December 15, 2014

Financial Contracting with Enforcement Externalities

Lukasz A. Drozd and Ricardo Serrano-Padial*

[Preliminary]

Abstract

Financial markets crucially rely on the development of an infrastructure dedicated to the enforcement of contracts. Here we study the effects of limited enforcement capacity on financial contracting by proposing a new theory of costly state verification. In our model the principal contracts with a population of entrepreneurs, who borrow to finance risky projects under limited liability. To sustain incentives to repay debt, the principal must build enforcement capacity ex ante, which determines state verification efforts ex post. Our theory sheds new light on such phenomena as credit crunches and the link between enforcement infrastructure accumulation, economic growth and political economy frictions.

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

JEL codes: D82, D84, D86, 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, Marzena Rostek, Joel Sobel, Marek Weretka and audiences at Philadelphia Fed, St. Louis Fed, UW-Madison and the 2013 Fall Midwest Theory Conference for their comments. Drozd acknowledges the financial support of the Rodney White Center at the Wharton School of the University of Pennsylvania. All remaining errors are our own.

1 Introduction

Financial contracting under asymmetric information often involves some form of costly state verification. Such a need arises when contracts are contingent on events or actions that can only be observed at a cost by the uninformed party. For instance, in financing of risky investment projects, it is often desirable to make repayment conditional on the success of the project. However, to sustain incentives or preserve the liquidation value of the project lenders must incur costs to learn the true outcome of reportedly failed projects. In applied contexts, the idea of costly state verification has proven fruitful to relate such credit market imperfections as incompleteness of contracts, borrowing constraints or credit spreads to economic fundamentals, thereby improving our understanding of the interaction between financial markets and the economy.[1]

In this paper we enrich this standard framework by considering frictions that limit the ability of the principal to sustain state verification strategies independently of the collective action taken by agents in the economy. Specifically, in our model the principal contracts with a population of agents rather than with a single agent. Agents are entrepreneurs who take loans to finance risky investment projects under limited liability, and the heterogeneous outcome of their projects can only be observed by the principal at a cost. To ensure repayment and preserve the liquidation value of reportedly failed projects, the principal must build enforcement capacity ex ante. Such a capacity may constrain her ability to verify agents' project returns ex post, implying that the collective action taken by agents, i.e., the default rate on loans, crucially influences the probability that a defaulting agent will be subject to state verification ex post and thus her incentives to default. We refer to this effect as an *enforcement externality* and study its implications in the context of financing of risky investment.

[1] Costly state verification framework was introduced by Townsend (1979). The basic idea has been further developed by Gale and Hellwig (1985) and the ensuing literature. It has been successfully employed in large scale DSGE models following the work of Bernanke et al. (1999).

Apart from the need to build enforcement capacity, the incentive problem faced by the principal in our model is fairly standard. Since agents may lack funds when their project fails and are protected by limited liability, loan contracts must admit the possibility of default. Default implies that the principal may liquidate the project at a dead-weight cost. The default option introduces a moral hazard problem because the potential lack of enforcement allows agents to take over part of the liquidation value of the project. Contracts maximize agents' ex ante utility subject to a zero profit condition in the competitive loan market. Accumulation of enforcement capacity is governed by a trade off between principal's own consumption of initial resources and entrepreneurs' utility. The latter feature allows us to explore the effects of political economy distortions or other imperfections affecting principal's incentives to build enforcement capacity.

The presence of enforcement capacity raises two important questions. The first question is how it affects the economy's response to aggregate shocks. This question is broadly relevant for the theory of credit crunches. In this respect, the key novel implication of our model is that a shock that initially raises the default rate in the economy, by lowering expected state verification probability ex post, adversely affects incentives to default of all agents in the economy. As a consequence, the initial shock is amplified by the constraint. Importantly, the amplification mechanism is driven by strategic complementarities associated to enforcement externalities, which lead to the *contagion* of default among agents whose intrinsic propensity to default differs due to heterogeneity of project returns. In particular, default of prone-to-default agents endogenously triggers default of some agents who are less prone to default, rendering the principal's preferred allocation fragile to shocks: since maintaining spare capacity is socially costly, and so is equilibrium featuring a high default rate, the principal reacts to shocks by issuing less credit to entrepreneurs so as to prevent a clustering of defaults. Accordingly, credit provision is highly sensitive and shocks that raise default rate on pre-existing loans can lead to credit crunches in the market for new loans.

