Theoretical Economics 14 (2019), 887–925 1555-7561/20190887

Disputes, debt, and equity

Alfred Duncan
Department of Eonomics, University of Kent

Charles Nolan
Department of Economics, University of Glasgow

We show how the prospect of disputes over firms' revenue reports promotes debt financing over equity. This is demonstrated in a costly state verification model with a risk-averse entrepreneur. The prospect of disputes encourages incentive contracts that limit penalties and avoid stochastic monitoring, even when the lender can commit to stochastic monitoring. Consequently, optimal contracts shift from equity toward standard debt. In short, when audit signals are weakly correlated with true incomes, standard debt contracts emerge as optimal; if audit signals are highly correlated with true incomes, optimal contracts resemble equity. When audit costs are sufficiently high, stochastic monitoring may be optimal. Optimal standard debt contracts under imperfect audits are shown to reproduce key empirical facts of U.S. firm borrowing.

Keywords. Microeconomics, costly state verification, external finance, leverage.

JEL classification. D52, D53, D82, D86.

Introduction and overview

What form should an optimal external finance contract take to handle the prospect of audit errors? Our answer is that standard debt contracts are often optimal. Thus, we propose a theory of debt and limited liability based on inaccurate auditing.

We conduct the analysis in a costly state verification environment that typically implies that equity contracts are optimal. However, that conclusion is shown to rest crucially on the efficacy of the audit technology. We introduce wrongful penalties through an imperfect audit technology although our results carry over to other situations where the lender or bankruptcy court might erroneously dispute the borrower's revenue report.[1]

In the model we present, a contract is an enforceable agreement between a risk-averse borrower and a risk-neutral lender covering the amount borrowed (leverage), an

Alfred Duncan: a.j.m.duncan@kent.ac.uk
Charles Nolan: Charles.Nolan@glasgow.ac.uk
The authors would like to thank Andrew Clausen, John Hardman Moore, Joel Sobel, Jonathan Thomas, Yiannis Vailakis, and conference and seminar participants for helpful comments and discussions. We are also very grateful for detailed and helpful comments from the three referees. This is a substantially revised version of an earlier paper of the same title. All errors are our own.

[1]That audits may be less than perfect in practice appears to be widely acknowledged and documented. See Bazerman et al. (2002).

Available at http://econtheory.org. https://doi.org/10.3982/TE2574

audit strategy dependent on the state of nature declared by the borrower, and payoffs to the parties given the declared state of nature and the findings of any audit. Perceived misreporting may be penalized and truth-telling may be rewarded. Whether the audit is *perfect*—capable without exception of identifying the true underlying state—can make all the difference to the form of the optimal contract.

The literature currently suggests that parties to an external finance contract ought to employ stochastic audits and levy rewards/penalties liberally. Stochastic audits save on resources, leaving more for consumption, and payments and penalties can be used both to reward truth-telling and to penalize misreporting. Moreover, as long as consumption is perfectly observed, sometimes ever-increasing rewards and penalties may be used to sustain truth-telling while economizing on costly audits—sometimes even if those audits are imperfect, as in Bolton (1987).[2] However, when audits are costly and income is private information (so that final consumption cannot be known with certainty by all parties), we show that penalties are necessarily limited. That being so, the usefulness of audits rises and rises decidedly so in low states. We explain these points now in a little more detail.

With perfect audits, the optimal contract usually employs stochastic audits with large penalties. That way, the parties to the contract conserve on the cost of audits while ensuring truth-telling; penalizing those who misreport and rewarding those who do not. The level of borrowing complements these decisions: higher leverage, other things constant, boosts the borrowers expected payoff but is accompanied by larger penalties for misreporting. The upshot is that the contract delivers substantial risk-sharing: That is, the borrower's consumption is relatively insensitive, and the lender's return is relatively sensitive, to reported income. Such a contract has key properties of equity finance.

Imperfect audits complicate things. Stochastic audits and large penalties run the risk of penalizing truth-telling agents when project returns are indeed low, while higher leverage, ceteris paribus, increases both the expectation and the variance of borrowers' consumption. Risk-averse borrowers at the margin fret more about the cost of having a truthful low report overturned than they welcome the prospect of an overturned high report. The optimal contract, therefore, has lower equilibrium borrowing relative to the perfect audits case and lower penalties.[3] Moreover, audits are employed only when the desire for risk-sharing is especially intense, i.e., in sufficiently bad states of the world. In other states, the borrower absorbs marginal income risk, avoiding costly, potentially erroneous audits. Such a contract—auditing and risk-sharing with the lender only in sufficiently bad states—is akin to a standard debt contract.

Imperfect audits, deterministic incentive regimes, and leverage

When audits are perfect but costly, there are limits on how much auditing is desirable. Below, Theorem 1, a benchmark result in the literature, shows that the application of

[2]We explain in more detail later why our result differs from Bolton (1987).

[3]Our model is static and our penalties are just units of the consumption good, but this reasoning holds under alternative settings where other enforcement schemes such as nonpecuniary penalties or exclusion may be applied. When disputes are possible, even honest borrowers prefer the penalties for dishonesty to be smaller, all else equal.

perfect audits with certainty is wasteful; resources are saved by constructing a noisy audit signal using a lottery. Hence, the entrepreneur's consumption is protected from random fluctuations in firm value in a way that resembles equity finance. Theorem 1 thus indicates why deterministic audit schemes *promote* the use of "imperfect" audits.

In the model with costly, imperfect audit technology, entrepreneurs may experience one of a number of out-turns for their project, and while stochastic monitoring may be optimal when audit costs are high, things are very different when they are sufficiently low. Stochastic auditing and a penalty may deter some, but not all, from misreporting. In other words, the deterrence of marginally fraudulent reports does not imply the deterrence of major fraudulent reports; the single-crossing property is absent. So while low-income agents might easily be deterred from misreporting, any decrease in audit probability might encourage high-income agents fraudulently to default.

Thus, with penalties limited, higher audit probability may be part of the response, and even as the probability of audit approaches 1, the marginal benefit of auditing can remain strictly positive. Increases in audit probability facilitate a given insurance plan with smaller penalties, reducing the costs of wrongful errors. So, in low states, where the marginal utility of the entrepreneur is high, the insurance benefits from auditing outweigh the cost. The optimal audit probability following low reports is 1, resembling a standard debt contract. That is the key new result proved in Theorem 2. Strictly speaking, Theorem 2 shows that debt is the globally optimal contract when audit cost is low enough. Later in the paper, optimal standard debt contracts under imperfect audits are shown to reproduce key empirical facts of U.S. firm borrowing.

So, introducing imperfect audits encourages deterministic audit regimes. Moreover, the interaction between leverage and costly, imperfect auditing helps underpin the finding that deterministic incentive schemes can be optimal. Note that leverage and audit probability are similar in that higher leverage increases the expectation and the variance of consumption, as noted above; so too does a decrease in audit probabilities. Audit costs and quality also play an important role in determining optimal leverage. When audit costs are low, optimal leverage is such as to permit large gains from insurance or auditing. This is what Gale and Hellwig (1985) also find in their seminal paper. When audit quality is low, the cost of enforcement increases quickly in leverage, restricting optimal leverage below the perfect audits case (Propositions 4 and 6).

1. Literature

Equity finance typically allows issuers to reduce repayments or dividends in bad times while reductions in the value of assets are shared between borrowers and lenders. Debt finance is more rigid. Debts are only reduced or discharged in bankruptcy, which follows large falls in income or asset values. So, surely it would be better if there was less debt and more equity?

Townsend (1979) was first to propose an explanation for the prevalence of debt contracts. He shows that when a risk-averse borrower's income is costly to verify, a standard debt contract is superior either to a strict debt contract, where repayments are constant across states, or a standard equity contract, where repayments are proportional to the borrower's income. The difficulty with the equity contract is that to ensure the borrower does not misreport income, the investor needs to undertake a costly audit regardless of the report. A superior contract prescribes audits and risk-sharing only following sufficiently low reports, when the borrower's marginal utility and sensitivity to risk are highest. If the borrower's income is sufficiently high, they make a fixed repayment and absorb any remaining income risk at the margin. Such a contract is the standard debt contract that is widespread in personal and corporate loan markets.

Townsend's analysis constrained agents to deterministic auditing regimes. However, he suggested a better contract might employ a stochastic auditing schedule (Townsend 1979, Section 4). This conjecture was confirmed by Border and Sobel (1987) and Mookherjee and Png (1989), who also showed agents should be rewarded following verified truthful reports. In sum, deterministic audits of low reports are unnecessary for contract enforcement, and stochastic reports of moderate and high reports are worthwhile when the costs of audit are low. The resulting optimal contracts resemble equity finance—repayments are contingent on marginal fluctuations in income even across relatively high states.

However, that risk-sharing comes at a cost that is not captured in the benchmark model. So as to ensure truth-telling when the probability of audit is low, audits that contradict the borrower's report can result in penalties far larger than the amount borrowed. If that audit technology were to contradict a *truthful* report, then the prospect of sizable wrongful penalties might render such contracts unacceptable to the borrower. Indeed, even if the entrepreneur were merely to *fear* that audits may not be perfect or that their truthful report may be disputed by the lender or bankruptcy court, they would likely balk at a contract that leaves open the prospect of large penalties following disputed reports. In short, equity-like contracts provide more insurance across states, but may exacerbate already bad situations for a borrower. Hence the motivation of this paper.

Above we noted the link between optimal leverage and audit costs. Gale and Hellwig (1985) also studied the effects of audit costs and risk aversion on leverage in a costly state verification model with perfect and deterministic audits. Our analysis permits stochastic audit regimes, and finds alternative interactions between leverage and the contracting environment: leverage has a dramatic impact on the nature of the efficient contract in our model, and it is the joint determination of leverage and incentive regime that encourages debt contracts in our framework. Finally, Bolton (1987) presents a principal–agent model where a risk-averse agent's effort is unobservable without undertaking an audit. He shows that even when audits are imperfect and the principal can commit, the maximal deterrence may be optimal. That finding would seem to suggest that the sorts of contracting results just mentioned (the desirability of stochastic audits with sizable penalties) is robust to the incorporation of imperfect audits and risk-averse agents. However, we find that Bolton's (1987) results do not go through and we explain below, in Section 5, why that is the case.

1.1 *Further features of optimal contracts*

As in Townsend's (1979) original analysis, our model motivates an endogenous form of limited liability.[4] Following default, the optimal repayment is determined by both reported income and the revealed audit signal. The severity of repayment following a disputed low report varies with the parameterization of the model: under some parameterizations, as in our numerical example, there is considerable loan relief even following disputed reports.

In our model, when the audit technology is relatively accurate, the costs associated with wrongful penalties and audit errors decline. Optimal contracts involve a high degree of risk sharing, with larger penalties and lower audit probabilities. Essentially, these contracts resemble equity even if the resource costs of audit are significant. In this sense, our framework nests both equity-like contracts and debt-like contracts as optimal under various configurations. In Section 6, we draw on Herranz et al. (2015) among others and show that under plausible parameterizations, the model can explain both debt and equity contracts. We show that audit quality is the prime determinant as to which type of contract is optimal. Moreover, the predicted debt contracts generate equilibrium relationships between interest rates, default probabilities, and leverage ratios that are broadly consistent with empirical estimates.

1.2 *Commitment*

Like most studies, this paper considers an environment where the lender is able to commit ex ante to an incentive regime that is wasteful ex post. That commitment may indicate a concern for reputation or delegation to a specialized auditor or bankruptcy court as in Melumad and Mookherjee (1989). Krasa and Villamil (2000) investigate what happens when lenders cannot commit to costly audits. That lack of commitment means the revelation principle does not hold; borrowers in equilibrium misreport their income with positive probability. It turns out that lack of commitment means that deterministic audits may be a feature of the optimal contract. Audits can occur only if the expected value of penalties levied following audits exceeds the audit costs. If true for a particular reported income, then this report will be audited with certainty. In short, for Krasa and Villamil (2000), the ability to commit implies equity-like contracts are preferable, whereas for us it does not.

1.3 *Road map*

The rest of the paper is set out as follows. Section 2 lays out the model environment and the nature of the auditing technology. Section 3 establishes key features of efficient contracts. In Section 4, we present the perfect audits benchmark (and Theorem 1). Section 5 explores the imperfect audits case and establishes that debt contracts may be globally

[4]In applications of Townsend's framework with risk neutrality, including Gale and Hellwig (1985) and Bernanke et al. (1999), liability is limited only by the inability to pay; the lender simply takes everything upon default.

optimal (Theorem 2). Section 6 provides numerical analysis of the general model, showing that debt is the globally optimal contract when audit quality is low enough. Moreover, the model generates empirically plausible equilibrium relationships between interest spreads, default probabilities, and leverage. Section 7 offers concluding remarks. Appendix A contains formal arguments and proofs. Appendix B contains a figure that illustrates the results of the numerical example described in Section 6.

2. The environment

We study a one period problem of a risk-averse and credit constrained entrepreneur. The assumption of risk aversion is supported by empirical evidence. The most relevant study for our analysis is Herranz et al. (2015), who estimate a median coefficient of relative risk aversion of 1.5 for small business owners in the U.S. Federal Reserve Survey of Small Business Finances. It is important to note that the assumption of risk aversion for entrepreneurs does not directly translate into risk-averse firm decision making. The entrepreneur's exposure to and appetite for the firm's risk taking is dependent on the risk-sharing capacity of the firm's capital structure.

The entrepreneur has access to a special technology offering high returns that are uncorrelated with other projects undertaken in the economy. The outcome of the project is initially private information to the entrepreneur, limiting the sharing of risk between the entrepreneur and a financial intermediary. Contract repayments are enforceable, but can only be conditioned on public information. The public information available to condition contracts includes any message sent by the entrepreneur, m, and any audit signal produced by the audit technology, σ.[5] The entrepreneur makes a take-it-or-leave-it contract offer to the financial intermediary, who is well diversified and operating in a perfectly competitive market. An optimal contract maximizes the entrepreneur's expected utility.[6]

2.1 *The entrepreneur*

The entrepreneur enjoys consumption at the end of the period according to $\mathcal{U}(x) : X \to \mathbb{R}$, where $\mathcal{U}'$, $-\mathcal{U}'' > 0$ and X is a closed, right unbounded interval of real numbers.[7] The entrepreneur brings wealth α into the period. Combining the entrepreneur's wealth α with the net funds transferred from the financial intermediary b, the project produces the consumption good according to stochastic gross return $(\alpha + b)\theta$. The revenue shock θ is drawn from a discrete distribution, $\theta \in \Theta$, where Θ is a set of possible, distinct values of the shock θ occurring with nonzero probability. By convention, we order the values of Θ as $\{\theta_1, \theta_2, \dots, \theta_n\}$, where $\theta_i > \theta_j$ if and only if and only if $i > j$. The unconditional

[5]In what follows lowercase Greek letters indicate exogenous variables and parameters, lowercase Latin letters denote endogenous or decision variables, and uppercase calligraphic letters indicate mappings or functions.

[6]An alternative formulation would be to maximize the profits of the financial intermediary subject to satisfaction of some participation constraint of the entrepreneur. This distinction has no effect on the main results of this paper.

