efforts $e < f$. The assumption of MLRP is pervasive in the literature of economics of information, and in particular in the principal-agent literature. The intuition behind it is that the higher the observed level of output, the more likely it is to come from a distribution associated with a higher effort level.

An important property of MLRP is that distributions that satisfy it also satisfy *first order stochastic dominance* (FOSD). For instance, [10] proved that $\sum_{s'=1}^{s} \pi_e^{s'} \geq \sum_{s'=1}^{s} \pi_f^{s'}$ for all states s and efforts $e < f$. A simple consequence of this that plays an important role in our derivations is that probabilities for the highest outcome S are sorted in increasing order with respect to efforts; i.e., $\pi_e^S \leq \pi_f^S$ for $e \leq f$. Note that in the case of two outcomes, MLRP and FOSD are equivalent.

2.2 Worst-Case Welfare Loss

The goal of a social planner is to choose the effort level e that maximizes the social welfare, defined as $u_e^{SW} \triangleq \pi_e y - c_e$, the sum of the welfare of the principal and the agent. The social planner is not concerned about wages, since risk neutrality ensures that wages are a pure transfer of wealth between the principal and the agent. Thus, the optimal social welfare is given by $u^{SO} \triangleq \max_{e\in\mathcal{E}}\{u_e^{SW}\}$. We denote the set of first-best efficient efforts by $\mathcal{E}^{SO} \triangleq \arg\max_{e\in\mathcal{E}}\{u_e^{SW}\}$. For analytical tractability, we will assume that the harder the agent works, the higher the social welfare in the system. In the two-outcome case, this assumption can be relaxed. In the general case, we believe that our results continue to hold without it.

Assumption 1. *The sequence of prevailing social welfare under increasing efforts is non-decreasing; i.e.,* $u_e^{SW} \leq u_f^{SW}$ *for all efforts* $e \leq f$.

For a given instance of the problem, we quantify the inefficiency of an effort e using the ratio of the social welfare under the socially-optimal effort to that under e. The main goal of the paper is to compute the worst-case welfare loss for arbitrary instances of the problem. This is defined as the smallest upper bound on the efficiency of a second-best optimal effort, which is commonly referred to as the *Price of Anarchy*[1] [9]. Therefore, the worst-case welfare loss, denoted by ρ, is defined as

$$\rho = \sup_{\pi,y,c} \frac{u^{SO}}{\min_{e\in\mathcal{E}^P} u_e^{SW}}, \tag{9}$$

[1] Actually, the price of anarchy for a maximization problem such as the one we work with in this article is often defined as the inverse of the ratio in (9). We do it in this way so ratios and welfare losses point in the same direction.

where the supremum is taken over all valid instances as described at the beginning of this section. Of course, the previous ratio for an arbitrary instance of the problem is at least one because the social welfare of an optimal solution cannot be smaller than that of an equilibrium, guaranteeing that $\rho \geq 1$. Next, we state the main result of our article that shows that under MLRP the worst-case welfare loss is bounded above by the number of efforts, and that this bound is tight.

Theorem 1. *Suppose that MLRP holds. Then, in the risk-neutral principal-agent problem with limited liability, the worst-case welfare loss ρ is exactly E.*

2.3 Preliminaries

In this section, we consider the principal's problem and reformulate it in a way that is more amenable to understand its properties, which will be useful to prove our worst-case bounds. The dual of MPLP(e), displayed in (5)-(8), is given by

$$\max_{p \in \mathbb{R}^E} \sum_{f \neq e} (c_f - c_e) p_f - c_e p_e \tag{10}$$

$$\text{s.t.} \sum_{f \neq e} (\pi_f^s - \pi_e^s) p_f - \pi_e^s p_e \leq \pi_e^s \quad \forall s \in \mathcal{S}, \tag{11}$$

$$p \leq 0\,.$$

Here, p_e is the dual variable for the IR constraint (6), while p_f is the dual variable for the IC constraint (7) for effort $f \neq e$. Notice that the null vector $\mathbf{0}$ is dual-feasible, and hence the dual problem is always feasible. Furthermore, since we only consider feasible efforts the primal is also feasible and by strong duality we have that the solution to the dual program is z_e. Notice that summing constraints (11) over $s \in \mathcal{S}$ and using that $\sum_{s \in \mathcal{S}} \pi_f^s = 1$ for all $f \in \mathcal{E}$, we get that $p_e \geq -1$. We now state some useful results.

