

Decision, Risk and Operations
Working Papers Series

**The cost of moral hazard and limited liability
in the principal-agent problem**

F. Balmaceda, S. R. Balseiro, J. R. Correa, N. E. Stier-Moses

July 2010; revised October 2010
DRO-2010-06

THE COST OF MORAL HAZARD AND LIMITED LIABILITY IN THE PRINCIPAL-AGENT PROBLEM

F. BALMACEDA*, S.R. BALSEIRO[†], J.R. CORREA*, AND N.E. STIER-MOSES[†]

* *Departamento de Ingeniería Industrial, Universidad de Chile, Santiago, Chile*
{fbalmace,jcorrea}@dii.uchile.cl

[†] *Graduate School of Business, Columbia University, New York, USA*
{sbalseiro13,stier}@gsb.columbia.edu

ABSTRACT. In this paper we quantify the potential social-welfare loss due to the existence of limited liability in the principal-agent problem. The worst-case welfare loss is defined as the largest possible ratio between the social welfare when the agent chooses the effort that is optimal for the system and that of the sub-game perfect equilibrium of the game. Our main result establishes that under the monotone likelihood-ratio property and a limited liability constraint, the worst-case welfare loss (also known as the *Price of Anarchy*) is exactly equal to the number of efforts available.

1. INTRODUCTION

We analyze the classical principal-agent problem as put forward by Grossman and Hart (1983) for the case in which the agent is risk neutral and subject to limited liability. In this situation, a principal hires an agent to perform an action that makes him exert costly effort. The action influences the distribution of the output and cannot be observed by the principal. To provide an incentive so the agent performs an action with high effort,¹ the principal has to pay the agent more when the realization of the output suggests that the effort chosen by the agent was higher. This, together with limited liability, imposes a gap between the marginal cost of effort 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.²

The environment we analyze is of practical importance in number of settings and has been the workhorse to understand many interesting economic phenomena where incentives play a crucial role such as the theory of insurance under moral hazard (Spence and Zeckhauser, 1971), the theory of managerial firms (Alchian and Demsetz, 1972; Jensen and Meckling, 1979), optimal sharecropping contracts between landowners and tenants (Stiglitz, 1974), the efficiency wages theory (Shapiro and Stiglitz, 1984), financial contracting (Holmström and Tirole, 1997), and job design and multi-tasking (Holmström and Milgrom, 1991).

Instead of extending the principal-agent model with limited liability and providing a new application for it, this paper attempts to quantify the welfare loss implied by it. The reason for this is threefold. First, the nature of the informational problem in moral hazard models make them difficult to be estimated empirically and thus, empirically, the welfare loss are hard to quantify. Second, the main consequences of moral hazard are by now well understood and deeply rooted in the economics of information literature, thus the moral-hazard paradigm is ripe for a deeper

Date: July 2010. This version: Oct 2010.

¹We will use the terms actions and efforts interchangeably.

²When the participation constraint, rather than the limited-liability constraint, binds, providing incentives is costless since the agent cares only about the expected compensation and the participation constraint binds on his expected payoff.

analysis of the quantitative, rather qualitative, consequences of it. Third, if we, as economists, believe that the qualitative predictions of moral hazard are of practical importance, then we should be able to provide a measure of the potential welfare loss that may occur when moral hazard is not carefully dealt with in real life situations.

In order to quantify the maximum social-welfare loss due to the existence of moral hazard and limited liability, we rely on the concept of worst-case welfare loss, which quantifies the efficiency of a system when its players play a non-cooperative game versus choosing a socially-optimal solution. The idea of using worst-case analysis to study non-cooperative games was introduced by Koutsoupias and Papadimitriou (1999), and it is commonly referred to as the *Price of Anarchy* in the computer science literature (Nisan et al., 2007). The use of the Price of Anarchy as a metric of the welfare loss is widely applied in economics to problems such as in the study of competition and efficiency in congested markets (Acemoglu and Ozdaglar, 2007), games with serial, average and incremental cost sharing (Moulin, 2008), price and capacity competition (Acemoglu et al., 2009), and VCG mechanisms (Moulin, 2009). 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 is chosen—and that of the sub-game perfect equilibrium in which the principal offers the agent a contract and then the agent chooses effort. The worst ratio is with respect to the parameters that define an instance of the problem. 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 2, establishes that under the monotone likelihood-ratio property, 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.

We also study the worst-case welfare loss in an extension where there are multiple independent tasks, but the effort choice is restricted to two efforts only. 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.

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 respect to that arising when the agent chooses the socially-optimal effort.

In terms of the economic interpretation of our results, we believe that we can relate the upper bound on the welfare loss with the agent’s discretion, understood as the number of available actions to choose from. When there are just two possible actions available per task, we can think of the job as having little discretion relative to another job where a manager can freely choose one out of E possible actions. When the job requires the choice between two different actions the worst-case welfare loss is 2, while when the jobs demands the choice of one level of effort among E possibilities, the worst-case welfare loss is E . That is, the worst-case welfare loss increases with the agent’s discretion, and therefore moral hazard should be dealt with more carefully in jobs where the agent’s discretion is higher.

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. 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 (see Balmaceda 2010 for more details on this).

The papers closest to this one are Babaioff et al. (2006, 2009). They introduce 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. They show that this loss may be unbounded for technologies that exhibit complementarities between agents, while it can be bounded by a small constant for technologies that exhibit substitutabilities between agents. 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 rest of the paper is organized as follows. In Section 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 conclude with an example that shows that the worst-case bound is attained. In Section 4, we extend our results in several directions, while in Section 5, we study the two-effort-two-outcome model with multiple tasks. Section 6 concludes with some remarks and future directions of study. The proofs that are not central to the discussion can be found in the appendix.

2. THE PRINCIPAL-AGENT MODEL

2.1. The Basic Setup. We consider the basic principal-agent model with $E \geq 2$ effort levels and $S \geq 2$ outcomes (Grossman and Hart, 1983).³ The agent chooses an effort $e \in \mathcal{E} \triangleq \{1, \dots, E\}$, 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 $s \in \mathcal{S} \triangleq \{1, \dots, S\}$ 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 $y = \{y^1, \dots, y^S\}$. 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 $\pi_e = \{\pi_e^1, \dots, \pi_e^S\}$.

Because the agent's chosen effort e cannot be observed the principal, he can write a wage contract that depends only on the outcome y . The principal makes a take-it-or-leave-it offer to the agent that specifies a state-dependent wage schedule $w = \{w^1, \dots, w^S\}$. The contract is subject to a *limited liability* (LL) constraint specifying that the wage must be nonnegative in every possible state. 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 it satisfies the *individual rationality* (IR) constraint specifying that the contract must yield an expected utility to the agent greater than or equal to that of choosing the *outside option*. 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 his expected payoff $\pi_e w - c_e$; that is, the difference between the expected wage and the cost of the effort chosen.