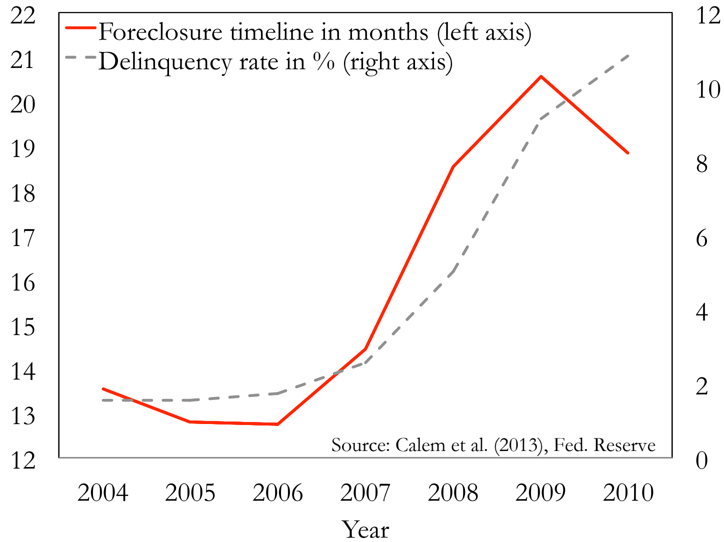

Figure 1: The U.S. foreclosure timeline (left axis) and mortgage delinquencies (right axis).

A good illustration of the above mechanism being at play in the data is the 2007-2009 wave of foreclosures in the U.S. mortgage market. As is clear from Figure 1, the enforcement infrastructure was stretched thin during this period, which created strong financial incentives to default due to significantly lengthened foreclosure timelines. In many troubled areas defaulting borrowers could enjoy rent-free consumption of housing for as long as 1-2 years.[2] In this context, the key prediction of our model is that an event that initially raises the default rate will result in severe tightening of credit availability in the new loans market.

The second question we address is how the presence of enforcement externalities affects the incentives of developing economies to build enforcement infrastructure so as to enhance the depth of their financial markets and this way spur entrepreneurship. This question is germane

[2]Foreclosure delays mostly stemmed from the backlog of filings in the court system and were specially acute in states with judicial foreclosure (Cordell et al., 2013). Foreclosure proceedings can be linked to state verification. While lenders and borrowers are bound by a contract that allows lenders to take over the property, only courts can verify whether such a contract is valid by issuing an appropriate order. Thus, the court infrastructure introduces a capacity constraint in the verification and enforcement of mortgages.

to the recent theory emphasizing the role of state capacity in economic development put forward by Besley and Persson (2010, 2009, 2011). In this context, our analysis shows how accounting for the strategic complementarities brought by limited capacity profoundly affects growth dynamics in this class of models by making incentives for capacity accumulation much more sensitive to political economy distortions. In particular, the aforementioned clustering effect can result in development traps, characterized by lack of investment in enforcement capacity and diversion of resources to principal's consumption.

From a technical point of view, the analysis of our model involves overcoming two important challenges. First, strategic complementarities may introduce multiple equilibria when entrepreneurs observe the principal's enforcement capacity. To address this issue, we resort to global games methods introduced by Carlsson and van Damme (1993), Morris and Shin (1998) and Frankel, Morris and Pauzner (2003) and select a unique equilibrium. Specifically, we show that when agents observe enforcement capacity with a small noise all equilibria except one are eliminated.

The second challenge we face is that heterogeneity of returns greatly complicates the characterization of the unique equilibrium. The lack of such a characterization would prevent us from analyzing the principal's choice of contracts and enforcement capacity. To address this problem, we build on the insightful work of Sakovics and Steiner (2012) on games with symmetric equilibria, and show how to characterize equilibrium when heterogeneity leads to asymmetric equilibria, as is the case in our model.[3] In doing so, we are also able to identify the conditions under which there exist a cluster of agents with heterogeneous (intrinsic) propensities to default that share the same default strategy.

Our paper provides two numerical applications of substantive interest. The first shows that,

[3]The model of Sakovics and Steiner (2012) allows for heterogeneous preferences but imposes preference restrictions that lead to symmetric equilibria in the complete information game. In the context of our model, this would imply that either all agents or none of them default. In contrast, equilibrium in our model typically involves some agents repaying and some defaulting.

in response to large but infrequent shocks, the enforcement capacity constraint may indeed be occasionally binding, as it is too costly to maintain residual enforcement capacity. In such a case, a credit crunch typically follows, implying a contraction of credit by up to 30-50% in our calibration. The key reason why this is happening is that, in the absence of the shock, enforcement externalities lead to a cluster of agents that are indifferent between defaulting and repaying at the optimal contract. Hence, if the principal ignores the shock when issuing loans there would be a wave of defaults, which can only be prevented by sharply reducing credit so as to lower entrepreneurs' propensity to default. Our second application shows that, due to the aforementioned clustering effect, political economy frictions can have a much larger impact on economic convergence of developing economies, compared to the case in which enforcement externalities are absent. In particular, we illustrate how a small increase in the principal preferences' for own consumption can lead to an abrupt switch to what we refer to as a *financial development trap* characterized by the absence of lending to entrepreneurs.

