

Optimal contracts under imperfect enforcement revisited*

Hans K. Hvide[†]

August 25, 2005

Abstract

We consider a financing game with costly enforcement based on Townsend (1979), but where monitoring is non-contractible and allowed to be stochastic. Debt is the optimal contract. Moreover, the debt contract induces creditor leniency and strategic defaults by the borrower on the equilibrium path, consistent with empirical evidence on repayment and monitoring behavior in credit markets.

Keywords: Costly state verification, debt contract, priority violation, strategic defaults.

JEL codes: D02, D82, G21, G33.

1 Introduction

Debt contracts are ubiquitous in financial markets. The classic Townsend (1979) considers a setting where a project is financed by an outside investor, and the subsequent cash flow is observable only to the borrower. Townsend shows that if monitoring is contractible and deterministic, debt contracts are optimal. Under the optimal contract, the investor

*Thanks to Tore Leite, Arne-Christian Lund, and Roland Strausz for comments.

[†]Department of Finance and Management Science, Norwegian School of Economics and Business, Helleveien 30, 5045 Bergen, Norway. Hvide is also affiliated with CEPR and IZA. Email: hans.hvide@nhh.no.

monitors (or liquidates) upon a default, and defaults by the borrower are driven purely by lack of liquidity.

This paper considers optimal contracts and repayment/monitoring behavior in a modified Townsend setting where monitoring is non-contractible and allowed to be stochastic. We find that the optimal contract is a debt contract. Under the optimal contract, the borrower strategically defaults (i.e., offers the investor a partial repayment even with sufficient funds to repay in full) and the investor is lenient towards defaults (i.e., less than certain monitoring upon defaults by the borrower) in equilibrium.

Our finding that debt is optimal adds to the literature exploring optimal contracting under repayment frictions. Our finding that debt incurs strategic defaults and creditor leniency fits well with the empirical corporate finance literature on repayment behavior in credit markets, such as Brown et al., (2003), Esty and Megginson (2003), and Davydenko (2005). For example, in a broad sample of firms, Davydenko (2005) finds that about 70% of defaulting firms are not liquidated.

The intuition behind our results can be understood from the following figure.

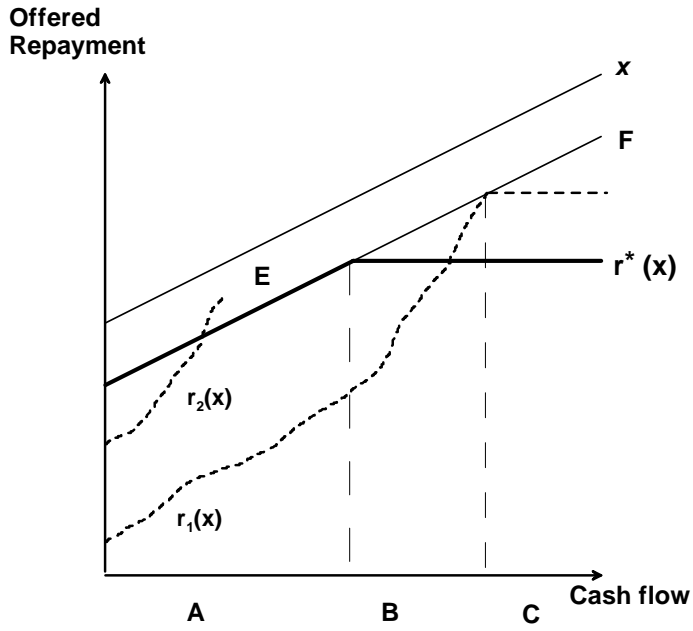


Figure 1

The figure depicts the cash flow x (known only to the manager) on the horizontal axis and the dollar amount offered in payment to the investor on the vertical axis. The bold line depicts the optimal repayment function $r^*(x)$ (how to implement $r^*(x)$ is discussed below). $r^*(x)$ follows the feasibility barrier **F** for x in region **A** and gives a constant payout in regions **B** and **C**. We now argue that $r^*(x)$ must beat alternative payment functions, such as $r_1(x)$, by having lower monitoring costs. To make a comparison between $r^*(x)$ and $r_1(x)$ interesting, assume that $r^*(x)$ and $r_1(x)$ induce the same aggregate payment to the investor. Note first that to induce any non-constant $r(x)$, the investor needs to more likely monitor the lower the payment. At the level of maximal payment the monitoring probability is zero. Given these observations, let us compare the monitoring costs for $r^*(x)$ and for $r_1(x)$ in the regions **A**, **B**, **C**. In **C**, the investor receives his maximal payout under both $r^*(x)$ and $r_1(x)$ and does not have incentives to monitor in either case. Judged from this region alone, $r^*(x)$ and $r_1(x)$ are equally good. In **B**, $r^*(x)$ offers the maximal payout, and incurs no monitoring, while $r_1(x)$ pays less than its maximal payout and therefore must imply some monitoring by the investor (if not, the manager would never offer the maximal payout). Therefore $r^*(x)$ beats $r_1(x)$ in **B**. In **A**, $r^*(x)$ must also beat $r_1(x)$, because $r^*(x)$ offers more than $r_1(x)$ (more precisely, $r^*(x)$ has a

lower proposed writedown than $r_1(x)$, where the proposed writedown equals the maximal payout subtracted the payment offer). Thus $r^*(x)$ dominates $r_1(x)$ in all regions **A**, **B**, and **C**, and must therefore beat $r_1(x)$. Now consider a payment scheme $r_2(x)$, which crosses the line $\mathbf{F} = x - c$ (where c is the cost of monitoring) and enters the area **E**, and note that it is not feasible. If the manager plays $r_2(x)$, it would be strictly optimal for the investor to not monitor following payments in **E**; by monitoring he gets at most $x - c$, while by accepting the payment offer he gets more. But then an equilibrium with $r_2(x)$ would unravel, and cannot exist.