³Later on, in Section 4, we relax some of the assumptions presented below.

The principal's problem consists on choosing a wage schedule w and an effort intensity e for the agent that solve the following problem:

$$\begin{aligned}
(1) \quad & u^P \triangleq \max_{e \in \mathcal{E}, w} \pi_e(y - w) \\
(2) \quad & \text{s.t. } \pi_e w - c_e \geq 0 \quad (\text{IR}) \\
(3) \quad & e \in \arg \max_{f \in \mathcal{E}} \{\pi_f w - c_f\} \quad (\text{IC}) \\
(4) \quad & w \geq 0. \quad (\text{LL})
\end{aligned}$$

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

$$u^P = \max_{e \in \mathcal{E}} \{\pi_e y - z_e\},$$

where z_e is the minimum expected payment incurred by the principal so that the agent accepts the contract and picks effort e . 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. Hence, $u^P = \max_{e \in \mathcal{E}} \{u_e^P\}$ and $\mathcal{E}^P = \arg \max_{e \in \mathcal{E}} \{u_e^P\}$.

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):

$$\begin{aligned}
(5) \quad & z_e = \min_{w \in \mathbb{R}^S} \pi_e w \\
(6) \quad & \text{s.t. } \pi_e w - c_e \geq 0 \\
(7) \quad & \pi_e w - c_e \geq \pi_f w - c_f \quad \forall f \in \mathcal{E} \setminus e \\
(8) \quad & w \geq 0.
\end{aligned}$$

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 schedule 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.

The Monotone Likelihood-Ratio Property. We assume that the probability distributions π_e satisfy the well-known *monotone likelihood-ratio property* (MLRP). That is, $\{\pi_e\}_{e \in \mathcal{E}}$ verifies

$$\frac{\pi_e^s}{\pi_f^s} \geq \frac{\pi_e^t}{\pi_f^t} \quad \text{for all states } s < t \text{ and efforts } e < f.$$

The assumption of MLRP is pervasive in the literature of economics of information, and in particular in the principal-agent literature. It ensures 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, Rothschild and Stiglitz (1970) proved that

$$\sum_{s'=1}^s \pi_e^{s'} \geq \sum_{s'=1}^s \pi_f^{s'} \quad \text{for all states } s \text{ and efforts } e < f.$$

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.

Worst-Case Welfare Loss. The goal of a social planner is to choose the effort level e that maximizes the social welfare $u_e^{\text{SW}} \triangleq \pi_e y - c_e$, defined as the sum of the utility of the principal $\pi_e y - z_e$ and that of the agent $z_e - c_e$. 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

$$u^{\text{SO}} \triangleq \max_{e \in \mathcal{E}} \{u_e^{\text{SW}}\}.$$

We denote the set of first-best efficient efforts by $\mathcal{E}^{\text{SO}} \triangleq \arg \max_{e \in \mathcal{E}} \{u_e^{\text{SW}}\}$. 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 removed. In the general case, we conjecture that our results continue to hold without it.

Assumption 1. The sequence of prevailing social welfare under increasing efforts is non-decreasing; i.e., $u_e^{\text{SW}} \leq u_f^{\text{SW}}$ 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* in the computer science literature⁴ (Nisan et al., 2007). Therefore, the worst-case welfare loss is defined as

$$(9) \quad \rho \triangleq \sup_{\pi, y, c} \frac{u^{\text{SO}}}{\min_{e \in \mathcal{E}^P} u_e^{\text{SW}}},$$

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 $\rho \geq 1$. The main result of our article states that, under MLRP and Assumption 1, the worst-case welfare loss is exactly E .

2.2. Preliminaries. Observe that the principal's problem can be reformulated 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

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

$$(11) \quad \begin{aligned} \text{s.t.} \quad & \sum_{f \neq e} (\pi_f^s - \pi_e^s) p_f - \pi_e^s p_e \leq \pi_e^s \quad \forall s \in \mathcal{S}, \\ & p \leq 0. \end{aligned}$$

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 the fact 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 simple results.

⁴Actually, 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.

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

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

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 . In other words,*

$$c_e = \min_{w \in \mathbb{R}^S} \{ \pi_e w \text{ s.t. (6) and (7)} \}.$$

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. \square

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

$$w^2 - w^1 \geq \frac{c_2 - c_1}{\pi_2^2 - \pi_1^2}.$$

Notice that both the numerator and denominator are nonnegative. Hence, the boundary of this constraint is given by a 45° line, as shown by Figure 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.

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 Figure 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 Figure 1b. This occurs either when the cost of working hard is too high, or the probability of a

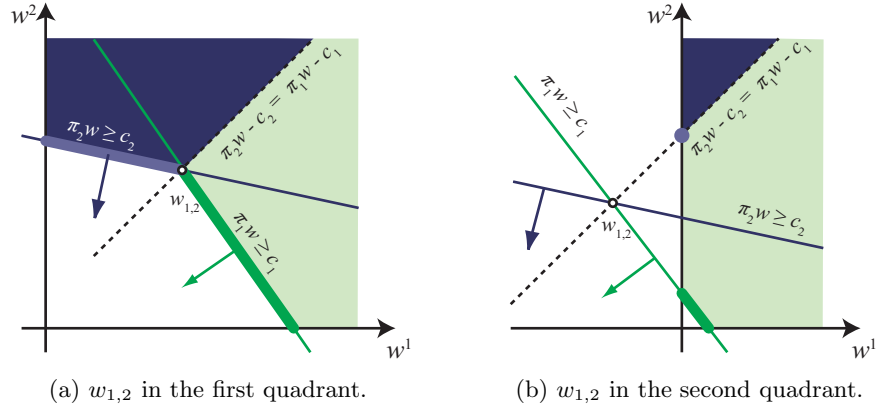


Figure 1: Feasible regions of MPLP(e) for $e \in \{1, 2\}$ (light and dark shade, respectively), according to the location of $w_{1,2}$. Optimal solutions are denoted with a bold point or segment, depending on whether they are unique or not. Arrows indicate the negative gradient of the objective function.

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

$$w_2 = \left(0, \frac{c_2 - c_1}{\pi_2^2 - \pi_1^2} \right).$$

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.⁵

The previous analysis will enable us to bound the worst-case welfare loss. Under Assumption 1, effort 2 is socially-optimal: $u_2^{\text{SO}} = u_2^{\text{SW}} \geq u_1^{\text{SW}}$. 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

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

where the inequalities follow, respectively, from Lemma 1, the principal's choice of $e = 1$, Assumption 1, and FOSD. Reshuffling terms, we have that $u_2^{\text{SW}} \leq 2u_1^{\text{SW}}$ 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.