Related literature.—Externalities associated with capacity constrained enforcement have been long recognized as important in the economic literature. However, our paper is the first to analyze such a feature in a costly state verification framework. To do so we develop global games methods for economies featuring agent heterogeneity, following the approach of Frankel, Morris and Pauzner (2003) and Sakovics and Steiner (2012). We build on the key insight from Sakovics and Steiner (2012) and extend it to derive new results for games in which payoff heterogeneity leads to asymmetric equilibrium exhibiting partial clustering of strategies. In a broader context, our paper is one of the first characterizing the effect of heterogeneity on the fragility of equilibrium outcomes in global games.

Very few papers consider the consequences of limited enforcement capacity in the context of financial contracting. A notable exception is the work of Arellano and Kocherlakota (2008), who stress the role of exogenous limits to asset liquidation during sovereign crisis events. Their paper points out potential for multiplicity of equilibria implied by the fact that during crises more firms may need to be liquidated. It also provides evidence highlighting the relevance of such frictions. Another less closely related study is the paper by Bond and Rai (2009), who look at compliance in the micro-finance context of group lending. In contrast to our paper, however, they study a symmetric information game in which a homogeneous population of borrowers may stage speculative attack on the exogenously fixed enforcement ability of a single lender. The effect of capacity constraints has been also studied in the literature on tax evasion and crime control, which we do not discuss here in the interest of brevity.

2 Model

Our model specifies a game between a single principal and a population of ex ante identical agents (entrepreneurs) of measure one. Agents need resources to finance risky investment projects. The principal has deep pockets and provides funding, but must earns non-negative expected profits from the repayment of loans. In order to do so, she accumulates *enforcement capacity* ex ante and verifies or *monitors* agent projects ex post. Monitoring sustains incentives and preserves the liquidation value of the projects.

We restrict the contractual space to debt contracts subject to limited liability. Accordingly, in the case of default the principal can seize agents' projects and otherwise the agent keeps the project. Importantly, project liquidation is inefficient, as it involves a dead-weight loss. The principal only obtains the full liquidation value of the project when she monitors it. In the absence of any enforcement by the principal defaulting agents capture part of their project liquidation value.

Projects yield a random return $w \in [0, \infty)$. The returns become known after investing and are private information to the agents. The output of the project is given by $w(y + b)$, where

y is entrepreneurs' own equity and b is the loan from the principal. Returns w are distributed according to a continuous distribution function F, which is common knowledge. The associated density function is denoted by f.

Agents maximize their own profit and are risk neutral. The principal maximizes expected utility of the agents subject to a zero profit condition, her own internal valuation of ex ante resources invested in enforcement infrastructure, and the enforcement technology.

In what follows, we lay out the timing of the game, and specify actions and payoffs.

2.1 Timing of Events, Payoffs, and Actions

The game between the principal and the agents has three stages that occur sequentially:

1. Capacity accumulation stage. In the first stage, the principal first decides how to allocate her exogenous resources R between building enforcement capacity X at a cost $c(X)$ and own consumption g, which may represent the provision of public goods or, alternatively, a diversion of resources to principal's own consumption. X represents the mass of projects the principal will be able to monitor. As will be clear later, g determines the opportunity cost of using ex ante resources to invest in enforcement capacity. Accordingly, the principal faces the following constraints:

$$R = c(X) + g, \tag{1}$$

and

$$g \geq 0. \tag{2}$$

After the principal chooses X and g, an aggregate random shock absorbs an amount $s \in [0, \bar{s}]$ of her enforcement capacity. The shock can be interpreted as an increase in the capacity needed to enforce pre-existing contracts. As a result, the residual enforcement capacity to enforce new contracts is given by $X(s) = X - s$.

2. Contracting stage. During the contracting stage $X(s)$ is fixed and cannot be changed by the principal. Each entrepreneur applies for a loan with the principal to invest in a project. The loan is described by a tuple $(b(s), \bar{w}(s))$, where b is the loan amount and $\bar{w}(s)(y + b(s))$ is the repayment amount, so agents who repay get to consume $(y + b(s))(w - \bar{w}(s))$.

