State Verification under Limited Capacity

Lukasz A. Drozd and Ricardo Serrano-Padial*

Abstract

Under asymmetric information enforcement of financial contracts often involves some form of costly state verification. Here we enrich the basic theory of costly state verification by considering frictions that limit the adjustment of enforcement capabilities in the short-run. Specifically, in our model a principal contracts with a population of agents rather than a single agent. To sustain incentives, she must build up enforcement capacity in order to verify their state ex post. We study how this friction affects credit provision and shock propagation. Our analysis sheds new light on such economic phenomena as credit crunches or the link between the accumulation of enforcement infrastructure and economic growth, as recently emphasized by the theory of State Capacity (Besley and Persson, 2009, 2010).

Keywords: costly state verification, state capacity, financial accelerator, credit crunch, global games, uniform selection, coordination

JEL codes: D82,D84,D86,E22,E32,G21,O16,O17,O43.

* Drozd: The Wharton School of the University of Pennsylvania. Serrano-Padial: Department of Economics, University of Wisconsin-Madison. We thank Dean Corbae, Patrick Kehoe, Antonio Penta, Marzena Rostek, Joel Sobel, Marek Weretka and audiences at UW-Madison and the 2013 Fall Midwest Theory Conference for comments. Drozd acknowledges the financial support from the Rodney White Center at the Wharton School of the University of Pennsylvania. All remaining errors are our own.
1 Introduction

Under asymmetric information financial contracting often necessitates some form of costly state verification. Such a need arises when it is desirable to make contracted terms dependent on events that can only be observed by the uninformed party at a cost. For instance, in financing of risky investment projects, it is often necessary to make repayment conditional on the success of the project. However, to avoid opportunistic behavior, lenders must verify the outcome of the project in the case of at least some agents who report low returns. The idea of costly state verification has helped explain such key features of the credit markets as incompleteness of contracts, credit spreads, borrowing constraints, and relate them to economic fundamentals.

In this paper, we enrich the basic theory (Townsend, 1979; Gale and Hellwig, 1985) by considering frictions that limit the adjustment of enforcement capabilities in the short-run. Specifically, in our model a principal contracts with a population of agents, and to sustain incentives, must build enforcement capacity ex ante to verify their state ex post. Our approach thus contrasts with the standard framework, which assumes that the probability of state verification that each individual agent faces can be sustained by the principal independently of the collective actions taken by other agents in the economy.

Formally, in our model the ex-post probability $P$ of state verification faced by agents reporting low returns is subject to a constraint:

$$P \leq \frac{X}{1 - \psi(P)},$$

where $X$ is the measure of agents whose state the principal can verify, i.e. her enforcement capacity, and $1 - \psi(P)$ is the fraction of agents who report low returns in response to $P$.

Under the standard approach to costly state verification, such a constraint is absent, implying that economy’s enforcement infrastructure $X$ can be always flexibly adjusted so as to ensure that
verification probability $P$ can be chosen independently of what agents report (which determines $\psi(P)$). In contrast, in our model those reports crucially affect $P$ and thus individual incentives. Accordingly, such a capacity constraint introduces an enforcement externality that affects credit provision.

The presence of such a constraint brings about two important questions that we study in this paper. The first question is how an economy that slowly accumulates enforcement infrastructure grows out of such a constraint, and how it affects its transitional dynamics. This is particularly relevant to the recent theory of State Capacity by Besley and Persson (2009, 2010), who emphasize the role of political economy frictions in the development of a State enforcement infrastructure that guarantees the well-functioning of credit markets and taxes.

The second question is how, since maintaining spare capacity is costly, enforcement capacity can become binding in response to shocks. In this context, our model offers a new theory of credit crunches, which is supported by the recent data. As Figure 1 forcefully illustrates, the enforcement infrastructure dealing with mortgage defaults was overwhelmed during the recent financial crisis, which greatly affected the financial incentives of homeowners to default due to severely lengthened foreclosure timelines—implying an option of “free renting” for an extended period of time.\footnote{Foreclosure timeline is the length of time between initial mortgage delinquency and completion of foreclosure. In some states it increased up to three years. Extended foreclosure timelines enable mortgage defaulters to live in their homes without making housing payments, thus providing an implicit transfer and liquidity benefit. For further information refer to Calem et al. (2013).}

The paper presents a general theory of enforcement of binary contracts under asymmetric information. We show how the presence of ex post capacity constraints changes the predictions of existing models by affecting both ex ante investments in capacity and the choice of contracts. To answer the above questions, the paper develops two applications of the theory. In the first application we examine the relationship between building a nation’s enforcement infrastructure
and economic development in a model of investment in State capacity a la Besley and Persson (2009). Our analysis shows that enforcement externalities represent a substantial drag for economic growth by reducing the size of credit markets and impairing the ability of the government to raise revenue to invest in enforcement capacity buildup. Moreover, they reduce the effectiveness of initial investments in capacity, making economic development much more sensitive to political economy frictions. In the second application we revisit the celebrated financial accelerator model of Bernanke et al. (1999) in which firms need external financing to invest in production. In this context, capacity constraints link future credit provision to current default rates, creating a new channel for the propagation of credit shocks powerful enough to generate credit crunches.

Apart from the negative impact of enforcement externalities on compliance rates, the main economic insight of our theoretical analysis is that small changes in capacity can lead to substantial jumps in compliance. In the two applications we consider this brings up the need for the principal to build precautionary capacity or, whenever this is inefficient, to choose contracts

Figure 1: The U.S. foreclosure timeline (left axis) and mortgage delinquencies (right axis).
that provide little incentives to deviate. The latter effect leads to underdeveloped credit markets and low government revenue in our first application, and it causes large credit contractions due to increases in the default rate of existing loans in our second application. A key technical challenge in our framework is that enforcement externalities introduce the possibility of multiple equilibria, which can limit the applicability of the theory. We resolve equilibrium indeterminacy using contagion style arguments from the global games literature and propose a new method to analytically characterize the uniquely selected equilibrium. Building on existing selection results (Frankel et al., 2003; Sakovics and Steiner, 2012), our equilibrium characterization can handle virtually unrestricted heterogeneity patterns, making the proposed framework widely applicable to study enforcement frictions in a variety of settings.

In terms of related literature, very few papers consider the consequences of limited enforcement in the context of financial contracts. One exception is the paper by Bond and Rai (2009), who study compliance in the micro-finance context of group lending. Their paper features no asymmetric information and focuses on a speculative attack on the fixed enforcement ability of the lender. Arellano and Kocherlakota (2008) highlight strategic complementarities in default and the possibility of multiple equilibria when there is limited ability to liquidate assets in a model featuring two firms. Their approach and applications are different from ours. Rather than focusing on equilibrium selection under general heterogeneity and on credit provision, they investigate whether the bad equilibrium can lead to sovereign debt crises.

More broadly, our paper is related to the literature that studies tax compliance. Among the few papers studying coordination among taxpayers, Bassetto and Phelan (2008) consider the effect of tax auditing capacity and the associated multiplicity of self-fulfilling equilibria. In this context, Sanchez Villalba (2012) applies a global games approach to a specific setup with identical taxpayers. His focus is on identifying the optimal auditing probability. In contrast, we provide a general framework for enforcement with limited capacity and apply it to the study of
enforcement externalities and their effect on the endogenous choice of capacity and contracts/tax rates.

The rest of the paper is organized as follows. In Section 2 we start from a stripped down version of the two applications to put our theory of enforcement in context. Section 3 and 4 present the theory and analytic results. Section 5 revisits the examples by deriving the implications of the theory. Section 6 provides a glimpse of the fully-fledged applications and Section 7 concludes.

2 Motivating Example

We begin by laying a stylized model of investment with credit that isolates the common features of the aforementioned applications and helps describe the impact of enforcement externalities in an intuitive way.

The economy is populated by a principal and a continuum of entrepreneurs of measure one with access to an investment opportunity. Entrepreneurs are risk neutral, have initial income $y > 0$ and can borrow from the principal to invest. There are two periods, indexed by $t = 0, 1$. At $t = 0$ the principal chooses how much enforcement capacity $X \in [0, 1]$ to build at a cost $c(X)$ and provides credit to entrepreneurs, which they use to invest in a risky project. The goal of the principal is to maximize entrepreneurs’ aggregate payoff net of capacity costs, subject to the constraint that she must earn zero aggregate profits from credit contracts. After these choices are made, entrepreneurs invest all their income $y$ plus the amount borrowed $b$ and the first period ends.

At $t = 1$ each entrepreneur privately learns the outcome of her project, which can be good ($w = 1$) or bad ($w = 0$). The probability that the project fails is $p$. If the project is unsuccessful, which happens with probability $p$, it yields zero returns. A successful project yields $RK$, where
$R > \frac{1}{1-p}$ is per dollar return and $K = y + b$ is the capital invested.\footnote{Returns above $1/(1-p)$ imply that expected returns are positive.} Credit contracts are given by $(C(\delta), \delta)$, where $\delta(w) \in \{0, 1\}$ is the default rule—with 0 meaning default—and $C(\delta)$ the contract terms as a function of default. Specifically, failed entrepreneurs are allowed to default, while successful ones are supposed to repay the loan plus interest. That is, $\delta = w$ and $C(\delta) = (b, rb\delta)$, where $b$ is the size of the loan and $rb\delta$ is the repayment amount.

After project returns are realized agents simultaneously choose whether to repay ($a = 1$) or default ($a = 0$). Upon observing default choices, the principal can verify (henceforth monitor) the project returns of at most a mass $X$ of agents and seize the project returns of any defaulting entrepreneur found to have a successful project. Accordingly, the ex post utility of entrepreneurs $u(a, w, m)$, with $m = 1$ indicating being monitored and $m = 0$ otherwise, is given by

$$u(a, w, m) = \begin{cases} R(y + b)w - rb & a = 1 \\ R(y + b)w & a = 0 \& m = 0 \\ 0 & a = 0 \& m = 1. \end{cases}$$

This minimal setup allows analyzing the impact of enforcement externalities on credit provision and capacity choice. We will focus on addressing two different questions.

- **Limited capacity.** The first question, in the spirit of our application on State capacity, is to understand how limited enforcement resources affect investment and growth once enforcement externalities are taken into account. In the context of the example, this can be done by placing an exogenous upper bound on the choice of $X$ and comparing credit provision with and without enforcement externalities. In the general, dynamic version of the application, such upper bound is endogenously given by the ability of the government to collect taxes in previous periods, which depends on its past capacity to enforce them.
and also involves a coordination game among taxpayers.

- *Shock propagation.* The second question deals with the economic impact of short-run variation in enforcement needs. The simplest way to analyze this issue is to introduce a stochastic shock $\zeta$ to enforcement capacity, which hits the economy after the principal has chosen $X$ but before issuing credit contracts. That is, the principal has residual capacity $X - \zeta$ to enforce newly issued contracts. Such shocks can be thought of as representing the default rate of a pool of pre-existing loans—e.g., sub-prime mortgages at the onset of the financial crisis. Our second application considers a richer setup where project returns $w$ are heterogeneously distributed and contract default rules are endogenously chosen as in Bernanke et al. (1999).

In order to address these questions we first need to characterize the enforcement game played by the principal and the agents. The next section introduces an enforcement model that allows for general distributions of agent types (project returns in this example) and is applicable to the enforcement of any binary mechanism in which a principal faces a large set of agents.

## 3 Theory

We consider a principal who engages in a contracting problem with a set of heterogeneous agents. It is assumed that the contracts have already been signed and the principal only allocates previously accumulated enforcement capacity to enforce compliance with the contracted terms. In what follows next, we begin our analysis with the description of the implied ex post enforcement game. Since the ex ante problem that endogenously determines the fundamentals of the game is specific to a particular application of the game, its description is relegated to Section 5.

Compared to the above example, our general framework exhibits generic binary contracts $(C(\delta), \delta)$ and arbitrary distributions of agent types. Specifically, in the ex post enforcement game
the principal deals with a measure one of non-atomistic agents grouped into a finite set of types $w \in \mathcal{W}$, each with a positive mass. Types are private information of the agents, but have a publicly known distribution $F$; the associated probability measure is denoted by $f$. Agents of each type $w$ are uniformly indexed by $\omega \in [0, f(w)]$.

The principal’s goal is to enforce a binary contractual arrangement $(\mathcal{C}(\delta), \delta)$, where $\mathcal{C}(\delta)$ represents contract terms contingent on $\delta : \mathcal{W} \to \{0, 1\}$, which is a mapping from types to a prescribed binary action that the agent of a given type should take.

After privately learning their types, agents simultaneously choose a binary action $a \in \{0, 1\}$. We say that an agent with type $w$ complies with the contract if she chooses action $a = \delta(w)$. We assume that the mass of agents with $\delta(w) = 0$ is positive and less than one. While the principal does not observe agent types, only their actions, she has the ability to verify the type of at most a measure $X$ of agents. The choice of $X$ is done before types are realized and actions are taken, Hence, $X$ is fixed at the enforcement stage.

Given the contract, agent payoffs—denoted by $U$—are assumed to depend on her type $w$, the action $a$ she chooses, as well as whether she has or has not been subject to state verification by the principal—denoted by a binary indicator $m \in \{0, 1\}$. The utility function satisfies several conditions stated below. They guarantee that, first, state verification incentivizes compliance, and second, that the contractual prescription of taking action $a = 1$ is what agents find costly to do—i.e. agents would like to choose $a = 0$ whenever possible.

**Assumption 1.** $U(a, w, m)$ satisfies for all $w \in \mathcal{W}$

(i) $U(a, w, 0) \geq U(a, w, 1)$ for all $a \in \{0, 1\}$;

(ii) $U(0, w, 0) > U(1, w, 0)$; and $U(0, w, 1) > U(1, w, 0)$ if $\delta(w) = 0$;

(iii) $U(\delta(w), w, 1) > U(1 - \delta(w), w, 1)$. 

Conditions (i) says that agents weakly prefer not to be monitored. Condition (ii) implies that agents with $\delta(w) = 1$ prefer $a = 0$ if they expect to be monitored with sufficiently low probability, while agents with $\delta(w) = 0$ always prefer action $a = 0$, regardless of monitoring probabilities. Finally, (iii) means that agents are better off complying whenever they expect to be monitored with sufficiently high probability. It is easy to check that payoffs (2) in the example satisfy Assumption 1.

The above preferences imply that the principal will never verify agents who report action $a = 1$, since only those with $\delta(w) = 1$ may find optimal to do so. Thus, what is relevant for agents in our model is their belief about the probability of state verification when they choose action $a = 0$. We denote this belief by $P \in [0, 1]$. Given this belief—and privately realized types—agents choose $a$ to maximize their expected utility given by $(1-a)(PU(a, w, 1) + (1-P)U(a, w, 0)) + aU(a, w, 0)$.

To induce compliance with the contract, after observing agents actions the principal chooses the monitoring probability $P$ faced by agents who choose $a = 0$, subject to the constraint that she can verify at most a measure $X$ of agents.\(^3\) Agents are aware of this and anticipate that $P$ will depend on the aggregate compliance of agents with $\delta(w) = 1$ when the constraint binds. Since agents with $\delta(w) = 0$ always choose $a = 0$, the mass of agents choosing $a = 1$ is

$$\psi := \sum_{w \in W} \int_{\omega \in [0, f(w)]} a(w, \omega) d\omega,$$

where $a(w, \omega)$ denotes the equilibrium strategy chosen by an agent of type $w$ and index $\omega$. Accordingly, the measure of agents choosing $a = 0$ is $1 - \psi$, leading to capacity constraint

$$P(1 - \psi) \leq X. \quad (3)$$

\(^3\)We assume that the principal cannot condition verification on agents’ indices. We find such a possibility unrealistic in any relevant applications of the model.
We next characterize equilibrium in the enforcement game by focusing on two cases. In the first case we assume that agents perfectly observe $X$ and show that it leads to self-fulfilling beliefs and multiplicity. To address the shortcoming of having multiple equilibria, we then consider an selection mechanism based on global games methods. In the global game version of our model agents privately observe a noisy signal of $X$, which exhibits a unique equilibrium. We characterize the resulting unique equilibrium by taking signal noise to zero. As a result, our refinement implies identical finite order beliefs regarding $P$ as in the game without noise. However, in the limit case there is no common knowledge of $X$. We also show that o equilibrium selection is determined by the fundamentals of the game such as $F$ and $X$ and independent of the noise structure.