⁵This might not be the case if the limited liability constraint requires $w^2 \geq \ell$, where ℓ is large. This will be discussed later.

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 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.

Theorem 1 characterizes 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 the principal's utility when a dominated effort is chosen 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, $w_{e,f} = (0, \dots, 0, (c_e - c_f)/(\pi_e^S - \pi_f^S))$, which is a nonnegative vector. As a consequence of the risk neutrality of the agent, the optimal wages in Theorem 1 reward the agent only when the realized outcome is the highest.

Theorem 1. *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*

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

Moreover, the optimal wage w_{e_r} corresponding to effort e_r is $w_{e_r, e_{r-1}}$ if $r > 1$ and $(0, \dots, 0, c_{e_1}/\pi_{e_1}^S)$ 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 the agent's utility $z_e - c_e$, and that dominated efforts violate this order.

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

Relevance is central to the analysis of the principal-agent problem. Because the utility of the agent increases with his effort intensity, the principal has no incentive to implement any effort higher than e^{SO} , defined as the smallest of the socially-optimal efforts. Indeed, for any effort $f > e^{\text{SO}}$, Corollary 1 implies that

$$u_{e^{\text{SO}}}^P - u_f^P = u^{\text{SO}} - u_f^{\text{SW}} + (z_f - c_f) - (z_{e^{\text{SO}}} - c_{e^{\text{SO}}}) \geq 0.$$

Hence, efforts higher than e^{SO} provide a suboptimal utility to the principal and can be disregarded. Furthermore, combining Assumption 1 and Corollary 1, we obtain that 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,

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

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

Notice that the previous proposition together with Theorem 1 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 2. *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 $u^{\text{SO}} = u_E^{\text{SW}}$. 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 1 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 1 imply that for any effort $e > 1$,

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

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

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

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 $E u_1^{\text{SW}} \geq u_E^{\text{SW}}$. \square

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. As a consequence, the social welfare of subgame perfect equilibrium is guaranteed to be at least $1/E$ of the socially-optimal welfare. 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 $\pi_e = (1 - \varepsilon^{E-e}, \varepsilon^{E-e})$ for $e \in \mathcal{E}$. The probability distributions are such that effort E guarantees a successful outcome with probability one, while the lower efforts intensities generate a failed outcome with high probability. Clearly, these distributions verify that $\pi_1^2 \leq \dots \leq \pi_E^2$, 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 the agent's utility is $z_e - c_e = e - 1$ for all $e \in \mathcal{E}$. Since we need all efforts to be relevant, we have that $z_e =$

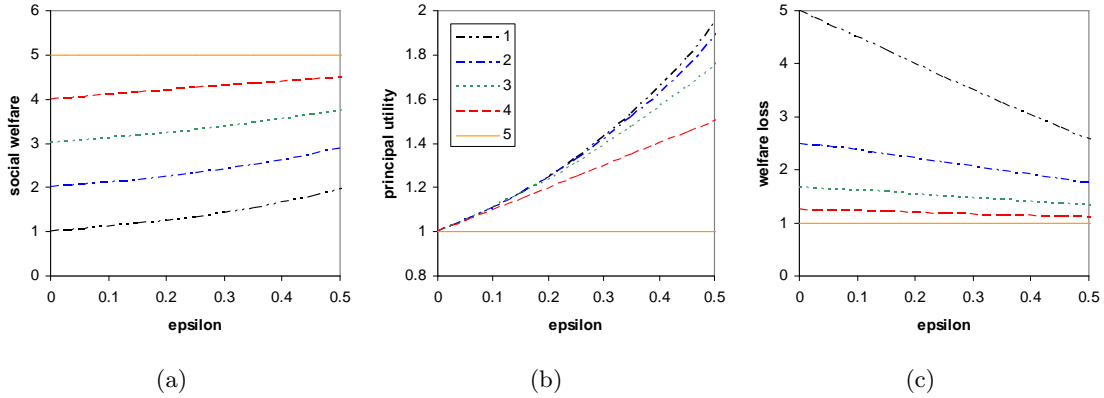


Figure 2: Example with 2 outcomes and 5 efforts. The 5 lines in each plot represent the 5 efforts. Plot (a) shows that the social welfare for effort e converges to e as $\varepsilon \rightarrow 0^+$, and that they are sorted in increasing order. Plot (b) shows that the principal's utilities converge to 1 for all efforts, and that they are sorted in decreasing order. Plot (c) plots the welfare loss $u^{\text{SO}}/u_e^{\text{SW}}$, which converges to $5/e$. Since the principal chooses $e = 1$, the worst-case welfare loss is 5.

$(c_e - c_{e-1})\pi_e^S / (\pi_e^S - \pi_{e-1}^S)$. We obtain $c_{e-1} = c_e \varepsilon - (e-1)(1-\varepsilon)$ for $e = 2, \dots, E$. Notice that this implies that $w_{e+1}^2 - w_e^2 = 1/\varepsilon^{E-e}$, where $w_e = (0, (c_e - c_{e-1})/(\pi_e^2 - \pi_{e-1}^2))$ 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 $u_e^{\text{SW}} = e + \sum_{i=1}^{E-e} \varepsilon^i$, 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 $u_1^{\text{SW}} \leq \dots \leq u_E^{\text{SW}}$ and the principal's utilities satisfy $u_1^P \geq \dots \geq u_E^P$, so it is optimal for the principal to implement effort 1.

The welfare loss corresponding to this instance is given by $u_E^{\text{SW}}/u_1^{\text{SW}} = E/(1 + \sum_{i=1}^{E-1} \varepsilon^i)$, which converges to E as $\varepsilon \rightarrow 0^+$. Therefore, Theorem 2 is tight because we found a series of instances converging to a matching upper bound. To summarize, we plot both utilities and their ratios in Figure 2.

4. ROBUSTNESS

In this section we look at the robustness of our results. We show that our results generalize to: first, arbitrary (potentially negative) costs for any effort, and an outside option with nonzero utility; second, more general limited liability constraints and the effect of guaranteeing a minimum output (in this context, we provide more accurate bounds that depend on some other characteristics of the instance); third, the case of two efforts and an arbitrary number of outputs without the MLRP assumption; fourth, the case of two outcomes when we dispense of Assumption 1 (in this case, the sequence of social welfare utilities is unimodal, and any effort violating that order is infeasible); and fifth, the case in which we adapt the worst-case bounds provided earlier to the utility of the principal rather than social welfare.