This argument gives intuition for why the optimal payment function $r^*(x)$ is flat for high cash flows and pays "as much as possible" for low cash flows. It does not take much imagination to guess that $r^*(x)$ can be implemented with a debt contract, which is indeed also the case. Note that the argument also gives intuition for why strategic defaults occur in equilibrium. Interpreting (as it turns out correctly) the maximal payout of $r^*(x)$ as the contractual debt obligation, we see that the borrower defaults on debt in the upper interval of **A**, where he in fact has sufficient funds to repay debt in full but chooses not to do so.

Several papers have modified the Townsend (1979) basic assumptions that monitoring is contractible and deterministic. Townsend (1979) showed with an example that a (non-debt) contract with stochastic monitoring dominates a debt contract with deterministic monitoring. Border & Sobel (1987), Mokherjee and Png (1989), and Boyd & Smith (1994) show that optimal contracts under these conditions, i.e., stochastic but non-contractible monitoring, tend to involve some forgiveness of the contractual obligation, but there is no guarantee that the optimal contract will be debt-like. In an important paper, Krasa & Villamil (2000) allow for stochastic monitoring and assume that monitoring is non-contractible. Their solution is surprisingly similar to Townsend's in that debt contracts are optimal, and the equilibrium payment and monitoring behavior is also essentially the same. Krasa-Villamil (2000) require equilibrium contracts to be "time consistent" (their equation (1.4)), meaning that optimal contracts need to be immune to renegotiation at the interim stage (after the entrepreneur has made a payment offer but before the investor has decided whether to monitor). Time consistency implies deterministic monitoring in equilibrium; if the investor is indifferent between monitoring and not monitoring for some

payment offer, as must be the case for stochastic monitoring to occur in an equilibrium, there would be mutual gains from "bribing" the investor to refrain from it.¹ Such bribes are possible through rewriting the contract. We study a closely related problem to Krasa & Villamil (2000) in that we allow stochastic monitoring and assume that monitoring is non-contractible. We differ, however, in not requiring time consistency (we rule out interim renegotiation of contracts) and as a result get stochastic monitoring under the optimal contract. The other difference to Krasa & Villamil (2000) is that we require the payment to be continuous in the underlying cash flow. This assumption means that we can employ differentiation techniques to solve our problem. Other related papers include Hvide & Leite (2005) who derives the optimal mix of debt and equity in a setting without commitment and allowing for random monitoring. Their pure debt equilibria have the same structure as the equilibria in the present paper, but Hvide & Leite (2005) does not derive optimal contracts. Gale & Hellwig (1989) analyze a similar payment game to in the present paper and derive necessary conditions for the existence of signalling equilibria that are broadly consistent with the equilibria of the present paper, but do not derive optimal contracts. In Section 1, we set up the model and Section 2 contains the results. Section 3 concludes.

2 Model

There are two risk-neutral agents, an entrepreneur and an investor. The penniless entrepreneur is endowed with a project that requires I units of funding to yield the cash flow x . The cash flow is stochastic with density $h(\cdot)$ defined on $X = [x_L, x_H]$. In return for providing I , the investor gets a claim on x . This claim is a function $f : X \rightarrow \mathfrak{R}$. We make the feasibility restriction $f(x) \leq x, \forall x \in X$, and denote the set of contracts satisfying it for F .² After being funded, x is generated and observed only by the entrepreneur. The entrepreneur makes a payment offer r to the investor, where the payment

¹The time consistency requirement of Krasa-Villamil (2000) implies that monitoring can occur in equilibrium only when the entrepreneurs lacks liquidity. In other words, time consistency rules out strategic defaults.

²Note that we depart slightly from Krasa & Villamil (2000) in allowing the claim only to depend on x , and not on the payment offer r . This has no impact on our results.

function $r(x)$ is a mapping $r : X \rightarrow \mathfrak{R}$ with the restriction $r \leq x$. We consider deterministic³ and absolutely continuous payment functions $r(x)$, which implies that its derivative $r'(x)$ exists almost everywhere. The set of payment functions satisfying these criteria is denoted by R . The investor accepts or rejects the offer r based on his posterior beliefs h' . If the investor accepts, he receives r , and the manager gets the residual $x - r$. If the investor rejects/monitors, he receives a payoff y according to the written contract, i.e., $y = \min[f(x), x - c]$, and the manager gets the residual. Note that implicit in this formulation the cost of monitoring c is taken from the firm's cash flow (our results do not depend upon this assumption). The investor's accept probability function is a mapping $P : \mathfrak{R} \rightarrow [0, 1]$. To ensure sufficient liquidity to cover the monitoring cost, we assume that $c \leq x_L$. To make the problem interesting, we finally assume that an $r(x)$ that gives a constant payout on X falls short of making the investor willing to participate.

Let $e = 1$ if the investor rejects/monitors and $e = 0$ if the investor accepts an offer. The payoff functions π_i , where $i = I, E$ are then given by,

$$\begin{aligned}\pi_E &= (1 - e)(x - r) + e(x - y) = x - (1 - e)r - ey \\ \pi_I &= (1 - e)r + ey\end{aligned}\tag{1}$$

For a given strategy tuple $\langle r(x), P(r) \rangle$ the expected payoffs are given by,

$$\begin{aligned}E\pi_E &= \int_X [P(r(x))(x - r) + (1 - P(r(x)))(x - y - c)]dH \\ E\pi_I &= \int_X [P(r(x))r + (1 - P(r(x)))y]dH\end{aligned}\tag{2}$$

The investor participation constraint emerges from setting $E\pi_I = I$. The basic trade-offs are as follows. The manager makes a payment offer to the investor trading off the gains from cash diversion with cost of an increased probability of monitoring (and hence reducing the net payoff via reducing the cash flow). The investor follows a monitoring

³There are technical problems in defining mixed strategies for a continuous type space. Barring such problems, we conjecture that a mixed repayment strategy is not consistent with equilibrium (in contrast to in Persons, 1997, which operates with a finite type space). The intuition is that a continuous X pins down a unique accept probability function $P(\cdot)$, which in turn makes only one repayment offer optimal for given $\langle f(x), x \rangle$. Martimort & Stole (2002) makes a similar observation in a different context.

strategy that balances off the cost of monitoring against the possible gain from detecting a diversion attempt by the manager. We focus on Perfect Bayesian equilibria (PBE) of the payment game. That a tuple $\langle r(x), P(r), h, h' \rangle$ is a PBE means that a) $P(r)$ is optimal play by the investor given his posterior beliefs h' , b) The manager anticipates the investor's behavior and chooses r to maximize his payoff, and c) The investor's posterior beliefs are formed using Bayes' rule whenever possible.