Lemma 1. *The social welfare is at least the principal's utility; i.e., $u_e^{SW} \geq u_e^P$ for all efforts $e \in \mathcal{E}$.*

Proof. Notice that since z_e solves MPLP(e), we have that $z_e \geq c_e$ for all $e \in \mathcal{E}$. Thus, $\pi_e y - z_e \leq \pi_e y - c_e$. □

The next result stresses the importance of the agent's limited liability in the model. It is a well-known result that we state for the sake of completeness. Without the LL constraint (4), it is optimal for the principal to implement the socially-optimal effort and he captures the full social surplus, leaving no utility to the agent. As a consequence, the worst-case welfare loss is 1 meaning that, albeit unfair to the agent, the contract is efficient.

Lemma 2. *If the principal and the agent are risk-neutral and there is no limited liability constraint, the minimum expected payment z_e incurred by the principal when inducing a feasible effort e is c_e, that is, $c_e = \min_{w \in \mathbb{R}^S} \{\pi_e w$ s.t. (6), (7)$\}$.*

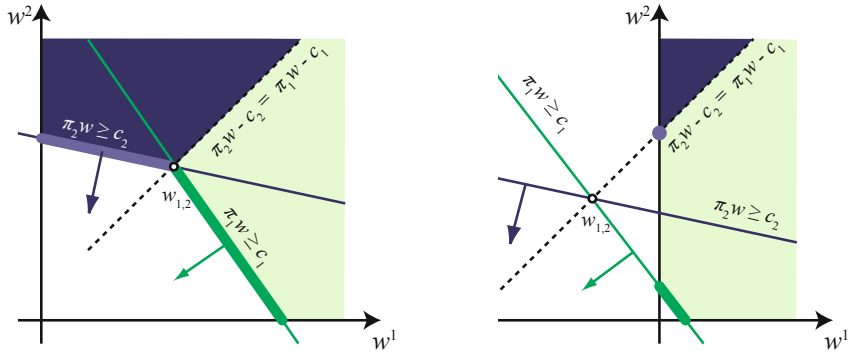
Proof. Since the effort e is feasible there exists a vector w satisfying (6) and (7). Assume for a contradiction that (6) is not tight and consider $w' = w - \mathbf{1}\varepsilon$, where $\mathbf{1}$ is the all-ones vector. Clearly w' still satisfies (7) so we can select ε so that the objective function is smaller and (6) is still feasible. □

3 Bounding the Welfare Loss

3.1 The Case of Two Efforts and Two Outcomes

In this section we look at the case with 2 efforts (such as shirk and work) and 2 states (such as fail and success), and show that the worst-case welfare loss is at most 2. This simple case is a useful exercise to gain intuition and improve the understanding of the general case. First, we provide a geometric characterization of the minimum-cost wage schedule implementing a given effort level, and compute the associated expected payments. Then, we proceed to bound the worst-case welfare loss.

Consider MPLP(2), corresponding to the agent working hard. The feasible set of wages is defined by the IR, IC and LL constraints. The IC constraint (7) ensures that the agent prefers effort 2 over 1, which can also be written as $w^2 - w^1 \geq (c_2 - c_1)/(\pi_2^2 - \pi_1^2)$. Notice that both the numerator and denominator are nonnegative. Hence, the boundary of this constraint is given by a $45°$ line, as shown by Fig. 1 which plots the feasible regions for the two efforts. The IC constraint for $e = 1$ is the same with the inequality reversed. An implication of FOSD is that the IR constraint for effort 1 is steeper than that for effort 2.



(a) $w_{1,2}$ in the first quadrant. (b) $w_{1,2}$ in the second quadrant.