4.1. Arbitrary Costs and Reservation Utility. In some situations the agent may *extract* utility when performing a low effort. For example, exerting a low effort may give the agent more time to perform other activities to his personal benefit. Hence, in this section we consider that efforts may have negative personal costs for the agent. Note also that this extension will be of use when the utility of the outside option is nonzero. In this more general situation, we still get a worst-case welfare loss exactly equal to the number of efforts.

In the characterization of relevant efforts and their optimal wage schedules given by Theorem 1, the expected payment for the lowest relevant effort becomes $z_{e_1} = \max\{c_{e_1}, 0\}$. The rest of the proof is identical except for the inductive step. When we construct the solution for a new relevant effort in MPLP and decrease the wage corresponding to the state S , we may now reach either the IR constraint (6), an IC constraint (7), or the LL constraint (8). The latter prevents the expected payment from being negative. Since all the remaining results hold, our worst-case bound also does.

According to the IR constraint (6), the agent accepts the contract if the expected utility is greater than or equal to the utility of the outside option, which we denote by \underline{u} . So far we considered that $\underline{u} = 0$. To allow it to take arbitrary values, we set $c'_e = c_e + \underline{u}$ for all $e \in \mathcal{E}$, and obtain an instance with zero reservation utility and costs shifted by \underline{u} . Since we argued that our worst-case bound holds for negative costs, it also holds for arbitrary utilities for the outside option. Finally, the bound is tight even if we set the value of \underline{u} arbitrarily because in the instance of Section 3.3 all costs diverge to infinity.

4.2. Arbitrary Minimum Wages and Minimum Outputs. Employment law, union contracts or an agent with bigger negotiating power may demand that the wage be at least a minimum wage $\ell \in \mathbb{R}$, regardless of the state of the system. In this case, the principal solves

$$u^P(\ell) = \max_{e \in \mathcal{E}, w} \{ \pi_e(y - w) \text{ s.t. (2), (3), and } w \geq \ell \mathbf{1} \},$$

where the LL constraint now has a lower bound of ℓ instead of 0. If we denote the principal's maximum expected utility when effort e is implemented under a minimum wage of ℓ by $u_e^P(\ell)$, the set of optimal efforts for the principal is $\mathcal{E}^P(\ell) \triangleq \arg \max_{e \in \mathcal{E}} \{u_e^P(\ell)\}$.

A consequence of setting a minimum wage is that the payment to the agent might be greater than the output of the firm for some realizations of the state. Under these wage restrictions, the principal might not find it advantageous to engage in the activity because labor could be too expensive. To avoid this we introduce an IR constraint for the principal that consists in assuming that $y \geq v\mathbf{1}$, where $v \in \mathbb{R}$ denotes the minimum output. In other words, we only consider instances in which the firm's output cannot be too small so there is a guarantee for sufficient revenues. In the following proposition, we prove that the worst-case welfare loss is still E , as long as the minimum wage is less than or equal to the minimum output.

Proposition 2. *If the minimum wage ℓ is smaller than or equal to the minimum output of the principal v , the worst-case welfare loss is exactly E . Otherwise, it is unbounded.*

Proof. To prove the first part, assume that we are given an instance $I = (\pi, y, c)$ that satisfies the restrictions given by ℓ and v . Applying the change of variables $w' = w - \ell\mathbf{1}$, one obtains a new instance $I' = (\pi, y - \ell\mathbf{1}, c - \ell\mathbf{1})$ in which both outputs and costs are reduced by ℓ . Since, negative costs do not modify the worst-case welfare loss, all the previous results hold because the outputs of the modified instance are nonnegative.

We prove the second part for $v = 0$ and an arbitrary ℓ . Using the previous change of variables, we can adapt this example to any $v < \ell$. The following is an instance with 2 efforts and 2 states and arbitrarily-bad welfare loss. Let $\pi_1 = (1/2, 1/2)$, $c_1 = \ell$, $\pi_2 = (1/3, 2/3)$, $c_2 = 7\ell/6$, and $y = (\ell/6, 7\ell/4)$. Clearly, in the bad state we have that $y_1 < \ell$. Solving the problem, we have $u_1^{\text{SW}} = u_1^P = -\ell/24$, $u_2^{\text{SW}} = \ell/18$, and $u_2^P = -4\ell/9$. Since the principal would get a negative utility under both efforts, he does not find it advantageous to sign a contract with the agent and uses his outside option. Hence, the social welfare is 0 while the socially-optimal outcome is that the agent does the task exerting effort 2, making the welfare loss equal to infinity. \square

The following proposition analyzes the impact of parameters ℓ and v in the worst-case welfare loss by making the functional dependence explicit. In addition, we parametrize the welfare loss with the difference between the highest and smallest social welfare. When that difference is large

the bound is close to E as before. Interestingly, smaller differences guarantee smaller losses. The latter is relevant because the variability of the output in practical applications is not unbounded and is reasonably easy to estimate. Hence, the following formula allows us to establish a more practical measure of the loss in the principal-agent problem.

Theorem 3. *If we parametrize the worst-case welfare loss ρ with ℓ , v and the maximum difference of social utility $M \triangleq \max |u_E^{SW} - u_1^{SW}|$, then*

$$\rho(\ell, v, M) = \frac{ME + (E - 1)(v - \ell)}{M + (E - 1)(v - \ell)}.$$

Proof. We restrict ourselves to instances satisfying $|u_E^{SW} - u_1^{SW}| \leq M$. Applying the change of variables used in the proof of Proposition 2, we restrict our consideration to instances that satisfy $y \geq v'\mathbf{1}$, where $v' \triangleq v - \ell$. In order to obtain the upper bound for the worst-case welfare loss, we repeat the steps of the proof of Theorem 2, except for (13) where from $y \geq v'\mathbf{1}$, the fact that $\pi_e \mathbf{1} = 1$, and the bounded difference of social utilities we have

$$\begin{aligned} u_1^{SW} &\geq u_e^{SW} + c_{e-1} - \pi_{e-1}^S \frac{\pi_e - \pi_{e-1}}{\pi_e^S - \pi_{e-1}^S} y = u_e^{SW} - u_{e-1}^{SW} + \frac{\pi_{e-1} \pi_e^S - \pi_e \pi_{e-1}^S}{\pi_e^S - \pi_{e-1}^S} y \\ &\geq u_e^{SW} - u_{e-1}^{SW} + v' \geq u_e^{SW} - u_{e-1}^{SW} + \frac{v'}{M} (u_E^{SW} - u_1^{SW}). \end{aligned}$$