The implementation problem can be formulated as,

Problem 1 (3)

$$\text{Max}_{\langle r(\cdot), P(\cdot) \rangle} E\pi_E$$

$$\text{s.t. } E\pi_I = I$$

$$r(x) \in R$$

$$f(x) \in F$$

Strategies and beliefs are PBE

Problem 1 amounts to finding the payment function and monitoring probabilities that maximize the expected utility of the entrepreneur given the incentive constraints. Problem 1 is equivalent to finding a contract $f(x)$ that minimizes the expected monitoring (verification) cost $V = \int_X (1 - P(\cdot)) dH$ subject to the investor's participation constraint. Let us define a debt contract as,

$$f^D(x) = \min(x, d) \tag{4}$$

This contract entitles the investor to the full cash flow up to a point d , and then a constant payout.

3 Analysis

The main result of the paper is as follows.

Theorem 1 *(i) The optimal contract is a debt contract. (ii) Under the optimal contract, the investor is lenient with defaults and the manager defaults strategically.*

We prove the theorem in several steps. Since the investor cannot precommit to a monitoring strategy, the revelation principle does not hold, and we have to apply a more indirect method of proof.⁴ The strategy of the proof is to solve a simplified version of Problem 1 and then show that the solution also solves Problem 1.

First some definitions. Let $\Gamma(f)$ be the set of PBE induced by a contract $f(x) \in F$. We say that the payment function $r(x)$ is *inducible* (implementable) if there exists $f(x) \in F$ such that $r(x)$ is contained in $\Gamma(f)$. Let \tilde{x} be some arbitrary constant on X and denote by B (where $B \subset R$) the set of payment functions satisfying (i) $r'(x) > 0$ for $x \in [x_L, \tilde{x}]$ and (ii) $r'(x) = 0$. B contains all continuous payment functions that are either strictly increasing on X or initially strictly increasing and then flat.

Now define Problem 1' as Problem 1 except $r(x) \in B$ is substituted in for $r(x) \in R$ in (3). We start out by solving Problem 1' and then show that the solution to Problem 1' is also a solution to Problem 1. The method we use to solve Problem 1' is the standard one of first finding the cheapest way to induce an arbitrary $r(x) \in B$ and then find the optimal $r(x)$. We first note the following.

Lemma 1 *For any $r(x) \in B$ to be inducible, it must satisfy $r(x) \leq x - c, \forall x \in X$.*

Proof. The proof is by contradiction. Let us assume that $r(x) \in R$ is strictly increasing on X . Suppose that there exists a contract $\hat{f}(x) \in F$ that induces a $r(x)$ with $r(x) > x - c$ on some interval $X' = [x_1, x_2]$. Since $\hat{f}(x) \leq x$ by feasibility we must have that $r(x) > \hat{f}(x) - c$ on X' . But then the investor would accept offers on $[r(x_1), r(x_2)]$ with probability 1. In that case, the entrepreneur never offers more than $r(x_1)$ on X' , which contradicts the assumption that $r(x)$ is strictly increasing. To extend the proof to the case where $r(x)$ is flat at the top is straightforward and omitted. ■

Lemma 1 limits the set of feasible payment functions in B to lie below the line **F** in Figure 1.

Lemma 2 *For any inducible $r(x) \in B$, (i) The contract $f^*(x) = r(x) + c$ induces $r(x)$. (ii) The associated accept probability function is $P(r) = e^{\frac{r-r(x_H)}{c}}$. (iii) $f^*(x)$ is the cheapest*

⁴Bester & Strausz (2001) show that a modified version of the revelation principle holds under limited commitment. Since we operate in a setting with a continuous type space, their results do not immediately apply.

way to induce $r(x)$.

Proof. Fix $r(x) \in B$ and suppose that the contract is $f^*(x) = r(x) + c$. Since $r(x) \leq x - c$ by Lemma 1, clearly $f^*(x) \in F$. We show that $f^*(x)$ induces $r(x)$. Note that if the manager adheres to $r(x)$, the investor is indifferent to monitoring or not since $r = y$. Given that the manager plays $r(x)$, any $P(r)$ is therefore consistent with optimal play by the investor (given h' appropriately defined). We now construct $P(r)$ such that the manager does not have incentives to deviate from $r(x)$. Since the ensuing $P(r)$ is unique, we thereby prove both (i) and (ii). For given x , the expected payoff for the manager from offering r equals,

$$U_E(r) = P(r)(x - r) + (1 - P(r))(x - f^*(x)) \quad (5)$$

Differentiating with respect to r , we get

$$\begin{aligned} U'_E(r) &= P'(r)(x - r) - P - P'(r)(x - f^*(x)) \\ &= P'(r)(f^*(x) - r) - P \end{aligned} \quad (6)$$