4 Equilibrium Characterization

We begin our analysis by recasting agent types so as to conveniently describe the kind of heterogeneity that is relevant for the game, namely heterogeneity in the verification probability $P$ that makes agents indifferent between taking actions 0 and 1. The following lemma is instrumental to obtain such a reformulation:

Lemma 1. For each $w \in W$ such that $\delta(w) = 1$ there exists a unique $\theta(w) \in (0,1)$, given by

$$\theta(w) = \frac{U(0, w, 0) - U(1, w, 0)}{U(0, w, 0) - U(0, w, 1)},$$ (4)

such that an agent with type $w$ chooses $a = 1$ if she believes that $P$ is greater than $\theta(w)$ and chooses $a = 0$ if she believes $P < \theta(w)$. Moreover, all types $w \in W$ such that $\delta(w) = 0$ strictly prefer $a = 0$ over $a = 1$, regardless of their monitoring beliefs.

The lemma provides a mapping from intrinsic types $w \in W$ to what we refer as induced
indifference types \( \theta(w) \). Indifference types embed all the relevant information about agents’ preferences over actions since they represent points of indifference across actions in the space of types \( \mathcal{W} \). Consequently, the set of indifference types \( \theta(w) \) associated to agents with \( \delta(w) = 1 \) and its distribution in the population is all we need to solve the game. We denote such a set by \( \Theta \), which has a smallest element \( \underline{\theta} > 0 \) and a largest one \( \overline{\theta} < 1 \) by Lemma 1.

In what follows next, we thus switch to the space of indifference types. To implement this conversion we note the following. First, the distribution \( F \) of intrinsic types induces some distribution \( G \) over indifference types, with associated probability mass function \( g \). Second, we assign types \( w \) with prescribed \( \delta(w) = 0 \) an indifference type strictly greater than one. Hence, \( 1 - G(1) > 0 \) represents the mass of these agents in the overall population and \( G(1) > 0 \) represents the mass of agents with \( \delta(w) = 1 \). In this context, \( X \) and \( G \) represent the relevant economic fundamentals as far as enforcement is concerned.

In our example the application of (4) to entrepreneur payoffs (2) yields indifferent types \( \theta(0) > 1 \)—default is a dominant strategy for unsuccessful entrepreneurs—and \( \theta(1) = \frac{r - b}{Ry + b} \) for successful entrepreneurs. Indifferent types in this example follow a binary distribution with \( G(\theta) = 0 \) if \( \theta < \theta(1) \) and \( G(\theta) = 1 - p \) if \( \theta \in [\theta(1), 1] \). The indifference type of successful entrepreneurs shows that a higher leverage \( b/(y + b) \) or a higher interest rate \( r \) makes default more attractive. This may leave a capacity constrained principal with no other way to incentivize repayment but to limit credit provision (lower \( b \)). That is, limited capacity introduces a credit constraint. As we show in Section 5, the presence of enforcement externalities can severely tighten credit constraints.

\(^4\)Note that \( G \) is a mixture of a distribution with finite support \( \Theta \subset [\underline{\theta}, \overline{\theta}] \), where \( 0 < \underline{\theta} \leq \overline{\theta} < 1 \) and a point mass above one that pools all the agents with prescribed action \( \delta(w) = 0 \).
4.1 Equilibria under Common Knowledge

We begin our analysis of the game under the assumption that $X$ is common knowledge. In such a case, given the strategy of other players $a(w, \omega)$, an agent’s belief about monitoring probability $P$ is given by

$$ P = \min \left\{ \frac{X}{1 - \psi}, 1 \right\}, $$

which follows from (3). The equation highlights that the strategy of other agents crucially influences the equilibrium enforcement probability $P$, creating a feedback loop between compliance of others and incentives to comply.

The analysis of the equilibrium effects of such enforcement externalities is challenged by the fact that self-fulfilling beliefs may lead to multiple equilibria. Intuitively, if a sufficiently high fraction of agents believes that fewer agents will comply, it may well justify such a decision in equilibrium. Because of this, the predictions of the model cannot be determined within the model, which limits what we can learn from it.

To illustrate how self-fulfilling beliefs can lead to multiple equilibria, consider again the example. Figure 2 plots the relationship between agents’ beliefs about $P$ and the actual fraction of agents being monitored, which is given by the RHS of (5). To pin down $\psi$ note that, by Lemma 1, the best response of each agent is to comply if her indifference type $\theta$ is less than $P$ and to not do so if $\theta > P$. That is, $\sum_{\theta < P} g(\theta) \leq \psi \leq G(P)$. Hence, equilibria in the figure are given by beliefs $P$ at which $\min \left\{ \frac{X}{1 - G(P)}, 1 \right\}$ crosses or touches the diagonal.

Note that the actual fraction monitored is a step function of $P$ due to the discreteness of $G$. Hence, multiplicity of equilibria arises when this function crosses or touches the diagonal at the jump point $\theta(1)$. Intuitively, this means that when all agents of type $\theta(1)$ coordinate on switching from $a = 0$ to $a = 1$ the monitoring probability they face jumps above $\theta(1)$, justifying such a switch, while if they don’t switch they face $P < \theta(1)$ rationalizing the choice of $a = 0$. 

12
Accordingly, there is one equilibrium in which no successful entrepreneur complies and one with full compliance. The figure also shows the presence of a partial compliance equilibrium in which successful entrepreneurs are indifferent between repaying and defaulting since the mass of repaying agents $\psi$ is chosen to satisfy $\theta(1) = X(1 - \psi) = P$.

Our first result derives the necessary and sufficient condition for multiplicity of equilibria for general $G$. The condition boils down to the existence of a type $\theta \in \Theta$ at which $\frac{X}{1 - G(\theta)}$ crosses or touches the diagonal, thus generalizing the above example with binary $G$. A caveat is that when $G$ has a rich support, e.g., when $G$ is a discrete approximation of a continuous distribution, multiplicity may just be a by-product of such a discretization and involve small differences in compliance affecting the behavior of only one type in $\Theta$. Accordingly, we also provide conditions for the existence of multiplicity involving changes in behavior of more than one type across equilibria, which we refer to as global multiplicity. These conditions not only
justify our later application of an equilibrium selection mechanism, but more crucially they
tells us when our selection mechanism will lead to big changes in compliance in response to
small changes in enforcement capacity.\footnote{Intuitively, if there are multiple equilibria, the equilibrium selection mechanism selects one of the equilibria of the original game depending on the value of the fundamentals. If the value of fundamentals changes, the selected equilibrium may well be different.}

Define $G^{-}(\theta)$ as $\limsup_{\theta' \uparrow \theta} G(\theta')$, and note that it is given by $\sum_{\theta' < \theta} g(\theta')$ when $\theta \in \Theta$ and $G(\theta)$ otherwise.

**Definition 1.** Let $\underline{X} = \min_{\theta \in \Theta} \theta (1 - G(\theta))$ and $\bar{X} = \max_{\theta \in \Theta} \theta (1 - G^{-}(\theta))$.\footnote{Note that: $\underline{X} > 0$ and $\bar{X} < \bar{X} < 1$, since $\bar{\theta} > 0$ and $\bar{\theta} < 1$.}

**Theorem 1.** For all $\underline{X} < X < \bar{X}$ there exists

1. Multiplicity iff $\theta (1 - G(\theta)) \leq X \leq \theta (1 - G^{-}(\theta))$ for some $\theta \in \Theta$; and,

2. Global multiplicity iff $\theta (1 - G^{-}(\theta)) \geq X \geq \theta' (1 - G(\theta'))$ for some $\theta, \theta' \in \Theta$ with $\theta < \theta'$.

If $X < \underline{X}$ then $\psi = 0$ in equilibrium, while $\psi = G(1)$ for all $X > \bar{X}$.

Proofs are relegated to the appendix. As mentioned, the first condition implies the fraction monitored crossing/touching the diagonal at some type in the support of $G$, i.e., $\frac{X}{1 - G^{-}(\theta)} \leq \theta \leq \frac{X}{1 - G(\theta)}$. In the same vein, the condition for global multiplicity implies either the existence of a second crossing at some type $\theta' \in \Theta$ or that the only crossing happens at $\theta < \bar{\theta}$ and the fraction monitored stays above the diagonal until it hits one. The intuition is similar to the one provided in the binary case: a coordinated switch to $a = 1$ of several types leads to a realized monitoring probability above their types, thus rationalizing such a switch.

Figure 3 illustrates the result in a slightly different way than in the above example, namely by plotting the fixed point equation $\theta = \frac{X}{1 - G(\theta)}$ as the intersection of capacity $X$ (red line) and the function $\theta (1 - G(\theta))$ (blue line). Such graphical representation makes it easier to see...
how the set of equilibria changes with enforcement capacity. Equilibrium compliance rates are given by $G(\theta)$, with $\theta$ being the type—not necessarily in $\Theta$—at which the lines intersect or by $\theta = 1$ if the blue line hits $\theta = 1$ below $X$, indicating that the capacity constraint is not binding when all agents comply. The three equilibria on the left involve low compliance and, importantly, small changes in compliance across them since they only affect the behavior of a single type. In contrast, the two equilibria on the right exhibit much higher compliance rates. These two sets of equilibria illustrate the existence of global multiplicity, which is associated to the overall shape of $\theta(1 - G(\theta))$ being non-monotonic.

Figure 3: Multiplicity of Equilibria Under Common Knowledge.
4.2 Selection Mechanism

Multiplicity of equilibria relies on the assumption that agents have common knowledge of $X$. As the global games literature suggests, even infinitesimal uncertainty in this respect can induce large strategic uncertainty about the equilibrium action of others, potentially eliminating equilibrium multiplicity. We next show that our game admits such a selection and provide an analytic characterization of the unique equilibrium in the limit, i.e., when uncertainty vanishes.

The global game version of our enforcement game is identical to the complete information game except that each agent of type $\theta \in \Theta$ receives a signal $x = X + \nu \eta$ where $0 < \nu < \bar{\nu} := \min \{\theta(1 - G(1)), 1 - \bar{\theta}\}$ is a scale factor and $\eta$ is a random variable, independent of $X$, with distribution $H_\theta$.\(^7\) We assume that $H_\theta$, which is allowed to vary by type, has support $[-1/2, 1/2]$ and a continuous density bounded away from zero in its support. Noise terms $\eta$ are i.i.d. for agents with the same type, and also independently distributed across types. Before receiving the signals, agents do not have any prior knowledge of $X$ and thus their prior is the uniform distribution on $[0, 1]$.\(^8\) That is, an agent with signal $x$ and type $\theta$ believes that $X = x - \nu \eta$ with $\eta \sim H_\theta$.

In this environment our goal is to characterize equilibrium as $\nu$ goes to zero, i.e., as signal noise vanishes. We look for strategy profiles that survive iterated strict dominance. But first we establish that there exists a unique equilibrium of the game:

**Theorem 2.** The game has an essentially unique equilibrium.\(^9\) Equilibrium strategies are characterized by cutoffs $k(\theta)$ on signal $x$, such that all agents of type $\theta$ choose action $a = 1$ if $x \geq k(\theta)$ and $a = 0$ otherwise.

---

\(^7\)The upper bound on $\nu$ is helpful to show uniqueness of equilibrium by ensuring that the boundary issues associated with signals close to 0 or 1 only arise when capacity is such that all agents have a dominant strategy.

\(^8\)Our results do not hinge upon the uniform prior assumption. As Frankel et al. (2003) show, equilibrium selection arguments work in the limit as signal error goes to zero since any well-behaved prior will be approximately uniform over the small range of $X$ that are possible given an agent’s signal.

\(^9\)In the sense that equilibrium strategies may differ in zero probability events.
The uniqueness result follows from the fact that games with strategic complementarities feature a smallest and largest Nash equilibrium, both exhibiting monotone strategies (Milgrom and Roberts, 1990), and from a modified version of the usual translation argument showing that there can be at most one equilibrium in monotone strategies (see e.g. Frankel et al., 2003).

The basic idea behind this argument is as follows. Let \( k = (k(\theta))_{\theta \in \Theta} \) denote a profile of monotone or cutoff strategies and let \( P_\theta[\cdot|k; k(\theta)] \) and \( E_\theta[\cdot|k; k(\theta)] \) be the probability and expectation conditional on an agent of type \( \theta \) receiving a signal equal to her cutoff \( x = k(\theta) \) and on agents following strategy profile \( k \). Then, observe the following. First, for \( k \) to be an equilibrium profile it must satisfy the following set of indifference conditions:

\[
E_\theta[P|k; k(\theta)] = \theta \quad \forall \theta \in \Theta. \tag{6}
\]

Second, note that shifting all agents’ thresholds by \( \Delta > 0 \), would have the following three effects on the expected monitoring probability of an agent receiving a signal exactly equal to her threshold (henceforth the threshold type): (i) there would be less compliance for any fixed profile of realized signals, which in isolation would drive up expected \( P \); (ii) the other agents would be expected to receive higher signals, which would drive up compliance levels and lower expected \( P \); and (iii) capacity would be expected to be higher overall, which would drive up expected \( P \). Now, it turns out that additivity of noise implies that (i) and (ii) exactly cancel out, and thus the overall effect is driven by (iii), implying that expected \( P \) goes up as thresholds are raised this way. While this applies to the case when thresholds shift by the same magnitude, the argument generalizes to asymmetric shifts, leading to a unique solution of the system (6).

Intuitively, uniqueness of the solution is brought about by the persistence of strategic uncertainty due to the presence of noise. This is reflected by the described above ‘translation argument’. Such an invariance of beliefs is also related to another property of equilibrium be-
beliefs, which turns out key to characterize equilibrium thresholds as noise vanishes. This property is known as the “Laplacian property” in symmetric global games with binary actions \( a \in \{0, 1\} \) (Morris and Shin, 2003), and has been extended to environments with a restricted form of preference heterogeneity by Sakovics and Steiner (2012). The Laplacian property essentially states that an agent receiving the threshold signal believes that the aggregate action in the game (i.e., the compliance rates of types in \( \Theta \)) is uniformly distributed on \([0, 1]\).

In environments where primitive types are heterogeneous but all agents have a common cutoff \( c \) on the aggregate action such that \( a = 1 \) is preferred by all agents if the aggregate action is above \( c \), Sakovics and Steiner (2012) show that the Laplacian property still holds ‘on average.’ That is, even though agents of different types may have close by different signal thresholds, the weighted average of their beliefs about the aggregate action is uniformly distributed in \([0, 1]\), with weights given by the mass of agents associated to each threshold. In essence, as Sakovics and Steiner (2012) explain, the event of receiving a threshold signal, i.e., \( \{x : k(\theta) = x, \theta \in \Theta\} \), is uninformative about the type of the agent. Accordingly, the Laplacian property applies to the belief about the aggregate action conditional on receiving a threshold signal but ‘unconditional’ on type. The authors refer to this property a the “belief constraint,” which we adopt. They use it to show that all agent strategies converge to a unique threshold strategy as the noise is taken to zero. This limit threshold is independent of the particular shape of noise distributions, a desirable property called uniform selection of equilibrium in global games (Frankel et al., 2003).

Unfortunately, under unrestricted heterogeneity of indifference types typically there no unique limit threshold. For instance, in the example of Figure 3, an agent with type \( \theta \) will be characterized by a threshold close to \( \overline{X} \), as otherwise her expected monitoring probability would be higher than \( \theta \), regardless of the behavior of other agents. In contrast, an agent with a higher type, say \( \overline{\theta} \), must face a much higher expected monitoring probability for her indifference condition to be satisfied. But then her threshold cannot be close to the one of type-\( \theta \). This is because
when $\nu$ is sufficiently small she faces monitoring probabilities—conditionally on her threshold signal—that are lower than $k(\bar{\theta}) + \nu/2$. In this context, were $k(\bar{\theta})$ close to $\bar{X}$, her expected monitoring probability would be much lower than $\bar{\theta}$.