[7]That is, either $X = [\underline{X}, +\infty)$ for some $\underline{X} \in \mathbb{R}$ or $X = (-\infty, +\infty)$.

probability of revenue draw θ_i is denoted $\Delta(\theta_i)$, with $\sum_{i=1}^{n} \Delta(\theta_i) = 1$. By construction, $\Delta(\theta_i) \in (0, 1]\ \forall i \in \{1, 2, \dots, n\}$. The operator $\Delta(\cdot)$ is used throughout this paper to generate probability measures over its arguments.

The entrepreneur can send a public signal indicating the realized state of their project. As we study direct truth-telling mechanisms, we can restrict the message space as follows: message m is drawn from $M = \{m_1, m_2, \dots, m_n\}$, where a message of m_i is interpreted as a declaration that the entrepreneur has received revenue shock θ_i. As the revenue shock θ_i is initially only observed by the entrepreneur, entrepreneurs may have an incentive to misreport (that is, to report m_i when the true revenue is θ_j for some $j \neq i$). Under any truth-telling mechanism, equilibrium messaging obeys the conditional probability distributions $\Delta(m_i|\theta_i) = 1\ \forall i \in \{1, 2, \dots, n\}$. The message m is modeled as a revelation of assets held by the entrepreneur. It is assumed that the entrepreneur can hide assets, but cannot hypothecate assets. Formally, agents can reveal message m_i if and only if their true revenue shock is greater than or equal to θ_i.[8]

2.2 *The financial intermediary*

There exists a well diversified financial intermediary who can make credible commitments to future actions.[9] Any contract involving the entrepreneur and the financial intermediary is small from the perspective of the financial intermediary's balance sheet. The entrepreneur's technology shock θ is uncorrelated with other shocks in the economy and the returns of other assets/liabilities of the financial intermediary's balance sheet. It follows that the financial intermediary is risk neutral with respect to claims contingent on the entrepreneur's individual specific technology shock θ.

Financial intermediaries operates in a perfectly competitive market. Their opportunity cost of funds is given by ρ; any contract offering an expected return on possibly state contingent loans exceeding ρ is acceptable. This condition is formalized below in Constraint 3. The opportunity cost of funds could be thought of as some combination of the interest rate paid by a risk-free bond, the interest rate paid by the intermediary to his/her deposit holders, and the intermediary's administrative costs.

The following two assumptions ensure that there are positive, finite gains from trade between the entrepreneur and the financial intermediary.

Assumption 1. Expected project returns exceed the financial intermediary's opportunity cost of funds, $\sum_{\theta \in \Theta} \Delta(\theta)\theta > \rho$.

Assumption 2. In the lowest state, project returns are lower than the financial intermediary's opportunity cost of funds, $\theta_1 < \rho$.

[8]The revelation of assets during the message reporting relaxes the constraint set of the optimal contract problem.

[9]Efficient contracts typically require commitment on behalf of the financial intermediary. One might think of this as sustained either through the intermediary's concern for his/her reputation or through delegation to a specialist bailiff or auditor as in Melumad and Mookherjee (1989).

Assumption 1 ensures that there are economic gains from diverting resources to the entrepreneur's project, even when the entrepreneur has access to a deposit facility at the bank that yields a risk-free return equal to the bank's opportunity cost of funds, ρ. Assumption 1 is strong enough to ensure that $b > -\alpha$. Assumption 2 specifies that the entrepreneurs' projects are risky. In bad states, a project yields lower returns than the risk-free asset. Assumption 2 is a necessary but not sufficient condition for the existence of finite leverage optimal contracts.

2.3 *Audits*

There exists an audit technology $\mathcal{T}(\kappa, \mathcal{S}(\theta))$ that is characterized by an audit cost parameter κ and a mapping $\mathcal{S}(\theta)$ from realized revenues θ to distributions of audit signals σ. The resource cost of an audit is the product $\kappa(\alpha + b)$; that is, audit costs are linearly increasing in the assets controlled by the entrepreneur. Following an audit, the audit technology produces a signal $\sigma \sim \mathcal{S}(\theta)$. This signal σ is assumed to be drawn from a discrete set of potential audit signals denoted by Σ. The action to undertake an audit is common knowledge and so is the signal provided, σ. The entrepreneur knows if (s)he has been audited and, if so, knows the result of the audit. The audit technology is exogenous: we do not allow agents to choose between competing technologies.[10]

The signal produced by the audit technology maps from the space of realized shocks θ as follows: if there is no audit, the audit signal is the empty set, $\sigma = \varnothing$; if there is an audit, the audit signal is drawn from the set Σ. The cardinality of the set of possible signals $|\Sigma|$ does not necessarily equal the cardinality of the set of possible revenue outturns $|\Theta|(= n)$. Also, for now, we do not require an ordering of the elements of the set of possible signals, Σ. The probability of revenue shock θ conditional upon audit signal σ is denoted $\Delta(\theta|\sigma)$.

An *audit strategy* is a mapping from messages to audit probabilities and is denoted $\mathcal{Q} : m \to [0, 1]$. Under truth-telling mechanisms, we can restrict our attention to reports $m \in \{m_1, m_2, \dots, m_n\}$ and we denote $q_i = \mathcal{Q}(m_i)$. It is assumed that an audit strategy can be agreed and committed to ex ante. Audit strategies are defined in contracts, and implemented ex post by the financial intermediary. The probability of the couplet (σ, θ) conditional on the probability of audit q is denoted $\Delta(\theta, \sigma|q)$.

The following definition specifies what is meant by the terms "perfect audit" and "imperfect audits."

Definition 1. (i) Imperfect Audits. An audit technology is *imperfect* if and only if there exists some couplet $(\theta_i, \sigma) \in \Theta \times \Sigma$ such that $\Delta(\theta_i, \sigma|1) > 0$ and $\Delta(\theta_i|\sigma) \in (0, 1)$.

(ii) Perfect Audits. An audit technology is *perfect* if and only if for all couplets $(\theta_i, \sigma) \in \Theta \times \Sigma$ the following statement holds: if $\Delta(\theta_i, \sigma|1) > 0$, then $\Delta(\theta_i|\sigma) \in \{0, 1\}$.

[10] This would be an interesting extension of the model. In particular, it may be the case that the optimal auditing technology used to enforce equity contracts differs from that used to enforce debt contracts. This may help us understand the coexistence of debt and equity finance issued by individual firms.

Audits are imperfect since in some states, following the audit, the signal can merely indicate how likely a particular revenue draw was. Note that under perfect audits, it is possible that multiple audit signals perfectly predict a single revenue shock. That is, it is possible that there exist two distinct signals $\sigma, \sigma' \in \Sigma$ with the property that for some θ_i, $\Delta(\theta_i, \sigma|1) > 0$ and $\Delta(\theta_i, \sigma'|1) > 0$, and $\Delta(\theta_i|\sigma) = \Delta(\theta_i|\sigma') = 1$. In this example, the revenue shock θ_i does not predict a unique audit signal with certainty (by Bayes' law, $\Delta(\sigma|\theta_i) \in (0, 1)$). However, the audit signals σ, σ' do predict a specific revenue shock θ_i with certainty. It is the latter that matters for contract enforcement.

3. Contracts

We have already described two elements of financial contracts: b is the amount of real resources transferred from the financial intermediaries to entrepreneurs at the beginning of the period to invest in the project; $\mathcal{Q}$ is the audit strategy that specifies the probabilities that audits will occur conditional upon messages m announced by the entrepreneur following the private realization of the revenue shock θ.

The third element of a financial contract is the *repayment function*. The repayment function maps message and audit signal pairs to real transfers of resources from the entrepreneur to the financial intermediary at the end of the period, $\mathcal{Z}: M \times \sigma \to \mathbb{R}$. This repayment does not need to be strictly positive.[11] Under truth-telling direct mechanisms, we denote $\mathcal{Z}(m_i, \sigma)$ by $z_i(\sigma)$. The fourth element of the financial contract is the consumption allocation function. The consumption function maps the revenue state, the message, and the audit signal to final consumption of the entrepreneur $\mathcal{X}: \Theta \times M \times \sigma \to \mathbb{R}$. The consumption allocation function is denoted by $\mathcal{X}(\theta, m, \sigma)$, where the ordering of the tuple (θ, m, σ) replicates the timing of the model; first the entrepreneur receives revenue shock θ, then reports message m, before the audit signal σ is revealed. We typically focus our analysis on audit strategies $\mathcal{Q}$, borrowing b, and repayment allocations $\mathcal{Z}$; the consumption allocations $\mathcal{X}$ are uniquely determined by the repayment allocations and borrowing by the budget constraint, which is (1) below.

Definition 2. A *contract* is a tuple $\mathcal{C} = (b, \mathcal{Q}, \mathcal{Z}, \mathcal{X})$ that is agreed at time zero and is common knowledge. A contract is a combination of an amount of resources transferred from the financial intermediary to the entrepreneur for investment, b, an audit strategy, $\mathcal{Q}$, a repayment function $\mathcal{Z}$, and a consumption allocation function $\mathcal{X}$.

The motivation for this paper is the search for environments where optimal contracts resemble standard debt contracts, which we define as follows.

Definition 3. We specify the following two benchmark classes of debt contracts.

(a) A *noncontingent debt* contract is a contract with constant repayments across all states and messages $z_i(\sigma) = z_j(\sigma') \; \forall m_i, m_j \in M, \sigma, \sigma' \in \Sigma$.

[11] As pointed out by Bolton (1986), restricting repayments to be nonnegative would not affect the optimal allocations, subject to an adjustment to the model to allow for the entrepreneur to hold collateral in the form of risk-free deposits.

(b) A *standard debt* contract has the following two properties: (i) the contract specifies a constant repayment when either the entrepreneur's message is equal to or above some threshold m_k,

$$z_i(\sigma) = z_j(\sigma') \quad \forall m_i, m_j \in \{m_k, m_{k+1}, \dots, m_n\}, \sigma, \sigma' \in \Sigma;$$

(ii) reports below the threshold are audited, $q_i = 1\ \forall i < k$.

We refer to the reporting of a message m_i where $i < k$ as *default*.

Note that debt contracts in our model do not restrict the entrepreneur to zero consumption following default. In fact, in the examples that we consider, entrepreneurs enjoy strictly positive consumption in all circumstances, even following a default. This positive consumption could represent income already paid to the entrepreneur during the life of the project.

3.1 *Constraints*

Contracts in our framework are subject to four classes of constraints. The first class of constraints is budget constraints.

Constraint 1 (Budget). *State contingent budget constraints are specified as*

$$\mathcal{X}(\theta_i, m_j, \sigma) = (\alpha + b)\theta_i - \mathcal{Z}(m_j, \sigma) \quad \forall (m_i, \sigma, \theta_j) \in M \times \Sigma \times \Theta. \tag{1}$$

Constraint 2 (Audit probability). *Audit probability constraints are specified as*

$$\mathcal{Q}(m) \geq 0 \quad \forall m \in M \tag{2}$$

$$1 - \mathcal{Q}(m) \geq 0 \quad \forall m \in M. \tag{3}$$

Constraint 3 (Financial intermediary's participation). *The participation constraint is specified as*

$$\sum_{m \in M} \Delta_f(m) \sum_{\sigma} \Delta_f\big(\sigma | m, \mathcal{Q}(m)\big) \mathcal{Z}(m, \sigma) \geq b\rho + \sum_{m \in M} \Delta_f(m) \mathcal{Q}(m)(\alpha + b)\kappa, \tag{4}$$

where the probability measures $\Delta_f(\cdot)$ are constructed from the financial intermediary's information set.

Constraint 4 (Incentive compatibility). *Contracts are referred to as* incentive compatible *if and only if the following constraints hold:*

$$m_i \in \arg\max_{m} \sum_{\sigma} \Delta\big(\theta_i, \sigma | \mathcal{Q}(m)\big) \mathcal{U}(\theta_i, m, \sigma) \quad \forall i \in \{1, 2, \dots, n\}. \tag{5}$$

We typically refer to (1) as $\mathrm{BC}_{i,j,\sigma}$. Equation (1) shows the revenue received by the entrepreneur from his/her project. Following the repayment $\mathcal{Z}(m, \sigma)$, the remainder available for the entrepreneur to consume is $\mathcal{X}(\theta, m, \sigma)$. We frequently use the shorthand notation $\mathcal{U}(\theta, m, \sigma) := U(\mathcal{X}(\theta, m, \sigma))$ and $\mathcal{U}'(\theta, m, \sigma) := U'(\mathcal{X}(\theta, m, \sigma))$.

The second class of constraints is the set of bounds on audit probabilities.

The left hand side of (4) captures the expected repayment constructed from the financial intermediary's information set Ω_f. The right hand side reflects the opportunity cost of the intermediary's funds, ρ, and probable audit costs. Under any contract, the financial intermediary must forecast the entrepreneurs messaging strategy to form an expectation of repayment. The revelation principle holds in our setting. This means that there exists an optimal contract under which the entrepreneur weakly prefers truthfully to reveal his/her true θ in all states. We refer to contracts that induce truth-telling as *truth-telling contracts*. Under any truth-telling contract, $\Delta(m_i|\theta_i) = 1\ \forall i \in \{1, 2, \ldots, n\}$. Therefore, under truth-telling we can rewrite Constraint 3 as

$$\sum_{\sigma} \Delta(\theta_i, \sigma|q_i) z_i(\sigma) - b\rho - \sum_{i=1}^{n} \Delta(\theta_i) q_i (\alpha + b)\kappa \geq 0. \tag{6}$$

We typically refer to (6) as PC.

It is useful to deconstruct the Incentive Compatibility constraint into a set of constraints that compare individual pairs of reports. That is, a contract is incentive compatible if for all state pairs (θ_i, θ_j), an entrepreneur who receives true return θ_i weakly prefers to report m_i over m_j. This pairwise formulation of the Incentive Compatibility constraint is equivalent to (5) and is formalized by

$$\begin{aligned}&\sum_{\sigma} \Delta(\theta_i, \sigma|q_i)\mathcal{U}(\theta_i, m_i, \sigma) - \sum_{\sigma} \Delta(\theta_i, \sigma|q_j)\mathcal{U}(\theta_i, m_j, \sigma) \geq 0 \\ &\forall i \in \{1, 2, \ldots, n\}, j < i.\end{aligned} \tag{7}$$

We typically refer to (7) as $\text{ICC}_{i,j}$ for truthful report m_i and misreport m_j. There are two challenges to enforcement in the model under imperfect audits. First, penalties may be wrongfully applied to truth-telling agents. Second, even when the audit signal has identified a fraudulent report, the financial intermediary may remain uncertain of the true income of the misreporting agent, and may therefore be unable to impose a maximal penalty. So we ensure that all penalty repayment allocations are payable by any agent who may be charged this allocation. In other words, for Constraint 4 to be well defined, it must be the case that consumption allocations both on and off the equilibrium path are in the domain of the utility function:

$$\mathcal{X}(\theta_i, m_j, \sigma) \in X \quad \forall \big\{(i, j, \sigma)|\Delta(\theta_i, \sigma|q_j) > 0 \wedge j \leq i\big\}.$$

Of course, this becomes redundant when the domain of the utility function is unbounded below (that is, when $\text{Dom}(\mathcal{U}) = \mathbb{R}$), in which case any agent is able to make any repayment regardless of assets.