Fig. 1. Feasible regions of MPLP(e) for $e \in \{1, 2\}$ (light and dark shade, respectively), according to the location of $w_{1,2}$. Optimal solutions are denoted with a bold point or segment, depending on whether they are unique or not. Arrows indicate the negative gradient of the objective function.

It will be useful to introduce the point $w_{1,2}$, defined as the intersection point between the IC constraint and the IR constraints for both efforts. This point is given by

$$w_{1,2} = \left(\frac{c_1\pi_2^2 - c_2\pi_1^2}{\pi_2^2 - \pi_1^2}, \frac{c_1\pi_2^2 - c_2\pi_1^2}{\pi_2^2 - \pi_1^2} + \frac{c_2 - c_1}{\pi_2^2 - \pi_1^2} \right).$$

The second component of this vector is nonnegative and larger than the first component because $c_2 \geq c_1$, $\pi_2^2 \geq \pi_1^2$, and $\pi_1^1 \geq \pi_2^1$.

If $w_{1,2}$ lies in the first quadrant, as in Fig. 1a, the situation is very similar to the case without liability constraints discussed earlier. Indeed, the wages $w_{1,2}$ are optimal because they satisfy all constraints and minimize the objective of MPLP. This implies that the optimal expected payment is equal to the effort's cost, and because of Assumption 1 the principal chooses $e = 2$ leaving the agent with zero surplus. The case of greater interest is when $w_{1,2}$ lies in the second quadrant, as in Fig. 1b. This occurs either when the cost of working hard is too high, or the probability of a good outcome when working hard is too low. In this case, the incentive compatible wage schedule that induces participation at the lowest cost for the principal does not satisfy the limited liability constraint. Thus, the optimal solution, attained at the intersection of the IC constraint and the vertical axis, is $w_2 = \big(0, (c_2 - c_1)/(\pi_2^2 - \pi_1^2)\big)$. The minimum expected payment for effort 2 is $z_2 = \pi_2^2(c_2 - c_1)/(\pi_2^2 - \pi_1^2)$, which is strictly larger than c_2 because the IR constraint is not binding, leaving the agent with a positive rent. The analysis for effort 1 is simpler. Under the assumption of nonnegative costs, any point that is nonnegative and for which the IR constraint is binding is optimal and attains the value c_1. Thus, the minimum expected payment equals the effort's cost, and the agent obtains zero surplus.[2]

The previous analysis will enable us to bound the worst-case welfare loss. Under Assumption 1, effort 2 is socially-optimal: $u^{SO} = u_2^{SW} \geq u_1^{SW}$. If the second-best optimal effort is 2, the worst-case welfare loss is 1. So we consider that it is second-best optimal to induce effort 1; i.e, $u_1^P \geq u_2^P$. Since the principal prefers effort 1, it must be that $z_2 > c_2$. Hence, $w_{1,2}$ must lie in the second quadrant, and $z_2 = (c_2 - c_1)\pi_2^2/(\pi_2^2 - \pi_1^2)$. Then, we have that

$$u_1^{SW} \geq u_1^P \geq u_2^P = \pi_2 y - z_2 = u_2^{SW} + c_2 - \pi_2^2 \frac{c_2 - c_1}{\pi_2^2 - \pi_1^2} = u_2^{SW} + c_1 - \pi_1^2 \frac{c_2 - c_1}{\pi_2^2 - \pi_1^2}$$
$$\geq u_2^{SW} + c_1 - \pi_1^2 \frac{(\pi_2 - \pi_1)y}{\pi_2^2 - \pi_1^2} \geq u_2^{SW} + c_1 - \pi_1 y = u_2^{SW} - u_1^{SW}, \quad (12)$$

where the inequalities follow, respectively, from Lemma 1, the principal's choice of $e = 1$, Assumption 1, and FOSD. Reshuffling terms, we have that $u_2^{SW} \leq 2u_1^{SW}$ from where the optimal social welfare cannot be better than twice the social welfare under the effort chosen by the principal. We conclude that the worst-case welfare loss is at most the number of efforts.