Nonetheless, our key insight is that—as the next lemma states—the belief constraint applies to any subset of types: the ‘average belief’ of agents with types in a subset $\Theta' \subseteq \Theta$ about the proportion of those agents in that subset choosing $a = 1$ is uniformly distributed in $[0, 1]$.

Let $\varphi(X, k, \Theta')$ be the proportion of agents with types in $\Theta'$ choosing $a = 1$ when capacity is $X$, agents follow strategy profile $k$ and receive signals $x = k(\theta')$ for all $\theta' \in \Theta'$. That is,

$$\varphi(X, k, \Theta') = \frac{1}{\sum_{\Theta'} g(\theta')} \sum_{\Theta'} \left(1 - H_{\theta'} \left(\frac{k(\theta') - X}{\nu}\right)\right) g(\theta').$$  \tag{7}$$

**Lemma 2** (Belief Constraint). If $k$ is the unique equilibrium profile then, for any subset $\Theta' \subseteq \Theta$ and any $y \in [0, 1]$,

$$\frac{1}{\sum_{\Theta'} g(\theta')} \sum_{\Theta'} P_{\theta} \left(\varphi(X, k, \Theta') \leq y | k, k(\theta)\right) g(\theta) = y.$$  \tag{8}$$

The above result is instrumental to characterize equilibrium thresholds as $\nu$ goes to zero, and thus to engage in comparative statics and solve the game. In particular, it allows to derive closed-form solutions for the expected monitoring probabilities for different types and thus implies tractable indifference conditions from which we can obtain $k$. We state this remarkable result in its full generality below. In stating this result we refer to a partition $\Phi = \{\Theta_1, \cdots, \Theta_I\}$ of $\Theta$ as being monotone if $\max \Theta_i < \min \Theta_{i+1}$, $i = 1, \cdots, I - 1$, and denote the lowest and highest elements of $\Theta_i$ by $\theta_i$ and $\bar{\theta_i}$, respectively.

**Theorem 3.** In the limit, as $\nu \to 0$, the equilibrium profile $k$ is given by a unique monotone partition $\Phi = \{\Theta_1, \cdots, \Theta_I\}$ and a unique vector $(k_1, \cdots, k_I)$ satisfying the following conditions:
(i) \( k(\theta) = k(\theta') = k_i \) for all \( \theta, \theta' \in \Theta_i \).

(ii) \( k_i < k_{i+j} \) for all \( i = 1, \cdots, I - 1 \) and \( j = 1, \cdots, I - i \).

(iii) \( \bar{\theta}_i (1 - G(\bar{\theta}_i)) \leq k_i \leq \theta_i (1 - G(\bar{\theta}_i)) \) for all \( i = 1, \cdots, I \).

(iv) \[ \int_0^1 \min \left\{ \frac{k_i}{1 - G(\bar{\theta}_i) + (1 - y) \sum_{\Theta_i} g(\theta)}, 1 \right\} dy = \frac{\sum_{\Theta_i} \theta g(\theta)}{\sum_{\Theta_i} g(\theta)} \text{ for all } \Theta_i \in \Phi. \]

**Corollary 1.** Condition (iv) can be written as

\[
\sum_{\Theta_i} \theta g(\theta) = \begin{cases} 
  k_i \log \frac{1 - G(\bar{\theta}_i)}{1 - G(\bar{\theta}_i)} & \text{if } k_i \leq 1 - G(\bar{\theta}_i) \\
  k_i \left( \log \frac{1 - G(\bar{\theta}_i)}{k_i} + 1 \right) - (1 - G(\bar{\theta}_i)) & \text{if } k_i > 1 - G(\bar{\theta}_i).
\end{cases}
\]

The intuition why it is possible to obtain closed-form solutions of indifference conditions (6) is as follows. On one hand, the beliefs of an agent with type \( \theta \in \Theta_i \) about capacity conditional on receiving her threshold signal \( k_i \) collapse on \( k_i \) as signal noise vanishes. In addition, the agent perfectly learns in the limit that all agents with thresholds lower than \( k_i \) are complying, since their signals are very close to \( k_i \)—and thus above their respective thresholds—when noise is sufficiently small, while those with higher thresholds are not complying. These two facts combined imply that the agent faces a monitoring probability equal to \( \min \left\{ \frac{k_i}{1 - G(\bar{\theta}_i) + (1 - y) \sum_{\Theta_i} g(\theta)}, 1 \right\} \), where \( y \) represents the fraction of agents with threshold \( k_i \) choosing to comply. On the other hand, her beliefs about \( y \) conditional on receiving signal \( k_i \), which are needed to derive her expected monitoring probability, do not collapse as noise vanishes and, furthermore, they are a complicated function of others’ thresholds and noise distributions. However, if we average the indifference conditions of types in \( \theta \in \Theta_i \), the belief constraint generalized by the lemma above tells us that the average belief about \( y \) of threshold types is uniform. Hence, these average indifference conditions can be solved and—combined with some monotonicity restrictions to ensure that incentives are satisfied (condition (iii) in Theorem 3)—the limit thresholds can be found.
Our last theoretical result simply states that pooling of types in equilibrium is naturally tied to the earlier discussion of global multiplicity of equilibria in the common knowledge game. This is illustrated by Figure 4.

**Corollary 2.** There is at least one $\Theta_i \in \Phi$ containing more than one type if and only if there exists global multiplicity for some $X$ in the common knowledge game.

The reason why agents with very different indifference types may pool on the same threshold lies on the shape of $\theta(1 - G(\theta))$. Whenever there is a sufficient mass of types with relatively high $\theta$, $\theta(1 - G(\theta))$ will drop substantially at each of these high types, making its overall shape ‘singled-peaked’ as in Figure 3. What does this shape means in terms of incentives to comply once agents use threshold strategies? It means that, since $k(\theta)$ is monotone in $\theta$ by agents’ indifference conditions, as $X$ is raised some types start to switch from $a = 0$ to $a = 1$. In particular, as the more prevalent types are switched compliance rates jump significantly and so does $P = X/(1 - \psi) = X/(1 - \psi)$, providing strong incentives for higher types to switch at the same time, which leads to a cascade effect of simultaneous switches.

This argument also hints at what equilibrium is selected in the limit: roughly, the one that requires the ‘least’ amount of coordination among agents for them to find their equilibrium action optimal. Naturally, at low capacities incentives are strong for most types to not comply, unless most agents comply, so the selected equilibrium will exhibit very low compliance rates. As capacity gets raised incentives for complying become stronger, placing lower demands on compliance rates for $a = 1$ to be a best response for most types. This is illustrated by the cutoffs in Figure 4.

We finish this section by highlighting two comparative statics consequences of the above results, which play a key role in the applications that we consider next. The first one is the fact that compliance rates respond in a highly non-linear fashion to changes in capacity. Figure
Figure 4: Limit Thresholds

4 provides an example: when capacity is increased from $k_5$ to just below $k_6$ compliance rates remain unaffected, even though a higher compliance level could be achieved in the common knowledge version of the game. However, when the capacity reaches $k_6$ equilibrium jumps from low to full compliance. Note that these effects are closely related to the earlier analysis of multiplicity of equilibrium.

The second consequence is that, compared to the case of flexible capacity, the presence of enforcement externalities leads to the need to build up precautionary capacity. Specifically, under flexible capacity the capacity needed to guarantee a given compliance level, say $G(\theta)$ for some
\(\theta\), is \(\min_{\theta' \geq \theta} \theta' (1 - G(\theta'))\), given that the principal can commit to monitor with intensity \(\theta' \geq \theta\). In contrast, a strictly higher capacity is typically needed to guarantee compliance of at least \(G(\theta)\) due to enforcement externalities, resulting in either unused capacity ex post or suboptimally high monitoring probabilities. This is because—by condition (iii) in Theorem 3—if \(\theta \in \Theta_i\) then the capacity needed to ensure compliance level of at least \(G(\theta)\) is \(k_i \geq \bar{\theta}_i (1 - G(\bar{\theta}_i)) \geq \min_{\theta' \geq \theta} \theta' (1 - G(\theta'))\).

5 Implications of the Coordination Game

Equipped with the equilibrium characterization of Theorem 3 we can apply our enforcement theory to study the effects of short-run rigidities in enforcement capacity. We now address the two questions put forward in our simple example of Section 2 before developing our two applications in the next section.

We use a two-stage approach in each of the problems below. First, we recast our equilibrium and its implied compliance rates in terms of primitive types by expressing \(\theta\), \(G\) and \(g\) in terms of \(w\) and \(F\). The second step is to embed equilibrium compliance rates and ex post payoffs into the principal’s problem of choosing \(X\) and contract terms.

Given contract \((b, r)\), recall that indifference types are \(\theta(0) > 1\) and \(\theta(1) = \frac{r b}{R y + b}\), with \(G(\theta(1)) = 1 - p\). Since there is only one indifference type in \(\Theta\), the application of Corollary 1 is straightforward, implying that type \(\theta(1)\) agents repay the loan if \(X \geq k\) and default otherwise, where \(k\) satisfies

\[
\frac{r}{R y + b} \frac{b}{R y + b} = \begin{cases} 
-k \log p & \text{if } k \leq p \\
(1 - \log k) - p & \text{if } k > p.
\end{cases}
\]

Several things are worth noting. First, the principal must choose capacity and contracts \((b, r)\) so that \(k \leq X\), otherwise every entrepreneur defaults and the principal cannot break even. Second,
the principal sets \( r = \frac{1}{1-p} \) so as to earn zero expected profits. To see why notice that expected profits when \( X \geq k \) are given by \((1-p)rb - b\), which are non-negative if \( r \geq 1/(1-p) \) regardless of the size of the loan \( b \). But then, a higher \( r \) both increases the incentive to default by increasing \( \theta(1) \) and lowers entrepreneurs' aggregate payoff. Finally, since higher \( b \) leads to higher aggregate payoffs given that (per dollar) project returns are constant, the principal chooses the highest \( b \) such that \( k \leq X \). But since the LHS and the RHS of (10) are increasing in \( b \) and \( k \), respectively, \( b \) will be chosen so that \( k = X \). Combining these facts, we have that \( b \) and \( X \) satisfy

\[
\frac{1}{R(1-p)} \frac{b}{y+b} = \begin{cases} 
-X \log p & X \leq p \\
X(1 - \log X) - p & X > p.
\end{cases}
\] (11)

5.1 Limited capacity

The first issue we address is the impact of enforcement externalities when the principal cannot set capacity too high. The simplest way to illustrate it is to compare the optimal provision of credit (i.e., borrowing level \( b \)) for different \( X \) with and without externalities. The former is given by (11). Since the LHS is increasing in \( b \), equation (11) represents a borrowing constraint induced by limited capacity.

Credit provision without externalities is found as follows. The principal commits to monitor with probability \( \theta(1) \), in which case successful entrepreneurs repay and the ex post capacity needed to sustain such a monitoring probability is \( X = p\theta(1) \). Hence, without enforcement externalities borrowing is given by

\[
\frac{1}{R(1-p)} \frac{b}{y+b} = \begin{cases} 
\frac{X}{p} & X < p \\
\infty & X \geq p.
\end{cases}
\] (12)
The comparison between (11) and (12) yields two observations. First, credit constraints are much tighter when externalities are present. The simplest way to see this is to compare the RHS of both expressions. If \( X \geq p \) then borrowing is unconstrained under flexible capacity while the RHS of (11) is less than one. For \( X < p \) the RHS of (11) is \(-p \log p\) as large as the RHS of (12), which is a strictly concave function of \( p \), equal to zero at \( p = 0 \) and \( p = 1 \) and reaches a maximum of \( e^{-1} \approx 0.368 \). One way to interpret this number is that enforcement externalities reduce the ‘effectiveness’ of enforcement by over 60%: the capacity under no externalities needed to yield the same credit constraint is at most 37% the capacity under externalities. The second observation is that enforcement externalities become stronger as the failure rate \( p \) goes down. This is because \( p \) represents the mass of non-strategic agents, i.e., those choosing default regardless of default rates in the economy. Hence, a high \( p \) implies that strategic complementarities are small and vice versa. As a consequence, while credit constraints in the absence of externalities vanish in the limit as \( p \) goes to zero, they are always binding in the presence of them. It is worth pointing out that the limit case is similar to the speculative borrower attack analyzed by Bond and Rai (2009) in which there is no asymmetric information about project returns.

Figure 5 illustrates the potentially large effect of enforcement externalities on borrowing. Under reasonable parameter values—\( R = 1.1 \), \( p = 0.05 \) and \( y = 1 \)—limited capacity severely limits borrowing and investment. As we show in our first application, this not only limits economic growth in the short run but also can distort the principal’s incentives to build future capacity. The reason is that, if agents value public good services and discount the future the principal may find optimal to spend scarce tax revenues on providing public goods given that capacity investment will be initially ineffective in relaxing credit constraints.
Figure 5: The Effect of Capacity Constraint on Borrowing. (Blue line denotes capacity constrained economy, red (dashed) line denotes a counterfactual commitment case in which agents are unaware of the constraint and principal can commit to a verification strategy.)

5.2 Shock Propagation

Now consider our second question. The principal must build $X$ at a cost $c(X)$ before learning the shock $\zeta$ that yields $X - \zeta$ as the effective capacity to enforce new loans, which are issued after the shock realization. Borrowing levels contingent on the shock $b_\zeta$ will then be chosen according to credit constraint (11) but replacing $X$ by $X - \zeta$. To keep the planner from offering infinite credit in the absence of a shock, we impose an exogenous bound on how much capital can be invested in the project, which translates into a limit $\bar{b}$ on borrowing.\footnote{Otherwise, the principal can finance any finite capacity cost by lump sum taxes on entrepreneurs.} Accordingly, the principal chooses $X$ to maximize expected aggregate payoffs net of capacity costs, that is, she

\begin{equation}
X = \text{arg max}_X \mathbb{E}[\pi(X) - c(X)]
\end{equation}
solves

\[
\max_{x, b_x \leq \bar{b}} \mathbb{E} [(1 - p)R(y + b_r)] - c(X) \tag{13}
\]

s.t. \[\frac{1}{R(1-p)} \frac{b_r}{y + b_r} \leq \begin{cases} 
0 & X - \zeta \leq 0 \\
-(X - \zeta) \log p & 0 < X - \zeta \leq p \\
(X - \zeta)(1 - \log(X - \zeta)) - p & X - \zeta > p.
\end{cases} \forall \zeta.
\]

To illustrate how shocks to capacity can severely affect credit provision consider the following simple example. Capacity costs are linear \(c(X) = 0.2X\); there is a single shock \(\zeta = 0.1\) (e.g., a pike in the default rate of pre-existing loans) that hits with probability \(p = 0.02\) and the borrowing limit is \(\bar{b} = 2\). As above, initial income is \(y = 1\), returns are \(R = 1.1\) and the failure rate is \(p = 0.5\). In this example, since it is costly to build capacity and the shock hits with low probability, the principal finds optimal to sets capacity to \(X = 0.32\) so that, in the absence of the shock, the borrowing constraint is only binding at \(b_0 = \bar{b} = 2\). When the shock hits there is a credit crunch: borrowing drops by 43% to \(b_{0.1} = 1.13\). In contrast, without externalities the principal chooses \(X = 0.13\), which is high enough to prevent the shock from having any impact on credit and investment.

6 Applications

6.1 Origins of State Capacity

Our first application follows closely Besley and Persson (2009), within which we use our theory to differently endogeneize the key constraints that link enforcement capacity to economic growth. To keep our analysis simple, we abstract from the political economy considerations in the original
papers, though the implications of our model in this respect will be clear.

The interesting feature of this particular model is that—since enforcement capacity is the key factor limiting the rate of economic growth—the constraints are binding at all times, as in our example with limited capacity. Consequently, enforcement constraints have long-run implications for deterministic dynamics.