Before continuing, we define the concept of a feasible contract, and the active set of constraints. We use this below in the analysis of the optimization problem. A contract $\mathcal{C}$ is *feasible* if and only if the constraints $\text{BC}_{i,j,\sigma}$, PC, and $\text{ICC}_{i,j}$ are satisfied for all i, j, σ.

Let $\mathcal{C}$ be a feasible contract. The *active set* of constraints, denoted $\mathcal{A}(\mathcal{C})$, is the set of binding constraints,

$$\begin{aligned} \mathrm{BC}_{i,j,\sigma} \in \mathcal{A}(\mathcal{C}) & & \\ \mathrm{PC} \in \mathcal{A}(\mathcal{C}) & \iff & \mathrm{PC} = 0 \\ \mathrm{ICC}_{i,j} \in \mathcal{A}(\mathcal{C}) & \iff & \mathrm{ICC}_{i,j} = 0. \end{aligned}$$

3.2 *Optimal contracts*

Optimal contracts are formalized by Programme 1. We assume that the optimal contract maximizes the entrepreneurs' utility conditional on participation of the financial intermediary and truth-telling. Note that there may be welfare maximizing contracts that are not classed as optimal under this definition, because they do not induce truth-telling as a dominant reporting strategy. By the revelation principle, we know that there exist welfare maximizing contracts that do induce truth-telling, and our definition of optimality restricts our attention to these truth-telling contracts.

Programme 1. *A contract is* optimal *if and only if it maximixes the entrepreneur's utility subject to feasibility*

$$\max_{\mathcal{C}} \sum_{i,\sigma} \Delta(\theta_i, \sigma | q_i) \mathcal{U}(\theta_i, m_i, \sigma)$$

subject to (1), (2), (3), (6), and (7).

The first order conditions of Programme 1 do not always provide sufficient, or even necessary, conditions for optimal contracts. Sufficiency is broken as a result of the nonconvexity of the constraint set. The necessity of the first order conditions of Programme 1 is also not guaranteed; when there exists some i such that $q_i \in \{0, 1\}$, there can exist optimal contracts that do not satisfy the first order condition for audit probability q_i. These results are formalized as follows.

Remark 1. The first order conditions of Programme 1 specify neither sufficient nor necessary conditions for optimal contracts.

The proof of Remark 1 along with the proofs of all other results are contained in Appendix A. Constraints 1 and 2 are affine in the choice variables $\mathcal{Z}$, $\mathcal{Q}$, $\mathcal{X}$, and b. Constraints 3 and 4 are nonconvex.[12] Focussing on the subproblem of determining utility

[12]Furthermore, in contrast with Grossman and Hart (1983) and Bolton (1987), rewriting the problem in terms of utility allocations does not convexify the subproblem of optimizing repayment allocations and borrowing conditional upon a given audit strategy. To see this, consider the substitution $\mathcal{X}(\theta, m, \sigma) = \xi(\mathcal{U}(\theta, m, \sigma))$, where $\xi = \mathcal{U}^{-1}(\mathcal{X}(\theta, m, \sigma))$. The function ξ is convex, and after substitution appears in the budget constraints (1), which become nonconvex. In Programme 1, the budget constraints are equality constraints: they both constrain the consumption of truth-telling agents from above, and constrain the punishments imposed on misreporting agents from below. The introduction of any nonlinearity into the budget constraints breaks the convexity of these constraints.

allocations and the initial transfer of resources to the entrepreneur, we can derive necessary optimality conditions. These optimality conditions are presented next.

Proposition 1. *Let $\mathcal{C}$ be a globally optimal contract that satisfies the following necessary conditions.*

Repayments $\mathcal{Z}(m_i, \sigma)$ satisfy

$$\frac{\mathcal{U}'(\theta_i, m_i, \sigma)}{\lambda} = \frac{1}{1+\sum_k \mu_{i,k}} \left[1 + \sum_j \mu_{j,i} \frac{\Delta(\theta_j|\sigma)}{\Delta(\theta_i|\sigma)} \frac{\mathcal{U}'(\theta_j, m_i, \sigma)}{\lambda}\right] \tag{8}$$

for all (m_i, σ). The initial transfer of resources to the entrepreneur b satisfies

$$\sum_i \Delta(\theta_i)\big[\theta_i - q(\theta_i)\kappa\big] - \rho = \sum_{i,\sigma} \Delta(\theta_i, \sigma|q_i)\left[\sum_j \mu_{j,i} \frac{\Delta(\theta_j|\sigma)}{\Delta(\theta_i|\sigma)} \frac{\mathcal{U}'(\theta_j, m_i, \sigma)}{\lambda}(\theta_j - \theta_i)\right] \tag{9}$$

for all $i \in \{1, 2, \dots, n\}$ and

$$\lambda > 0, \quad \mu_{j,i} \geq 0\, \forall i, j, k.$$

Equation (8) presents the optimality condition for repayment $z_i(\sigma)$. The left hand side is the ratio of the marginal utility of consumption in state (θ_i, σ) to the shadow cost of the participation constraint, λ. This ratio is a marginal rate of substitution from expected consumption marginal utility to realized consumption marginal utility and is equal to 1 under perfect information. This marginal rate of substitution is decreasing in $\sum_j \mu_{i,j}$, the sum of shadow costs of binding incentive constraints in which the repayment $z_i(\sigma)$ enters in the first term of the constraint (7). In other words, this term captures the cost of ensuring that an agent who earns θ_i reports m_i truthfully. All else equal, this cost of ensuring truthful reporting is high when the marginal utility of consumption is low relative to expected marginal utility.

The term $\sum_j \mu_{j,i} \frac{\Delta(\theta_j|\sigma)}{\Delta(\theta_i|\sigma)} \frac{\mathcal{U}'(\theta_j, m_i, \sigma)}{\lambda}$ captures the cost of ensuring that agents who earn θ_j do not report m_i. The sum $\sum_j \mu_{j,i}$ is not dependent on σ and captures the shadow costs of binding incentive constraints in which repayments $z_i(\cdot)$ enter in the second term of the constraint (7). Conditional upon the signal σ, these incentive costs are increasing in the marginal likelihood ratio $\frac{\Delta(\theta_j|\sigma)}{\Delta(\theta_i|\sigma)}$. When this marginal likelihood ratio is close to unity, the signal σ leaves the lender unable to detect misreporting by agents who earn θ_j and report m_i. The final term, $\frac{\mathcal{U}'(\theta_j, m_i, \sigma)}{\lambda}$, captures the marginal increase in utility of a misreporting agent who receives θ_j and reports m_i, conditional upon the audit signal σ. This marginal utility is normalized by the shadow cost of the participation constraint, λ. All else equal, when this marginal utility is low, a marginal increase in $z_i(\sigma)$ has a small increase in the value of misreporting m_i for an agent who receives θ_i and, therefore, has a small impact on incentive compatibility.

Equation (9) captures the optimality condition for the initial amount of resources transferred from the financial intermediary to the entrepreneur, b. The left hand side captures the net contribution of a marginal increase in b to expected consumption.

The first term $\sum_i \Delta(\theta_i)\theta_i$ is the expected per unit increase in revenue. The second term $\sum_i \Delta(\theta_i)q(\theta_i)\kappa$ is the expected increase in audit costs and ρ is the financial intermediary's opportunity cost of funds. The right hand side captures the incentive costs of an increase in b. The term $\sum_{i,\sigma} \Delta(\theta_i, \sigma|q_i)$ takes the expectation over states (θ_i, σ), conditional upon the audit strategy $\mathcal{Q}(m)$. The sum $\sum_j \mu_{j,i}$ captures the shadow costs of the binding incentive constraints relating to entrepreneurs receiving revenue θ_j who weakly prefer reporting m_j over m_i. These shadow costs are scaled by three factors: First, the marginal likelihood ratio $\frac{\Delta(\theta_j|\sigma)}{\Delta(\theta_i|\sigma)}$ captures the ability of the lender to identify misreporting agents who earn θ_j and report m_i; second, the term $\frac{\mathcal{U}'(\theta_j,m_i,\sigma)}{\lambda}$ captures the marginal increase in the value of an agent who receives θ_j and misreports m_i conditional upon audit signal σ; third, the term $(\theta_j - \theta_i)$ captures the rate at which a rise in borrowing increases the absolute risk across revenue states.

Corollary 1. *Let $\mathcal{C}$ be a globally optimal contract. The repayment terms of contract $\mathcal{C}$ satisfy the following conditions:*

(i) A repayment allocation $\mathcal{Z}(m_i, \sigma)$ is contingent on the audit signal σ only if the message m_i enters on the right hand side of a binding incentive constraint. That is,

$$\exists \sigma, \sigma' \text{ s.t. } \mathcal{Z}(m_i, \sigma) \neq \mathcal{Z}\big(m_i, \sigma'\big) \quad \Longrightarrow \quad \exists \theta_j \text{ s.t. } \mathrm{ICC}_{j,i} \in \mathcal{A}(\mathcal{C}).$$

(ii) Under nonincreasing absolute risk aversion, if the incentive constraint $\mathrm{ICC}_{i,j}$ is binding, then repayments are strictly increasing in the conditional likelihood ratio of θ_j with respect to θ_i. That is,

$$A'(x) \leq 0 \wedge \mathrm{ICC}_{j,i} \in \mathcal{A}(\mathcal{C}) \wedge \frac{\Delta(\theta_j|\sigma')}{\Delta(\theta_i|\sigma')} > \frac{\Delta(\theta_j|\sigma)}{\Delta(\theta_i|\sigma)} \quad \Longrightarrow \quad \mathcal{Z}\big(m_i, \sigma'\big) > \mathcal{Z}(m_i, \sigma),$$

where $A(x) := \frac{-\mathcal{U}''(x)}{\mathcal{U}'(x)}$ is the Arrow–Pratt measure of absolute risk aversion.

Corollary 1 formalizes two properties of optimal repayments. First, any repayment that appears only on the left hand side of binding incentive constraints—this includes the highest repayment θ_n—is not contingent on audit signals. Increases in the uncertainty of consumption following report i decrease expected utility for agents earning θ_i and also decrease the incentive for these agents to report income truthfully. These costs can be compensated for if the increased uncertainty deters false reports from agents with true income greater than i. But if all incentive constraints $\mathrm{ICC}_{j,i}$ are nonbinding, then there is no benefit attainable from uncertainty in repayments following report m_i. Second, if a repayment allocation is present on the right hand side of a binding incentive constraint, repayments are increasing in the conditional likelihood ratio of misreporting j with respect to truthfully reporting i. The conditional likelihood ratio is the marginal rate of transformation of consumption for misreporting agents with respect to truth-telling agents. When the conditional likelihood ratio is high, the disincentive effect of high repayments is large relative to the direct effect on expected utility.

Proposition 2. *Let $\mathcal{C} = (\mathcal{Q}, \mathcal{Z}, \mathcal{X}, b)$ be a globally optimal contract. Contract $\mathcal{C}$ has the following properties:*

(i) The financial intermediary's participation constraint is binding: PC $\in \mathcal{A}(\mathcal{C})$.

(ii) The highest possible report is never audited: $\mathcal{Q}(m_n) = 0$.

(iii) There exists a binding incentive constraint: $\exists i, j$ *s.t.* $\mathrm{ICC}_{i,j} \in \mathcal{A}(\mathcal{C})$.

Let the audit strategy $\mathcal{Q}^*(m)$ be taken as given and let the probabilities of all revenue states be positive conditional upon any audit signal $(\Delta(\theta_i|\sigma) > 0\ \forall \theta_i, \sigma)$.

(iv) Conditional upon $\mathcal{Q}^*$, under the optimal allocation of repayments $\mathcal{Z}$ and borrowing b, there exists a binding incentive constraint $\exists i, j$ *s.t.* $\mathrm{ICC}_{i,j} \in \mathcal{A}(\mathcal{C})$.

Proposition 2 characterizes restrictions on the active set of constraints under optimal contracts. Part (i) states that the financial intermediary's participation constraint is binding; the entrepreneur's utility can be increased if the financial intermediary's participation constraint is not binding. Part (ii) states that audits of the highest reports are always wasteful; this follows from Corollary 1(i). Part (iii) states that any optimal contract features a binding incentive constraint; any deviation from this would mean that costly audits are being undertaken wastefully. Part (iv) goes further. Under the assumption that any message–audit signal pair could be consistent with truth-telling, there exists a binding incentive constraint when repayments and leverage are chosen optimally, regardless of the given audit strategy. The condition $(\Delta(\theta_i|\sigma) > 0\ \forall \theta_i, \sigma)$ is a stronger condition than imperfect audits, and severely restricts the ability of the financial intermediary to detect misreporting agents.

4. Perfect audits

In the Introduction, we stated that the interaction between leverage and costly, *imperfect* audits underpins the optimality of deterministic contracts. Before establishing that and other results, it is insightful to analyze the case of *perfect* audits. Theorem 1 restates Mookherjee and Png's (1989) finding that debt contracts are not optimal. Moreover, we go on to show that optimal leverage is unbounded for reasonable parameter values.

Theorem 1 (Mookherjee and Png 1989, Proposition 1). *Under perfect audits* $(\delta(\theta_i|\sigma_k) \in \{0, 1\}\ \forall i, k)$, *any optimal contract without certain immiseration following truthful reports* $(\sum_\sigma \Delta(\theta_i, \sigma|q_i)\mathcal{U}(\theta_i, m_i, \sigma) > \underline{\mathcal{U}}\ \forall i)$ *cannot include certain auditing of any report. That is,* $\mathcal{Q}(\theta) < 1\ \forall \theta$.

We provide a proof of Theorem 1 similar to Mookherjee and Png (1989). However, we place a somewhat different interpretation on it that proves useful in the next section. The proof works by starting with a contract with certain auditing, $\mathcal{Q}(\theta_i) = 1$ for some i. We show that the marginal audit following report m_i can be replaced by a lottery that replicates the repayments following audits for truth-telling agents receiving true return

θ_i, while retaining incentive compatibility and relaxing the lender's participation constraint. Our proof shows the value of weak audit signals and provides some intuition over their role in promoting standard debt contracts. Under perfect audits, an agent forced to audit with certainty conditional upon a given report m_i would wish to reduce costs by weakening the quality of the audit signal, even via a naïve strategy of introducing a lottery between the true audit signal and an uninformative signal.

So by Theorem 1, one sees that for any fixed level of audit costs, optimal contracts under perfect audits do not apply certain audits following any report ($q_i^* < 1$). Proposition 3 explores how this result relates to audit costs. As audit costs fall, we would expect optimal audit probabilities to increase. Proposition 3 states that even as audit costs approach 0, optimal audit probabilities approach strictly interior values. It follows that certain audits and standard debt contracts cannot be explained by low audit costs.

Proposition 3. *Under perfect audits, as audit costs approach* 0*, optimal audit probabilities approach values strictly less than* 1*:* $\lim_{\kappa\to 0+} q_i^* \in [0, 1)$.