Summing over $e > 1$ we conclude that $u_1^{SW} (E + (E - 1)v'/M) \geq u_E^{SW} (1 + (E - 1)v'/M)$, obtaining the bound. To show that this bound is tight consider the same instance of Section 3.3, but with $z_e - c_e = (e - 1)M/(E - 1)$, and $y = (v', v' + w_E^2 + M/(E - 1))$. Social utilities are now $u_e^{SW} = v' + M(e + \sum_{f=1}^{E-e} \varepsilon^f)/(E - 1)$ for $e = 1, \dots, E - 1$. Notice that the difference of social utilities is upper bounded by M , and converges to this upper bound as $\varepsilon \rightarrow 0$. Then, the worst-case welfare loss for this family of instances is

$$\lim_{\varepsilon \rightarrow 0^+} \frac{u_E^{SW}}{u_1^{SW}} = \lim_{\varepsilon \rightarrow 0} \frac{E + \frac{E-1}{M}v' - \varepsilon^{E-1}}{1 + \frac{E-1}{M}v' + \sum_{f=1}^{E-1} \varepsilon^f} = \frac{ME + (E - 1)v'}{M + (E - 1)v'}. \quad \square$$

4.3. Relaxing MRLP. We now consider the case of 2 efforts and an arbitrary number of states, and show that the main result holds without MRLP. Surprisingly, the optimality conditions of the problem that characterizes the optimal wages suffice. Although with MRLP optimal contracts only paid positive wages for the highest output, this could happen for other states without MRLP. Indeed, the positive wage corresponds to the state that has the highest likelihood ratio, as formalized by the next proposition.

Proposition 3. *In the case with two non-identical efforts, the minimum expected payment when the principal induces the highest effort is*

$$z_2 = \max \left\{ c_2, (c_2 - c_1) \frac{\pi_2^\sigma}{\pi_2^\sigma - \pi_1^\sigma} \right\},$$

where $\sigma \triangleq \arg \max_{s \in \mathcal{S}} \pi_2^s / \pi_1^s$ is the state that achieves the highest likelihood ratio.

Proof. Consider the dual of MPLP(2). Without MRLP, constraints associated to $s < S$ are not redundant. Instead, we will show that all constraints are dominated by the one associated to σ , and that the dual problem is bounded. Recalling that the dual constraints (11) can be written as $(1 - \pi_1^s / \pi_2^s)p_1 + p_2 + 1 \geq 0$ for $s \in \mathcal{S}$, observe that these constraints can be represented by lines across point $-\mathbb{1}_2$ that are normal to vector $(1 - \pi_1^s / \pi_2^s, 1)$. Then, by definition, the constraint corresponding to σ dominates those of the other states.

To prove that the dual problem is bounded, we proceed by contradiction. Recall that the objective function is $(c_1 - c_2)p_1 - c_2p_2$, which has negative coefficients. The unboundedness implies

that $\pi_2^s \leq \pi_1^s$ for all $s \in \mathcal{S}$ because otherwise a feasible direction cannot exist since $p \leq 0$. Hence, π_1 and π_2 must coincide because they are probability vectors, which contradicts the fact that efforts are non-identical. Finally, the feasible region is a triangle with vertices $\{\mathbf{0}, -\mathbb{I}_2, -\mathbb{I}_1\pi_2^\sigma/(\pi_2^\sigma - \pi_1^\sigma)\}$. Evaluating the objective in all extreme points and using strong duality, we obtain the optimal expected payment. \square

Repeating the steps of Theorem 2, with the exception that the optimal contract now pays a positive wage when the realized state is σ , we obtain that the upper bound for the worst-case welfare loss is 2.

4.4. Relaxing the Monotonicity of Social Welfare with Respect to Effort Intensity. It is important to note that Assumption 1 was not used to characterize the minimum expected payments of the principal in Theorem 1 and Corollary 1. Although the assumption was used in Proposition 1, in fact, the proposition also holds under the following weaker assumption.

Assumption 2. The sequence of prevailing social welfare $\{u_e^{\text{SW}}\}_{e=1}^E$ is *unimodal*.

To prove that unimodality is sufficient, first recall that Corollary 1 implied that the principal has no incentive to implement any effort higher than the smallest socially-optimal effort e^{SO} . Hence, without loss of generality the analysis can be restricted to efforts $\{1, \dots, e^{\text{SO}}\}$. Because the subsequence satisfies Assumption 1, Proposition 1 follows literally without modification, from where we conclude that all results in Section 3.2 hold and the worst-case welfare loss is at most E .

In the case of two outcomes and an arbitrary number of efforts, it turns out that the sequence $\{u_e^{\text{SW}}\}_{e=1}^E$ is always unimodal. Indeed, take any two consecutive feasible efforts e and $e+1$, such that $u_e^{\text{SW}} \leq u_{e+1}^{\text{SW}}$. To conclude that the sequence is unimodal, we will show that effort $e-1$ verifies $u_{e-1}^{\text{SW}} \leq u_e^{\text{SW}}$. The condition $u_e^{\text{SW}} \leq u_{e+1}^{\text{SW}}$ can be expressed as

$$\frac{c_{e+1} - c_e}{\pi_{e+1}^S - \pi_e^S} \leq (-1, 1)y.$$

Because effort e is feasible we must have that $w_{e,e-1} \leq w_{e+1,e}$. Otherwise, the polyhedron of wage schedules that implement effort e would be empty. The last inequality can be written as

$$\frac{c_e - c_{e-1}}{\pi_e^S - \pi_{e-1}^S} \leq \frac{c_{e+1} - c_e}{\pi_{e+1}^S - \pi_e^S}.$$

Putting the two inequalities together, we obtain that $u_{e-1}^{\text{SW}} \leq u_e^{\text{SW}}$, and thus Assumption 2 holds. Thus, in the case of two outcomes, only the FOSD assumption is needed to upper bound the worst-case welfare loss by E .

4.5. A Different Worst-Case Bound. We finish this section by considering an alternative worst-case bound that takes the point of view of the principal. Specifically, we focus on the impact of the limited liability constraint in the principal's utility. We denote the optimal principal's utility when the agent is not subject to the LL constraint (4) by u^{PnoLL} . With this, the worst-case bound from the perspective of the principal is the ratio of the principal's optimal utility when the limited-liability constraint is enforced to the principal's optimal utility when it is not present. In other words, we compute $r = \sup_{\pi, y, c} \{u^{\text{PnoLL}}/u^P\}$. We will show that under MLRP, the worst-case utility loss is bounded by the the number of efforts, and that this bound is essentially tight.