### 6.1.1 Environment

The setup involves an application of our game four times on the timeline of the model, twice across each period. Figure 6 summarizes the timing of events and outlines the key decisions within the model. In this context, the key implication that we will be interested in studying is how enforcement capacity affects economic growth across the two periods.

![Figure 6: Timing of events in Application 2.](image)

The economy is populated by a unit measure of risk neutral agents and a benevolent principal, referred to as the State thereafter. The time horizon is two periods. The periods are indexed as 0 and 1, indexed by \( t \).
There are two goods in the economy: a private good and a public good. The provision of the public good is exclusively financed by the State from taxes. Tax collection must be enforced by the State alongside private contracts; enforcement of legal contracts and taxes is subject to legal capacity $X$ and fiscal capacity $X'$, respectively; both are accumulated in the previous periods by the State and fixed thereafter. The initial period zero value is given.

Agents’ preferences are linear in consumption: $u(c_t, g_t) = c_t + \alpha g_t$, where $c_t$ and $g_t$ are private and public good consumption. ($\alpha > 0$ is the relative weight of the public good.)

At the beginning of each period a fraction $\sigma < 1/2$ of agents has access to a risky project and they become borrowers or entrepreneurs. Projects yield a return $r$ per unit of invested capital with probability $1 - p$ and zero with probability $p$. Expected returns are greater than 1, i.e., $r > 1/(1 - p)$. All agents begin the period with some exogenous endowment $y$ that they can consume, lend to other agents through credit markets or invest directly in their own project.

In each period there exists a credit market in which agents with a project (henceforth borrowers) can borrow from agents without projects (lenders). In order to borrow, an agent must pledge a fraction $c_t \leq 1$ of her end-of-the-period gross income $w_t$ as collateral. After borrowing and investing, entrepreneurs privately observe their stochastic type $w_t$ and decide whether to repay the loan or default.

The State provides legal capacity $X_t \in [0, 1]$ which facilitates enforcement of credit contracts in the economy. In particular, capacity allows the principal to monitor up to measure $\sigma X_t$ of defaulting borrowers. If a defaulting borrower is monitored lenders, she can seize her collateral $c_t w_t$. Accordingly, if the amount borrowed is $b_t$ at interest rate $R_t$, pre-tax income in period $i$ for agents with access to the project is given by $y_t(1, w_t, \cdot) = w_t - R_t b_t$, $y_t(0, w_t, 0) = w_t$ and $y_t(0, w_t, 1) = (1 - c_t) w_t$. It is clear that the induced preference relation satisfies Assumption 1 iff $R_t b_t < c_t r (y_0 + b_t)$, where the last term represents income earned by the entrepreneur when the project is successful. We assume/show that this is the case. Note that by risk neutrality
borrowers invest all their initial income $y$ in the project.

Agents must pay $\tau_t y_t$ in taxes to the State, where $\tau_t$ is the proportional tax rate imposed on agents’ income $y_t$. Tax rate $\tau$ is a choice variable of the State (government). In each period, and after learning her income, agents decide whether or not to pay taxes. Similarly, the State accumulates and later uses its fiscal capacity $X'_t \in [0, 1]$ to enforce collection of taxes, which allows to audit at most measure $X'_t$ of agents in the economy.

We assume that agents decide whether to pay taxes of not. The State can audit a fixed measure of agents constrained by her fiscal capacity. if a taxpayer is audited her income is revealed to the State. If an agent is found to owe taxes she is required to pay them and incurs a penalty proportional to her actual income, $\phi y_t$. Let $a = 1$ denote the action of paying due taxes and $a = 0$ denote otherwise. If the agent pays her taxes her utility is unaffected by an audit. Given these assumptions, we note that the after-tax income is given by $y^\tau_t(1, y_t, \cdot) = (1 - \tau_t)y_t$, $y^\tau_t(0, y_t, 0) = y_t$, and $y^\tau_t(0, y_t, 1) = (1 - t_i - \phi)y_t$. It is clear that the implied preferences (monotonic in income) also obey Assumption 1.

The State in period zero sets the tax rate $\tau_0$ and allocates the revenue from taxes collected into provision of public goods and accumulation of future legal and fiscal capacity $(X_1, X'_1)$. The cost of augmenting legal and fiscal capacity is governed by convex functions $c(X_1 - X_0)$ and $c_r(X_{r,1} - X_{r,0})$, respectively, as in Besley and Persson (2009). Tax revenue in period two finances the provision of public goods.

We are now ready to state the objective of the State. To this end, we assume that the State is benevolent and makes choices to maximize agents’ expected utility from both periods, subject to a balanced budget constraint in each period. Since the shock of the entrepreneur is binary, it should be clear at this point that legal and fiscal capacity both will induce a constraint how high taxes can be (to induce compliance) and how much money the principal can give to entrepreneurs so that they comply—as otherwise nobody would comply and there would be no
way to satisfy the budget constraint. We show that this is the case in the next section formally, but to complete the description of the setup without otherwise burdensome notation, here we take it as given to state the problem solved by the principal. Denoting the discount factor by $\delta_t$ and setting $\delta_0 = 1$, the State (principal) it is given by:

$$\max \left\{ \sum_{t=0}^{1} \delta_t \left( (1 - t_t) [\sigma y + (1 - \sigma)(1 - p)(y + b_t(r - 1/(1 - p))] + \alpha g_t \right) \right\}$$

s.t.

$$\tau_t \leq \tau_t(X'_t), \quad t = 0, 1,$$

$$b_t \leq b_t(X_t), \quad t = 0, 1,$$

$$c_X(X_1 - X_0) + c_X'(X'_1 - X'_0) + g_0 - \tau_0 [\sigma y + (1 - \sigma)(1 - p)(y + b_0(r - 1/(1 - p))] \leq 0,$$

$$g_1 - \tau_1 [\sigma y + (1 - \sigma)(1 - p)(y + b_1(r - 1/(1 - p))] \leq 0.$$

### 6.1.2 Equilibrium Characterization

In order to apply the results obtained in the theory section, we first need to derive the indifference types induced by $w_t$. To this end, we let $0 \leq b_t \leq \bar{b}$ denote the amount borrowed by an agent with access to the project, where $\bar{b} + y$ represents the highest amount that can be invested on a project. To keep the derivation of equilibrium in the credit market simple, we assume that $\bar{b} < y(1 - \sigma)/\sigma$.\(^{11}\) Note that by risk neutrality, entrepreneur’s investment level is $y + b_t$. Consequently, entrepreneur’s income before repayment/default decisions is either $w_t = 0$ (project fails) or $w_t = r(y + b_t)$ (project succeeds).

We are now ready to derive indifference types of entrepreneurs. First, we note that the indifference type of unsuccessful entrepreneurs is $\theta_{0,t} > 1$. Second, we observe that the indifference type of a successful entrepreneur, denoted by $\theta_{1,t}$, is determined by the indifference condition

---

\(^{11}\)The condition ensures that the aggregate demand for funds is always lower than the total available funds.
\[ w_t - R_t b_t = w_t - \theta_{1,t} c_t w_t, \] which leads to \( \theta_{1,t} = \frac{R_t b_t}{c_t r (y + b_t)}. \) We also note that \( g(\theta_1) = 1 - p. \)

Applying Theorem 3 and Corollary 1, we immediately obtain the (limit) threshold for successful entrepreneurs from equation (9):

\[
(1 - p) \theta_{1,t} = \begin{cases} 
-k_{1,t} \log p & \text{if } k_{1,t} \leq p \\
-k_{1,t} (\log k_{1,t} - 1) - p & \text{if } k_{1,t} > p.
\end{cases}
\]

(14)

(It is clear that this analytic solution makes the model particularly tractable, as otherwise it would have to be solved numerically.)

In equilibrium in the credit market, we must have \( X_t \geq k_{1,t}, \) as otherwise all entrepreneurs would default and lenders could not break even. Furthermore, since \( k(1 - \log k) \) is an increasing function, we know that the equilibrium debt contract is characterized by a triple \((R_t, b_t, c_t)\) that satisfies:

\[
X_t \geq -\frac{1 - p}{\log p} \frac{R_t b_t}{c_t r (y + b_t)} \quad \text{if} \quad \frac{R_t b_t}{c_t r (y + b_t)} \leq -\frac{p \log p}{1 - p}, \quad \text{and}
\]

\[
X_t (1 - \log X_t) \geq p + (1 - p) \frac{R_t b_t}{c_t r (y_0 + b_t)} \quad \text{otherwise}.
\]

To solve for the contract explicitly, we note the following. First, since there are more lenders than borrowers, and given borrowers borrow up to \( \bar{b}, \) competition in the credit market results in zero expected profits. This implies that lenders charge gross interest rate \( R_t = 1/(1 - p) < r. \) Second, we observe that it must be true that \( c_t > R_t / r, \) as otherwise \( \theta_{1,t} > 1 \) and all entrepreneurs would default (which cannot be the case). Given these two properties—since risk neutral entrepreneurs want to borrow as much as possible to invest in the project—we know that entrepreneurs will pledge as much collateral as possible, i.e. choose \( c_t = 1, \) so as to relax the above constraints. (This is because the RHS is increasing in \( b_t \) and decreasing in \( c_t \).)
whenever $R_t/r < c_t < 1$.) Finally, for the same reason, the above constraints hold with equality in equilibrium as long as $b_t \leq \bar{b}$ does not bind. Accordingly, from the above inequalities, we obtain the following borrowing constraints as a function of legal capacity $X$:

$$b_t(X_t) = \begin{cases} 
-(r \log p)X_t & X_t \leq p \\ 
1 + (r \log p)X_t & X_t > p \\
\end{cases}$$

It is easy to check that the RHS in the above expressions are increasing in $X_t$, reaching infinity as the denominator goes to zero. Given this, we define $X_{\text{max}}$ as the capacity satisfying $b_t(X_{\text{max}}) = \bar{b}$ and note for what follows next that the government will never set legal capacity above $X_{\text{max}}$.

The characterization of the fiscal constraints is analogous to the above and is relegated to the appendix.

### 6.1.3 Findings

The key implication of our model is that the capacity constraints induce convexity of borrowing constraints and fiscal constraint. In the case of legal capacity, this important result is summarized below. (Since they are analogous, we relegate the discussion of the fiscal capacity to the appendix.)

**Proposition 1.** If $X_{\text{max}} \leq p$ then $b_t(X_t)$ is convex in $[0, X_{\text{max}})$ and concave in $(X_{\text{max}}, \infty)$. If $X_{\text{max}} > p$ there exists $p \leq \hat{X} \leq X_{\text{max}}$ such that $b_t(X_t)$ is convex in $[0, p)$, concave in $(p, \hat{X})$, and convex in $(\hat{X}, X_{\text{max}})$.

To highlight the implications of the above result, consider a counterfactual commitment case as a reference point. This case assumes that agents in the economy are unaware of the constraint and act ‘as if’ the principal’s capacity constraint allowed her to fully respond to any coordinated deviation and thereby stick to the previously announced (committed) verification strategy $P$. 

33
Setting \( c_t = 1 \), it is easy to see that this ad hoc case yields the following borrowing constraint:

\[
b_t^c(X_t) = \frac{(1 - p)rX_t}{p - (1 - p)r}y.
\]

The comparison of the two constraints leads to the following striking result. This result shows that, in the presence of enforcement externalities, borrowing constraints are not only tighter but, more importantly, get relaxed far more slowly as capacity grows.

**Proposition 2.** *In the range of* \( X_t \) *such that* \( b_t^c(X_t) < \bar{b}, \) *\( b_t^c(X_t) \) is convex, greater than* \( b_t(X_t), \) *and steeper than* \( b_t(X_t).\)

The intuition behind the above result is quite simple. By making compliance levels insensitive to increases in legal and fiscal capacity, enforcement externalities can severely slow down development because of fewer projects being started. Through the budget constraint, its effects unravel into a dynamic conundrum. This is because low production today lowers tax revenue and thus cripples accumulation of future enforcement capacity.

It is worth highlighting that the constraints do not vanish as the project failure rate \( p \) goes down to zero. This is in contrast to our counterfactual commitment case, in which case enforcement can be ensured by setting \( X' = \sigma p \) and \( X = p, \) and thus is essentially costless when \( p \) is low enough. The intuition why this is not the case in our model is as follows. A lower \( p \) reduces the *ex post* use of capacity, and by doing so it exacerbates enforcement externalities by reinforcing complementarity between agents’ actions.

To illustrate the potential quantitative effects associated with enforcement externalities, we next consider a numerical example. The parameter values are as follows:

**Example 1.** *Initial income is 1. The discount factor is* \( \delta = 0.95 \) *and the preference for public good is* \( \alpha = 1.1. \) *Mass of entrepreneurs is* \( \sigma = 0.4. \) *Project return is* \( r = 1.1 \) *with a 5% failure probability. Initial fiscal capacity is* \( \tau_1 = 0.01. \) *Capacity costs are quadratic:* \( c_{X'}(\Delta X') = 10(\Delta X)^2 \)
and $c_X(\Delta X) = 10(\sigma \Delta X)^2$. Finally, tax penalties are 10% of taxpayer income.

In this example, we solve numerically the principal’s problem and plot the solution for different levels of $(X_0, X'_0)$. As has been described, we assume that the principal maximizes the discounted average payoff in the economy by choosing, for each period $t = 1, 2$, tax rates $\tau_t \in [0, 1]$, borrowing levels $b_t \in [0, \bar{b}]$, and public good provision $g_t \geq 0$. Crucially, the principal needs to choose period one capacity $(X_1, X'_1)$ so that the borrowing and tax rate constraints are satisfied and the budget is balanced.

Figure 7 presents the results by comparing the growth implications of our model across the two period to the implications of analogously parameterized commitment case (as discussed above). As we can see, the counter-factual economy grows rapidly, whereas the economy with fixed capacity only operates at full steam when initial legal capacity is close to being non binding. As is clear from the figure, the disparity in growth rates far exceeds what one could possibly expect a priori.

### 6.2 Financing of Investment and Credit Crunches

Our second application is inspired by the well known financial accelerator framework developed by Bernanke et al. (1999). Using this environment, we explore a novel and distinct channel that can result in credit crunches in response to large but infrequent aggregate shocks.

#### 6.2.1 Environment

The economy is comprised of a continuum of ex ante identical, risk neutral entrepreneurs with initial wealth $N$ and a benevolent principal whose sole objective is to maximize output in this economy. There are three periods indexed by 0, 1 and 2. Periods 0 and 1 describe the ex ante problem that endogenously defines the inputs into our game and period 3 pertains to the game
Figure 7: The Effect of Capacity Constraint on Growth Rate. (Blue line denotes capacity constrained economy, red (dashed) line denotes a counterfactual commitment case in which agents are unaware of the constraint and principal can commit to a verification strategy.)

itself—onto which this setup will map. Figure 8 summarizes timing of events.

In period 1 each entrepreneur gains access to her respective investment project. The quality of the project is stochastic and determines entrepreneur’s type $w$. The type is realized (late) in period and is private information of each entrepreneur. The return from the project is given by $wRK$, where $K$ is capital invested into the project and $R$ is the average return on capital (projects) in the economy. It is assumed that before learning their type entrepreneurs can borrow $B$ from the principal. Entrepreneurs thus invest $K = B + N$ in their project.

The contractual space is restricted to a binary debt contracts featuring an option to default.\footnote{While we restrict the contractual space to binary contracts, Gale and Hellwig (1985) show they are optimal in a principal-agent setting without capacity constraints. We conjecture that under reasonable conditions binary contracts are still optimal when we add a borrowing constraint induced by limited capacity.}

Specifically—just as in the original formulation—external financing contracts are characterized by a threshold $\overline{w}$ and capital level $K$ such that the agent is supposed to pay $wRK$, if $w \geq \overline{w}$,
and otherwise she can default (renege) on $B$. In the latter case, if a defaulting entrepreneur is subjected to state verification by the lender (principal), it is assumed that the lender can seize the project net of (sunk) liquidation costs. These costs are proportional to returns and given by $\mu wRK$. The entrepreneur’s payoffs are given by $U(1, w, \cdot) = (w - \overline{w})RK$, $U(0, w, 0) = wRK$ and $U(0, w, 1) = 0$, and they satisfy the requirements of Assumption 1.