Proposition 3 shows that the cost of contract enforcement under perfect audits is low. Within costly state verification models, leverage acts as a substitute for reductions in auditing: an increase in leverage increases both the expected consumption and the consumption risk of agents. The following proposition shows the importance of this substitutability for optimal contracts.

Proposition 4. *If audits are perfect* ($\delta(\theta_i|\sigma_k) \in \{0, 1\}\ \forall i, k$) *and sufficiently inexpensive* ($\kappa < \mathbb{E}(\theta) - \rho$)*, and projects enjoy constant returns to scale, optimal external finance contracts do not exist; limiting external finance contracts approach infinite leverage and infinite entrepreneurial consumption.*

So parameter values are important for equilibrium out-turns and later we discuss this more fully. For now we note that estimates of the direct costs of auditing range between 0.01 and 0.06 of assets.[13] While it is not straightforward to estimate the marginal return on assets for entrepreneurs, we can obtain estimates of the average return on assets for entrepreneurs: Herranz et al. (2015) find a mean annual gross return on assets of 1.30 for small firms in the 1993 Survey of Small Business Finances, and a median annual gross return on assets of 1.09 for the same sample. They also estimate the lenders' annualized gross opportunity cost of funds at 1.012. Our model predicts that these values are not consistent with optimal leverage under perfect audits.

5. Imperfect audits

The previous section showed that under perfect audits, it is never optimal to audit a single report m_i with certainty (Theorem 1). In proving this result, we construct a welfare improving perturbation that bundles the resource-costly perfect audit signal with a

[13] See, for example, Warner and Gruber (1977), Weiss (1990), and Altman (1984).

lottery; this bundle can be thought of as a lower cost but imperfect audit technology. Under the perturbed contract, the new imperfect audit technology is applied with certainty following report m_i.

Indeed, under imperfect audits, certain auditing of an individual report can be optimal. Optimal contracts can even take the form of standard debt (Theorem 2). Even if the audit probability is high, any further increase in the audit probability does increase the set of feasible consumption allocations available to the entrepreneur. Under imperfect audits, increasing the probability of audit defrays the incentive costs of contract enforcement more widely. This increases risk-sharing across states.

Assumption 3. The probabilities of all revenue states are positive conditional upon any audit signal, $(\Delta(\theta_i|\sigma) > 0, \forall \theta_i, \sigma)$.

Within this section, we assume that, conditional upon any given audit signal, all revenue states are possible (Assumption 3). This assumption is sufficient to prove the existence of optimal contracts (Proposition 5), the existence of optimal debt contracts (Theorem 2), and the presence of an upper bound on leverage (Proposition 6).

Proposition 5. *Optimal contracts exist.*

In contrast to the case of perfect audits (Proposition 4), Proposition 5 implies interior optimal leverage and consumption allocations. Proposition 5 may be contrasted with the principal–agent model studied by Bolton (1987, Proposition 4). Bolton showed that in the principal–agent model, imperfect signals are not sufficient to deter maximal penalties following overturned reports: optimal contracts do not exist, as any contract can be improved upon by lowering consumption closer to zero following overturned reports. With hidden income, driving down consumption payoffs to induce truth-telling eventually loses some of its effectiveness compared to principal–agent models in which consumption allocations are specified directly. Hidden income cushions the marginal incentive effects of high contractual repayments.[14]

Theorem 2. *When κ is sufficiently small, optimal debt contracts are standard debt contracts.*

Theorem 2 presents a sufficient condition for standard debt to be optimal under imperfect audits. Standard debt contracts combine the certain auditing of low reported revenues with no audits of high reported revenues. Certain auditing of report m_j requires that the following incentive constraint adapted from (7) must be binding for some θ_i:

$$\sum_{\sigma} \Delta(\theta_i, \sigma|q_i)\mathcal{U}(\theta_i, m_i, \sigma) \geq \sum_{\sigma} \Delta(\theta_i, \sigma|q_j = 1)\mathcal{U}(\theta_i, m_j, \sigma). \tag{10}$$

[14] Within our proof, the equation that captures this distinction is (14). When there is hidden income $((\alpha + b^s)(\theta_j - \theta_i))$, (14) holds and the limit of the sequence is zero. When there is no hidden income, this limit may be nonzero.

Under perfect auditing, the signal σ reveals the true income θ_i with certainty. The sum on the right hand side of (10) becomes a single value $\mathcal{U}(\theta_i, m_j, \sigma = \theta_i)$ and this value can be set equal to or arbitrarily close to the lower support of the range of the utility function $\mathcal{U}$ without harming any truth-telling agent. This means that the incentive constraint will be slack for any allocation that does not immiserate with certainty agents who truthfully report m_i.[15]

Under imperfect auditing, the lottery on the right hand side of (10) remains, and repayment allocations $\mathcal{Z}(m_j, \sigma)$ may affect both truth-telling and misreporting agents. Utility allocations on the right hand side, $\mathcal{U}(\theta_i, m_j, \sigma)$, are bounded below by what is acceptable ex ante to entrepreneurs who know they may be punished for misreporting ex post. In addition, repayments are bounded by the inability of the audit technology to reveal the true income of misreporting agents, even when the audit signal can detect whether they have misreported their income.[16]

This explains how the marginal value of audits can remain positive as the probability of audit approaches 1. At the same time, lower quality audits increase the costs of enforcing risk-sharing contracts. We show in our proof contained in Appendix A and in the numerical simulations studied in Section 6 that optimal contracts under imperfect audits reduce the probability of audit to 0 across high revenue states, where the value of risk-sharing is low; optimal contracts increase the probability of audit to 1 across low revenue states, where the value of risk-sharing is high.

We saw in Section 4 that when audits are perfect and sufficiently inexpensive, the optimal leverage ratio is infinite. Consider (9), the first order necessary condition for borrowing b. The right hand side of the equation captures the incentive costs of increasing leverage. When audits are perfect, sufficiently high audit probabilities leave all incentive constraints slack and all Kuhn–Tucker multipliers $\mu_{j,i} = 0$. When audits are imperfect, deterministic audit strategies do not guarantee slack incentive constraints. Proposition 6 provides sufficient conditions under which optimal leverage is bounded above.

Proposition 6. *Nonnegative consumption in all truth-telling states requires that leverage is bounded above as*

$$\frac{\alpha + b}{\alpha} \leq \frac{\rho}{\rho - \left(\theta_1 - \sum_i \Delta(\theta_i)\mathcal{Q}(\theta_i)\kappa\right)}.$$

[15]It is possible to make (10) hold by introducing an exogenous limited liability constraint. This effectively imposes a lower bound on the support of the range of $\mathcal{U}$. However, for the incentive constraint to bind with certain auditing, the consumption allocations of truth-telling agents would also need to be constrained by the limited liability constraint.

[16]In the example we use to prove Theorem 2, the first channel is important: entrepreneurs write contracts that limit ex post penalties, as these penalties are wrongfully applied to truth-tellers with positive probability. In the numerical example we present in Section 6, the second channel is important: the audit technology can correctly identify misreporting agents with positive probability. The optimal standard debt contracts do not need to punish truth-tellers, but they are unable to impose maximal penalties on high income misreporting agents, as the audit technology does not precisely reveal the true income of misreporting agents.

Proposition 6 shows the effect of imperfect audits on optimal leverage. Under imperfect audits, the possibility of low out-turns has an important effect on the limits of leverage. As leverage increases, revenue in the low state approaches 0 by Assumption 2. Positive probabilities of revenue states conditional upon any audit signal comprise a stronger condition than imperfect audits, and imply that any "penalty" repayment allocation is occasionally wrongfully applied to truth-telling agents under contracts with positive probability auditing.

6. Numerical example

In this section, we simulate a numerical version of the model that approximates the general model developed earlier and allows us to compare the predictions of the model with stylized facts about small business borrowing. It also serves to highlight the main trade-offs that determine the choice between debt and equity. The main finding of our numerical exercise is that when audit signals are weakly correlated with true incomes, standard debt contracts emerge as optimal. When audit signals are highly correlated with true incomes, optimal contracts resemble equity.

Audits play two important roles in costly state verification contracts. First, as emphasized by Townsend (1979), audits of low reports allow contract enforcers to punish misreporting agents with high penalties. Second, as emphasized by Border and Sobel (1987), audits allow contract enforcers to reward truth-telling agents with low repayments after truthful reports; truth-tellers want to be audited. Our numerical example shows both of these mechanisms at work. When audits are high quality, even high reports are audited, and verified truthful reports are rewarded with low repayments. The second mechanism requires interior audit signals: when audit strategies are deterministic, there is no meaningful distinction between repayments following an unaudited report and repayments following an audit consistent with the reported income. In our computed example, the first mechanism is dominant when the audit signal has low quality. Audit strategies are deterministic, with certain audits following low reports. Audit signals are used to punish misreporting agents. The second mechanism emerges as the audit signal quality improves. High reports are audited and truthful reports are rewarded with low repayments relative to unaudited reports.

6.1 *Solution strategy*

The model is solved numerically for an example with 10 revenue states. For the main numerical simulations, we hold the amount of resources transferred to the entrepreneur b constant to focus on the link between optimal audit strategies and audit quality. This leaves us with $n + n(l + 1)$ choice variables to be determined, where n is the number of revenue states and l is the number of signal states.[17] It follows that in the perfect audits version of our model, there are $10 + 10(10 + 1) = 120$ choice variables. Given the

[17]The audit strategy $\mathcal{Q}(m)$ comprises $|M| = |\Theta| = n$ elements; the repayment strategy $\mathcal{Z}(m, \sigma)$ comprises $|M| \times |\Sigma + 1| = n(l + 1)$ states, where the $l + 1$ possible signals include the l signals drawn from the audit technology and the uninformative, null signal $\{\varnothing\}$ drawn when no audit occurs.

nonconvexity of Programme 1, the large number of choice variables poses a challenge for combining speed with numerical stability. The algorithm we describe below takes advantage of parallel processing when we sample over candidate audit strategies.

The solution algorithm splits the problem into two loops. The inner loop solves for the optimal repayment allocations $\mathcal{Z}$ conditional upon a candidate solution for the audit strategy $\mathcal{Q}$. This inner loop is solved using a feasible directions search algorithm. The outer loop solves for the optimal audit strategy $\mathcal{Q}$ using a combination of global sampling and local perturbation routines. The global sampling routine provides assurance that the solutions found are globally optimal, while the local perturbation routine improves precision.

6.2 *Functional forms*

Gross returns θ are drawn from a discrete uniform grid unif$\{\theta_1, \theta_{10}\}$ with 10 draws. The probabilities $\pi(\theta_i)$ are drawn from the binomial distribution $B(9, p)$. The return draw can be considered to be a function of smaller positive and negative draws, consistent with standard interpretations of the binomial distribution as the sum of "good" draws from an urn with replacement. The auditor is assumed to observe a subset of the individual draws that determine gross returns. Audits are perfect when the auditor observes all individual draws and thereby can observe θ with certainty. Audits are imperfect when the auditor observes a strict subset of the individual draws that determine gross returns.[18] Initial net worth α is a scaling parameter, and is set at 100.

The entrepreneur's coefficient of relative risk aversion $(\frac{-c\mathcal{U}''(c)}{\mathcal{U}'(c)})$ is assumed to be constant, with utility taking the functional form $\mathcal{U}(c) = \frac{c^{1-\gamma}}{1-\gamma}$. Under this functional form and parameterization, there is no optimal contract under perfect auditing; allocations approach arbitrarily close to first-best allocations. When $\gamma \geq 1$, $\lim_{c\to 0^+} \mathcal{U}(c) = -\infty$. Given any audit strategy with strictly positive audits in all states ($\mathcal{Q}(m)$ subject to $\mathcal{Q}(m) > 0\ \forall m$), maximal penalties ensure that all incentive compatibility constraints are relaxed, regardless of allocations and the level of individual audit probabilities. It follows that we can always perturb $\mathcal{Q}$ to reduce audit probabilities in all states, retaining incentive compatibility and expected utility while relaxing the participation constraint of the financial intermediary.

6.3 *Calibration*

Table 1 presents our preferred calibration of the model. For the matched parameters, the calibration procedure minimizes the distance from the model parameters to the available empirical estimates. For the unmatched parameters, the calibration procedure selects parameters that generate predicted values for optimal borrowing, probability of default, and credit spreads that are close to the empirical estimates. The probability of default is defined as the probability of a revenue state occurring that elicits a positive

[18] For example, let gross returns be affected by both input costs and input quality. It may be the case that the audit signal σ could reveal input costs with certainty, but not input quality.

TABLE 1. Calibration.

Description	Data	Model
Unmatched parameters		
Return bounds, $\{\theta_1, \theta_{10}\}$		$\{0.58, 1.62\}$
Distribution parameter, p		0.435
Number of signals, $\lvert\Sigma\rvert$		3
CRRA coefficient, γ		0.5
Matched parameters		
Opportunity cost, ρ	1.01	1.012
Audit costs, κ	(0.01, 0.06)	0.012
Matched predicted moments		
Borrowing, b^*	44.9	45.0
Probability of default	0.040	0.047
Credit spread	0.039	0.017

probability of audit under the optimal audit strategy. The credit spread is defined as the ratio of the repayment following the highest revenue report to the amount borrowed, b^*.

Data sources are as follows. Borrowing is taken from the U.S. Federal Reserve Z1 tables.[19] Audit costs are set within the bounds of empirical estimates of the direct cost of bankruptcy as a share of firm assets between 0.01 and 0.06.[20] The financial intermediary's opportunity cost of funds and the probability of default are drawn from the estimates of Herranz et al. (2015). The interest rate spread is constructed from interest expense and liability data captured by the U.S. Federal Reserve Z1 tables. From these figures, we can construct approximate annualized loan interest rate spreads of 3.51% and 4.35%, where the former estimate constructs the opportunity cost of funds ρ from financial institutions' interest expense and the latter estimate equates the opportunity cost of funds to the London Interbank Offered Rate.[21]

6.4 *Results*

RESULT 1. *The model predicts standard debt contracts as globally optimal contracts when the quality of the audit technology is low.*

Panels (A), (B), and (C) of Figure 1 present the importance of audit quality for the optimality of standard debt contracts. To make the comparison as straightforward as possible and to allow comparison with the perfect audits case, the amount of borrowing has been held constant across audit strategies for panels (A), (B), and (C) at 45, which is

[19]B.104 Balance Sheet of Nonfinancial Noncorporate Business, FL114104006 Debt as a percentage of net worth, 2016 Q4.

[20]See Warner and Gruber (1977), Weiss (1990), and Altman (1984).

[21]All values are for 2015 (USD billions). Nonfinancial noncorporate business: interest paid, 244.0 (FA116130001); total loans, 4,746.8 (FL114123005). Financial business: interest paid, 1,447.1 (FA796130001); total liabilities, 88,676.2 (FL794194005); 12 month LIBOR 0.79% (St. Louis Federal Reserve FRED Database: USD12MD156N).

the optimal level of borrowing for the calibrated model. When the audit technology produces a weak signal, standard debt contracts are optimal. When the number of signal states is 1, the audit signal is uninformative and the optimal contract is noncontingent debt with no auditing. When the number of signal states is 2 or 3, the optimal contract is standard debt, auditing low reports with certainty and not auditing any moderate or high reports. As the number of signal states increases, optimal contracts start to resemble equity contracts. When the audit signal is perfect, contracts are able to implement allocations arbitrarily close to the first-best allocations.