For $r(x)$ to be optimal play by the manager, it must be a local maximum for all x ,

$$P'(r)c - P(r) = 0 \quad (7)$$

The unique solution to this differential equation (barring the trivial solution $P(x) = 0$) is $P(r) = K \frac{r}{e^c}$, where K is an integration constant. Invoking the corner condition $P(r(x_H)) = 1$ (the investor accepts the maximal offer with probability 1) and simplifying,

$$P(r) = e^{\frac{r - r(x_H)}{c}} \quad (8)$$

$P(r)$ is increasing and convex in r . Note that the associated monitoring probability $1 - P$ lies between zero and one for all r . To show that adhering to $r(x)$ is a global optimum

for the manager, observe that $P'(r) = \frac{P(r)}{c}$. Substituting into $U'_E(r)$,

$$\begin{aligned} U'_E(r) &= P'(r)(f^*(x) - r) - P \\ &= P(r)\left[\frac{f^*(x) - r}{c} - 1\right] \\ &= P(r)[f^*(x) - c - r]/c \end{aligned} \tag{9}$$

This expression is negative for $r > f^*(x) - c$ and positive for $r < f^*(x) - c$. Hence $r(x) = f^*(x) - c$ is a global optimum for the manager. To complete the proof of (i) and (ii), we need to construct beliefs that support this separating equilibrium. The prior of the investor is that x follows $h(x)$. For an offer r on the equilibrium path, the investor's posterior beliefs h' are degenerate at $r + c$ for $r < r(x_H)$, and unrestricted for $r = r(x_H)$. These posterior beliefs are obviously consistent with the manager's strategy. We do not need to restrict the investor's posterior beliefs for offers outside the interval $[r(x_L), r(x_H)]$; for any posterior beliefs with support X it will be optimal for the investor to accept $r > r(x_H)$ and optimal to reject $r < r(x_L)$. We have then proved (i) and (ii).

We now need to show (iii) that there are no cheaper ways to induce an arbitrary $r(x) \in B$. We show that a contract $\hat{f}(x)$ where $\hat{f}(x) \neq r(x) + c$ for some interval(s) on X must be suboptimal. We initially assume that $r(x)$ is strictly increasing on X . Now, since $\hat{f}(x) \neq r(x) + c$ for some interval(s) on X , there must exist constants x_1 and x_2 such that $\hat{f}(x) > r(x) + c$ or $\hat{f}(x) < r(x) + c$ for $x \in X' = [x_1, x_2] \subset X$. For convenience assume that $\hat{f}(x) = f^*(x)$ for $x \notin X'$ (the logic of the proof is the same if this condition does not hold). First let $\hat{f}(x) < r(x) + c$ on X' . This implies that $r(x) > \hat{f}(x) - c$ on X' and the investor would set $P(r) = 1$ for $r \in [r(x_1), r(x_2)]$. But in that case the manager would offer $r(x_1)$ for all $x \in X'$, which is inconsistent with $r(x)$ being strictly increasing. Now let $\hat{f}(x) > r(x) + c$ on X' . Then the investor would set $P(r) = 0$ for $r \in [r(x_1), r(x_2)]$, since $y > r$. For the manager to have incentives to follow $r(x)$ for $x \in [x_L, x_1]$ it follows immediately that $P(r) = 0$ for $r \in [r(x_L), r(x_1)]$. Now consider the interval $[x_2, x_H]$. Since $\hat{f}(x) = f^*(x)$ for $x \in [x_2, x_H]$, by the same argument as in the first part of the proof we must have that $P(x) = e^{r(x) - r(x_H)}$ for $x \in [x_2, x_H]$. Let us now compare the monitoring costs induced by $\hat{f}(x)$ with the monitoring costs induced by $f^*(x)$, assuming that $\hat{f}(x)$ induces $r^*(x)$. For $x \in [x_2, x_H]$, the accept probability is the

same for every x , and the expected monitoring cost of $\hat{f}(x)$ and $f^*(x)$ on $[x_2, x_H]$ must be the same. For $x \in [x_L, x_2]$, however, the monitoring costs induced by $\hat{f}(x)$ must be strictly higher than the monitoring costs by $f^*(x)$, since $f^*(x)$ induces investor lenience while under $\hat{f}(x)$ the investor monitors with probability 1 for $x \in [x_L, x_2]$. It follows immediately that $f^*(x)$ dominates $\hat{f}(x)$, and consequently $f^*(x)$ is the optimal contract to induce $r(x)$. We initially made the assumption that $\hat{f}(x) = f^*(x)$ for $x \notin X'$. The proof in the case where $\hat{f}(x) \neq f^*(x)$ on more than one interval is a simple extension and omitted. We also initially made the assumption that $r(x)$ is strictly increasing on X . The proof of the case where $r(x)$ is flat in the upper region is also a simple extension, and omitted. ■

We have shown that $f^*(x) = r(x) + c$ is the optimal contract to induce a $r(x) \in B$ that satisfies the feasibility condition in Lemma 1. Equipped with Lemma 1 and Lemma 2 we can replace Problem 1' with an equivalent and more manageable Problem 1''. Let us denote by B' the set of payment functions in B that satisfies the condition in Lemma 1.

Problem 1'' (10)

$$\begin{aligned} \text{Max}_{\langle r(\cdot) \rangle} & \int e^{r(x) - r(x_H)} dH \\ \text{s.t.} & E\pi_I = I \\ & r(x) \in B' \\ & f(x) \in F \end{aligned}$$

Strategies and beliefs are PBE

To obtain Problem 1'' from Problem 1', we have substituted in $r(x) \in B'$ for $r(x) \in B$ by Lemma 1, and $P(r) = e^{\frac{r - r(x_H)}{c}}$ by Lemma 2. Moreover, since Lemma 2 enables us to map $r(x)$ into $P(r)$, observe that we now maximize over only $r(x)$ instead of over $\langle r(\cdot), P(\cdot) \rangle$ as in Problem 1'. Informally speaking, Problem 1'' is the problem depicted in Figure 1. Now define $D = [x_L, m + c)$ and $E = [m + c, x_H]$. Clearly $D \cup E = X$ and $D \cap E = \emptyset$.