There is an aggregate lender (principal) who wishes to maximize entrepreneurs net income, less costs of sustaining compliance. In period zero she decides how much enforcement capacity $X$ to build—which costs her $c(X)$—and is fixed throughout.

The principal starts with some fixed measure of pre-existing (lateral) projects on her balance sheet, which may become non-performing at the onset of period one (right before she signs contracts with entrepreneurs). We assume that an aggregate shock $\zeta$ governs the performance of her existing portfolio, and to simplify our analysis, this shock simply soaks up a fraction of principal’s enforcement capacity $X$, implying that she is left with residual enforcement capacity $X - \zeta$ to enforce new contracts. Our goal will be to analyze how the realization of $\zeta$ affects the principal’s choice of contract terms $\overline{w}$ given enforcement game in period three.

We are now ready to formally lay out the principal’s problem. Her goal is to maximize

---

13This setup is equivalent to a one in which there is a large number of lenders who compete in a Bertrand fashion.
expected income of entrepreneurs net of liquidation and capacity costs $c(X)$. She does so by choosing capacity (ex ante problem) taking into account that, after learning $\zeta$, she issues credit contracts to maximize aggregate entrepreneur payoffs subject to an interim budget balance constraint and to the equilibrium compliance rates $\psi(X-\zeta)$ determined in the ex post coordination game (interim problem). We define first the interim problem and its associated value function $V(X, \zeta)$. The monitoring probability in equilibrium is $P_\zeta = \min \left\{ \frac{X-\zeta}{1-\psi(X-\zeta)}, 1 \right\}$. Let $\zeta$ be distributed according to $Z(\zeta)$, with support $Z$ and denote $(\bar{w}_\zeta, K_\zeta)$ the contract chosen by the planner when the mass of non-performing loans is $\zeta$. Accordingly, after learning $\zeta$ she solves for any given $X$

$$V(X, \zeta) = \max_{\bar{w}_\zeta \geq 0, K_\zeta \geq N} E(a(w-\bar{w})RK_\zeta + (1-a)(1-P_\zeta)wRK_\zeta)$$

s.t. $E_\zeta \left( 1_{w \geq \bar{w}_\zeta} RK_\zeta + 1_{w < \bar{w}_\zeta} (1-\mu)wRK_\zeta P_\zeta \right) - (K_\zeta - N) \geq 0,$

where the first term of the budget constraint (inside the expectation) is revenue from non-defaulting borrowers, the second term is revenue from liquidating the projects of monitored defaulting entrepreneurs, and the last term is just the aggregate amount borrowed (we assume zero cost of funds). Given this value function, the ex ante problem simply involves maximizing the expected value function net of capacity costs:

$$\max_{X \in [0,1]} E V(X, \zeta) - c(X). \quad (16)$$

Finally, the expectations involved in the maximization problem can be expressed in line with those developed by Bernanke et al. (1999). There it is shown that $K_\zeta$ is an increasing function of $\bar{w}_\zeta$. Accordingly, the implicit constraint on $\bar{w}_\zeta$ limits how much credit is provided to entrepreneurs. Such constraint is what propagates credit shocks by linking default rates of pre-
existing loans to the provision of future loans. Specifically, a drop in the residual capacity left
to monitor new loans caused by a high realization of $\zeta$ tightens the constraint on $\tilde{w}$ and hence
leads to less borrowing.

6.2.2 Equilibrium Characterization

In order to solve the principal’s problem we need to pin down compliance rates for each possible
contract and residual capacity. We do so by recasting our equilibrium characterization and
its implied compliance rates in terms of primitive types, i.e., entrepreneur returns $w$. First,
we make the simplifying assumption that return distribution $F$ has a continuous density with
support $(0, \infty)$ and, furthermore, that the conditions in Theorem 3 apply to continuous $G$.

Such assumption greatly simplifies the analysis and helps provide clean intuition. In addition,
since $G$ can always be approximated by a discrete distribution and payoff functions are nicely
behaved the results derived here roughly apply to ‘nearby’ economies with discrete type spaces.

We first derive the indifference types of entrepreneurs. Recall that contracts are given by
a threshold $\tilde{w}$ on idiosyncratic returns, investment level $K$ and default decision $\delta(w)$ such that
the agent pays back $\tilde{w}RK$ to the principal if $w \geq \tilde{w}$ (i.e., $\delta(w) = 1$ ) and defaults ($\delta(w) = 0$) if
$w < \tilde{w}$. Under no capacity restrictions Gale and Hellwig (1985) show that the optimal incentive
compatible contract takes this form when lenders seize up the project upon default, as is the
case here.$^{15}$

An agent with returns $w$ that faces monitoring probability $\theta(w)$ gets a payoff from defaulting
equal to $(1 - \theta(w))wRK$, whereas she gets $(w - \tilde{w})RK$ when she pays back the loan. Hence, her
indifference type is $\theta(w) = \frac{\tilde{w}}{w}$. The distribution of indifference types can then be expressed as

$^{14}$The support $(0, \infty)$ makes the derivations below easier but it may violate the requirement that there is a
positive lower bound $\theta$ on indifference types. Nonetheless, it is clear that $F$ can be finely approximated by a
truncated distribution or by a discrete one that takes care of this issue.

$^{15}$Note that these contracts can be expressed in terms of standard defaultable loans $(b, r, \delta(w))$, where $b$ is the
size of the loan and $r$ the gross interest rate. In the case of no default, payments are $rb = \tilde{w}R(b + N)$, implying
$b = \frac{\tilde{w}RN}{r - \tilde{w}R}$, which is increasing in $\tilde{w}$. 

39
G(θ(w)) = 1 − F(w) and g(θ(w)) = f(w)\frac{w^2}{\bar{w}}. We impose the following restrictions on F: \( \mathbb{E}w = 1 \) and \( wf(w)/F(w) \) is strictly decreasing in \( w \), greater than one at \( w \) close to zero and converges to zero as \( w \to \infty \). The latter condition is satisfied by distributions such as the log-normal—used in Bernanke et al. (1999). It implies that \( \theta(1 − G(\theta)) \) is single-peaked and, as a result of Theorem 3 leads to the following characterization of equilibrium thresholds.

**Proposition 3.** If \( \theta(1 − G(\theta)) \) is single-peaked in \((0, 1]\) then equilibrium thresholds are increasing and equal to

\[
k(\theta) = \begin{cases} 
\theta(1 − G(\theta)) & \theta \leq \theta^* \\
\theta^*(1 − G(\theta^*)) & \theta \geq \theta^*,
\end{cases}
\]  

(17)

with \( \theta^* \) being the unique solution in \([\theta, 1)\) to

\[
\int_{\theta^*}^{1} \theta dG(\theta) = \theta^*(1 − G(\theta^*))(1 − \log \theta^*) − (1 − G(1)).
\]  

(18)

The existence of a common threshold for types above threshold \( \theta^* \) is brought about by the monotonicity restriction (iii) and the single-peakedness of \( \theta(1 − G(\theta)) \), and it is the continuous approximation of thresholds in Figure 4. This implies that, when the mass of agents with types above \( \theta^* \) is sufficiently high, the principal will typically issue contracts so as to ensure the capacity is at least \( k(\theta^*) \)—otherwise she will not break even due to widespread default. Hence, the possibility of widespread defaults introduces an implicit constraint to the principal’s problem. We obtain the constraint by expressing the above result in terms of returns—that is, we substitute for \( \theta \) and \( G(\theta) \). Let \( w_{\text{max}} := \arg\max_w F(w)/w \) and also define \( \tilde{w}(X) \) as the unique solution to \( X = \frac{\tilde{w}}{w}F(w) \) in \([w_{\text{max}}, \infty)\).\(^{16}\) Our final result characterizes compliance levels for any contract.

\(^{16}\)Since \( wf(w)/F(w) \) is strictly decreasing, we must have that \( \frac{\tilde{w}}{w}F(w) \) is strictly decreasing in \((w_{\text{max}}, \infty)\) and thus the solution to \( X = \frac{\tilde{w}}{w}F(w) \) in \([w_{\text{max}}, \infty)\), if it exists, must be unique.
Proposition 4. Given any contract \((\bar{w}, K, \delta(w))\) with \(\delta(w) = 1\) if \(w \geq \bar{w}\), there exists \(w^* > w_{\text{max}}\) such that the compliance rate is

\[
\psi(X) = \begin{cases} 
1 - F(\hat{w}(X)) & X < \frac{\bar{w}}{w^*} F(w^*) \\
1 - F(\bar{w}) & X \geq \frac{\bar{w}}{w^*} F(w^*)
\end{cases}
\] (19)

where \(w^* = \bar{w}\) if \(\bar{w} > w_{\text{max}}\) and, when \(\bar{w} < w_{\text{max}}\), \(w^*\) is the unique solution in \((w_{\text{max}}, \infty)\) to

\[
\int_{\bar{w}}^{w^*} \frac{\bar{w}}{w^*} f(w) dw = \frac{\bar{w}}{w^*} F(w^*) \left(1 - \log \frac{\bar{w}}{w^*}\right) - F(\bar{w}). \tag{20}
\]

Furthermore, \(\frac{\bar{w}}{w^*} F(w^*)\) is increasing in \(\bar{w}\).

6.2.3 Findings

The above result establishes that when contracts have a default threshold smaller than \(w_{\text{max}}\), enforcement externalities can severely reduce contract compliance. This is because the default rate jumps from \(F(\bar{w}) < F(w_{\text{max}})\) to \(F(\hat{w}(X)) > F(w^*)\) when capacity drops below \(\frac{\bar{w}}{w^*} F(w^*)\).

Given this, the planner will set \(\bar{w}\) below \(w_{\text{max}}\) when the return distribution places sufficient mass in \([0, w_{\text{max}}]\), as otherwise default rate would be too high. Consequently, enforcement externality induces a constraint on \(\bar{w}\) as a function of capacity \(X\).

To provide a numerical illustration of the effects generated by the model, consider the following parameter setting. The parameters values have been chosen roughly in line with those used in financial accelerator models—e.g., Bernanke et al. (1999) and Christiano et al. (2014), and imply that, in the absence of shocks, a default rate is 3% and investment is \(K = 1.5\).

Example 2. The aggregate return is \(R = 1.03\), \(F\) is lognormal with variance 0.28, initial wealth is \(N = 1\), liquidation costs are \(\mu = 0.12\), and capacity costs are linear \(c(X) = 0.055X\). Finally, the distribution of \(\zeta\) has three support points: \(\zeta = 0\) with probability \(p_0 = 0.88\), \(\zeta = 0.025\) with
probability \( p_1 = 0.1 \), and \( \zeta = 0.1 \) with probability \( p_2 = 0.02 \).

As it turns out after solving (16) numerically, in this example the principal will set capacity at \( X = 0.17 \), which—relative to the case of \( \zeta = 0 \)—leads to a 15% reduction in credit provision in the case of the intermediate shock \( \zeta = 0.025 \). In contrast, the high shock of \( \zeta = 0.1 \) triggers a credit crunch and nonlinear response of the economy, simply because the economy hits the endogenous enforcement constraint. In such a case, borrowing is 58% lower than in the case of no shock. Importantly, enforcement externalities account for about 4/5ths of this credit disruption. This follows from the fact that in the case we keep capacity fixed at \( X = 0.17 \) but shut down the externality (by letting the planner commit to a certain monitoring intensity) borrowing only drops by 12% following the same shock. Remarkably, the credit crunch arises without the need of convex capacity costs nor any change in economic fundamentals (i.e., \( R, N \) and \( F \)). Thus, our propagation mechanism is fully independent of and complementary to the considerations put forward in the financial accelerator literature; our conjecture is that the presence of the two mechanisms in the same model—something we abstract from here—would only reinforce the results. We are stunned by the magnitude of this result and find this mechanism likely to be in some cases more potent than the original balance sheet externality.

7 Conclusions

Thus far studies of contract enforcement under asymmetric information in macroeconomics and finance have almost exclusively focused on the costs of enforcement, abstracting from any other frictions that costly enforcement may possibly involve. This simplification—and the broad-based relevance of enforcement frictions—clearly calls for further analysis, both in terms of empirics and theory.

In this paper we have contributed to this agenda by relaxing one of the central assumptions
made in most enforcement models, namely the fact that the principal can commit to a particular verification strategy. We highlight that only a special type of enforcement technology is consistent with this assumption, i.e. technologies that do not involve any rigidities in the accumulation of enforcement infrastructure (in the short-run). If that’s not the case, we have demonstrated that potentially large strategic complementarities arise. We have studied their implications by developing a general framework and new tools, applicable to heterogeneous agents economies. We have provided examples illustrating the relevance of our analysis for questions of substantive interest.

A Proofs of Section 4

A.1 Common Knowledge Proofs

To characterize the set of equilibrium compliance levels, we need to find the set of all possible indifferent or marginal types associated to capacity $X$.

**Definition 2.** $\theta^* \in [0, 1]$ is a marginal type for capacity $X$ if it satisfies

$$
\theta^* = \frac{X}{1 - G(\theta^*)}
$$

or if $\theta^* = 1$ and $X > 1 - G^-(1)$. The set of marginal types associated to $X$ is $\Theta^*(X)$.

The next result pins down the set of equilibrium compliance levels.

**Theorem 4.** The set of equilibrium compliance levels associated to $X$ is given by

$$
\Psi(X) := \left\{ \psi : \psi = \begin{cases}
1 - X/\theta^* & G^-(\theta^*) < G(\theta^*), \\
G(\theta^*) & \text{otherwise}
\end{cases}, \quad \theta^* \in \Theta^*(X) \right\}.
$$

**Proof.** The proof is divided in two parts. First, we argue that any compliance level $\psi \in \Psi(X)$ can be sustained in equilibrium by constructing a profile of strategies in which all agents are best responding. Second, we show that there is no equilibrium exhibiting a compliance level $\psi' \notin \Psi(X)$.

To show that for each $\psi \in \Psi(X)$ we can find a profile of strategies constituting an equilibrium we follow the arguments in the main text. Let $\theta^*$ be the marginal type associated to $\psi$. First notice that if $\theta^* = 1$ and $X \geq 1 - G^-(1)$ then all agents with $\theta \leq 1$ choosing $a = 1$ constitutes an equilibrium with $\psi = G(\theta^*) = G(1)$. This is because, when all of them comply, they face
monitoring probability \( P = \min\{X/(1 - G(1)), 1\} \geq \min\{X/(1 - G^-(1)), 1\} = 1 \), and thus complying is a best response since \( \theta \leq P \) for all \( \theta \in \Theta \).

Next, consider the case of \( \theta^* \in \Theta \), which means that \( G^-(\theta^*) > G(\theta^*) \). Since \( \psi = X/\theta^* - 1 \), the monitoring probability faced by agents is \( P = \theta^* \) and thus all agents with type \( \theta^* \) are indifferent between \( a = 0 \) and \( a = 1 \). Accordingly, we can construct an equilibrium in which a mass \( X/\theta^* - (1 - G^-(\theta^*)) \) of agents with type \( \theta^* \) choose \( a = 1 \), with types below and above \( \theta^* \) choosing \( a = 1 \) and \( a = 0 \), respectively. Such equilibrium is well-defined since, by (21), the mass of type-\( \theta^* \) agents choosing \( a = 1 \) is positive and (weakly) less than the overall mass of type-\( \theta^* \) agents, given by \( G(\theta^*) - G^- (\theta) \).

Finally, if \( \theta^* \notin \Theta \) then \( G^-(\theta^*) = G(\theta^*) \) and thus \( \psi = G(\theta^*) \). We focus on the case of \( \theta^* < 1 \) since we’ve already dealt with \( \theta^* = 1 \). The monitoring probability is \( P = \theta^* \), given that \( X = \theta^*(1 - G(\theta^*)) \) by (21) and we can construct an equilibrium in which all agents with types \( \theta < \theta^* \) select \( a = 1 \) and all types above \( \theta^* \) choose \( a = 0 \), yielding compliance level \( G(\theta^*) \).