If an agent defaults on the loan the project is liquidated at a dead-weight liquidation cost proportional to the project's return, $(1-\mu)(y+b(s))w$, where $\mu < 1$. If the agent is monitored by the principal she gets nothing while the principal seizes the entire residual value of the project $\mu(y+b(s))w$. However, if the agent is not monitored, the principal retains only a portion $(1-\gamma)$ of the liquidation value and the agent gets the remaining fraction $\gamma \in [0, 1]$. These assumptions allow the model to capture in a reduced form the possibility that any defaulting agent eventually loses the project but that the lack of prompt enforcement actions allow him to temporarily enjoy some rents from the project. Note that when $\gamma = 1$ the agent is able to get away with the full liquidation value of the project when she is not monitored.

Contractual terms $(b(s), \bar{w}(s))$ are chosen so as to maximize the surplus from production, which we describe formally in the next section.

3. Enforcement stage. In the last stage of the game project returns are realized and privately observed by agents. Agents simultaneously decide whether to default or pay back $\bar{w}(y + b)$.

After observing action taken by the agents, the principal allocates her enforcement capacity to monitor those agents who choose to default. Given limited enforcement capacity, the probability of monitoring $P(s)$ faced by defaulting agents is subject to an *enforcement capacity constraint* given by:

$$P(s)\psi(P^e(s)) \leq X(s), \text{ all } s \in S, \tag{3}$$

where $\psi(P^e)$ denotes the fraction of agents who choose to default in state s when all expect P^e the monitoring probability faced by defaulting agents. The constraint simply states that the mass of agents who are monitored in equilibrium, $P(s)\psi(P^e(s))$, must no larger than the residual capacity $X(s)$.

We next derive best response functions of the agents. At the end of the section we define the problem of the principal. All proofs are relegated to the Appendix.

2.2 Objectives and Best Responses

Let the event of being monitored be random variable $m = 0, 1$, where $m = 1$ means the agent is monitored ex post. Furthermore, let $a = 0, 1$ denote the repayment decision of the agent, where $a = 1$ means repayment. Then, given contract $(b(s), \bar{w}(s))$, agents' utility is given by

$$u(a, w, m) := \begin{cases} (y + b(s))(w - \bar{w}(s)) & a = 1 \\ \gamma\mu(y + b(s))w & a = 0 \ \&\ m = 0 \\ 0 & a = 0 \ \&\ m = 1. \end{cases} \tag{4}$$

Given her belief $P^e(s)$, an agent's best response function $a(s, w)$ maximizes her expected utility, that is, it solves

$$a(s, w) := \underset{a}{\operatorname{argmax}}\{P^e(s)u(w, a, 1) + (1 - P^e(s))u(w, a, 0)\}. \tag{5}$$

It is easy to see that the above best response function implies a simple cutoff strategy w.r.t. her expected monitoring probability P^e.[4]

[4]Given the continuity of F we can assume without loss that an indifferent agent always chooses to repay.

Lemma 1. *The best response function of an agent with contract* $(b, \bar{w})$ *is*

$$a(s,w) = \begin{cases} 1 & \text{if } P^e(s) \geq \theta_{\bar{w}(s)}(w) \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where $\theta_{\bar{w}}(w) := 1 - \frac{1}{\mu\gamma}\left(1 - \frac{\bar{w}}{w}\right)$ *is agent's propensity to default.*

The best response function allows us to formally define the default rate in the economy when all agents share the same belief P^e,

$$\psi(P^e(s)) := \int_{\{w: a(s,w)=0\}} f(w)dw. \tag{7}$$

The goal of the principal is to maximize the sum of the expected aggregate utility of entrepreneurs and the provision of public goods g, weighted by $\alpha \geq 0$, which captures principal's own utility from using in alternative purposes ex ante resources that could be invested in enforcement infrastructure. In this context, α captures political economy distortions affecting the principal's choice. Formally, the principal chooses X, g, and $(b(s), \bar{w}(s))$ to maximize:

$$\alpha g + \mathbb{E}[\int_{\{w:a(s,w)=1\}} (y + b(s))(w - \bar{w}(s))dF + \mu(1 - P(s)) \int_{\{w:a(s,w)=0\}} \gamma(y + b(s))wdF], \tag{8}$$

subject to $R = c(X) + g$, $g \geq 0$, and

$$\begin{aligned} b(s) \quad \leq \quad & \int_{\{w:a(s,w)=1\}} (y + b(s))\bar{w}(s)dF + \\ & \mu P(s) \int_{\{w:a(s,w)=0\}} (y + b(s))wdF + \\ & \mu(1 - P(s)) \int_{\{w:a(s,w)=0\}} (1 - \gamma)(y + b(s))wdF \text{ for all } s. \end{aligned} \tag{9}$$

The objective function of the principal (8) integrates over the aggregate state s and idiosyncratic returns of all projects. The second and third term capture the expected utility of entrepreneurs who repay and default, respectively.