Lemma 3 *The solution to Problem 1'' is $r^*(x)$, where*

$$r^*(x) = \begin{cases} x - c & \text{for } x \in D \\ m & \text{for } x \in E \end{cases} \quad (11)$$

Proof. $r^*(x)$ follows the feasibility barrier $r(x) \leq x - c$ and then becomes flat for $x = m + c$ (as depicted in Figure 1). To prove (i), let $\hat{r}(x) \in B'$ be an arbitrary alternative payment function in B' that raises the same amount as r^* , i.e., $\int_X \hat{r}(x)dH = \int_X r^*(x)dH$ (recall that one example of an $\hat{r}(x)$ is $r_1(x)$ depicted in Figure 1). Recall that for an arbitrary $\hat{r}(x)$, its expected monitoring cost equals $\int_X c[1 - P(\hat{r}(x))]dH$. Let \hat{V} be the expected monitoring cost of $\hat{r}(x)$ and V^* be the expected monitoring cost of $r^*(x)$. To show that $r^*(x)$ solves Problem 1'' is equivalent to showing that $\hat{V} > V^*$. In the following we show that $\hat{V} \geq V^*$. To extend the proof to holding for strict inequality is straightforward and omitted. Denote the expected monitoring cost of $r^*(x)$ on D (E) for V_D^* (V) and the expected monitoring cost of $\hat{r}(x)$ on D (E) for \hat{V}_D (\hat{V}_E). By definition, $\hat{V}_D + \hat{V}_E = \hat{V}$ and $V_D^* + V_E^* = V^*$. $r^*(x) = r(x_H)$ for $x \in E$ implies $V_E^* = 0$ and therefore $V_E^* \leq \hat{V}_E$. It therefore suffices to prove that $V_D^* \leq \hat{V}_D$. Since $r^*(x) = x - c$ for $x \in D$, Lemma 1 implies $r^*(x) \geq \hat{r}(x)$ for $x \in D$. Recall from Lemma 2 that for an arbitrary $\hat{r}(x)$ we have $P(\hat{r}(x)) = e^{\frac{\hat{r}(x) - \hat{r}(x_H)}{c}}$. To show that $V_D^* \leq \hat{V}_D$ it is therefore sufficient to show that $\hat{r}(x_H) \geq r^*(x_H)$. Since $r^*(x) \geq \hat{r}(x)$ for $x \in D$, we have that $\int_D r^*(x)dH \geq \int_D \hat{r}(x)dH$. Therefore $\int_E \hat{r}(x)dH \geq \int_E r^*(x)dH$ must hold for the investor to be indifferent between $r^*(x)$ and $\hat{r}(x)$. But since $r^*(x) = 0$ for $x \in E$, there must exist a constant $\tilde{x} \in E$ such that $\hat{r}(x) \geq (\leq) r^*(x)$ for $x > (<) \tilde{x}$. Therefore $\hat{r}(x_H) \geq r^*(x_H)$. Note finally that by adjusting m we can satisfy any feasible investor participation constraint (it is easy to show that $r^*(x)$ maximizes the range of fundable projects). That completes the proof. ■

We have shown that $r^*(x)$ solves Problem 1'' and now show that $f^D(x)$ is the optimal contract inducing $r^*(x)$.

Lemma 4 *$f^D(x)$ induces $r^*(x)$ and is the cheapest way to induce it.*

Proof. That $f^D(x)$ induces $r^*(x)$ follows from Lemma 2, part (i). That $f^D(x)$ is optimal in inducing $r^*(x)$ follows from Lemma 2, part (iii). ■

We have showed that $r^*(x)$ is optimal in B and that a debt contract is optimal in inducing $r^*(x)$. It follows directly that the manager defaults strategically under the optimal contract since for $x \in [x_L, m + c]$ the contractual obligation is x , while the actual payment offer equals $x - c$. Also, the creditor is lenient towards defaults, since his monitoring probability $1 - P$ is less than unity for any equilibrium path default. We have therefore proven Theorem 1 under the limitation $r(x) \in B$. That payment functions not in B cannot be optimal is intuitively straightforward but formally quite tedious, and relegated to the appendix.

Lemma 5 *Any $r(x) \notin B$ must be dominated by $r^*(x)$.*

We have then proved Theorem 1. Let us consider an example to highlight the economic behavior induced by the optimal contract. All qualitative features of the example hold generally.

Example 1 *Let $c = 1$ and x be uniformly distributed on $[1, 2]$.*

The contract is $f^D(x) = \min(x, d)$ which implies that the manager plays $r(x) = \min(x - c, d)$. The manager defaults for $x \in [x_L, d + c)$, a purely strategic default for $x \in [d, d + c)$ and partly strategic, partly liquidity-based for $x \in [x_L, d)$. The creditor monitors according to $P(r) = e^{\frac{r - r(x_H)}{c}}$; the higher payment, the lower probability for monitoring, and the higher maximal amount, the more monitoring. Since $r(x_H) - r$ can be interpreted as a writedown proposal by the manager, the accept probability is decreasing and concave in the magnitude of the writedown proposal.⁵ Substituting in for $r(x)$ we get $P(r) = \min(e^{\frac{x - c - d}{c}}, 1)$. The investor's participation constraint simplifies to $\int_{x_L}^{d+c} (x - c)dx + \int_{d+c}^{x_H} mdx = I$, which substituting in for $c = 1$ and solving gives $d = 1 - \sqrt{1 - 2I}$. We can note that d is (increasing and) convex in I .

The maximum fundable amount is obtained for $D = x_H - c = 2 - 1 = 1$, in which case the investor's payoff becomes $\int_X (x - c)dH = \int_1^2 (x - 1)dx = 1/2$. Hence in this

⁵The intuition for concavity is that when r is low then $P(\cdot)$ is low and the gains from cheating is small simply because the probability of getting away with it is low. On the other hand the loss from cheating is proportional in $P'(\cdot)$. The only way to induce adherence to $r(x)$ is therefore for the cheating deterrence to be stronger the higher r , or in other words for $P'(\cdot)$ to be higher for higher r .

example any $I \in [0, .5]$ is obtainable from the investor. The literature on bankruptcy costs (e.g., Andrade & Kaplan, 1998) finds that bankruptcy costs are about 10-30% of bankrupt firms' value. Interpreting c as bankruptcy costs, the example generates expected bankruptcy costs $E(c/x|e = 1)$ within these bounds for $I \in [0.15, 0.37]$.