Result 2. *The model predicts an optimal leverage ratio that is consistent with empirical estimates; the optimality of standard debt is consistent with optimal leverage choices.*

Panel (F) displays the relationship between leverage and welfare for our calibration. This calibration generates the predictions that the optimal financial contract is a standard debt contract, and that optimal leverage ratios and default probabilities are very close to empirical estimates (Table 1). Panel (F) also presents the relationship between leverage and welfare restricting agents to noncontingent contracts. In contrast to optimal contracts, noncontingent contracts result in much lower optimal leverage and welfare gains from financial trade. This is the case despite the fact that the optimal contract involves only auditing and state contingency in 4.7% of out-turns (4.7% being the probability of default).

Panels (D) and (E) show for our calibration the relationships between welfare, optimal borrowing, and audit quality. Under our calibration, perfect audits ($|\sigma| = 10$) result in infinite leverage and infinite welfare for the entrepreneur. When the audit technology is noisy, improvements in the audit technology have large effects on optimal borrowing and when borrowing is chosen optimally, they have large effects on welfare.

7. Discussion

Standard debt contracts can be the optimal form of external finance contracts when contract enforcement is uncertain due to noisy audit signals. Supporting truth-telling under stochastic audit strategies requires large penalties. When there is no guarantee that these penalties are fairly applied, these contracts will not be acceptable to risk-averse entrepreneurs. The resulting efficient contracts audit, consequently, only on low reports, but likely audit low reports with certainty. As a result, only small penalties are required to ensure truth-telling in equilibrium. In fact, the penalty following a disputed report in an optimal debt contract is typically very close to fully repaying the debt.

Imperfect verification also implies other interesting properties of optimal contracts. For instance, it means that borrowers can only pass a limited amount of risk on to lenders, regardless of contracted audit strategies. Further, even when projects enjoy constant returns to scale and audits are relatively inexpensive, firm size and leverage is endogenously limited by the entrepreneur's risk preference.

While our main results center on the optimality of debt, contrary to the existing costly state verification literature, our model has implications for when equity-like contracts are optimal. In short, we find that when audit signals are weakly correlated with true incomes, standard debt contracts emerge as optimal. Alternatively, if audit signals are highly correlated with true incomes, optimal contracts resemble equity. Finally, the model developed here looks to be able to engage in interesting ways with empirical data, suggesting some interesting avenues for future research.

We conclude with two, more theoretical, observations. First, Harsanyi (1973) presents an isomorphism between games with mixed strategy Nash equilibria and "disturbed" games with stochastic payoffs and pure strategy Nash equilibria. Our findings resonate with Harsanyi's result; adding randomness to the audit signal encourages pure strategy auditing. However, there is no direct isomorphism between mixed audit strategies under perfect audits and deterministic audit strategies under imperfect audits in our model.

Second, the standard debt contracts we derived under imperfect monitoring enjoy an additional benefit that we did not formalize. When enforcement is certain, or near certain, incentive compatibility is not sensitive to the risk tolerance of the entrepreneur. That reduces the potential for adverse selection in two forms: first, the preferences of the entrepreneur may be unobservable; second, the entrepreneur may have access to hidden wealth. The presence of either of these sources of asymmetric information would make it more difficult to employ a stochastic incentive scheme.

Appendix A: Proofs

The proofs of results below are presented in the order in which they appear in the main text.

Proof of Remark 1. Nonsufficiency. In the working paper version of the current paper, we provide an example of multiplicity of locally optimal contracts, each of which satisfy the first order conditions of Programme 1 (Duncan and Nolan 2017).

Nonnecessity. Let $\mathcal{C}$ be a contract with some report m_i such that the audit probability following this report is equal to 0, $q_i = 1$. Consider the repayment $z_i(\varnothing)$. This repayment, $z_i(\varnothing)$, occurs with zero probability in expected utility, in the participation constraint and in the incentive compatibility constraint. Perturbing the value of $z_i(\varnothing)$ has no effect on expected utility or on feasibility. No first order condition exists for the repayment $z_i(\varnothing)$. However, the repayment $z_i(\varnothing)$ does enter with positive probability into the first order condition for q_i; the costs and benefits of a marginal decrease in q_i are dependent on the value of the repayment $z_i(\varnothing)$, which will be paid with positive probability after any such perturbation. Starting from contract $\mathcal{C}$, we can manipulate $z_i(\varnothing)$ in any way, retaining optimality and feasibility but violating the first order condition for q_i reported below:

$$\lambda\Delta(\theta_i)(\alpha+b)\kappa = \sum_{\sigma}\Delta_{q_i}(\theta_i,\sigma|q_i)\begin{bmatrix} \mathcal{U}(\theta_i,m_i,\sigma)+\lambda z_i(\sigma) \\ +\sum_j \mu_{i,j}\mathcal{U}(\theta_i,m_i,\sigma) \\ -\sum_j \mu_{j,i}\dfrac{\Delta(\theta_j|\sigma)}{\Delta(\theta_i|\sigma)}\mathcal{U}(\theta_j,m_i,\sigma)\end{bmatrix} + \nu_0(m_i) - \nu_1(m_i).$$

Alternatively, let $\mathcal{C}$ be a contract with the following property: the report–signal pair (m_i, σ) is such that (a) the repayment $\mathcal{Z}(m_i, \sigma)$ enters the second term in the incentive compatibility constraint (7) for an agent receiving revenue shock θ_j and considering misreporting m_i, and (b) is paid with probability 0 in truth-telling equilibrium (that is, (a) $\Delta(\theta_j, \sigma|\mathcal{Q}(m_i) > 0)$ and (b) $\Delta(\theta_i, \sigma|\mathcal{Q}(m_i)) = 0$). The first order condition for $\mathcal{Z}(m_i, \sigma)$ cannot be satisfied. If $\mathcal{C}$ is an optimal contract, the repayment $\mathcal{Z}(m_i, \sigma)$ will be sufficiently large such that the incentive constraint is not binding. Any further increases in the repayment $\mathcal{Z}(m_i, \sigma)$ will leave expected utility constant and will not affect feasibility. □

Proof of Proposition 1. To begin, we split Programme 1 into two distinct optimization problems. First, the entrepreneur determines the audit strategy $\mathcal{Q}$. Second, the entrepreneur determines the optimal repayment allocations $\mathcal{Z}$, consumption allocations $\mathcal{X}$, and amount of resources initially transferred to the entrepreneur b. The second stage of the problem is reported as Programme 2.

Programme 2. *We have*

$$\max_{\mathcal{Z},\mathcal{X},b} \sum_{i,\sigma} \Delta(\theta_i, \sigma|q_i)\mathcal{U}(\theta_i, m_i, \sigma)$$

subject to

$$(\alpha + b)\theta_i - \mathcal{Z}(m_j, \sigma) - \mathcal{X}(\theta_i, m_j, \sigma) = 0 \qquad (\phi_{i,j}(\sigma))$$

$$\sum_{i,\sigma} \Delta(\theta_i, \sigma|q_i)z_i(\sigma) - b\rho - \sum_{i=1}^{n} \Delta(\theta_i)q_i(\alpha + b)\kappa \geq 0 \qquad (\lambda)$$

$$\sum_{\sigma} \Delta(\theta_i, \sigma|q_i)\mathcal{U}(\theta_i, m_i, \sigma) - \sum_{\sigma} \Delta(\theta_i, \sigma|q_j)\mathcal{U}(\theta_i, m_j, \sigma) \geq 0 \qquad (\mu_{i,j}).$$

Contemporaneous utility $\mathcal{U}$ is a concave function of consumption allocations $\mathcal{X}$ and, therefore, Programme 2 is nonconvex. Programme 2 cannot be convexified via rewriting the program to solve for allocations of $\mathcal{U}$ directly.[22] Before applying the Kuhn–Tucker theorem to generate first order necessary conditions, we must first verify a constraint qualification. We verify the Mangasarian–Fromowitz constraint qualification (Mangasarian and Fromovitz 1967).

Definition 4. Given a point $\mathcal{C} \in \mathbb{R}^k$ and the active set $\mathcal{A}(\mathcal{C})$, let $\mathcal{I}(\mathcal{C})$ denote the indices of binding inequality constraints $g_i(\mathcal{C}) = 0$ and equality constraints $h_j(\mathcal{C}) = 0$. The Mangasarian–Fromowitz constraint qualification holds if and only if there exists a vector

[22]This substitution would generate a convex cost function $\mathcal{X}(\cdot) = \mathcal{U}^{-1}(\cdot)$. Recall that the budget constraints in our model are equality constraints, not inequality constraints. The introduction of convex cost function $\mathcal{X}$ would make the budget constraints nonconvex.

$d \in \mathbb{R}^k$ such that

$$\nabla g_i(\mathcal{C})'d > 0 \quad \forall i \in \mathcal{I}(\mathcal{C})$$
$$\nabla h_j(\mathcal{C})'d = 0 \quad \forall j \in \mathcal{I}(\mathcal{C}).$$

Lemma 1. *Programme 2 satisfies the Mangasarian–Fromowitz constraint qualification.*

Proof[23]. The active set $\mathcal{A}(\mathcal{C})$ of the subproblem consists of the budget constraints, the participation constraint, and the active incentive constraints (as the audit probabilities are predetermined, the upper and lower bounds on audit probabilities are not active in the subproblem, Programme 2). We start by reporting the gradients of the equality and inequality constraints:

∇BC:

$$\frac{\partial \mathrm{BC}_{i,j,\sigma}}{\partial b} = -\theta, \qquad \frac{\partial \mathrm{BC}_{i,j,\sigma}}{\partial \mathcal{Z}(m_j,\sigma)} = 1, \qquad \frac{\partial \mathrm{BC}_{i,j,\sigma}}{\partial \mathcal{X}(\theta_i, m_j,\sigma)} = 1;$$

∇PC:

$$\frac{\partial \mathrm{PC}}{\partial b} = -\sum_{\theta} \Delta(\theta) q(\theta)\kappa - \rho, \qquad \frac{\partial \mathrm{PC}}{\partial \mathcal{Z}(m,\sigma)} = \Delta\big(\theta, \sigma | q(\theta)\big);$$

$\nabla \mathrm{ICC}_{i,j}$:

$$\begin{aligned}
\frac{\partial \mathrm{ICC}_{i,j}}{\partial \mathcal{Z}(m_i,\sigma)} &= -\Delta\big(\theta_i, \sigma | q(m_i)\big)\mathcal{U}'(\theta_i, m_i, \sigma), \\
\frac{\partial \mathrm{ICC}_{i,j}}{\partial \mathcal{Z}(m_j,\sigma)} &= \Delta\big(\theta_i, \sigma | q(m_j)\big)\mathcal{U}'(\theta_i, m_j, \sigma) \\
\frac{\partial \mathrm{ICC}_{i,j}}{\partial b} &= \sum_{\sigma} \Delta\big(\theta_i, \sigma | q(m_i)\big)\theta_i \mathcal{U}'(\theta_i, m_i, \sigma) \\
&\quad - \sum_{\sigma} \Delta\big(\theta_i, \sigma | q(m_j)\big)\theta_i \mathcal{U}'(\theta_i, m_j, \sigma).
\end{aligned}$$

We wish to construct a vector d, specified by Definition 4. Now let $\mathcal{D} : \mathcal{C} \to \mathbb{R}$ be a mapping from the vector of contract terms in $\mathcal{C}$ to the reals. The elements of vector d are represented by the images of function $\mathcal{D}$, such that without loss of generality, $d_i = \mathcal{D}(\mathcal{Z}(m,\sigma))$ if and only if $\mathcal{C}_i = \mathcal{Z}(m,\sigma)$. We search for some function $\mathcal{D}$ such that

$$\sum_{c \in \mathcal{C}} \frac{\partial g_i}{\partial c}\mathcal{D}(c) > 0, \quad \text{and} \quad \sum_{c \in \mathcal{C}} \frac{\partial h_j}{\partial c}\mathcal{D}(c) = 0.$$

We construct $\mathcal{D}$ with three steps. First we set $\mathcal{D}(b)$, second we set $\mathcal{D}(\mathcal{Z})$, and finally we set $\mathcal{D}(\mathcal{X})$.

[23]The proof does rely on the impermissability of high messages following low revenue states. A more general proof that relaxes this assumption is available from the authors on request.

First, set

$$\mathcal{D}(b) = 0.$$

Second, we set the values associated with repayment allocations $\mathcal{D}(\mathcal{Z}(m_i, \sigma))$.

Choose some signal σ'' such that $\Delta(\theta_j, \sigma''|q_j) > 0$. Set $\mathcal{D}(\mathcal{Z}(m_j, \sigma'')) > 0$. We now have

$$\sum_{c\in\mathcal{C}} \frac{\partial \mathrm{PC}}{\partial c}\mathcal{D}(c) > 0.$$

This inequality is retained by ensuring nonnegativity of all $\mathcal{D}(\mathcal{Z}(m, \sigma))$.

Now, for $i \in (2, 3, \dots, n)$ and $j \in \{1, 2, \dots, i\}$, choose some signal $\sigma'_{i,j}$ such that $\Delta(\theta_i, \sigma'_{i,j}|q_j) > 0$. Set $\mathcal{D}(\mathcal{Z}(m_j, \sigma'_{i,j}))$ sufficiently high such that

$$\sum_{\sigma}\left[\frac{\partial \mathrm{ICC}_{i,j}}{\partial \mathcal{Z}(m_i, \sigma)}\mathcal{D}\big(\mathcal{Z}(m_i, \sigma)\big) + \frac{\partial \mathrm{ICC}_{i,j}}{\partial \mathcal{Z}(m_j, \sigma)}\mathcal{D}\big(\mathcal{Z}(m_j, \sigma)\big)\right] > 0.$$

Third and finally, we need to satisfy the Mangasarian–Fromowitz conditions for the budget constraints. These constraints are equality constraints; therefore, we set

$$\mathcal{D}\big(\mathcal{X}(\theta_i, m_j, \sigma)\big) = -\mathcal{D}\big(\mathcal{Z}(m_j, \sigma)\big).$$

This ensures

$$\sum_{c,\theta_i,m_j,\sigma} \frac{\partial \mathrm{BC}_{i,j,\sigma}}{\partial c}\mathcal{D}(c) = 0,$$

completing the proof. □

Lemma 1 allows us to apply the Kuhn–Tucker theorem to derive necessary conditions for Programme 2 using a first order approach.[24]

The first order necessary conditions for Programme 2 are

$$\begin{aligned} z_i(\sigma): \quad 0 = &-\Delta(\theta_i, \sigma|q_i)\mathcal{U}'(\theta_i, m_i, \sigma) + \lambda\Delta(\theta_i, \sigma|q_i) \\ &- \sum_j \mu_{i,j}\Delta(\theta_i, \sigma|q_i)\mathcal{U}'(\theta_i, m_i, \sigma) \\ &+ \sum_j \mu_{j,i}\Delta(\theta_j, \sigma|q_i)\mathcal{U}'(\theta_j, m_i, \sigma) \end{aligned} \tag{11}$$

$$\begin{aligned} b: \quad 0 = &\sum_{i,\sigma}\Delta(\theta_i, \sigma|q_i)\theta_i\mathcal{U}'(\theta_i, m_i, \sigma) - \lambda\left[\rho + \sum_{i=1}^{n}\Delta(\theta_i)q_i\kappa\right] \\ &+ \sum_{i,j}\mu_{i,j}\theta_i\left[\sum_{\sigma}\Delta(\theta_i, \sigma|q_i)\mathcal{U}'(\theta_i, m_i, \sigma)\right. \\ &\left. - \sum_{\sigma}\Delta(\theta_i, \sigma|q_j)\mathcal{U}'(\theta_i, m_j, \sigma)\right]. \end{aligned} \tag{12}$$

[24] See, for example, Bertsekas (1995).