To show that \( \psi \notin \Psi(X) \) can not arise in equilibrium, notice the following. Since all agents face the same monitoring probability when they choose \( a = 0 \), there must be a type \( \theta \in \Theta \) such that \( \psi \in [G^- (\theta), G(\theta)] \). We need to consider two cases: (i) \( \theta \notin \Theta^*(X) \), or (ii) \( \theta \in \Theta^*(X) \).

Case (i): if \( \theta \notin \Theta^*(X) \), either \( \theta = 1 \) and \( X < 1 - G(1) \) or \( X \notin \theta[1 - G(\theta), (1 - G^- (\theta))] \). The former can be divided into two sub-cases: (a) if \( X \geq \theta(1 - G(\theta)) \) then it must be that \( \psi \in \Psi(X) \) since \( G(\theta) = G(1) \) and hence we can find a marginal type \( \theta^* \in [\theta, 1) \) such that \( X = \theta^*(1 - G(\theta^*)) \); (b) if \( X < \theta(1 - G(\theta)) \) then \( \psi = G(1) \) cannot arise in equilibrium, since it requires full compliance of all types in \( \Theta \) but the monitoring probability would be strictly less than \( \theta \), so agents with this type would rather deviate and choose \( a = 0 \).

If \( X < \theta(1 - G(\theta)) \) then agents with type \( \theta \) with rather deviate and not comply since they face a monitoring probability less than \( \theta \). Finally, if \( X > \theta(1 - G^- (\theta)) \) we have two possible cases: if \( X \leq \theta'(1 - G^-(\theta')) \) where \( \theta' \) is the lowest type in \( \Theta \) above \( \theta \), then there is a marginal type \( \theta^* \in (\theta, \theta'] \) associated to \( \psi \) and thus \( \psi \in \Psi^*(X) \); if \( X > \theta'(1 - G^- (\theta')) \) then the monitoring probability is higher than \( \theta^* \) and agents of this type would rather deviate and comply.

Case (ii): if \( \theta \in \Theta^*(X) \) but \( \psi \notin \Psi(X) \) then it must be that \( \psi \neq 1 - X/\theta \) and \( G^-(\theta) > G(\theta) \). In such case, if \( \psi \in (G^- (\theta), G(\theta)) \) the above monotonicity argument implies that type-\( \theta \) agents are indifferent between \( a = 0 \) and \( a = 1 \) but they face a monitoring probability different from \( \theta \), a contradiction.

\[ \square \]

**Proof of Theorem 1.** We first show uniqueness when \( X < \bar{X} \) and \( X > \bar{X} \). We then identify necessary and sufficient conditions for multiplicity of compliance rates. Finally, we pin down necessary and sufficient conditions for global multiplicity.

First note that \( \theta(1 - G(\theta)) \) is continuous and increasing at all \( \theta \notin \Theta \), zero at \( \theta = 0 \), and jumps down at any \( \theta \in \Theta \). Hence, the set of local minimizers is given by \( \{0\} \cup \Theta \). Given this, if \( X < \bar{X} \), marginal types must be below the lowest type in \( \Theta \) and hence \( \psi(X) = G(\theta^*) = 0 \). Likewise, if \( X > \bar{X} \), marginal types are above the highest type in the support of \( G \), implying \( \psi(X) = G(1) \).

Second, if \( X \in \bigcup_{\theta \in \Theta} [\theta(1 - G(\theta)), \theta(1 - G^- (\theta))] \) there is multiplicity of compliance rates. To see why, note that there is at least one type \( \theta^* \in \Theta \) involving \( G^-(\theta^*) \leq \psi \leq G(\theta^*) \) with one inequality being strict. But then there must be at least another marginal type involving a
lower compliance level if $\psi > G^-(\theta^*)$ or higher if $\psi < G(\theta^*)$. The existence of a lower marginal type comes from the fact that $\psi > G^-(\theta^*)$ implies $X < \theta^*(1 - G^-(\theta^*))$. But then, given the properties of $\theta(1 - G(\theta))$, there must exist $\theta' < \theta^*$ such that $X = \theta'(1 - G(\theta'))$ with associated compliance level $\psi' \leq G(\theta') \leq G^-(\theta^*) < \psi$ by Lemma 4. A symmetric argument applies to the existence of a higher marginal type when $\psi < G(\theta^*)$.

The only if part directly follows from Theorem 4 and the properties of $\theta(1 - G(\theta))$: if $X \notin \bigcup_{\theta \in \Theta} [\theta(1 - G(\theta)), \theta(1 - G^-(\theta))]$ then there is at most one marginal type at which $X$ and $\theta(1 - G(\theta))$ intersect. If there were two types then it has to be that the one of them belongs to $\Theta$ since $\theta(1 - G(\theta))$ is strictly increasing outside $\Theta$. But that would imply that $X \in \bigcup_{\theta \in \Theta} [\theta(1 - G(\theta)), \theta(1 - G^-(\theta))]$, a contradiction.

Finally, we show that global multiplicity arises only when the conditions in the theorem are met. First note that if the condition does not hold there is at most one marginal type in $\Theta$ for any capacity $X$: if $\theta^* \in \Theta$ is a marginal type then by (21) we have that $X \geq \theta^*(1 - G(\theta^*)) > \theta(1 - G^-(\theta))$ for all $\theta \in \Theta$ such that $\theta < \theta^*$, and $X \leq \theta^*(1 - G^-(\theta)) < \theta(1 - G(\theta))$ for all $\theta \in \Theta$ such that $\theta > \theta^*$. But then, since $\theta(1 - G(\theta))$ is strictly increasing in $[0, 1] \setminus \Theta$, the only (two) other marginal types possible must be above the largest element of $\Theta$ lower than $\theta^*$ and below the smallest element of $\Theta$ bigger than $\theta^*$. This proves the only if part, since for any $X$ either there are no marginal types in $\Theta$, in which case there is a unique equilibrium, or there all the marginal types fall strictly between two adjacent types in $\Theta$.

The if part is straightforward. If there exists $\theta, \theta' \in \Theta$ with $\theta > \theta'$ and $\theta'(1 - G^-(\theta')) \geq \theta(1 - G(\theta))$ then $\theta$ and $\theta'$ are both marginal types of any capacity $X \in [\theta(1 - G(\theta)), \theta'(1 - G^-(\theta'))]$. But then Theorem 4 implies that there exist $\psi, \psi' \in \Psi(X)$ such that $\psi \leq G^-(\theta')$ and $\psi' \geq G^-(\theta) \geq G(\theta')$, with at least one of these inequalities holding strictly. That is, we can’t find a type $\hat{\theta} \in \Theta$ satisfying $\Psi(X) \subseteq [G^-(\theta), G(\theta)]$.

A.2 Equilibrium Selection Proofs

Proof of Theorem 2. The proof logic is as follows. First, we argue that the set of equilibrium profiles has a largest and a smallest element, each involving monotone strategies. Second, we show that there is at most one equilibrium in monotone strategies (up to discontinuities in cutoff strategies). But this implies that the equilibrium is essentially unique (i.e., up to discontinuities).

The existence of a smallest and largest equilibrium profile in monotone strategies follows from existing characterization results on supermodular games, e.g. Milgrom and Roberts (1990) and Vives (1990). Consider the game in which we fix the profile $x$ of signal realizations and agents choose actions $\{0, 1\}$ after observing their own signals. It is straightforward to check that the game satisfies the conditions of Theorem 5 in Milgrom and Roberts (1990), which states that the game has a smallest and largest equilibrium. That is, there exist two equilibrium strategy profiles, $\mathbf{a}(x)$ and $\overline{\mathbf{a}}(x)$ such that any equilibrium profile $\mathbf{a}(x)$ satisfies $\mathbf{a}(x) \leq \mathbf{a}(x) \leq \overline{\mathbf{a}}(x)$. Moreover, if we fix the strategy profile of all agents except agent $t$, the difference in expected payoff from choosing $a = 0$ versus $a = 1$ is increasing in $x$, since compliance levels are the same across signal profiles while $X$ is higher in expectation the higher the signal agent $t$ receives, thus implying a higher expected monitoring probability. That is, expected payoffs exhibit increasing differences w.r.t. $x$, and Theorem 6 in Milgrom and Roberts (1990) applies: $\mathbf{a}(x)$ and $\overline{\mathbf{a}}(x)$ are
non-decreasing functions of \( x \). But since an agent’s strategy can only depend on her own signal, this means that \( a(x, \theta) \) is a cutoff strategy for all \( \theta \in \Theta \).

To show that there is at most one equilibrium in monotone strategies we make use of the following two lemmas. Let \( k + \Delta = (k(\theta) + \Delta)_{\theta \in \Theta} \), while \( k \) and \( \bar{k} \) represent the profile of cutoffs associated to the smallest and largest equilibrium, respectively.

**Lemma 3.** If \( k \) is a profile of equilibrium strategies then \( k(\theta) \in [(\theta - \nu/2)(1 - G(1)), \theta + \nu/2] \) for all \( \theta \in \Theta \).

*Proof.* Note that \( k \) is an equilibrium if it solves (6). Note also that the value of \( X \) conditional on \( x \in [\nu/2, 1 - \nu/2] \) is at least \( x - \nu/2 \). Given this and the fact that monitoring probability is given by (5) we have that, if \( k(\theta) \in [\nu/2, 1 - \nu/2] \),

\[
\mathbb{E}_{\theta}[P|k;k(\theta)] \geq \mathbb{E}_{\theta}[X|k;k(\theta)] \geq k(\theta) - \nu/2.
\]

But this implies that \( \mathbb{E}_{\theta}[P|k;k(\theta)] > \theta \) when \( k(\theta) > \theta + \nu/2 \), a contradiction. A similar logic rules out \( k(\theta) > 1 - \nu/2 \) given that \( \mathbb{E}_{\theta}[X|k;x] \) is monotone in \( x \) and that \( \bar{\theta} < 1 - \nu \). Likewise, when \( k(\theta) \in [\nu/2, 1 - \nu/2] \),

\[
\mathbb{E}_{\theta}[P|k;k(\theta)] \leq \mathbb{E}_{\theta}\left[ \frac{X}{1 - G(1)} \bigg| \frac{k(\theta)}{1 - G(1)} \right] \leq \frac{k(\theta) + \nu/2}{1 - G(1)},
\]

which, using a symmetric argument, yields the above lower bound on \( K(\theta) \). \( \square \)

**Lemma 4.** If \( k \) is an equilibrium profile then \( \mathbb{E}_{\theta}[P|k;k(\theta)] < \mathbb{E}_{\theta}[P|k + \Delta;k(\theta) + \Delta] \) for all \( \Delta > 0 \) and all \( \theta \in \Theta \) such that \( k(\theta) + \Delta \leq \bar{k}(\theta) \).

*Proof.* First note that the pdf of \( X \) conditional on an agent of type \( \theta \) receiving signal \( x \in [\nu/2, 1 - \nu/2] \) is given by \( h_{\theta}\left(\frac{x - X}{\nu}\right) \). Also notice that an agent of type \( \theta' \) does not comply if she receives a signal \( x < k(\theta) \) and thus, the fraction of type-\( \theta' \) agents not complying when capacity is \( X \) is given by \( H_{\theta'}\left(\frac{k(\theta') - X}{\nu}\right) \). Since \( \nu \leq \theta(1 - G(1)) \) and, by Lemma 3, \( k(\theta) \geq (\theta - \nu/2)(1 - G(1)) \) we have that \( k(\theta) \geq \nu/2 \). Likewise, \( k(\theta) + \Delta \leq \bar{k}(\theta) \leq 1 - \nu/2 \) by Lemma 3 and the fact that
\[ \nu \leq 1 - \bar{\theta}. \] Hence, we can obtain the following inequality by a well-defined change of variables:

\[
\mathbb{E}_\theta[P(X)|k; k(\theta)] = \int_{-1/2}^{1/2} \min \left\{ \frac{X}{1 - G(1) + \sum_{\theta'} H_{\theta'} \left( \frac{k(\theta') - X}{\nu} \right) g(\theta')}, 1 \right\} \cdot h_\theta \left( \frac{k(\theta) - X}{\nu} \right) dX \\
< \int_{-1/2}^{1/2} \min \left\{ \frac{X + \Delta}{1 - G(1) + \sum_{\theta'} H_{\theta'} \left( \frac{k(\theta') - X}{\nu} \right) g(\theta')}, 1 \right\} \cdot h_\theta \left( \frac{k(\theta) - X}{\nu} \right) dX \\
= \int_{-1/2}^{1/2} \min \left\{ \frac{X'}{1 - G(1) + \sum_{\theta'} H_{\theta'} \left( \frac{k(\theta') - X'}{\nu} \right) g(\theta')}, 1 \right\} \cdot h_\theta \left( \frac{k(\theta) + \Delta - X'}{\nu} \right) dX' \\
= \mathbb{E}_\theta[P|k + \Delta; k(\theta) + \Delta].
\]

The inequality is strict since \( k \) being an equilibrium profile means that \( \mathbb{E}_\theta[P|k; k(\theta)] = \theta < 1 \). Accordingly, monitoring probabilities, conditional on \( x = k(\theta) \), are less than one for a positive measure of \( X \in [x - \nu/2, x + \nu/2] \) and hence expected monitoring probabilities go up strictly when capacity increases by \( \Delta \).

Equipped with Lemma 4 we know argue that \( k = \tilde{k} \). Assume, by way of contradiction, that \( \tilde{k}(\theta) > \bar{k}(\theta) \) for some \( \theta \in \Theta \). Denote \( \hat{\theta} = \arg \max_{\theta \in \Theta}(k(\theta) - \bar{k}(\theta)) \) and \( \hat{\Delta} = \bar{k}(\hat{\theta}) - \bar{k}(\hat{\theta}) \). By Lemma 4, we have that

\[ \hat{\theta} = \mathbb{E}_\theta[P|k; k(\hat{\theta})] < \mathbb{E}_\theta[P|k + \hat{\Delta}; \bar{k}(\hat{\theta})] \leq \mathbb{E}_\theta[P|k; \bar{k}(\hat{\theta})] = \hat{\theta}, \]

where the last inequality comes from the fact that compliance levels at \( \tilde{k} \) are higher, and so are monitoring probabilities conditional on \( x = \bar{k}(\hat{\theta}) \), than at \( k + \hat{\Delta} \geq \tilde{k} \).

**Proof of Lemma 2.** The result follows directly from the proof of Lemma 1 in Sakovics and Steiner (2012). To see why note first that Lemma 3 guarantees that threshold signals and thus the ‘virtual signals’ defined in their proof fall in \([\nu/2, 1 - \nu/2]\), which is needed for their belief constraint to hold. Second, it is straightforward to check that all the arguments and results in their proof hold unmodified if we condition all the probability distributions used in the proof on the event \( \theta \in \Theta' \) and focus on the aggregate action of agents with types in \( \Theta' \), rather than the aggregate action in the population.\(^{17}\)

\(^{17}\)When thresholds do not fall within \( \nu \) of each other, the distribution \( \tilde{F} \) of virtual errors \( \bar{n} \) need not be strictly increasing and thus its inverse may not be well-defined. Defining \( \tilde{F}^{-1}(u) = \inf\{\tilde{n} : \tilde{F}(\tilde{n}) \geq u\} \) takes care of this issue and ensures that the proof of Lemma 2 in Sakovics and Steiner (2012) applies to the general case.
Proof of Theorem 3. From Theorem 2 we know that for each $\nu > 0$ there exists essentially a unique equilibrium, which is in monotone strategies. Let $k^\nu(\theta)$ represent the equilibrium threshold of type-$\theta$ agents associated to $\nu > 0$, with $k^\nu$ denoting the equilibrium cutoff profile. The first step of the proof is to show that $k^\nu$ uniformly converge as $\nu \to 0$ and identify the set of indifference conditions that pin down the limit equilibrium. Let

$$A_\theta(y|k', \Theta') := P_\theta \left( \varphi(X, k', \Theta) \leq y | k', k^\nu(\theta) \right)$$

denote the strategic belief of an agent of type $\theta \in \Theta'$ when she receives her threshold signal $x = k^\nu(\theta)$.