We can calculate the gain in utility for the manager from defaulting strategically by playing $r^*(x)$ rather than adhering to the written contract by playing $r(x) = \min(x, d)$. For $x \in [x_L, d)$ the gain equals $(x - r)P(r(x)) = ce^{\frac{x-x_H}{c}}$, which increases in x . For $x \in [d, d+c)$ the gain equals $(d-r)P(r(x)) = (d-x+c)P(r(x)) = (d-x+c)e^{\frac{x-x_H}{c}}$ which decreases in x . Therefore the expected gain for the manager is concave and maximized for $x = d$. The economic implication is that under the optimal contract, the expected priority violation is maximized when the manager is closer to solvency, which is consistent with empirical evidence from Betker (1995).

Finally, as noted by Hvide & Leite (2005), the interest rate, $d/I - 1$, as spread over the riskfree rate increases in (i) the funding requirement I , and in (ii) the riskiness of the cash flow, under the optimal contract. An implication is that riskier firms on average face less lenient investors in default, as seen from the fact that a greater d implies a higher monitoring probability.

4 Discussion and conclusion

In this section we summarize and then discuss two crucial assumptions.

We have solved for optimal contracts and payment behavior in a modified Townsend setting where monitoring is non-contractible and stochastic monitoring is allowed. We found that the optimal contract is a debt contract. Under the optimal contract, the borrower strategically defaults on his payment obligation and the investor is lenient towards defaults. Our finding that debt is optimal adds to the literature exploring optimal contracts under repayment frictions in financial markets. Our findings that debt incurs strategic defaults and creditor leniency fit well with the corporate finance literature on repayment behavior of debt in real financial markets. In fact influential papers such as Anderson & Sundaresan (1996) and Mella-Barall & Perraudin (1997) argue that strategic defaults are a main reason for why observed risk premia on debt exceeds that implied by

the hugely influential Merton (1974) debt valuation model. We therefore believe that our solution match real-world contracts and repayment behavior fairly well.

We have made two main departures from Krasa & Villamil (2000); that renegotiation of contracts is not allowed (i.e., we do not require "time consistency") and that payment is continuous in the underlying cash flow. One way to defend our no-renegotiation assumption is that renegotiation would impose an ex-ante cost – through restricting the feasible utilities or simply through lawyer fees or delays in coming to agreement in the interim – that the parties impede in the initial contract. A different type of defense is the Segal & Whinston (2002) argument that which commitment assumptions to employ should be motivated by which assumptions produce the more realistic solution. Given the emphasis on strategic defaults and investor leniency in the corporate finance literature we feel that our assumptions fare well in this respect. The second departure from Krasa & Villamil (2000) is our requirement that the payment function is continuous, which allows for use of differentiation techniques.⁶ A simple "trembling" argument defends the continuity requirement. Suppose that the manager may tremble slightly when making his payment offer, so that the actual payment offer equals a distorted version of the equilibrium payment. This might be because of bounded rationality by the manager (or by the investor), because of rounding, or because of random mistakes by an intermediary. The difference to the payment game without trembles is that all payments occur with positive density in equilibrium. For sufficiently small trembles, $r^*(x)$ is still the optimal continuous payment function. More interestingly, there cannot exist an equilibrium with a discontinuous $r(x)$. To see why, suppose that $r(x)$ is continuous except in (countably) many points. Consider one of these points and label it $y \in X$. Then $r(y^-) > r(y^+)$, where "–" topscript denotes left limit and "+" topscript denotes right limit. Suppose that a payment offer $r(y^-) + \delta$ is observed, where δ is small relative to $r(y^+) - r(y^-)$. Then the investor will conjecture that the offer was made by a type on the interval $[x_L, y^-]$. But in that case, for sufficiently small trembles, the investor will strictly prefer to accept the offer $r(y^-) + \delta$. But then the offer $r(y^-)$ will never be made in equilibrium and an equilibrium with a discontinuous

⁶If we allow for discontinuous schemes, we would have an identical implementation problem to them, except for their added time consistency constraint. Obviously, then, we can obtain a higher entrepreneur payoff than Krasa-Villamil (2000).

payment function cannot exist.

5 References

Anderson, R. W. and S. Sundaresan (1996). Design and Valuation of Debt Contracts. *Review of Financial Studies*, 9, 37-68.

Andrade, G. & S. N. Kaplan (1998). How Costly is Financial (Not Economic) Distress? Evidence from Highly Leveraged Transactions that Became Distressed. *Journal of Finance*, 53, 1443-1493.

Bester, I. and R. Strausz (2001). Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case. *Econometrica*, 69, 1077-1099.

Betker, B. L. (1995). Management's Incentives, Equity's Bargaining Power, and Deviations from Absolute Priority in Chapter 11 Bankruptcies. *Journal of Business*, 68, 161-183.

Border, K. & J. Sobel (1987). Samurai Accountant: A Theory of Auditing and Plunder. *Review of Economic Studies*, 54, 525-40.

Boyd, J. H. and B. D. Smith (1994). How Good are Standard Debt Contracts? Stochastic versus Nonstochastic Monitoring in a Costly State Verification Environment. *Journal of Business*, 67, 539-561.

Brown, D. T. (1989). Claimholder Incentive Conflicts in Reorganizations: The Role of Bankruptcy Law. *Review of Financial Studies*, 2, 109-123.

Brown, D. T., B. A. Ciochetti, and T. J. Riddiough (2003). Theory and Evidence on the Resolution of Financial Distress. Mimeo, University of Florida, University of North Carolina, and University of Wisconsin. Available at www.ssrn.com.