By Bayes' theorem, we can rewrite $\Delta(\theta_i, \sigma|q_j) = \Delta(\sigma(q_j)|\theta_i)\Delta(\theta_i) = \Delta(\theta_i|\sigma(q_j))\Delta(\sigma(q_j))$. It follows that we have

$$\frac{\Delta(\theta_i, \sigma|q_j)}{\Delta(\theta_j, \sigma|q_j)} = \frac{\Delta(\theta_i|\sigma)}{\Delta(\theta_j|\sigma)}.$$

We can now divide (11) by $\Delta(\theta_i, \sigma|q_i)$ and rearrange to obtain

$$\frac{\mathcal{U}'(\theta_i, m_i, \sigma)}{\lambda} = \frac{1}{1 + \sum_j \mu_{i,j}} \left[1 + \sum_j \mu_{j,i} \frac{\Delta(\theta_j|\sigma)}{\Delta(\theta_i|\sigma)} \frac{\mathcal{U}'(\theta_j, m_i, \sigma)}{\lambda}\right]. \tag{8}$$

Use (11) to eliminate the terms

$$\sum_{i,\sigma} \Delta(\theta_i, \sigma|q_i)\theta_i \mathcal{U}'(\theta_i, m_i, \sigma) + \sum_{i,j} \mu_{i,j}\theta_i \sum_\sigma \Delta(\theta_i, \sigma|q_i)\mathcal{U}'(\theta_i, m_i, \sigma)$$

from (12):

$$\begin{aligned} 0 = {} & \sum_i \theta_i \sum_\sigma \lambda\Delta(\theta_i, \sigma|q_i) - \lambda\left[\rho + \sum_{i=1}^n \Delta(\theta_i)q_i\kappa\right] \\ & + \sum_i \theta_i \sum_\sigma \sum_j \mu_{j,i}\Delta(\theta_j, \sigma|q_i)\mathcal{U}'(\theta_j, m_i, \sigma) \\ & - \sum_i \theta_i \sum_\sigma \sum_j \mu_{i,j}\Delta(\theta_i, \sigma|q_j)\mathcal{U}'(\theta_i, m_j, \sigma). \end{aligned}$$

Relabelling the summands in the third term (swapping the is and js) allows us to simplify

$$\begin{aligned} 0 = {} & \sum_i \theta_i \sum_\sigma \lambda\Delta(\theta_i, \sigma|q_i) - \lambda\left[\rho + \sum_{i=1}^n \Delta(\theta_i)q_i\kappa\right] \\ & + \sum_{i,j} \sum_\sigma \Delta(\theta_i, \sigma|q_j)\mu_{i,j}(\theta_j - \theta_i)\mathcal{U}'(\theta_i, m_j, \sigma). \end{aligned}$$

The first term is just the product of the shadow cost of the participation constraint λ and expected returns.[25] Collecting terms in λ, we obtain (9):

$$\lambda\left[\sum_i \Delta(\theta_i)(\theta_i - q_i\kappa) - \rho\right] = \sum_{i,j} \sum_\sigma \Delta(\theta_i, \sigma|q_j)\mu_{i,j}(\theta_i - \theta_j)\mathcal{U}'(\theta_i, m_j, \sigma).$$

□

Proof of Corollary 1. (i) Part (i) follows from inspection of (8). If $\mathrm{ICC}_{j,i} \notin \mathcal{A}(\mathcal{C})$ $\forall j$, then (8) becomes

$$\mathcal{U}'(\theta_i, m_i, \sigma) = \frac{\lambda}{1 + \sum_j \mu_{i,j}},$$

which is constant for all σ.

[25] We have $(\sum_i \theta_i \sum_\sigma \Delta(\theta_i, \sigma|q_i) = \sum_{i,\sigma} \Delta(\theta_i, \sigma|q_i)\theta = \sum_i \Delta(\theta_i)\theta)$ by the law of total probability.

(ii) Consider revenue shock θ_i such that there exists some shock θ_j with the property that $\mathrm{ICC}_{j,i} \in \mathcal{A}(\mathcal{C})$. Start with the first order necessary condition for repayment allocations $\mathcal{Z}(m_i, \sigma)$, equality (8), which is repeated for convenience:

$$0 = -\Big(1 + \sum_k \mu_{i,k}\Big)\mathcal{U}'(\theta_i, m_i, \sigma) + \lambda + \sum_j \mu_{j,i} \frac{\Delta(\theta_j|\sigma)}{\Delta(\theta_i|\sigma)} \mathcal{U}'(\theta_j, m_i, \sigma).$$

We are interested in seeing how the repayment allocation $\mathcal{Z}(m_i, \sigma)$ must respond to fluctuations in the marginal likelihood ratio, $\Gamma_i^j(\sigma) := \frac{\Delta(\theta_j|\sigma)}{\Delta(\theta_i|\sigma)}$. Holding i constant, the remaining terms in μ, λ are held constant as we vary the repayment signal σ. Take the total derivative with respect to Γ and the repayment allocation $\mathcal{Z}(m_i, \sigma)$:

$$\begin{aligned} 0 = {} & \Big(1 + \sum_k \mu_{i,k}\Big)\mathcal{U}''(\theta_i, m_i, \sigma)\, d\mathcal{Z}(m_i, \sigma) \\ & - \sum_j \mu_{j,i} \Gamma_i^j(\sigma) \mathcal{U}''(\theta_j, m_i, \sigma)\, d\mathcal{Z}(m_i, \sigma) \\ & + \sum_j \mu_{j,i} \mathcal{U}'(\theta_j, m_i, \sigma)\, d\Gamma_i^j(\sigma). \end{aligned}$$

Divide through by $(1 + \sum_k \mu_{i,k})\mathcal{U}'(\theta_i, m_i, \sigma) (= \lambda + \sum_j \mu_{j,i} \frac{\Delta(\theta_j|\sigma)}{\Delta(\theta_i|\sigma)} \mathcal{U}'(\theta_j, m_i, \sigma)$ by (8)) and substitute $A(x) = \frac{-\mathcal{U}''(x)}{\mathcal{U}'(x)}$:

$$\begin{aligned} & \left[A(\theta_i, m_i, \sigma) - \frac{\sum_j \mu_{j,i} \Gamma_i^j(\sigma)}{\lambda + \sum_j \mu_{j,i} \Gamma_i^j(\sigma)} A(\theta_j, m_i, \sigma) \right] d\mathcal{Z}(m_i, \sigma) \\ & = \frac{\sum_j \mu_{j,i}}{1 + \sum_k \mu_{i,k}} \frac{\mathcal{U}'(\theta_j, m_i, \sigma)}{\mathcal{U}'(\theta_i, m_i, \sigma)}\, d\Gamma_i^j(\sigma). \end{aligned}$$

By the assumption of nonincreasing absolute risk aversion, $A(\theta_j, m_i, \sigma) \le A(\theta_i, m_i, \sigma)$. The multipliers μ, λ and the likelihood ratios Γ are all nonnegative; therefore, the difference $A(\theta_i, m_i, \sigma) - \frac{\sum_j \mu_{j,i} \Gamma_i^j(\sigma)}{\lambda + \sum_j \mu_{j,i} \Gamma_i^j(\sigma)} A(\theta_j, m_i, \sigma)$ and the ratio $\frac{\sum_j \mu_{j,i}}{1 + \sum_k \mu_{i,k}} \frac{\mathcal{U}'(\theta_j, m_i, \sigma)}{\mathcal{U}'(\theta_i, m_i, \sigma)}$ are strictly positive. It follows that $\frac{d\Gamma_i^j(\sigma)}{d\mathcal{Z}(m_i, \sigma)} > 0$, completing the proof. □

Proof of Proposition 2. (i) Let the financial intermediary's participation constraint be nonbinding. Let σ be a signal with the property that $\Delta(\theta_n, \sigma|q_n) > 0$. We can reduce the repayment $\mathcal{Z}(m_n, \sigma)$, which tightens the currently slack participation constraint, increases expected utility, and relaxes the incentive constraints $\mathrm{ICC}_{n,i}$ for $i \in \{1, 2, 3, \dots, n-1\}$.

(ii) The highest possible report m_n appears only on the left hand side of incentive constraints. By Corollary 1, the optimal repayment allocation $\mathcal{Z}(m_n, \sigma)$ is not contingent on the audit signal σ. It follows that we can remove all audits of reports m_n without affecting expected utility or the incentive constraints, but relaxing the participation constraint for any positive audit cost.

(iii) Let $\mathcal{C}$ be a contract with no binding incentive constraints.

Case (a). If all audit probabilities are equal to 0, $\mathcal{Q}(m) = 0\, \forall m$, then we can verify the proposition directly from the first order condition for $\mathcal{Z}$. When no audits are undertaken, incentive compatibility requires $\mathcal{X}(\theta_i, m_i, \varnothing) = \mathcal{X}(\theta_j, m_j, \varnothing) + (\alpha + b)(\theta_i - \theta_j)$ by (7). If no incentive constraints are binding, then the first order condition for $\mathcal{Z}$ states that $\mathcal{U}'(\theta_i, m_i, \varnothing) = \lambda = \mathcal{U}'(\theta_j, m_j, \varnothing)$, a contradiction.

Case (b). If there exists a positive audit probability for some report $\mathcal{Q}(m_i) > 0$, then we can reduce this audit probability, replacing the foregone audits with a lottery that replicates the distribution of the true audit technology conditional upon truth-telling. For details, see the proof of Theorem 1.

(d) If repayment allocations are not contingent on audit signals, then we can apply the same argument from part (iii)(a).

If repayment allocations are contingent on audit signals, then there exist i, σ, σ' such that $\mathcal{X}(\theta_i, m_i, \sigma) \neq \mathcal{X}(\theta_i, m_i, \sigma')$. It follows that $\mathcal{U}'(\theta_i, m_i, \sigma) \neq \mathcal{U}'(\theta_i, m_i, \sigma')$. This is consistent with the first order condition for repayment allocations (8) only if there exists some j such that $\mu_{j,i} > 0$. □

Proof of Theorem 1. Assume audits are perfect with strictly positive cost ($\Delta(\theta|\sigma) \in \{0, 1\}$ and $\kappa > 0$). Let $\mathcal{C}$ be a truth-telling contract with the feature that there is some report m_i that is followed by a certain audit, $q_i = 1$. Assume that under contract $\mathcal{C}$, any truth-telling agent enjoys some strictly positive probability of avoiding immiseration: $\forall i, \sum_\sigma \Delta(\theta_i, \sigma|q_i)\mathcal{U}(\theta_i, m_i, \sigma) > \underline{\mathcal{U}}$.

Perturbation 1. Consider the set of repayment allocations contingent on reported income m_i that occur with zero probability under truth-telling. Set all of the repayment allocations in this set equal to their corresponding maximal repayments:

$$\text{if } \Delta(\theta_i|\sigma) = 0 \quad \text{and} \quad \Delta(\theta_j|\sigma) = 1, \quad \text{then } \mathcal{Z}'(m_i, \sigma) = (\alpha + b)\theta_j.$$

Following Perturbation 1, any agent who misreports θ_i is immiserated with certainty. By assumption, truth-telling agents are not immiserated with positive probability. It follows that the incentive constraints $\mathrm{ICC}_{j,i}$ are nonbinding for all j.

Perturbation 2. Following the reported message m_i, employ the audit technology with probability $(1 - s)$. With probability s, draw a signal σ_s from the distribution $\sigma_s \sim \Sigma(\theta_i)$, that is, let $\Delta(\sigma_s|\theta_i) = \Delta(\sigma|\theta_i)$ for all values of σ_s and σ. Retain the same repayment schedule, but now condition repayments on either σ or σ_s, whichever is drawn, $\mathcal{Z}''(m_i, \sigma \cup \sigma_s) = \mathcal{Z}'(m_i, \sigma)$.

The new signal has no resource cost. However, the new strategy saves audit costs worth $\Delta(\theta_i)s(\alpha + b)\kappa$. This savings relaxes the participation constraint of the financial intermediary. Agents who truthfully report message m_i experience no change in their conditional distribution of repayments and, therefore, no change in expected utility. This also implies that for agents receiving return θ_i, their opportunity cost of misreporting remains constant, and the perturbation does not violate any incentive constraint of the form $\mathrm{ICC}_{i,j}$. Following Perturbation 1, incentive constraints of the form $\mathrm{ICC}_{j,i}$ are slack before Perturbation 2, and remain slack for sufficiently small s.

Taken together, Perturbations 1 and 2 relax a financial intermediary's participation constraint while maintaining incentive compatibility and expected utility. We can continue to apply Perturbations 1 and 2 to each report that motivates certain auditing under the initial contract $\mathcal{C}$ until all audit probabilities are strictly below 1. □

Proof of Proposition 3. We prove the proposition by solving for the contract that implements full insurance with the lowest possible total audit cost. When audit costs are 0, the full insurance, first-best, allocation is feasible and optimal. However, when audit costs are 0, excessive audits do not incur welfare costs. When audit costs are strictly positive but approaching 0, optimal repayment allocations approach the first-best allocations, but audit strategies ration audits to reduce resource costs.