**Lemma 5.** There exists a unique partition $\{\Theta_1, \cdots, \Theta_I\}$ and a set of thresholds $k_1 < k_2 < \cdots < k_I$ such that, as $\nu \to 0$, for all $\theta \in \Theta_i$, $i = 1, \cdots, I$, $k^\nu(\theta)$ uniformly converges to $k_i$. Moreover, thresholds $k = (k_1, \cdots, k_I)$ are the solution to the set of limit indifference conditions

$$\int_0^1 \min \left\{ \frac{k_i}{1 - \sum_{\cup_{j \leq i} \Theta_j} g(\theta') + (1 - y) \sum_{\Theta_i} g(\theta')}, 1 \right\} dA_\theta(y|k, \Theta_i) = \theta, \quad \forall \theta \in \Theta_i, \forall i, \quad (22)$$

where $A_\theta(y|k, \Theta_i)$ represents the strategic beliefs of type-$\theta$ agents in the limit and satisfies the belief constraint (8).

See proof below.

Equipped with this set of indifference conditions we next prove that the partition of types is monotone and that thresholds satisfy (iii) and (iv) in the theorem.

We show that the partition of types must be monotone by way of contradiction. Assume that there are two types $\theta > \hat{\theta}$ such that $\theta \in \Theta_i$ and $\hat{\theta} \in \Theta_m$ with $m > i$. Since $\hat{\theta} < 1$ the monitoring probability when no agent with type in $\Theta_m$ complies is strictly less than one, i.e., $1 - \sum_{\cup_{j < m} \Theta_j} g(\theta') < 1$. Otherwise (22) would be violated. In addition, $m > i$ implies that $k_m > k_i$ and that $\sum_{\cup_{j \leq i} \Theta_j} g(\theta') \leq \sum_{\cup_{j < m} \Theta_j} g(\theta')$. Combining all this and the fact that the LHS in (22) is bounded below by $1 - \sum_{\cup_{j < m} \Theta_j} g(\theta')$ and bounded above by $\min \left\{ \frac{k_i}{1 - \sum_{\cup_{j < m} \Theta_j} g(\theta')}, 1 \right\}$, lead to the following contradiction

$$\theta \leq \min \left\{ \frac{k_i}{1 - \sum_{\cup_{j \leq i} \Theta_j} g(\theta')}, 1 \right\} < \frac{k_m}{1 - \sum_{\cup_{j < m} \Theta_j} g(\theta')} \leq \hat{\theta}.$$

The monotonicity of the type partition implies that $1 - \sum_{\cup_{j \leq i} \Theta_j} g(\theta') = 1 - G(\hat{\theta}_i)$. Given this, it is straightforward to check that the above bounds on the LHS of (22) lead to condition (iii) in the theorem.
Finally, in order to obtain condition (iv) from (22) we make use of the belief constraint in the limit, which can be written as

$$
\frac{1}{\sum_{\Theta_i} g(\theta)} \sum_{\Theta_i} A_\theta(y|k, \Theta_i)g(\theta) = y. \tag{23}
$$

Multiplying both sides of (22) by $\frac{g(\theta)}{\sum_{\Theta_i} g(\theta)}$ and summing over $\theta \in \Theta_i$ we get

$$
\int_0^1 \min \left\{ \frac{k_i}{1 - G(\bar{\theta}) + (1 - y)\sum_{\Theta_i} g(\theta')}, 1 \right\} d \left( \frac{1}{\sum_{\Theta_i} g(\theta)} \sum_{\Theta_i} A_\theta(y|k, \Theta_i)g(\theta) \right) = \frac{\sum_{\Theta_i} \theta g(\theta)}{\sum_{\Theta_i} g(\theta)}.
$$

Finally, using the belief constraint (23) to substitute for the last term in the LHS yields condition (iv):

$$
\int_0^1 \min \left\{ \frac{k_i}{1 - G(\bar{\theta}) + (1 - y)\sum_{\Theta_i} g(\theta')}, 1 \right\} dy = \frac{\sum_{\Theta_i} \theta g(\theta)}{\sum_{\Theta_i} g(\theta)}. \tag{24}
$$

Proof of Lemma 5. To prove convergence we first need to introduce some notation. For sufficiently small $\nu$, partition the set of types into subsets $\Theta_i$ of types as follows: (i) if we order the signal thresholds of all types, any adjacent thresholds that are within $\nu$ of each other belong to the same subset; and (ii) $j > i$ implies that the thresholds associated to types in $\Theta_j$ are higher than those associated to $\Theta_i$—by at least $\nu$. Also, let $Q_\nu^\theta(z|k^\nu, y) := \mathbb{P}_\theta (X \leq z|k^\nu, k^\nu(\theta), \varphi(X, k^\nu, \Theta_i) = y)$ represent the beliefs about capacity of an agent of type $\theta \in \Theta_i$ conditional on receiving her threshold signal $k^\nu(\theta)$ and on the event that the fraction of agents with types in $\Theta_i$ choosing $a = 1$ is equal to $y$.

Note that a type-$\theta$ agent receiving signal $x = k^\nu(\theta)$ knows that all agents with types in $\Theta_j$ are complying if with $j < i$, and not complying if $j > i$. Also, the support of $Q_\nu^\theta(\cdot|k^\nu, y)$ must lie within $[k^\nu(\theta) - \nu/2, k^\nu(\theta) + \nu/2]$. Given this, by the law of iterated expectations, her expected monitoring probability conditional on $x = k^\nu(\theta)$ can be written in terms of her strategic belief as follows:

$$
\mathbb{E}_\theta(P|k^\nu; k^\nu(\theta)) = \int_{k^\nu(\theta) - \nu/2}^{k^\nu(\theta) + \nu/2} \int_0^1 \min \left\{ \frac{z}{1 - \sum_{\cup_{j \neq i} \Theta_j} g(\theta') + (1 - y)\sum_{\Theta_i} g(\theta')}, 1 \right\} dQ_\nu^\theta(z|k^\nu, y) dA_\theta(y|k^\nu, \Theta_i). \tag{25}
$$

In addition, notice that we can always express $\mathbb{E}_\theta(P|k^\nu; k^\nu(\theta))$ in terms of the threshold signal $k^\nu(\theta)$ and relative threshold differences $\Delta_\nu = (k^\nu(\theta') - k^\nu(\theta))/\nu$. Importantly, as Sakovics and Steiner (2012) emphasize, strategic beliefs only depend on the relative positions of thresholds
\(\Delta_{\Theta_i} = \{\Delta_{\nu}\}_{\nu \in \Theta_i}\). That is, keeping \(\Delta_{\Theta_i}\) fixed, \(A_\Theta(y|k^\nu, \Theta_i)\) does not change with \(\nu\). This directly implies that strategic beliefs satisfy the belief constraint when \(\nu = 0\).

Fix \(k^\nu(\theta) = k_i\) for some \(\theta \in \Theta_i\) and also fix \(\Delta_{\Theta_i}\), for all \(i = 1, \cdots, I\) and all \(\nu\) sufficiently small. By fixing relative differences, the partition \(\{\Theta_i\}_1\) still satisfies the above definition, and thus is independent of \(\nu\). We are going to show that indifference condition \(E_\theta(P|k^\nu; k^\nu(\theta)) = \theta\) is approximated by the limit condition in the lemma for \(\nu\) sufficiently small.

Note that the inner integral in (25) is bounded below by \(\min\left\{\frac{k^\nu(\theta) - \nu/2}{1 - \sum_{\nu \leq \theta_j} g(\theta') + (1 - y) \sum_{\Theta_i} g(\theta')}, 1\right\}\). Hence,

\[
\int_0^1 \min\left\{\frac{k_i - \nu/2}{1 - \sum_{\nu \leq \theta_j} g(\theta') + (1 - y) \sum_{\Theta_i} g(\theta')}, 1\right\} dA_\Theta(y|k^\nu, \Theta_i) \leq \min\left\{\frac{k_i + \nu/2}{1 - \sum_{\nu \leq \theta_j} g(\theta') + (1 - y) \sum_{\Theta_i} g(\theta')}, 1\right\} dA_\Theta(y|k^\nu, \Theta_i).
\]

The first term in these integrals is Lipschitz continuous. In addition, the next lemma shows that \(dA_\Theta(y|k^\nu, k^\nu(\theta))\) is bounded for all \(\nu\).

**Lemma 6.** \(0 \leq \frac{\partial A_\Theta(y|k^\nu, k^\nu(\theta))}{\partial y} \leq \sum_{\Theta_i} g(\theta') \frac{g(\theta)}{g(\theta')}\) for all \(\theta \in \Theta_i\) and all \(y\) in the support of \(A_\Theta(\cdot|k^\nu, k^\nu(\theta))\).

See proof below.

Hence, the LHS and the RHS of (26) uniformly converge to each other as \(\nu \to 0\), leading to limit indifference conditions (24). Note also that \(k^\nu(\theta) \in [-\nu/2, 1 + \nu/2]\) and, keeping \(\Theta^\nu(\theta)\) fixed, \(\Delta_{\nu} \in [-1, 1]\). Hence, the solutions to \(E_\theta(P|k^\nu; k^\nu(\theta)) = \theta\) lie in a compact set.\(^{19}\)

Accordingly, we can find \(\nu\) so that indifference conditions are within \(\varepsilon\) of the limit condition for all \(\nu < \nu\), leading to their solutions being in a neighborhood of \(k\).

**Proof of Lemma 6.** Let \(\varphi^{-1}(y, k^\nu, \Theta_i)\) be the inverse function of \(\varphi(X, k^\nu, \Theta_i)\) w.r.t. \(X\). Since the latter function is increasing in \(X\) as long as \(0 < \varphi(X, k^\nu, \Theta_i) < 1\), \(\varphi^{-1}\) is well defined and

\(^{18}\)This is straightforward to check. First, if we substitute \(X = k^\nu(\theta) - \nu\eta\) (since agents with type \(\theta\) get her threshold signal) and \(k(\theta') = \nu \Delta_\theta + k^\nu(\theta)\) into (5), we find that \(\varphi(X, k^\nu, \Theta_i)\) only depends on \(\Delta_\Theta_i\). But this means that \(A_\Theta(y|k^\nu, \Theta_i)\) only depends on \(\Delta_\Theta_i\), since \(H_\Theta\) is independent of \(\nu\).

\(^{19}\)If \(\Theta^\nu(\theta)\) and \(\Theta^\nu_\nu(\theta)\) are not kept fixed for some \(\theta\) when \(\nu\) is very small then \(E_\theta(P|k^\nu; k^\nu(\theta))\) would be discontinuous at some \(\nu\), implying a violation of the indifference condition.
increasing in such a range of capacities. Since the signal of an agent of type $\theta$ satisfies $x = X + \nu\eta$ we can express her strategic belief as

$$A_\theta(y|k^\nu, \Theta_i) = \mathbb{P}_\theta \left( \varphi^{-1}(y, k^\nu, \Theta_i) \geq k^\nu(\theta) - \nu \eta \right) = 1 - H_\theta \left( \frac{k^\nu(\theta) - \varphi^{-1}(y, k^\nu, \Theta_i)}{\nu} \right).$$

Differentiating w.r.t. $y$ yields

$$\frac{\partial A_\theta(y|k^\nu, \Theta_i)}{\partial y} = \frac{1}{\nu} h_\theta \left( \frac{k^\nu(\theta) - \varphi^{-1}(y, k^\nu, \Theta_i)}{\nu} \right) \frac{\partial \varphi^{-1}(y, k^\nu, \Theta_i)}{\partial y}$$

$$= \frac{h_\theta \left( \frac{k^\nu(\theta) - \varphi^{-1}(y, k^\nu, \Theta_i)}{\nu} \right)}{\sum_{\Theta_i} \frac{1}{g(\theta') \sum_{\Theta_i} h_{\theta'} \left( \frac{k^\nu(\theta') - \varphi^{-1}(y, k^\nu, \Theta_i)}{\nu} \right) g(\theta')}}.$$

For all $y \in (0, 1)$ we must have $h_\theta \left( \frac{k^\nu(\theta) - \varphi^{-1}(y, k^\nu, \Theta_i)}{\nu} \right) > 0$ since $h_\theta$ is bounded away from zero in its support. Hence, the last term is positive and weakly lower than $\frac{\sum_{\Theta_i} g(\theta')}{g(\theta)}$. □

**Proof of Corollary 1.** We just need to solve the integral in condition (iv) of Theorem 3. Since the integrand is increasing in $y$, the min operator applies only when $\frac{k_i}{1 - G(\bar{\theta}_i)} > 1$. Hence we need to consider two cases: $k_i \leq 1 - G(\bar{\theta}_i)$ and $k_i > 1 - G(\bar{\theta}_i)$. In the first case, we can drop the min operator and integrate to get

$$\int_0^1 \frac{k_i}{1 - G(\bar{\theta}_i) + (1 - y) \sum_{\Theta_i} g(\theta)} dy = \left. \frac{k_i}{\sum_{\Theta_i} g(\theta)} \log \left( 1 - G(\bar{\theta}_i) + (1 - y) \sum_{\Theta_i} g(\theta) \right) \right|_0^1$$

$$= -\frac{k_i}{\sum_{\Theta_i} g(\theta)} \log \frac{1 - G(\bar{\theta}_i)}{1 - G(\bar{\theta}_i)},$$

which yields the condition in the corollary for $k_i \leq 1 - G(\bar{\theta}_i)$ after multiplying both sides of (iv) by $\sum_{\Theta_i} g(\theta)$.

When $k_i > 1 - G(\bar{\theta}_i)$ the integral in (iv) becomes

$$\int_0^{1 - \frac{k_i - (1 - G^-(\bar{\theta}_i))}{\sum_{\Theta_i} g(\theta)}} \frac{k_i}{1 - G(\bar{\theta}_i) + (1 - y) \sum_{\Theta_i} g(\theta)} dy + \int_{1 - \frac{k_i - (1 - G^-(\bar{\theta}_i))}{\sum_{\Theta_i} g(\theta)}}^1 \frac{k_i}{1 - G^-(\bar{\theta}_i)} dy$$

$$= -\frac{k_i}{\sum_{\Theta_i} g(\theta)} \log \left( 1 - G(\bar{\theta}_i) + (1 - y) \sum_{\Theta_i} g(\theta) \right) \bigg|_0^{1 - \frac{k_i - (1 - G^-(\bar{\theta}_i))}{\sum_{\Theta_i} g(\theta)}} + \frac{k_i - (1 - G^-(\bar{\theta}_i))}{\sum_{\Theta_i} g(\theta)}$$

$$= -\frac{k_i}{\sum_{\Theta_i} g(\theta)} \log \frac{k_i}{1 - G^-(\bar{\theta}_i)} + \frac{k_i - (1 - G^-(\bar{\theta}_i))}{\sum_{\Theta_i} g(\theta)}.$$
Multiplying both sides of (iv) by \(\sum_{\Theta_i} g(\theta)\) and rearranging yields the condition in the corollary for \(k_i > 1 - G(\bar{\theta}_i)\).

\[ \theta \]

Proof of Corollary 2. First consider the ‘only if’ part and assume that there exist two types \(\theta > \theta'\) with the same threshold \(k_i\). We need to show that the complete information game has multiple equilibria. Condition (iii) in the 3 implies that \(\theta(1 - G(\theta)) \leq \theta'(1 - G^{-}(\theta'))\), which is a violation of the necessary and sufficient condition for the non-existence of between-type multiplicity stated in Theorem 1. Given this reasoning, the ‘if’ part is immediate: if \(\theta(1 - G(\theta)) > \theta'(1 - G^{-}(\theta'))\) for all \(\theta > \theta'\) then condition (iii) can only be satisfied if \(\Theta_i\) is a singleton for all \(i\).

\[ \theta \]

B Proofs of Section 6.2

Proof of Proposition 3. Given the continuity of \(G\), condition (iii) in Theorem translates into \(k(\theta) = \theta(1 - G(\theta))\) whenever there is only one type associated to threshold \(k(\theta)\). In addition, it also implies that \(k(\theta)\) is increasing and continuous. Hence, if there is an interval of types \([\bar{\theta}_i, \bar{\theta}_i]\) with the same threshold \(k_i\), we must have that \(k_i = \hat{\theta}_i(1 - G(\hat{\theta}_i))\). Finally, since \(\theta(1 - G(\theta))\) is first increasing and then decreasing, condition (iii) implies that \(k(\theta)\) is constant in at most one interval, which has upper bound \(\theta_1 = 1\). To see why there is only one ‘pooling threshold’, notice that condition (iii) requires that \(k_i = \hat{\theta}_i(1 - G(\hat{\theta}_i)) \geq \bar{\theta}_i(1 - G(\bar{\theta}_i))\). But, given the single-peakedness of \(\theta(1 - G(\theta))\) and the monotonicity of \(k(\theta)\), we must have that \(\theta(1 - G(\theta))\) is increasing at \(\bar{\theta}_i\) and decreasing at \(\hat{\theta}_i\). Accordingly, if such condition is satisfied by some interval \([\hat{\theta}_i, \bar{\theta}_i]\) we cannot find another interval satisfying the same condition that does not intersect with \([\hat{\theta}_i, \bar{\theta}_i]\). Finally, since \(\theta(1 - G(\theta))\) is strictly decreasing in \([\hat{\theta}_i, 1]\) the monotonicity of \(k(\theta)\) requires that \(\theta_1 = 1\).