3.2 The General Case

We now consider the general case of an arbitrary finite number of efforts and outcomes. Here, we need to study the primal and the dual of the MPLP simultaneously. As in the previous case, we first attempt to characterize the minimum expected payments for each effort level, and then prove that the worst-case welfare loss is bounded by E.

We saw earlier that in the case of 2 efforts both of them play a role in the worst-case bound. However, in the general case only some efforts will be *relevant*. There are some other efforts, referred to as *dominated*, that although feasible will not participate

[2] This might not be the case if the limited liability constraint requires $w^2 \geq \ell$, where ℓ is large. This is discussed in the full version of the paper.

in the analysis. Relevant efforts are always preferred to dominated efforts and thus the principal will choose just from among them. This is equivalent to discarding dominated efforts from any instance and does not affect the utilities of other efforts and the efficiency metric.

In Theorem 2, we characterize the relevant efforts. We do this by observing that effort E is always relevant. From this first relevant effort, we obtain a sequence inductively observing that for any relevant effort, in the optimal solution to MPLP only the IC constraint of another relevant effort is binding. Afterwards, we prove that when a dominated effort is chosen, the principal's utility is always dominated by that of a relevant effort (hence the name 'relevant'). As before, we define the wage vector $w_{e,f}$ as the intersection of IC constraints (7) for efforts e and f with the S axis. Hence, $w_{e,f} = \big(0, \ldots, 0, (c_e - c_f)/(\pi^S_e - \pi^S_f)\big)$, which is a nonnegative vector.

Theorem 2. *There exists a subsequence of relevant efforts, denoted by $\mathcal{R} = \{e_r\}_{r=1}^{R} \subseteq \mathcal{E}$ with $e_R = E$, such that the minimum expected payments for the principal are*

$$z_{e_1} = c_{e_1}, \quad \text{and} \quad z_{e_r} = \pi^S_{e_r} \frac{c_{e_r} - c_{e_{r-1}}}{\pi^S_{e_r} - \pi^S_{e_{r-1}}} \geq c_{e_r} \quad \text{for } r = 2, \ldots, R.$$

Moreover, the optimal wage w_{e_r} corresponding to effort e_r is $w_{e_r,e_{r-1}}$ if $r > 1$ and $(0, \ldots, 0, c_{e_1}/\pi^S_{e_1})$ if $r = 1$.

For a dominated effort $f \notin \mathcal{R}$, let $r(f) \triangleq \min\{e \in \mathcal{R} : e > f\}$ be the smallest relevant effort greater than f. The next corollary shows that relevant efforts are sorted with respect to $z_e - c_e$ and that dominated efforts violate this order.

Corollary 1. *Relevant efforts are sorted in non-decreasing order with respect to $z_e - c_e$; that is, $z_{e_r} - c_{e_r} \leq z_{e_{r+1}} - c_{e_{r+1}}$ for all $1 \leq r < R$. Moreover, $z_f - c_f \geq z_{r(f)} - c_{r(f)}$ for any dominated effort $f \notin \mathcal{R}$.*

Proof. For the first claim observe that $z_{e_r} - c_{e_r} = \pi_{e_r} w_{e_r} - c_{e_r} \leq \pi_{e_r} w_{e_{r+1}} - c_{e_r} = \pi_{e_{r+1}} w_{e_{r+1}} - c_{e_{r+1}} = z_{e_{r+1}} - c_{e_{r+1}}$, where the inequality follows from the fact that $w_{e_{r+1}}$ is feasible for MPLP(e_r) and that w_{e_r} is the optimal solution. The second equality holds because the IC constraint between efforts e_r and e_{r+1} is binding at $w_{e_{r+1}}$.