The incentive constraint for an agent receiving true return i and considering whether to report message j (7) is

$$\sum_{\sigma} \Delta(\theta_i, \sigma | q_i)\mathcal{U}(\theta_i, m_i, \sigma) \geq \sum_{\sigma} \Delta(\theta_i, \sigma | q_j)\mathcal{U}(\theta_i, m_j, \sigma).$$

Denote the full insurance consumption bundle by x^*. Under the full insurance allocation, agent i receives allocation x^* under truth-telling with certainty, so we can replace the left hand side with $\mathcal{U}(x^*)$. Following a false report of message m_j, agent i would receive consumption allocation $x^* + (\alpha + b)(\theta_i - \theta_j)$ if (s)he is not audited. This represents the full insurance allocation x^* received by any truth-telling agent j, plus the difference in revenue between out-turns i and j. If audited, an agent i who misreports m_j would have his/her true return revealed with certainty under perfect audits, and would be charged a maximal penalty. This agent would receive utility allocation $\mathcal{U}_{\min}$, equal to the lower support of the range of the utility function $\mathcal{U}$ if a lower support exists. If a lower support does not exist, $\mathcal{U}_{\min}$ can be set arbitrarily low. The incentive constraint can now be rewritten as

$$\mathcal{U}(x^*) = (1 - q_j)\mathcal{U}\big(x^* + (\alpha + b)(\theta_i - \theta_j)\big) + q_j\mathcal{U}_{\min}.$$

We can solve explicitly for the minimal audit probability necessary to attain incentive compatibility:

$$q_j = \frac{\mathcal{U}\big(x^* + (\alpha + b)(\theta_i - \theta_j)\big) - \mathcal{U}(x^*)}{\mathcal{U}\big(x^* + (\alpha + b)(\theta_i - \theta_j)\big) - \mathcal{U}_{\min}}.$$

This condition gives us the limiting optimal audit probability as audit costs approach 0. If the range of $\mathcal{U}$ is unbounded below, incentive compatibility can be attained with q_j arbitrarily close to 0. If the range of $\mathcal{U}$ is bounded below, we can attain incentive compatibility with strictly interior q_j. □

Proof of Proposition 4. To complete the proof, we use Proposition 1 to argue directly from the first order conditions (8) and (9). Set all audit probabilities equal to 1, $\mathcal{Q}(\theta) = 1$. Under perfect audits, this leaves $\Delta(\theta, \sigma) \in \{0, 1\}\ \forall(\theta, \sigma)$. It follows that the first order condition for repayments $\mathcal{Z}$ expressed as (8) can only be satisfied when $\mu_{i,j} = 0$ $\forall j, i$. If we substitute this result into (9), the right hand side is equal to 0. It follows that when $\kappa < \mathbb{E}(\theta) - \rho$, the first order condition for the borrowing cannot be satisfied: the marginal value of additional borrowing always exceeds the marginal cost. To show that entrepreneur consumption is also infinite, substitute this result into Constraints 1 and 3. □

Proof of Proposition 5. Let the value of any feasible contract $\mathcal{C}$ be denoted by the value function $\mathcal{V}$:

$$\mathcal{V}(\mathcal{C}) := \sum_{i,\sigma} \Delta\big(\theta_i, \sigma | q_i(\mathcal{C})\big)\mathcal{U}(\theta_i, m_i, \sigma|\mathcal{C}).$$

Denote the supremum of $\mathcal{V}$ by

$$\hat{\mathcal{V}} := \sup_{\mathcal{C}} \mathcal{V}(\mathcal{C}).$$

For any sequence of feasible contracts $\mathcal{C}^s$ with the property that $\mathcal{V}(\mathcal{C}^s) \to \hat{\mathcal{V}}$, we show that there exists $\hat{\mathcal{C}}$ with the properties that $\hat{\mathcal{C}}$ is feasible and that $\mathcal{V}(\hat{\mathcal{C}}) = \hat{\mathcal{V}}$. This result is sufficient to show that optimal contracts exist.

Recall that contract $\mathcal{C}$ is a tuple $(b, \mathcal{Q}, \mathcal{Z}, \mathcal{X})$. We limit our arguments to borrowing b and consumption $\mathcal{X}$, as the auditing strategy is defined over a compact set, and repayment $\mathcal{Z}$ is uniquely determined by repayments and borrowing in combination with the ex post budget constraints (Constraint 1).

(a) Lower limit of consumption.

Case 1: Utility is bounded below, $\mathcal{U}(0) = \underline{U}$ for some $\underline{U} \in \mathbb{R}$. When utility is bounded below, then the lower bound on consumption is feasible, and the set of feasible consumption allocations contains its lower limit points.

Case 2: Utility is unbounded below, $\lim_{x\to 0}\mathcal{U}(x) = -\infty$. Take a sequence of contracts with the property that $\mathcal{X}^s(\theta_i, m_i, \sigma) \to 0$ for some (θ_i, σ) such that $\Delta^s(\theta_i, m_i, \sigma) > 0$.

Let $\mathcal{U}_0 := \mathcal{U}(\alpha\rho)$. This is the value afforded an agent who deposits all of his/her wealth risk-free with the financial intermediary, $b = -\alpha\rho$. This value is finite, $\mathcal{U}_0 \in (-\infty, \infty)$. By continuity of the utility function, $\mathcal{V}(\mathcal{C}^s) \to \mathcal{V}^* \geq \mathcal{U}_0$. It follows that the product $\Delta^s(\theta_i, m_i, \sigma)\mathcal{U}(\mathcal{X}^s(\theta_i, m_i, \sigma))$ approaches a finite value. Since $\mathcal{U}(\mathcal{X}^s(\theta_i, m_i, \sigma)) \to -\infty$, the audit probability $\Delta^s(\theta_i, m_i, \sigma)$ approaches 0:

$$\Delta^s(\theta_i, m_i, \sigma) \to 0 \tag{13}$$

Let contract $\mathcal{C}^*$ be defined as follows. All contract terms in $\mathcal{C}^*$ are equal to those specified by $\mathcal{C}^s$ aside from

$$\mathcal{X}^*(\theta_i, m_i, \sigma) = (\alpha + b)\theta_i, \qquad \mathcal{Z}^*(m_i, \sigma) = 0.$$

By construction, this contract satisfies the ex post budget constraints (Constraint 1).

Consider the participation constraint. Under the new contract, this constraint is tightened by amount $\Delta^s(\theta_i, m_i, \sigma)\mathcal{Z}^s(m_i, \sigma)$. By (13), the audit probability $\Delta^s(\theta_i, m_i, \sigma) \to 0$; by the ex post budget constraints (Constraint 1), the repayment allocation $\mathcal{Z}^s(m_i, \sigma)$ is finite. It follows that the product $\Delta^s(\theta_i, m_i, \sigma)\mathcal{Z}^s(m_i, \sigma) \to 0$ and the participation constraint remains satisfied after the perturbation.

Consider a binding incentive compatibility constraint for an agent earning $\theta_j > \theta_i$. Under the new contract, this binding constraint is tightened by

$$\Delta^s(\theta_j, m_i, \sigma)\mathcal{U}\big((\alpha + b^s)(\theta_j - \theta_i) + \mathcal{X}^s(\theta_i, m_i, \sigma)\big).$$

The utility allocation $\mathcal{U}((\alpha + b^s)(\theta_j - \theta_i) + \mathcal{X}^s(\theta_i, m_i, \sigma)) \to \mathcal{U}((\alpha + b^s)(\theta_j - \theta_i)) \in (-\infty, \infty)$. The probability $\Delta^s(\theta_j, m_i, \sigma) \to 0$.[26] It follows that

$$\Delta^s(\theta_j, m_i, \sigma)\mathcal{U}\big((\alpha + b^s)(\theta_j - \theta_i) + \mathcal{X}^s(\theta_i, m_i, \sigma)\big) \to 0 \tag{14}$$

and, therefore, that the incentive compatibility constraint remains satisfied.

We have shown that the contract $\mathcal{C}^*$ is feasible and improves expected utility. But the new consumption bundle is now nonzero ($\mathcal{X}^*(\theta_i, m_i, \sigma) = (\alpha + b)\theta_i$); therefore, the limit of the sequence $\mathcal{C}^s$ does not specify a supremum.

(b) Upper limit of consumption. Consider a sequence of contracts with the property that $\mathcal{X}^s(\theta_i, m_i, \sigma) \to \infty$ for some (θ_i, σ) such that $\Delta^s(\theta_i, m_i, \sigma) > 0$. It follows by the ex post budget constraints (Constraint 1) that $\mathcal{Z}^s(m_i, \sigma) \to -\infty$.

The financial intermediary's participation constraint requires

$$\sum_{i,\sigma} \Delta^s(\theta_i, \sigma)\mathcal{Z}^s(m_i, \sigma) \geq b^s\rho + \sum_i \Delta(\theta_i)\mathcal{Q}^s(m_i)\kappa.$$

Borrowing b^s is bounded below by $-\alpha$. Expected audit costs $\sum_i \Delta(\theta_i)\mathcal{Q}^s(m_i)\kappa$ are bounded by $[0, \kappa]$. By part (a) of this proof, combined with the ex post budget constraints (Constraint 1), we also know that

$$\mathcal{Z}^s\big(m_j, \sigma'\big) \leq \big(\alpha + b^s\big)\theta_j \quad \forall j, \sigma'.$$

Together, we have

$$\Delta^s(\theta_i, \sigma)\mathcal{Z}^s(m_i, \sigma) \geq b^s\rho - \big(\alpha + b^s\big)\mathbb{E}(\theta_i)$$

$$\Delta^s(\theta_i, \sigma) \leq \frac{b^s\rho - \big(\alpha + b^s\big)\mathbb{E}(\theta_i)}{\mathcal{Z}^s(m_i, \sigma)}.$$

[26] Earlier we saw that $\Delta^s(\theta_i, m_i, \sigma) \to 0$. This must mean either that $q_i^s \to 0$ if σ is an audit signal or that $q_i^s \to 1$ if σ is the null signal. Either way, it follows that $\Delta^s(\theta_j, m_i, \sigma) \to 0\ \forall j$.

Consider the contract $\mathcal{C}^*$ defined with all terms equal to those specified by $\mathcal{C}^s$ aside from

$$\mathcal{X}^*(\theta_i, m_i, \sigma) = (\alpha + b)\theta_i, \qquad \mathcal{Z}^*(m_i, \sigma) = 0.$$

By construction, this contract satisfies the ex post budget constraints (Constraint 1).

The effect of this perturbation on the financial intermediary's participation constraint is positive, with expected repayments increasing by $\Delta^s(\theta_i, \sigma)(-\mathcal{Z}^s(m_i, \sigma))$.

Expected utility decreases by

$$\begin{aligned}\mathcal{V}^* - \mathcal{V}^s &= \Delta^s(\theta_i, \sigma)\big[\mathcal{U}\big((\alpha + b^s)\theta_i - \mathcal{Z}^s(m_i, \sigma)\big) - \mathcal{U}\big((\alpha + b^s)\theta_i\big)\big] \\ &\leq \big(b^s\rho - (\alpha + b^s)\mathbb{E}(\theta_i)\big)\frac{\mathcal{U}\big((\alpha + b^s)\theta_i - \mathcal{Z}^s(m_i, \sigma)\big) - \mathcal{U}\big((\alpha + b^s)\theta_i\big)}{\mathcal{Z}^s(m_i, \sigma)}.\end{aligned}$$

Concavity of the utility function implies that as $\mathcal{Z}^s(m_i, \sigma) \to -\infty$, the fraction

$$\frac{\mathcal{U}\big((\alpha + b^s)\theta_i - \mathcal{Z}^s(m_i, \sigma)\big) - \mathcal{U}\big((\alpha + b^s)\theta_i\big)}{\mathcal{Z}^s(m_i, \sigma)} \to 0.$$

Therefore, the perturbation delivers $\mathcal{V}^* = \hat{\mathcal{V}}$.

Now we turn to incentive compatibility. First we consider the incentive of an agent receiving θ_i truthfully to report his/her income; then we turn to an agent $\theta_j \neq \theta_i$. Conditional upon receiving shock θ_i, the expected utility of the agent decreases by

$$\begin{aligned}&\Delta^s(\sigma|\theta_i)\big[\mathcal{U}\big((\alpha + b^s)\theta_i - \mathcal{Z}^s(m_i, \sigma)\big) - \mathcal{U}\big((\alpha + b^s)\theta_i\big)\big] \\ &\quad= \frac{\Delta^s(\theta_i, \sigma)}{\Delta(\theta_i)}\big[\mathcal{U}\big((\alpha + b^s)\theta_i - \mathcal{Z}^s(m_i, \sigma)\big) - \mathcal{U}\big((\alpha + b^s)\theta_i\big)\big] \\ &\quad\leq \left(\frac{b^s\rho - (\alpha + b^s)\mathbb{E}(\theta_i)}{\Delta(\theta_i)}\right) \\ &\qquad\times \frac{\mathcal{U}\big((\alpha + b^s)\theta_i - \mathcal{Z}^s(m_i, \sigma)\big) - \mathcal{U}\big((\alpha + b^s)\theta_i\big)}{\mathcal{Z}^s(m_i, \sigma)}.\end{aligned}$$

Concavity of the utility function implies that as $\mathcal{Z}^s(m_i, \sigma) \to -\infty$, the fraction

$$\frac{\mathcal{U}\big((\alpha + b^s)\theta_i - \mathcal{Z}^s(m_i, \sigma)\big) - \mathcal{U}\big((\alpha + b^s)\theta_i\big)}{\mathcal{Z}^s(m_i, \sigma)} \to 0.$$

Therefore, the expected utility attained by truth-telling for an agent receiving shock θ_i is unchanged after the perturbation—incentive compatibility is retained.

For an agent receiving shock $\theta_j \neq \theta_i$, his/her expected repayment after reporting m_i is higher than before the perturbation, $\mathcal{Z}^*(m_i, \sigma) > \hat{\mathcal{Z}}(m_i, \sigma)$. Therefore, without loss of generality, the incentive compatibility constraint for agents receiving shock θ_j and considering report m_i is relaxed after the perturbation.

In sum, contract $\mathcal{C}^*$ is feasible, has finite consumption and delivers $\mathcal{V}^* = \hat{\mathcal{V}}$.

(c) Upper limit of borrowing. First note that borrowing b is bounded below by $-\alpha$, where $b=-\alpha$ is feasible and implies $\mathcal{V}=\mathcal{U}(\alpha\rho)$. So we are concerned with establishing a closed upper bound on borrowing b.

By Proposition 6, we know that nonnegative consumption implies the upper bound on leverage

$$\frac{\alpha+b}{\alpha} \leq \frac{\rho}{\rho-\left(\theta_1-\sum_i \Delta(\theta_i)\mathcal{Q}(\theta_i)\kappa\right)}.$$

Take a sequence of contracts with the property that $b^s \to \bar{b}$, where $\bar{b}$ is the upper limit of borrowing. By Proposition 6, this implies that there exists a state (θ_i, σ) where $\mathcal{X}^s(\theta_i, m_i, \sigma) \to 0$ and $\Delta(\theta_i, \sigma) > 0$. It follows that $\lim_{x\to 0}\mathcal{U}(x) = -\infty$. From here, we can apply the same arguments from part (a) of this proof to show that the sequence does not converge to a supremum, and, therefore, that an interior optimal borrowing allocation exists. □

Proof of Theorem 2. **Audit technologies.** Assumption 1 does not guarantee informativeness of audits. Indeed, it is permissible that there may exist states θ_i, θ_j with the property that $\forall\sigma$ subject to $\Delta(\theta_i, \sigma) > 0 \vee \Delta(\theta_j, \sigma) > 0$, $\Delta(\sigma|\theta_i) = \Delta(\sigma|\theta_j)$. If so, we say that the audit technology is *uninformative* with respect to revenue states j, i.

Let l denote the lowest state for which all the audit technology is uninformative with respect to state l and all states greater than l. If the audit signal were perfect, then l would equal n, the number of states; if the audit signal were independent of the revenue state, then l would equal 1.

We proceed to demonstrate that for κ sufficiently low, standard debt contracts with $q_i = 1$ for $i < l$ and $q_i = 0$ for $i \geq l$ are optimal. We start by demonstrating that there exists an optimal contract that takes the form of standard debt; then we show that standard debt is exclusively optimal for small increases in audit costs.