We finish the proof by showing that \(\theta_1 = \theta^*\), where \(\theta^*\) is the unique solution to (18) in \([\theta, 1]\). First note that \(k_1 \geq 1 - G(1)\) by condition (iii). Hence, using the continuous version of (9) and substituting for \(k_1 = \theta^*(1 - G(\theta^*))\) and \(\theta_1 = 1\) we obtain (18). To show that this equation has a unique solution in \([0, 1]\) we differentiate w.r.t. \(\theta^*\) the LHS of

\[ \theta^*(1 - G(\theta^*))(1 - \log \theta^*) - \int_{\theta^*}^1 \theta dG(\theta) = (1 - G(1)), \]

yielding \((- \log \theta^*) (1 - G(\theta^*) - \theta^*g(\theta^*))\). The first term in this expression is positive while the second term is the derivative of \(\theta^*(1 - G(\theta^*))\) w.r.t. \(\theta^*\), which is first positive then negative in \([\theta, 1]\). Hence, since the RHS is constant, the above expression has at most two solutions in \([0, 1]\). But notice that \(\theta^* = 1\) is always a solution so there is at most one solution in \([0, 1]\). Finally, the LHS is negative at \(\theta^* = 0\) and hence there exists a solution in \([0, 1]\).
Proof of Proposition 4. First, note that \( \theta(w)(1 - G(\theta(w))) = \frac{\bar{w}}{w} F(w) \) is single peaked when \( w f(w)/F(w) \) is strictly decreasing, greater than one at \( w \) close to zero and converges to zero as \( w \to \infty \). To see why notice that \( \frac{d}{dw} \left( \frac{\bar{w}}{w} F(w) \right) = \frac{w f(w) - F(w)}{w^2} \), with the numerator being first positive and then negative. Note however that, depending on \( \bar{w} \), the maximum may occur at some \( \theta \geq 1 \). Since \( \theta(w) \) is increasing in \( \bar{w} \) and \( \theta(\bar{w}) = 1 \) this will be the case whenever \( \bar{w} \geq w_{\max} \).

Accordingly, if \( \bar{w} \geq w_{\max} \) then \( \theta(1 - G(\theta)) \) is strictly increasing in \([0, 1]\), leading to thresholds \( k(\theta) = \theta(1 - G(\theta)) = \frac{\bar{w}}{w} F(w) \) by condition (iii) in Theorem 3. In this case compliance rates are \( \psi = G(k^{-1}(X)) = 1 - F(\bar{w}(X)) \) for all \( X < k(1) = F(\bar{w}) \) and, since agents with returns lower than \( \bar{w} \) always default regardless of monitoring probabilities, \( \psi(X) = 1 - F(\bar{w}) \) when \( X \geq K(1) \). These compliance rates coincide with (19) if we set \( w^* = \bar{w} \).

Now if \( \bar{w} < w_{\max} \) then Proposition 3 applies and expressions (19) and (20) are derived by just substituting \( \theta^* = \bar{w}/w^* \), \( g(\theta(w)) \) and \( G(\theta(w)) \) into (17) and (18) and computing compliance rates as above.

We finish by showing that \( \frac{\bar{w}}{w^*} F(w^*) \) is increasing in \( \bar{w} \). We do so by the following chain of facts: 1) \( 1 - G(1) = F(\bar{w}) \) is increasing in \( \bar{w} \); \( \theta^* \) is increasing in \( 1 - G(1) \); 3) \( \theta^*(1 - G(\theta^*)) = \frac{\bar{w}}{w^*} F(w^*) \) is increasing in \( \theta^* \) if \( \theta(1 - G(\theta)) \) is single-peaked in \([0, 1]\). Fact 1) is immediate and fact 3) follows from the proof of Proposition 3. Hence, we just need to show 2). To do so we implicitly differentiate (27) to obtain

\[
\frac{dLHS}{d\theta^*} \frac{d\theta^*}{d(1 - G(1))} = 1.
\]

In the proof of Proposition 3 we have already established that the LHS of (27) is increasing at \( \theta^* \) so we must have \( \frac{d\theta^*}{d(1 - G(1))} > 0 \).

\[\Box\]

C Characterization of Fiscal Constraints of Section 6.1

Here we characterize the effect of limited fiscal capacity on tax rates. We proceed in the same fashion as in the case of legal capacity. First we identify indifference types and then obtain the constraints on tax rates induced by the fixed in the short-run capacity \( X_t \). For simplicity, we assume that all lenders receive zero profits from lending.\(^{21}\)

Given that lenders make zero profits, the distribution of taxpayers in the population is as follows: a mass \( 1 - \sigma \) of lenders with income \( y_{1,t} = y \); a mass \( \sigma(1 - p) \) of successful entrepreneurs with income \( y_{2,t} = r(y + b_t) - b_t/(1 - p) \); and a mass \( \sigma p \) of unsuccessful entrepreneurs with income \( y_{3,t} = 0 \). Agents can either pay \( \tau y_t \) or fail to pay their taxes. Obviously the latter always pay zero, regardless of \( X_t \). To find the indifference type of a taxpayer with income \( y_{j,t}, j = 1, 2 \), recall that if an agent is audited and found not paying due taxes, she pays back owed taxes

\(^{21}\)This is the case, for instance, if there is a risk neutral intermediary that hedges all the idiosyncratic risk, as in the first application. This assumption has little effect on our results and only leads to a smaller set of distinct income types in equilibrium.
plus penalty \( \phi y_{j,t} \). Accordingly, her indifference type, denoted by \( \theta_{j,t}^r \), is given by the indifference condition \( (1 - \tau_t) y_t = y_t - \theta_t^r (\tau_t + \phi) y_t \). That is, \( \theta_{j,t}^r = \theta_t^r = \frac{\tau_t}{\tau_t + \phi} \) for \( j = 1, 2 \). Thus, the induced distribution of indifference types is binary: a mass \((1 - \sigma p)\) of agents have type \( \tau_t / (\tau_t + \phi) \) and a mass \( \sigma p \) of agents have a type above one. Hence, we can apply (14) by replacing \( \theta_t \) and \( p \) by \( \theta_t^r \) and \( \sigma p \), respectively. This yields the tax rate constraint

\[
\tau_t(X_t') = \begin{cases} 
\frac{- \log(\sigma p) X_t'} {1 - \sigma p + (\log(\sigma p)) X_t'} \phi & X_t' \leq \sigma p \\
\frac{X_t'(1 - \log X_t') - \sigma p}{1 - X_t'(1 - \log X_t')} \phi & \tau_t > \sigma p.
\end{cases}
\]  

(28)

Let \( X_{\text{max}}' \) be the lowest capacity satisfying \( \tau_t(X_{\text{max}}') = 1 \). We obtain similar results as those applying to borrowing constraints. Specifically, we find that the tax rate constraint is convex w.r.t. fiscal capacity.

**Proposition 5.** If \( X_{\text{max}}' \leq \sigma p \) then \( \tau_t(X_t') \) is convex in \([0, X_{\text{max}}']\). If \( X_{\text{max}}' > \sigma p \) there exists \( \sigma p \leq X' \leq X_{\text{max}}' \) such that \( \tau_t(X_t') \) is convex in \([0, \sigma p)\), concave in \((\sigma p, X')\), and convex in \((X', X_{\text{max}}')\).

Similarly, we introduce a counterfactual commitment case that assumes that agents are unaware of the constraint and act as if the principal’s capacity constraint allowed her to respond to coordinated deviation and stick to committed strategy. We find that in such a case tax rates can be higher and are more responsive to increases in initial capacity. Since the fiscal capacity required to enforce an auditing probability of at least \( \theta_t^r \) is \( X_t' \geq \sigma p \theta_t^r \), the counterfactual tax rate constraint is given by

\[
\tau_t^c(X_t') = \frac{X_t'}{\sigma p - \tau_t} \phi.
\]  

(29)

Obviously, this constraint does not bind for any \( X_t' \geq \sigma p \), since the government can enforce any tax rate less than one by committing to monitor with probability one any taxpayer who does not pay taxes.

**Proposition 6.** \( \tau_t^c(X_t') \) is convex in \([0, \sigma p)\), greater than \( \tau_t(X_t') \), and steeper than \( \tau_t(X_t') \).

### D Proofs of Section 6.1

**Proof of Proposition 1.** The second derivative of (15) w.r.t. \( X_t \) is given by

\[
\frac{d^2 b_t}{dX_t^2} = \begin{cases} 
\frac{2(r \log p)^2}{(1 + (r \log p) X_t)^2} y & X_t < p \\
\frac{-r (1 - r (X_t(1 - (1 - 2 \log X_t) \log X_t) - p))}{X_t (1 - r (X_t(1 - \log X_t) - p))^3} y & X_t > p.
\end{cases}
\]

First, notice that the denominator on the RHS of both cases must be positive, otherwise \( b_t(X_t) > \bar{b} \). This implies that \( d^2 b_t/dX_t^2 \) if \( X_t < p \) since both the numerator and denominator are positive.
Now consider the case of $X_t > p$. Since the denominator is positive, convexity is determined by the sign of the numerator. We prove the result by showing that, in the range of capacities $X_t > p$ for which the denominator is positive, the numerator crosses zero at most once and from below. This would imply that $d^2b_t/dX_t^2$ is negative to the left of the crossing point and positive to the right.

To show that the numerator either is above zero or crosses once and from below we first differentiate it w.r.t. $X_t$ to obtain $r(\log X_t)(3 + 2\log X_t)$. This expression is positive for $X_t < 1/e^{3/2}$ and negative for $X_t > 1/e^{3/2}$. That is, the numerator is single peaked. Next notice that if the denominator is zero then the numerator must be positive. This is because $X_t(1 - (1 - 2\log X_t)\log X_t) > X_t(1 - \log X_t)$ given that $\log X_t < 0$ for $X_t \in (0, 1)$. But, since the denominator is decreasing in $X_t$, this implies that in the relevant range of $X_t$, i.e., for those capacities leading to $0 < b_t(X_t) < \bar{b}$, the numerator never turns negative after becoming positive.

Let $\hat{X}$ be the capacity at which the numerator is zero. The above single crossing result implies that if $\hat{X} > p$ then $b_t(\hat{X})$ is concave in $(p, \hat{X})$. In addition, if $\hat{X} < X_{\max}$ then $b_t(\hat{X})$ is convex in $(\hat{X}, p_{i_{\max}})$.

**Proof of Proposition 2.** Convexity comes from the fact that the second derivative of $b_t^c(X_t)$ is given by

$$\frac{d^2b_t^c(X_t)}{dX_t^2} = \frac{2(1 - p)^2 pr^2}{(p - (1 - p)r X_t)^2} y > 0.$$

We next show that $b_t^c(X_t) > b_t(X_t)$ whenever $b_t^c(X_t) < \bar{b}$. Notice that a necessary condition for $b_t^c(X_t) < \bar{b}$ to hold for is $X_t < p$. This is because at $X_t = p$ the government can commit to monitor everyone with probability one. In such case, only non-successful entrepreneurs default, requiring a capacity equal to $p$. Thus we just need to compare borrowing constraints for $X_t \leq p$. Since $p/(1 - p) < -1/\log p$ for all $p \in (0, 1)$ we have that

$$b_t^c(X_t) = \frac{p X_t}{1 - p - r X_t} y > \frac{r X_t}{-\log p - r X_t} y = b_t(X_t).$$

To prove that $b_t^c$ is steeper than $b_t$ we just compute the derivatives and obtain

$$\frac{db_t^c(X_t)}{dX_t} = -\frac{(1 - p) pr}{(p - (1 - p) r X_t)^2} y = \frac{1}{X_t} \frac{b_t^c(X_t)}{1 - \frac{1 - p}{p} r X_t}$$

$$> \frac{1}{X_t} \frac{b_t(X_t)}{1 - (-\log p) r X_t} = -\frac{r \log p}{(1 + (r \log p) X_t)^2} y = \frac{db_t(X_t)}{dX_t},$$

where the inequality comes from the fact that $b_t^c(X_t) > b_t(X_t)$ and $(1 - p)/p > -\log p$ for all $p \in (0, 1)$. \qed
Proof of Proposition 5. Let $p' = \sigma p$. The second derivative of (28) w.r.t. $X'_t$ is given by

$$
\frac{d^2 \tau_t}{d X'^2_t} = \begin{cases}
\frac{2(\log p')^2}{(1-p' + (\log p')X'_t)^3} & X'_t < p' \\
-(1-p) (1 - X'_t(1 - (1 - 2 \log X'_t) \log X'_t)) & X'_t > p'.
\end{cases}
$$

First, notice that the denominator on the RHS of both cases must be positive, otherwise $\tau_t(X'_t) > 1$. This implies that $d^2 \tau_t/d X'^2_t$ if $X'_t < p'$ since both the numerator and denominator are positive.

Now consider the case of $X'_t > p'$. Since the denominator is positive, convexity is determined by the sign of the numerator. But then, it is straightforward—and therefore omitted—to check that the single crossing argument used in the proof of Proposition 1 applies here: the numerator crosses zero from below at most once and never crosses zero from above in $[p', X'_{\text{max}})$. This implies that the second derivative is concave in $(p', \hat{X}')$ and convex in $(\hat{X}', X'_{\text{max}})$, where $\hat{X}'$ satisfies $1 = \hat{X}'(1 - (1 - 2 \log \hat{X}') \log \hat{X}')$.

Proof of Proposition 6. Let again $p' = \sigma p$. Convexity comes from the fact that the second derivative of $\tau_t(X'_t)$ is given by

$$
\frac{d^2 \tau_t(X'_t)}{d X'^2_t} = \frac{2p'}{(p' - X'_t)^3} \phi > 0 \quad \forall X'_t < p'.
$$

We next show that $\tau^c_t(X'_t) > \tau_t(X'_t)$ whenever $X'_t < p'$. Since $p' < -(1 - p')/ \log p'$ for all $p' \in (0, 1)$ we have that

$$
\tau^c_t(X'_t) = \frac{X'_t}{(p' - X_t) \phi} > \frac{X'_t}{\frac{1-p'}{1- \log p'} - X'_t} \phi = \tau_t(X'_t).
$$

Likewise, we have that

$$
\frac{d \tau^c_t(X'_t)}{d X'_t} = \frac{p'}{(p' - X'^2_t) \phi} \phi = \frac{\tau_t(X'_t)}{X'_t} \frac{1}{1 - \frac{1}{p'} X'_t} > \frac{\tau_t(X'_t)}{X'_t} \frac{1}{1 - \frac{1}{1- p'} X'_t} = \frac{d \tau_t(X'_t)}{d X'_t}.
$$

References