Theorem 4. *Suppose that MLRP and Assumption 1 hold. Then, in the risk-neutral principal-agent problem with limited liability, $r = E$.*

Proof. According to Lemma 2, $u^{\text{PnoLL}} = u^{\text{SO}}$; hence, r is more easily expressed as $\sup_{\pi, y, c} \{u^{\text{SO}}/u^P\}$. From Lemma 1 we know that $u_e^{\text{SW}} \geq u_e^P$ for all efforts $e \in \mathcal{E}$. Thus $r \leq \rho = E$. Now, let $\mathcal{R} = \{e_r\}_{r=1}^R$ be the sequence of relevant efforts. By Proposition 1 we have that $z_{e_{r_1}} = c_{e_{r_1}}$, and then $u_{e_{r_1}}^P = u_{e_{r_1}}^{\text{SW}}$. Hence, following the proof of Theorem 2, we conclude that $r \leq E$. \square

The interpretation of this bound is that enforcing limited liability may be costly to the principal, but only to a certain extent, as discussed earlier for the main result. The previous proof provides another interpretation of the same ratio: it can be seen as the social welfare with and without limited liability constraints, and in that case the same conclusion applies to the system instead of to the principal.

5. MULTIPLE TASKS

Most principal-agent relationships are more complex than the one considered so far in the sense that an agent usually performs different tasks, each endowed with different actions. Thus, it is common to think of job complexity in terms of the number of tasks rather than number of actions available. In this section, we consider a principal-agent relationship with multiple tasks. In particular, we adopt the model proposed by Laux (2001) in which there are two effort levels for each task.

The principal is endowed with N identical and stochastically independent tasks. Each task has two possible outcomes, either success or failure. The corresponding payoffs for the principal are \bar{y} if the task is successful, and \underline{y} in the case of failure, with $\bar{y} > \underline{y}$. The agent can exert two efforts, either high or low. The high effort entails a cost c_h for the agent, while the cost of the low effort is c_l . Since a higher effort demands more work, $c_h > c_l$. Finally, we denote by p_h the probability of success when effort is high, and by p_l the probability of success when effort is low. MRLP implies that $p_h > p_l$ (the higher the effort, the greater the likelihood of success).

The principal hires an agent to perform the N tasks. Since tasks are identical, the principal offers a compensation that depends only on the number of tasks that end up being successful, denoted by $s \in \mathcal{S} = \{0, \dots, N\}$; the identity of each task is irrelevant. Hence, the agent is paid a wage w^s when s tasks turn out to be successful. The total revenue for the principal is thus $y^s = s\bar{y} + (N - s)\underline{y}$ for $s \in \mathcal{S}$. In view of the tasks' symmetric nature, the agent is indifferent between tasks and he is only concerned about the total number of tasks in which he exerts high effort. We define the *aggregated effort* $e \in \mathcal{E} = \{0, \dots, N\}$ as the number of tasks in which the agent works hard. Notice that, for notational simplicity, we adopt indices that start at zero for both efforts and states. We assume that effort costs are additive, and linear in the number of tasks. Hence, the aggregate costs for the agent are $c_e = ec_h + (N - e)c_l$ for $e \in \mathcal{E}$. Finally, note that the probability of having s successful tasks, given that the agent works hard on e tasks, is given by

$$\pi_e^s = \sum_{i=0}^s \binom{e}{i} p_h^i (1 - p_h)^{e-i} \binom{N-e}{s-i} p_l^{s-i} (1 - p_l)^{N-e-s+i},$$

where we assumed that $\binom{n}{k} = 0$ if $k > n$.

This model can be fully reduced to a principal-agent model with a single task, $N + 1$ states and $N + 1$ efforts. To map the multiple-task model into the model of Section 2 we show that the aggregate instance satisfies MLRP. Intuitively, this states that the larger the observed number of successful tasks, the more likely it is that the agent works hard in many tasks.

Lemma 3. *When the principal hires one agent to perform N identical and independent tasks, the distribution of the aggregated outcome satisfies MLRP.*

Therefore, by a simple application of Theorem 2, the worst-case welfare loss is upper bounded by $N + 1$. This bound, however, is not tight. Laux (2001) shows that when the manager exerts high effort in all tasks (aggregated effort N), the only binding constraint is the one in which the manager must have no incentive to choose low effort in the N tasks (aggregated effort 0). This implies that

only two aggregated efforts are relevant, namely N and 0 . Thus, the worst-case welfare loss is at most 2 and this bound is tight. The following result formalizes this discussion.⁶

Theorem 5. *In the principal-agent problem in which both players are risk-neutral and there are N identical and independent tasks, the only two relevant aggregated efforts are the ones in which: (i) the agent exerts high effort in each possible task (aggregated effort N), and (ii) the agent neglects all tasks (aggregated effort 0). Hence, the optimal wage schedule for aggregated effort N is*

$$w_N = N \frac{c_h - c_l}{p_h^N - p_l^N} \mathbb{I}_N.$$

Additionally, the worst-case welfare loss is 2.

Proof. From Theorem 1, aggregated effort N is always relevant. To prove that the next relevant aggregated effort in the sequence is 0, it suffices to show that $w_{N,0} \geq w_{N,e}$ for all $e = 1, \dots, N-1$. Letting $q = p_h/p_l > 1$, we can write

$$(14) \quad w_{N,e} = \frac{c_N - c_e}{\pi_N^S - \pi_e^S} = \frac{Nc_h - ec_h - (N-e)c_l}{p_h^N - p_h^e p_l^{N-e}} = \frac{c_h - c_l}{p_l^N} \frac{N-e}{q^N - q^e} = \frac{c_h - c_l}{p_l^N (q-1)} \left(\frac{\sum_{n=e}^{N-1} q^n}{N-e} \right)^{-1},$$

where the last equation follows from expressing $q^N - q^e$ as a geometric sum. The right-hand side is the multiplication of a positive coefficient and the reciprocal of an average of terms greater than one. Since $q > 1$, as e increases the smallest term is sequentially excluded from the average. Hence, the average is increasing in e , and $w_{N,e}$ is decreasing in e . Thus, $w_{N,0} \geq w_{N,e}$ for all $e = 1, \dots, N-1$, implying that the next and last relevant effort is 0, and $w_N = w_{N,0} \mathbb{I}_N$.

Finally, recall that in Theorem 2 the upper bound on the worst-case welfare loss is the number of relevant efforts. Because in this case $\mathcal{R} = \{0, N\}$, then $\rho \leq 2$. This bound is tight; to see that it suffices to take each task to be equal to the instance of Section 3.3 with $E = 2$. \square

6. CONCLUSIONS

This paper considers a worst-case approach to quantify the welfare loss that arises from the principal's impossibility to observe the agent's effort when there is limited liability. We have shown that the worst-case welfare loss exactly equals the number of efforts available to the agent, which suggests that the welfare loss in a principal-agent relationship depends on the agent's discretion understood as the set of efforts available to the agent.