Lemma 2. *Let $\kappa = 0$. Then there exists an optimal contract that takes the form of a standard debt contract with*

$$q(\theta_i) = \begin{cases} 0 & \forall i \geq l \\ 1 & \textit{otherwise}. \end{cases}$$

Proof of Lemma 2. By Proposition 5, an optimal contract exists. To see that optimal contracts exist with $q(\theta_i) = 0$ for $i \geq l$, refer to the first order condition for repayments (Proposition 1, (8)), which shows that if the audit signal σ is independent of the revenue state for revenue pair i, j, then repayments following reports m_i, m_j are independent of audit signals. It follows that repayments following reports for all states $i \geq l$ are not dependent on audit signals, and these signals can be discarded (or $q(\theta_i) = 0$) for all $i \geq l$.

For reports $i < l$, note that audits do not impose a resource cost and, therefore, can be employed with probability 1 at no cost. □

Lemma 3. *Following Lemma 2, consider an optimal standard debt contract with* $q(\theta_i) = 0$ $\forall i \geq l$ *and* 1 *otherwise. For each* $i < l$*, there is no optimal contract with* $q(\theta_i) < 1$.

Proof. By Proposition 1, (8), we know that the optimal repayment allocations following report i will be contingent on the audit signal σ if and only if there exists $j > i$ such that (i) the audit technology is informative with respect to revenue states i, j and (ii) the incentive compatibility constraint between the two revenue states is binding ($\mu_{j,i} > 0$).

(i) By construction of l, the audit technology is informative with respect to revenue states θ_{l-1}, θ_l. This means that the conditional distribution of audit signals differs across these two revenue states. Let θ_i be a revenue state with $i < l$. The audit technology must be informative with respect to either the pair θ_i, θ_{l-1} or the pair θ_i, θ_l, or both.

(ii) Consider Proposition 1, (8). Assumption 3 ensures that the ratio $\frac{\Delta(\theta_j|\sigma)}{\Delta(\theta_i|\sigma)}$ is finite valued for all σ. It follows that if the incentive constraints $\text{ICC}_{j,i}$ were not binding for all $j > i$, then the repayment mapping $\mathcal{Z}(m_i, \sigma)$ would not be contingent on σ. We also know that the signal m_l is not audited by Lemma 2. This generates a contradiction: if repayments are not contingent on audit signals for either reports m_l or m_i and $\theta_l \neq \theta_i$, then the incentive constraint $\text{ICC}_{l,i}$ would be binding for any optimal repayment profile.

Combining (a) and (b) with Proposition 1, we observe that the optimal repayment allocations under $q(\theta_i) = 1$ cannot be replicated with $q(\theta_i) < 1$ for $i < l$. Therefore, no optimal contract exists such that $q(\theta_i) < 1$. □

Proof of Theorem 2 (continued). Let $\mathcal{C}$ be an optimal contract with some strictly positive κ and at least one i for which $q_i \in (0, 1)$. If $i \geq l$, then, similarly to the case described in Lemma 2, optimal repayments are not contingent on the signal revealed following report m_i, and the audit probability q_i can be set equal to 0 without having any effect on expected utility or incentive compatibility, and relaxing the participation constraint. This generates a contradiction; the original contract $\mathcal{C}$ must not have been optimal.

For $i < l$, then we can decompose the welfare effects of an increase in q_i to 1 into two components: a risk sharing benefit and a budgetary cost, denoted δ_1 and δ_2 in units of expected welfare.

First, we perturb the contract to set $q_i' = 1$ for all $i < l$ and $q_i \in (0, 1)$. We then set consumption allocations optimally conditional upon the new audit strategy, holding expected consumption constant. Denote this updated consumption schedule x'. By Lemma 3, the updated consumption schedule (i) leaves consumption allocations contingent on audit signals, (ii) for each i, leaves the incentive compatibility constraints $\text{ICC}_{j,i}$ binding for some $j > i$, and (iii) cannot be replicated with a lower audit probability q_i. It follows that the adjustment in repayment schedule must result in a strictly positive welfare gain, $\delta_1 > 0$.

To obtain an upper bound on the welfare cost of the tightened participation constraint, consider the allocation $\underline{x}' := \min_{k,\sigma} \mathcal{X}'(\theta_k, m_k, \sigma)$. It is clear from Assumption 1 that the θ_k that satisfies this minimand is $\theta_k = \theta_1$. Consider a perturbation that funds the increase in audit costs by reducing the consumption allocation $\underline{x}'$. This perturbation relaxes any initially binding incentive constraint $\mathrm{ICC}_{j,1}$ for any j, and leaves all other binding incentive constraints unaffected. By Jensen's inequality, the welfare cost of this perturbation is bounded above by

$$\delta_2 < \sum_{i|q_i\in(0,1)} (\alpha+b)\pi_i(1-q_i)\kappa\mathcal{U}'\Bigl(\underline{x}' - \sum_{i|q_i\in(0,1)} \pi_i(1-q_i)\kappa\Bigr).$$

By Proposition 5, $\underline{x}'$ exists and is strictly positive; further, the utility allocation $\mathcal{U}(\underline{x}') \in (-\infty, \infty)$. Therefore, when κ is small, δ_2 exists, and as $\kappa \to 0$, we have $\delta_2 \to 0$. It follows that for sufficiently low κ, $\delta_2 < \delta_1$, and there exists a standard debt contract that is superior to the initial contract $\mathcal{C}$. □

Proof of Proposition 6. From the budget constraints (1), we have $\mathcal{X}(\theta_j, m_i, \sigma) = \mathcal{X}(\theta_i, m_i, \sigma) + (\alpha + b)(\theta_j - \theta_i)$. By assumption, $\Delta(\theta_i, \sigma|q_i) > 0\ \forall\theta_i, \sigma$. If consumption is nonnegative for all truth-telling states, $\mathcal{X}(\theta_i, m_i, \sigma) \geq 0\ \forall\theta_i, \sigma$, then it follows that the expectation on the right hand side of the incentive compatibility constraint $\mathrm{ICC}_{j,i}$ is bounded below as $\sum_\sigma(\theta_j, \sigma|q_i)\mathcal{U}(\theta_j, m_i, \sigma) \geq \mathcal{U}((\alpha+b)(\theta_j - \theta_i))$. Incentive compatibility requires

$$\sum_\sigma(\theta_j, \sigma|q_j)\mathcal{U}(\theta_j, m_j, \sigma) \geq \sum_\sigma(\theta_j, \sigma|q_i)\mathcal{U}(\theta_j, m_i, \sigma) \geq \mathcal{U}\bigl((\alpha+b)(\theta_j - \theta_i)\bigr),$$

which implies

$$\sum_\sigma(\theta_j, \sigma|q_j)\mathcal{X}(\theta_j, m_j, \sigma) \geq (\alpha+b)(\theta_j - \theta_i)$$

by Jensen's inequality. Using the budget constraint (1), we convert this inequality into an upper bound on repayments:

$$\sum_\sigma(\theta_j, \sigma|q_j)\bigl[(\alpha+b)\theta_j - \mathcal{Z}(m_j, \sigma)\bigr] \geq (\alpha+b)(\theta_j - \theta_i)$$

$$\sum_\sigma(\theta_j, \sigma|q_j)\mathcal{Z}(m_j, \sigma) \leq (\alpha+b)\theta_i.$$

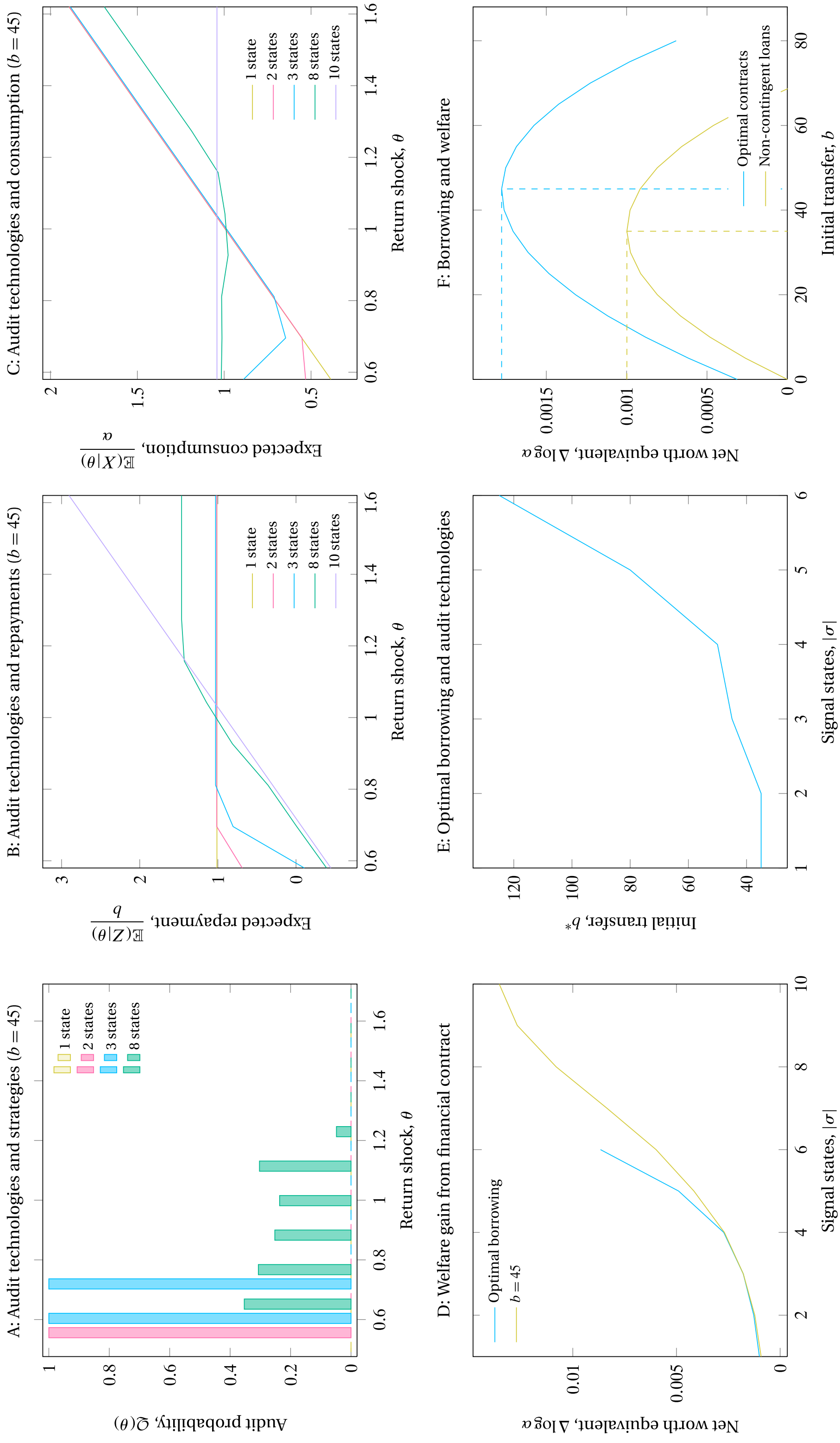
Now, substituting this into the participation constraint (6),

$$\sum_{i,\sigma}(\theta_i, \sigma|q_i)\mathcal{Z}(m_i, \sigma) \leq (\alpha+b)\theta_1$$

$$b\rho + \sum_{i=1}^{n}\Delta(\theta_i)q_i(\alpha+b)\kappa \leq (\alpha+b)\theta_1,$$

which can be rearranged to complete the proof. □

Appendix B: Figures



Notes: In panels (D) and (F), the net worth equivalent welfare gain is the log change in initial net worth α that would leave the entrepreneur indifferent between the optimal contract and financial autarky.

Figure 1. Numerical example.

References

Altman, Edward (1984), "A further empirical investigation of the bankruptcy cost question." *Journal of Finance*, 39, 1067–1089. [902, 907]

Bazerman, Max H., George Loewenstein, and Don A. Moore (2002), "Why good accountants do bad audits." *Harvard Business Review*, 80, 96–102. [887]

Bernanke, Ben S., Mark L. Gertler, and Simon Gilchrist (1999), "The financial accelerator in a quantitative business cycle framework." In *Handbook of Macroeconomics*, first edition, volume 1, Part C, Chapter 21 (J. B. Taylor and M. Woodford, eds.), 1341–1393, Elsevier. [891]

Bertsekas, Dimitri P. (1995), *Nonlinear Programming*. Athena Scientific, Belmont, Massachusetts. [912]

Bolton, Patrick (1986), *The Role of Contracts in Industrial Organisation Theory*. Ph.D. thesis, London School of Economics and Political Science. [895]

Bolton, Patrick (1987), "The principle of maximum deterrence revisited." University of California at Berkeley, Economics Working Paper No. 8749. [888, 890, 898, 903]

Border, Kim C. and Joel Sobel (1987), "Samurai accountant: A theory of auditing and plunder." *Review of Economic Studies*, 54, 525–540. [890, 905]

Duncan, Alfred and Charles Nolan (2017), "Disputes, debt and equity." Unpublished paper, School of Economics, University of Kent, SSRN 3069639. [909]

Gale, Douglas and Martin Hellwig (1985), "Incentive-compatible debt contracts: The one-period problem." *Review of Economic Studies*, 52, 647–663. [889, 890, 891]

Grossman, Sanford J. and Oliver D. Hart (1983), "An analysis of the principal-agent problem." *Econometrica*, 51, 7–45. [898]

Harsanyi, John C. (1973), "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points." *International Journal of Game Theory*, 2, 1–23. [909]

Herranz, Neus, Stefan Krasa, and Anne P. Villamil (2015), "Entrepreneurs, risk aversion, and dynamic firms." *Journal of Political Economy*, 123, 1133–1176. [891, 892, 902, 907]

Krasa, Stefan and Anne P. Villamil (2000), "Optimal contracts when enforcement is a decision variable." *Econometrica*, 68, 119–134. [891]

Mangasarian, Olvi L. and Stan Fromovitz (1967), "The Fritz John necessary optimality conditions in the presence of equality and inequality constraints." *Journal of Mathematical Analysis and Applications*, 17, 37–47. [910]

Melumad, Nahum D. and Dilip Mookherjee (1989), "Delegation as commitment: The case of income tax audits." *RAND Journal of Economics*, 20, 139–163. [891, 893]

Mookherjee, Dilip and Ivan Png (1989), "Optimal auditing, insurance, and redistribution." *Quarterly Journal of Economics*, 104, 399–415. [890, 901]

Townsend, Robert M. (1979), "Optimal contracts and competitive markets with costly state verification." *Journal of Economic Theory*, 21, 265–293. [889, 890, 891, 905]

Warner, Jerold B. and Martin J. Gruber (1977), "Bankruptcy costs: Some evidence." *Journal of Finance*, 32, 337–347. [902, 907]

Weiss, Lawrence A. (1990), "Bankruptcy resolution: Direct costs and violation of priority of claims." *Journal of Financial Economics*, 27, 285–314. [902, 907]

Co-editor Dilip Mookherjee handled this manuscript.

Manuscript received 28 June, 2016; final version accepted 8 January, 2019; available online 25 January, 2019.