Davydenko, S. A. (2005). When Do Firms Default? A Study of the Default Boundary, Mimeo, London Business School. Available at www.ssrn.com.

Gale, D. and M. Hellwig (1989). Repudiation and Renegotiation: The Case of Sovereign Debt. *International Economic Review*, 30, 3-31.

Hvide, H. K. & T. Leite (2005). Capital Structure under Costly Enforcement. Available at www.ssrn.com.

Krasa, S. and A. P. Villamil (2000). Optimal Contracts when Enforcement is a Decision Variable. *Econometrica*, 68, 119-134.

Martimort, D. and L. Stole (2002). Common Agency Equilibria with Discrete Mechanisms and Discrete Types. Mimeo, Chicago GSB. Available at <http://gsblas.uchicago.edu/papers/twotypes.html>.

Mella-Barral, P. and W. Perraudin (1997). Strategic Debt Service. *Journal of Finance*, 52, 531-556.

Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance*, 29, 449-70.

Mookherjee, D. and I. Png (1989). Optimal Auditing, Insurance, and Redistribution. *Quarterly Journal of Economics*, 104, 399-415.

Persons, J. C. (1997). Liars Never Prosper? How Management Misrepresentation Reduces Monitoring Costs. *Journal of Financial Intermediation*, 4, 269-306.

Segal, I. & M. D. Whinston (2002). The Mirrlees Approach to Mechanism Design with Renegotiation. *Econometrica*, 70, 1-45.

Townsend, R. M. (1979). Optimal Contracts and Competitive Markets with Costly State Verification. *Journal of Economic Theory*, 21, 265-293.

6 Appendix

Here we prove that any $r(x) \notin B$ must be dominated (Lemma 5). We split the proof into two parts. First we show that (i) $r(x) \notin B$ with $r(x) \leq x - c$ must be dominated, and then that (ii) $r(x) \notin B$ with $r(x) > x - c$ for some interval on X must be dominated.

To prove (i), denote a candidate payment function by $\hat{r}(x)$. First suppose that $\hat{r}(x)$ is weakly increasing. By the same construction as in Lemma 1, $\hat{r}(x)$ should be implemented by $f(x) = \hat{r}(x) + c$, and the only accept probability function consistent with $\hat{r}(x)$ being part of a PBE is $P(x) = e^{\frac{\hat{r}(x) - \hat{r}(x_H)}{c}}$. But then exactly the same dominance argument as in Lemma 1 shows that $r^*(x)$ dominates $\hat{r}(x)$. Suppose instead that $\hat{r}(x)$ is strictly decreasing on some interval(s). Again, by the same construction as in Lemma 1, the only accept probability function consistent with $\hat{r}(x)$ being part of a PBE is $e^{\frac{\hat{r}(x) - \hat{r}(x_u)}{c}}$, where $x_u = \arg \max_{\{x\}} \hat{r}(x)$. If $x_u = x_H$, the proof goes through by the same dominance

argument as before. Let us therefore suppose that $x_u < x_H$. We now construct an alternative payment function $\bar{r}(x)$ through modifying $\hat{r}(x)$ and show that $\bar{r}(x)$ constructed in a suitable manner dominates $\hat{r}(x)$. We assume for convenience that there exists x' so that $\bar{r}(x)$ reaches a local minimum for x' .⁷ We construct $\bar{r}(x)$ in two steps. In step 1, let $\bar{r}(x) = r(x_u) - \delta$ in an ϵ -neighborhood of x_u , labeled X_A . ϵ is small enough to guarantee that $\bar{r}(x)$ pays less than $\hat{r}(x)$ in X_A , and δ defined to ensure continuity of $\bar{r}(x)$ in the endpoints of X_A . In step 2, perform a similar modification of $\bar{r}(x)$ in a neighborhood of x_u , but now "shave" from below so that $\bar{r}(x)$ raises more than $\hat{r}(x)$. Formally, let $\bar{r}(x) = \hat{r}(x_u) + \psi$ in an ϵ -neighborhood of x_u labeled by X_B . ψ is defined to ensure continuity in the endpoints of X_B . Let now ϵ be such that the investor is indifferent between $\bar{r}(x)$ and $\hat{r}(x)$ (by the continuity of $\bar{r}(x)$ and $\hat{r}(x)$ such ϵ exists). Let us now compare the expected monitoring costs for $\bar{r}(x)$ and $\hat{r}(x)$. In X_A , the expected monitoring cost for $\bar{r}(x)$ is zero, while greater than zero for $\hat{r}(x)$. In X_B , the monitoring cost for $\bar{r}(x)$ is lower than for $\hat{r}(x)$, since the payment is higher for $\bar{r}(x)$. Finally, outside X_A and X_B , $\bar{r}(x)$ must also have a lower monitoring cost than $\hat{r}(x)$, since the maximal payout is higher for $\bar{r}(x)$ than for $\hat{r}(x)$. Hence $\bar{r}(x)$ beats $\hat{r}(x)$ and consequently any $\hat{r}(x)$ strictly decreasing on some interval(s) cannot be optimal.

We now need to show that (ii) $r^*(x)$ beats any $r(x) \notin B$ with $r(x) > x - c$ for some interval on X . Denote a candidate payment function of this type by $\hat{r}(x)$ and the set of such functions by \hat{R} , where $\hat{R} \subset R$. The optimal payment function in \hat{R} we denote by $\hat{r}^*(x)$. The strategy of the proof is to derive $\hat{r}^*(x)$ and then show that $r^*(x)$ beats $\hat{r}^*(x)$ by having lower monitoring costs. We first consider weakly increasing $\hat{r}(x)$ in steps 1-7.

Step 1. A weakly increasing $\hat{r}(x) \in \hat{R}$ must have a constant payout for $X' = [x_L, t]$, where t is some constant, since the same contradiction argument as eliminating $r_2(x)$ in Figure 1 would otherwise apply. It follows that to find $\hat{r}^*(x)$ we can restrict attention to

⁷If such a local minimum does not exist, $\hat{r}(x)$ must either be weakly increasing (in which case it is covered in the previous paragraph) or reach a local minimum for $x = x_H$. The proof extends readily to the latter case.