The principal-agent model in its different forms has been used to explain many contractual arrangements such as sharing contracts, insurance contracts, managerial contracts, political relationships and so on and so forth. In addition, it has been used to provide an economic theory of the firm and a theory of organizational forms. Our results show that in these cases and in many others the existence of an agency relationship with moral hazard may have nontrivial consequences in terms of welfare loss and thus the proper design of contracts and organizations to deal with moral hazard is of great practical importance.

Two concluding comments are useful: first, we leave open the question of how to relax Assumption 1 in the general case. We conjecture that the worst-case bounds do not change; and second, we leave as further work to consider a risk-averse agent. In the latter case the optimal contract is highly nonlinear, and thus its characterization in terms of the main parameters is a complex task. There is an exception to this, which is given by the linear agency model introduced in the literature by Holmström and Milgrom (1987).

⁶Our proof for the optimal wage structure exploits the results from Section 3, and hence it is somewhat simpler than the one given by Laux (2001).

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.

APPENDIX A. PROOFS

A.1. Proof of Theorem 1.

Proof. We construct the set \mathcal{R} backwards, starting from effort E . At each step, we guarantee that (i) the properties hold for efforts $\{e_r, \dots, e_R\}$, (ii) $w_{e_r, f} \geq w_{e_r}$ for all efforts $f > e_r$, and (iii) $w_{e_r, f} \leq w_{e_r}$ for all efforts $f < e_r$.

First, consider effort E . If $z_E = c_E$, then the principal gets the maximum possible social welfare for himself with E so he has no reason to chose any other effort, implying that $\mathcal{R} = \{E\}$ and that the wage vector $w_E = (0, \dots, 0, c_E/\pi_E^S)$ is optimal. To prove this, the IC constraint (7) together with MLRP give that $c_E - c_f \leq (\pi_E - \pi_f)w_E \leq (1 - \pi_f^S/\pi_E^S)\pi_E w_E$. Since there is a feasible solution that satisfies $\pi_E w = c_E$, we have that $c_f \pi_E^S \geq c_E \pi_f^S$, which proves that the IC constraints are valid for w_E . Then, w_E satisfies all constraints of MPLP(e) and attains the objective value c_E , proving its optimality. The only thing left to prove is (iii), which follows from the same inequality.

We now consider that $z_E > c_E$. The dual constraints (11) can be rewritten as

$$\sum_{e < E} \left(1 - \frac{\pi_e^S}{\pi_E^S}\right) p_e + p_E + 1 \geq 0 \quad \forall s \in \mathcal{S}.$$

From MLRP we know that $1 - \pi_e^S/\pi_E^S$ is increasing in s , and since $p \leq 0$ all constraints with $s < S$ are redundant. Thus, the extreme points are $\{\mathbf{0}, -\mathbb{I}_1 \pi_E^S/(\pi_E^S - \pi_1^S), \dots, -\mathbb{I}_{E-1} \pi_E^S/(\pi_E^S - \pi_{E-1}^S), -\mathbb{I}_E\}$, where \mathbb{I}_e is the unit vector corresponding to effort e . Evaluating the objective in all extreme points, we obtain the optimal dual value

$$(15) \quad z_E = \max_{e=1, \dots, E-1} \left\{ (c_E - c_e) \frac{\pi_E^S}{\pi_E^S - \pi_e^S} \right\}.$$

We let the next relevant effort, e_{R-1} , be the one attaining the maximum. Next, we show that the optimal solution to the primal problem is

$$w_E = w_{E, e_{R-1}} = \mathbb{I}_S \frac{c_E - c_{e_{R-1}}}{\pi_E^S - \pi_{e_{R-1}}^S}.$$

Observe that $\pi_E w_E = z_E \geq c_E$, so it satisfies the primal constraints (6) and (8). It also satisfies the IC constraints (7) because $(\pi_E - \pi_e)w_E \geq c_E - c_e$ follows directly from the optimality of effort e in (15). The latter, in addition, proves (iii). The optimality of w_E is immediate from strong duality for linear programming. Since (ii) is trivial in this case, that finishes the proof for the basic case.

For the inductive step, we consider problem MPLP(e_r), assuming that (i) the properties hold for $\{e_{r+1}, \dots, e_R\}$, (ii) $w_{e_{r+1}, f} \geq w_{e_{r+1}}$ for all $f > e_{r+1}$, and (iii) $w_{e_{r+1}, f} \leq w_{e_{r+1}}$ for all $f < e_{r+1}$.

The feasibility of $w_{e_{r+1}}$ for MPLP(e_{r+1}), and the equality $(\pi_{e_{r+1}} - \pi_{e_r})w_{e_{r+1}} = c_{e_{r+1}} - c_{e_r}$ imply that $w_{e_{r+1}}$ is feasible for MPLP(e_r). From $w_{e_{r+1}}$ we construct an optimal solution to MPLP(e_r) by decreasing its component S until we reach a face. If in this process we first reach the IR constraint, we have $z_{e_r} = c_{e_r}$ and $\mathcal{R} = \{e_r, \dots, e_R\}$, completing the proof similarly to the basic case. Otherwise, we let the next relevant effort e_{r-1} be that corresponding to the IC constraint that is reached first, which cannot be one corresponding to an effort larger than e_r . Indeed, we find the maximum ε such that $w_{e_{r+1}} - \varepsilon \mathbb{I}_S$ is feasible. The IC constraints for MPLP(e_r) can be rewritten as $w_{e_{r+1}}^S - \varepsilon \leq w_{e_r, f}^S$ for $f > e_r$ and $w_{e_{r+1}}^S - \varepsilon \geq w_{e_r, f}^S$ for $f < e_r$. Hence efforts $f > e_r$ do not impose a constraint for ε . Therefore, we have that

$$w_{e_r} = \mathbb{I}_S \max_{e < e_r} \left\{ \frac{c_{e_r} - c_e}{\pi_{e_r}^S - \pi_e^S} \right\}.$$

By construction w_{e_r} is feasible, it verifies (ii) and (iii), and it has the desired objective value. To conclude we consider the dual solution $p_{e_r} = -\mathbb{I}_{e_{r-1}} \pi_{e_r}^S/(\pi_{e_r}^S - \pi_{e_{r-1}}^S)$ and prove that w_{e_r} and p_{e_r} are an optimal primal-dual pair. Indeed, p_{e_r} is dual feasible because $p_{e_r} \leq 0$ and (11) hold (the latter because of MLRP).

Finally, the objective function value is $(c_{e_r} - c_{e_{r-1}})\pi_{e_r}^S / (\pi_{e_r}^S - \pi_{e_{r-1}}^S) = \pi_{e_r}^S w_{e_r}^S$, completing the last step of the proof. \square

A.2. Proof of Corollary 1.