For the second claim, let f be a dominated effort. If $f < e_{r_1}$, the result is trivial because $z_{e_{r_1}} - c_{e_{r_1}} = 0$. So, suppose that $e_r < f < e_{r+1}$. Using the dual of MPLP, as done previously, it is easy to observe that $p = -\mathbb{I}_{e_r} \pi^S_f / (\pi^S_f - \pi^S_{e_r})$ is dual feasible for effort f, and its objective value is $(c_f - c_{e_r})\pi^S_f/(\pi^S_f - \pi^S_{e_r}) = \pi^S_f w^S_{e_r,f}$, which by weak duality is a lower bound on z_f. Hence, $z_f \geq \pi^S_f w^S_{e_r,f} = \pi^S_{e_r} w^S_{e_r,f} + w^S_{e_{r+1},f}(\pi^S_f - \pi^S_{e_{r+1}}) + w^S_{e_{r+1}}(\pi^S_{e_{r+1}} - \pi^S_{e_r})$. Rearranging the terms, the last expression equals $z_{e_{r+1}} + c_f - c_{e_{r+1}} + \pi^S_{e_r}(w^S_{e_r,f} - w^S_{e_{r+1}}) \geq z_{e_{r+1}} + c_f - c_{e_{r+1}}$, where the inequality follows because $w_{e_r,f} \geq w_{e_{r+1}}$. Indeed,

$$w^S_{e_r,f} = \frac{c_f - c_{e_{r+1}}}{\pi^S_f - \pi^S_{e_r}} + \frac{c_{e_{r+1}} - c_{e_r}}{\pi^S_f - \pi^S_{e_r}} = w^S_{e_{r+1},f} \frac{\pi^S_f - \pi^S_{e_{r+1}}}{\pi^S_f - \pi^S_{e_r}} + w^S_{e_{r+1}} \frac{\pi^S_{e_{r+1}} - \pi^S_{e_r}}{\pi^S_f - \pi^S_{e_r}} \geq w^S_{e_{r+1}},$$

because $w_{e_{r+1},f} \leq w_{e_{r+1}}$ (this follows from Theorem 2) and $\pi^S_f - \pi^S_{e_{r+1}} \leq 0$. □

Relevance is central to the analysis of the principal-agent problem. Under Assumption 1, a social planner chooses effort E, a relevant effort, to maximize the social welfare. Furthermore, as a consequence of Corollary 1, there is always a relevant effort that is optimal for the principal.

Proposition 1. *There is always a relevant effort that is optimal for the principal; i.e.,* $\mathcal{E}^P \cap \mathcal{R} \neq \emptyset$.

Proof. We prove this claim by contradiction by supposing that no relevant effort is optimal for the principal. Let f be an optimal dominated effort, and consider the first next relevant effort $r(f)$. Using Corollary 1,

$$0 < u_f^P - u_{r(f)}^P = (\pi_f - \pi_{r(f)})y + z_{r(f)} - z_f \leq (\pi_f - \pi_{r(f)})y + c_{r(f)} - c_f = u_f^{SW} - u_{r(f)}^{SW},$$

which is a contradiction because Assumption 1 implies that f cannot have a larger social welfare than $r(f)$. □

Notice that the previous proposition together with Theorem 2 imply that the equilibrium of the principal-agent problem can be computed in $O(E^2+ES)$ time, instead of solving E linear programs. The quadratic term comes from finding the relevant efforts while the second term comes from evaluating the principal's utilities for all relevant efforts.

We are now in position to prove the main result.

Theorem 3. *Assume that MLRP and Assumption 1 hold. The worst-case welfare loss for the risk-neutral principal-agent problem with limited liability is at most* E.

Proof. Under Assumption 1, it is optimal for the system that the agent chooses effort E, so $u^{SO} = u_E^{SW}$. Furthermore, by Proposition 1 the optimal strategy for the principal is to implement a relevant effort $e \in \mathcal{R}$. Note that if we remove all efforts lower than e, a consequence of Theorem 2 is that u_f^P does not change for any effort $f > e$ and u_e^P may only increase. This is because after removing the lower efforts, z_e is reduced to c_e if they were not already equal. Notice also that a dominated effort cannot become relevant after removing the efforts lower than e. Therefore, this new instance has the same the worst-case welfare loss. Thus, we do not lose any generality if we consider that it is optimal for the principal to implement effort 1; i.e., $u_1^P \geq u_e^P$ for all $e \in \mathcal{E}$.