$\hat{r}(x)$ that are continuous approximations to $\hat{\rho}(x)$, where

$$\hat{\rho}(x) = \begin{cases} q & x \in [x_L, t] \\ x - c & \text{for } x \in [t, m + c] \\ m & x \in [m + c, x_H] \end{cases} \quad (12)$$

$\hat{\rho}(x)$ has constant payout q on X' , then follows $x - c$, and flattens at $x = m + c$.

Step 2. $\hat{r}^*(x)$ must induce the investor to monitor stochastically on X' : if it is strictly optimal for the investor to accept q then $\hat{r}(x)$ cannot be an equilibrium,⁸ and if it is strictly optimal for the investor to monitor with probability 1 then $\hat{r}(x)$ cannot be optimal. For the investor to monitor stochastically on X' we must have that,

$$\begin{aligned} \int_{X'} q dH &= \int_{X'} (f(x) - c) dH, \text{ which implies} \\ q(H(t) - H(x_L)) &= \int_{X'} f(x) dH - cH(t), \text{ which simplifies to} \\ q &= \int_{X'} f(x) dH / H(t) - c \end{aligned} \quad (13)$$

On the left hand side is what the investor gets if he accepts an offer q , and on the right hand side is what he expects to get if he monitors.

Step 3. For any choice of contract $f(x)$, equation (13) generates a function $q(t)$, where $q(x_L) = x_L - c$ by L'Hospitals rule. By a straightforward dominance argument, to find $\hat{r}^*(x)$ we want to pick the northernmost $q(t)$. This must arise from maximizing $\int_{X'} f(x) dH$ on X' with respect to $f(x)$, which is obtained by setting $f(x) = f^*(x) = x$ on X' .

Step 4. Substituting $f(x) = x$ back into (13), $\hat{r}^*(x)$ must satisfy

$$q = \int_{X'} x dH / H(t) - c \quad (14)$$

Note that $\int_{X'} x dH / H(t) = E(x|x \in X')$, where $E(x|x \in X')$ is the conditional mean of x

⁸Recall the assumption that the scheme $r(x) = q$ for $x \in X$ does not satisfy the investor's participation constraint. Therefore, candidate schemes with payout q for $x \in X'$ must have a higher payout for $x \notin X'$. But if $P(q) = 1$, the manager would offer q also for $x \notin X'$.

on X' . (14) implies that $E(x|x \in X') = q + c$, a fact that will be used in Step 7.

Step 5. Since $\int_{X'} x dH/H(t) - c = \int_{X'} (x - c) dH/H(t) = \int_{X'} r^*(x) dH/H(t)$, equation (14) implies that $\hat{r}^*(x)$ and $r^*(x)$ gives the same aggregate investor payoff on X' . It follows directly from (12) that $\hat{r}^*(x)$ and $r^*(x)$ must be identical on $x \in X'_c$ (where $X'_c = X/X'$) i.e., $\hat{r}^*(x) = r^*(x)$ for $x \in X'_c$. By Lemma 1, the accept probability and monitoring costs of $\hat{r}^*(x)$ and $r^*(x)$ are therefore also identical for $x \in X'$. To show that $r^*(x)$ beats $\hat{r}^*(x)$ it is therefore sufficient to show that $r^*(x)$ has a lower monitoring cost than $\hat{r}^*(x)$ on X' . This is equivalent to showing that $r^*(x)$ has a higher average accept probability than $\hat{r}^*(x)$.

Step 6. The average accept probability for $r^*(x)$ on X' equals $\int_{X'} P^*(x) dH/H(t)$, where $P^*(x) = e^{\frac{x-c-r^*(x_H)}{c}}$. Since $\hat{r}^*(x)$ has a constant payout on X' , its average accept probability simply equals $P^*(q) = e^{\frac{q-r^*(x_H)}{c}}$.⁹ We therefore need to show that

$$P^*(q) \leq \int_{X'} P^*(x) dH/H(t) \quad (15)$$

Step 7. We now show that (15) holds strictly, except in the non-generic case where it holds with equality. Note that the left hand side of (15) is unaffected by $h(x)$. It is therefore sufficient to show that (15) holds for the $h(x)$ that maximizes the right hand side subject to (13). Since both $r^*(x)$ and $\hat{r}^*(x)$ are linear in x , (13) will hold for any distribution that keeps $E(x|x \in X')$ constant, or in other words for any mean-preserving shift of $h(x)$ on X' . Since $H(t)$ is constant through mean-preserving shifts and $P^*(x)$ is convex in x , the right hand side of (13) is maximized by minimizing risk, i.e., putting an atom of the size $H(t)$ at the point $x = q + c$. Substituting into (15), we get the condition

$$e^{\frac{q-r^*(x_H)}{c}} \leq H(t) e^{\frac{q-r^*(x_H)}{c}} / H(t) = e^{\frac{q-r^*(x_H)}{c}} \quad (16)$$

which always holds. Hence we have shown that $r^*(x)$ dominates $\hat{r}^*(x)$ strictly except in the non-generic case where $h(x)$ is a non-generate distribution.

We finally need to eliminate $r(x)$ that have $r(x) > x - c$ on some interval *and* is

⁹This is where the requirement that $\hat{r}(x)$ is continuous bites. Allowing $\hat{r}(x)$ to be discontinuous in the point t would soften the incentive constraint of the manager, and decrease the monitoring probability for the offer q .

strictly decreasing on some (possibly different) interval. But this follows from the same type of argument as in part (i): for any such decreasing $r(x)$ we can construct an alternative payment function which pays more than $r(x)$ in the region where $r(x)$ is strictly decreasing and less in a region around the point where $r(x)$ is maximized, and show that this alternative payment function must dominate $r(x)$.