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 $\text{MPLP}(e_r)$ 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 $p = -\mathbb{I}_{e_r} \pi_f^S / (\pi_f^S - \pi_{e_r}^S)$ is dual feasible for effort f , and its objective value is $(c_f - c_{e_r})\pi_f^S / (\pi_f^S - \pi_{e_r}^S) = \pi_f^S w_{e_r, f}^S$, which by weak duality is a lower bound on z_f . Hence, $z_f \geq \pi_f^S w_{e_r, f}^S = \pi_{e_r}^S w_{e_r, f}^S + w_{e_{r+1}, f}^S (\pi_f^S - \pi_{e_r}^S) + w_{e_{r+1}}^S (\pi_{e_{r+1}}^S - \pi_{e_r}^S)$. Rearranging the terms, the last expression equals $z_{e_{r+1}} + c_f - c_{e_{r+1}} + \pi_{e_r}^S (w_{e_r, f}^S - w_{e_{r+1}}^S) \geq z_{e_{r+1}} + c_f - c_{e_{r+1}}$, where the inequality follows because $w_{e_r, f} \geq w_{e_{r+1}}$. Indeed,

$$w_{e_r, f}^S = \frac{c_f - c_{e_{r+1}}}{\pi_f^S - \pi_{e_r}^S} + \frac{c_{e_{r+1}} - c_{e_r}}{\pi_f^S - \pi_{e_r}^S} = w_{e_{r+1}, f}^S \frac{\pi_f^S - \pi_{e_{r+1}}^S}{\pi_f^S - \pi_{e_r}^S} + w_{e_{r+1}}^S \frac{\pi_{e_{r+1}}^S - \pi_{e_r}^S}{\pi_f^S - \pi_{e_r}^S} \geq w_{e_{r+1}}^S,$$

because $w_{e_{r+1}, f} \leq w_{e_{r+1}}$ (property (iii) in the proof of Theorem 1) and $\pi_f^S - \pi_{e_{r+1}}^S \leq 0$. \square

A.3. Proof of Lemma 3.

Proof. Let $\{X_e\}_{e \in \mathcal{E}}$ be a family of random variables, such that X_e is the random number of successes given that the agent works hard in e tasks. Then, X_e is the sum of e independent Bernoulli random variables with success probability p_h , and $N - e$ independent Bernoulli random variables with success probability p_l . Denote by $Y(p)$ a Bernoulli random variable with success probability p ; i.e., $\mathbb{P}(Y(p) = 1) = p = 1 - \mathbb{P}(Y(p) = 0)$. Hence, we may write X_e as

$$X_e = \sum_{f=1}^e Y_f(p_h) + \sum_{f=e+1}^N Y_f(p_l) = \sum_{f=1}^N Y_f(p_f(e)),$$

where the functions $\{p_f(e)\}_{f \in \mathcal{E}}$ equal p_l if $e < f$, and p_h otherwise. Notice that for all $f \in \mathcal{E}$ the functions $p_f(e)$ are non-decreasing in e . Ghurye and Wallace (1959) or more recently Huynh (1994) show that given any number of independent Bernoulli random variables Y_f with success probability $p_f(e)$ strictly between 0 and 1, and non-decreasing in e , then the sum $\sum Y_f$ has monotone likelihood-ratio with respect to e . \square

REFERENCES

- Acemoglu, D., K. Bimpikis, and A. Ozdaglar (2009). Price and capacity competition. *Games and Economic Behavior* 66(1), 1–26.
- Acemoglu, D. and A. Ozdaglar (2007). Competition and efficiency in congested markets. *Mathematics of Operations Research* 32(1), 1–31.
- Alchian, A. A. and H. Demsetz (1972). Production, information costs, and economic organization. *American Economic Review* 62(5), 777–95.
- Babaioff, M., M. Feldman, and N. Nisan (2006). Combinatorial agency. In *Proceedings of the 7th ACM Conference on Electronic Commerce (EC'06)*, Ann Arbor, MI, pp. 18–28. ACM Press, New York, NY.
- Babaioff, M., M. Feldman, and N. Nisan (2009). Free-riding and free-labor in combinatorial agency. In M. Mavronicolas and V. Papadopoulou (Eds.), *Proceedings of the 2nd International Symposium on Algorithmic Game Theory (SAGT'09)*, Volume 5814 of *Lecture Notes in Computer Science*, pp. 109–121. Springer.
- Balmaceda, F. (2010). Job design: Complexity, incentives and discretion. Working paper CEA-134, CEA, University of Chile.
- Ghurye, S. and D. Wallace (1959). A convolutive class of monotone likelihood ratio families. *Annals of Mathematical Statistics* 30(4), 1158–1164.
- Grossman, S. J. and O. D. Hart (1983). An analysis of the principal-agent problem. *Econometrica* 51(1), 7–45.
- Holmström, B. and P. Milgrom (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica* 55, 303–328.
- Holmström, B. and P. Milgrom (1991). Multi-task principal-agent analyses: Incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organization* 7, 24–52.
- Holmström, B. and J. Tirole (1997). Financial intermediation, loanable funds and growth. *The Quarterly Journal of Economics* 112(3), 663–691.
- Huynh, H. (1994). A new proof for monotone likelihood ratio for the sum of independent Bernoulli random variables. *Psychometrika* 59(1), 77–79.
- Jensen, M. C. and W. H. Meckling (1979). Rights and production functions: An application to labor-managed firms and codetermination. *The Journal of Business* 52(4), 469–506.
- Koutsoupias, E. and C. H. Papadimitriou (1999). Worst-case equilibria. In C. Meinel and S. Tison (Eds.), *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, Trier, Germany, Volume 1563 of *Lecture Notes in Computer Science*, pp. 404–413. Springer, Heidelberg.
- Laux, C. (2001). Limited liability and incentive contracting with multiple projects. *RAND Journal of Economics* 32, 514–526.
- Moulin, H. (2008). The price of anarchy of serial, average and incremental cost sharing. *Economic Theory* 36(3), 379–405.
- Moulin, H. (2009). Almost budget-balanced vcg mechanisms to assign multiple objects. *Journal of Economic Theory* 144(1), 96–119.
- Nisan, N., T. Roughgarden, É. Tardos, and V. V. Vazirani (2007). *Algorithmic Game Theory*. Cambridge University Press, Cambridge, UK.
- Rothschild, M. and J. E. Stiglitz (1970). Increasing risk: I. a definition. *Journal of Economic Theory* 2(3), 225–243.
- Shapiro, C. and J. E. Stiglitz (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review* 74(3), 433–44.
- Spence, M. and R. Zeckhauser (1971). Insurance, information, and individual action. *American Economic Review* 61(2), 380–87.
- Stiglitz, J. E. (1974). Incentives and risk sharing in sharecropping. *Review of Economic Studies* 41(2), 219–55.