To lower bound the total welfare of the lowest effort, u_1^{SW}, we proceed as in (12), working exclusively with relevant efforts. To simplify notation, in the remainder of this proof we drop the r subscript and assume that all efforts are relevant. Lemma 1 and Theorem 2 imply that for any effort $e > 1$,

$$u_1^{SW} \geq u_1^P \geq u_e^P = \pi_e y - z_e = u_e^{SW} + c_e - \pi_e^S \frac{c_e - c_{e-1}}{\pi_e^S - \pi_{e-1}^S} = u_e^{SW} + c_{e-1} - \pi_{e-1}^S \frac{c_e - c_{e-1}}{\pi_e^S - \pi_{e-1}^S}.$$

Since $u_e^{SW} \geq u_{e-1}^{SW}$ implies that $c_e - c_{e-1} \leq \pi_e y - \pi_{e-1} y$, the last expression is bounded by

$$u_e^{SW} + c_{e-1} - \frac{\pi_{e-1}^S}{\pi_e^S - \pi_{e-1}^S}(\pi_e - \pi_{e-1})y \geq u_e^{SW} + c_{e-1} - \pi_{e-1} y = u_e^{SW} - u_{e-1}^{SW}, \quad (13)$$

where the inequality in (13) follows from MLRP because $\pi_{e-1}\pi_e^S \geq \pi_e \pi_{e-1}^S$. Summing over $e > 1$ and rearranging terms we conclude that $E u_1^{SW} \geq u_E^{SW}$. □

This result shows that when the agent is covered against unfair situations in which he has to pay money to the principal even after having invested the effort, the fact that the principal induces the agent to implement the effort of his choice instead of a socially-optimal one is costly for the system. Indeed, the welfare loss due to limited liability and the impossibility of observing the effort exerted by the agent is bounded by the number of efforts. If we are willing to accept *the number of efforts* as a metric of the complexity of a principal-agent relationship, then the cost of coordination in the system is bigger for more complex relationships.

3.3 A Tight Instance

To wrap-up this section we construct a family of instances with 2 outcomes and E efforts whose worst-case welfare loss is arbitrarily close to the bound of E.

Fixing $0 < \varepsilon < 1$, we let the probabilities of the outcomes associated to each effort be $\pi_e = \left(1 - \varepsilon^{E-e}, \varepsilon^{E-e}\right)$ for $e \in \mathcal{E}$. Clearly, these distributions verify that $\pi_1^2 \leq \ldots \leq \pi_E^2$, and thus they satisfy MLRP. (Recall that in the case of two outcomes MLRP and FOSD are equivalent.)

Furthermore, we let $c_E = \varepsilon^{-E}$, and then set the remaining efforts so that $z_e - c_e = e - 1$ for all $e \in \mathcal{E}$. Since $z_e = (c_e - c_{e-1})\pi_e^S/(\pi_e^S - \pi_{e-1}^S)$, we obtain $c_{e-1} = c_e\varepsilon - (e-1)(1-\varepsilon)$ for $e = 2, \ldots, E$. Notice that this implies that $w_{e+1}^2 - w_e^2 = 1/\varepsilon^{E-e}$, where $w_e = \left(0, (c_e - c_{e-1})/(\pi_e^2 - \pi_{e-1}^2)\right)$ is the optimal solution to MPLP(e). Finally, let the output be $y = (0, w_E^2 + 1)$. One can prove inductively that the social utility is $u_e^{SW} = e + \sum_{i=1}^{E-e} \varepsilon^i$, and that principal's utility is $u_e^P = \sum_{i=0}^{E-e} \varepsilon^i$, for $e \in \mathcal{E}$. Hence, the instance fulfills Assumption 1 because $u_1^{SW} \leq \ldots \leq u_E^{SW}$ and the principal's utilities satisfy $u_1^P \geq \ldots \geq u_E^P$, so it is optimal for the principal to implement effort 1.

The welfare loss corresponding to this instance is given by $u_E^{SW}/u_1^{SW} = E/(1 + \sum_{i=1}^{E-1} \varepsilon^i)$, which converges to E as $\varepsilon \to 0^+$. Therefore, Theorem 3 is tight because we found a series of instances converging to a matching lower bound.

4 Generalizations of the Basic Model

The results we have provided hold true for generalizations of the basic problem introduced in Sect. 2. First, the main result is valid when the agent can incur arbitrary (potentially negative) costs for any effort, and when the utility for the outside option is arbitrary (so far it was assumed to be zero). Second, more general limited liability constraints and imposing a minimum output do not have an impact in the worst-case bounds presented earlier. In this context, we can provide more accurate bounds that depend on some other characteristics of the instance. Third, MRLP is not needed for the case of two efforts. All results remain valid without it. Fourth, considering the problem from the perspective of the principal, we can show how to adapt the worst-case bounds provided earlier and express them with respect to the principal's payoff. Fifth, in the case with two outcomes we relax Assumption 1 by showing that the sequence of social welfare utilities is unimodal, and that any effort violating that order is infeasible. Finally, when the principal hires an agent to perform multiple identical and independent tasks that follow the two-effort-two-outcome model, we can show the the worst-case welfare loss is independent of the number of tasks and equal to 2.

Acknowledgements. The research of the first and third authors was supported in part by FONDECYT through grants 1100267 and 1090050, respectively. Part of this work was done while the fourth author was visiting Universidad de Chile supported by the Millennium Institute on Complex Engineering Systems.

References

[1] Alchian, A.A., Demsetz, H.: Production, information costs, and economic organization. American Economic Review 62(5), 777–795 (1972)
[2] Babaioff, M., Feldman, M., Nisan, N.: Combinatorial agency. In: Proceedings of the 7th ACM Conference on Electronic Commerce (EC 2006), Ann Arbor, MI, pp. 18–28. ACM Press, New York (2006)
[3] Babaioff, M., Feldman, M., Nisan, N.: Free-riding and free-labor in combinatorial agency. In: Mavronicolas, M., Papadopoulou, V.G. (eds.) SAGT 2009. LNCS, vol. 5814, pp. 109–121. Springer, Heidelberg (2009)
[4] Grossman, S.J., Hart, O.D.: An analysis of the principal-agent problem. Econometrica 51(1), 7–45 (1983)
[5] Holmström, B., Milgrom, P.: Multi-task principal-agent analyses: Incentive contracts, asset ownership and job design. Journal of Law, Economics and Organization 7, 24–52 (1991)
[6] Holmström, B., Tirole, J.: Financial intermediation, loanable funds and growth. The Quarterly Journal of Economics 112(3), 663–691 (1997)
[7] Koutsoupias, E., Papadimitriou, C.H.: Worst-case equilibria. In: Meinel, C., Tison, S. (eds.) STACS 1999. LNCS, vol. 1563, pp. 404–413. Springer, Heidelberg (1999)
[8] Jensen, M.C., Meckling, W.H.: Rights and production functions: An application to labor-managed firms and codetermination. The Journal of Business 52(4), 469–506 (1979)
[9] Nisan, N., Roughgarden, T., Tardos, É., Vazirani, V.V.: Algorithmic Game Theory. Cambridge University Press, Cambridge (2007)
[10] Rothschild, M., Stiglitz, J.E.: Increasing risk: I. a definition. Journal of Economic Theory 2(3), 225–243 (1970)
[11] Shapiro, C., Stiglitz, J.E.: Equilibrium unemployment as a worker discipline device. American Economic Review 74(3), 433–444 (1984)
[12] Spence, M., Zeckhauser, R.: Insurance, information, and individual action. American Economic Review 61(2), 380–387 (1971)
[13] Stiglitz, J.E.: Incentives and risk sharing in sharecropping. Review of Economic Studies 41(2), 219–255 (1974)