The last constraint is an interim zero profit condition, i.e., the expected profit from loans must be non-negative. This constraint can be interpreted as a result of Bertrand competition between lenders in the credit market or, alternatively, a principal who offer best terms to the agents but must break even on average. The revenue (right-hand side) comprises loan payments as well as revenue from project liquidation. As is clear from the constraint, by monitoring agents who default, the principal prevents agents from taking a portion of the liquidated projects.[5]

2.2.1 Technical Assumptions and Parameter Restrictions

We make two assumptions regarding the distribution of types, F:

1. We restrict attention to a family of functions F that obey the restriction: $\dfrac{F(w)}{wf(w)}$ increasing and $\lim_{w\downarrow 0} \dfrac{F(w)}{wf(w)} < 1$.

   The usual distribution functions, such as log-normal or Pareto, satisfy this condition. Our results are general and we derive them in the appendix without this assumption.

2. To prove our results, we work with a discrete distribution F such that each type w has positive mass. However, to facilitate the exposition, we assume that F is an arbitrarily fine discrete approximation of a continuous distribution function satisfying the above restriction and state our results by taking the limit of the approximation to zero. In other words, we present the continuous distribution analogue of results that apply to economies with discrete F. In doing so, when we introduce noise we implicitly assume that in the

[5] Consider our earlier example of the foreclosure glut in the U.S. mortgage market. Delays implied by extended foreclosure timelines have benefited borrowers by allowing them to live in the property for an extended period of time. This benefit was at the expense of lenders who could invest funds obtained from selling foreclosed properties and additionally were exposed to losses due to deferred maintenance.

model with continuous F the law of large numbers holds by type as is the case in the discrete version.[6]

Finally, we assume that liquidation costs are high enough relative to the expected rate of return to prevent unbounded loans levels sustained by a default rate of 100% in the absence of capacity constraints. Specifically, we require that $\mu \mathbb{E} w < 1$.

2.3 Equilibrium of the Enforcement Game

We focus first on the analysis of the game associated to the enforcement stage of the model (stage 3). At this stage, g, X, s, $\bar{w}(s)$ and $b(s)$ are all fixed. To ease notation we drop references to s.

2.3.1 Multiplicity of equilibria under common knowledge

We begin our analysis under the assumption of common knowledge of X. In such a case agents know exactly the monitoring capacity X of the principal and the equilibrium monitoring probability P. Thus, given the strategy of other players $a(w)$, the expected monitoring probability P^e each agent faces when choosing $a = 0$ is

$$P^e = P = \min\left\{\frac{X}{\psi(P)}, 1\right\}. \tag{10}$$

The above equation reveals that the strategy of other agents crucially influences P^e, creating a feedback loop between incentive to default and the choices of other agents. This feature is central to our analysis and we refer to it as the *enforcement externality*. In particular, it introduces strategic complementarities, making our game supermodular.

The presence of enforcement externality brings about the possibility of multiple self-fulfilling

[6] We ignore the measurability issues associated with having a continuum of agents.

equilibria. Multiplicity limits how much we can learn from our model, as its predictions are effectively determined by external factors rather than the economic fundamentals in the model. In the next section we show that this result is not a robust prediction of our environment and is eliminated by a simple refinement. However, we first characterize what happens in its absence.

To this end, we observe that there are up to three equilibrium in our model. The first equilibrium is an efficient equilibrium under which all agents who can repay—those with $w \geq \bar{w}$, do not default. This equilibrium requires that $F(\bar{w}) \leq X$. To see why, note that if only agents with $w < \bar{w}$ default, the default rate is $F(\bar{w})$ and so $F(\bar{w}) \leq X$ implies $P^e = 1$ and hence that everyone with $w \geq \bar{w}$ is best responding in equilibrium.

However, there may also exist two other *inefficient* equilibria. In these equilibria some agents with $w > \bar{w}$ default. These equilibria are sustained by a binding enforcement capacity constraint.[7]

To pin down inefficient equilibria, consider a candidate equilibrium in which some agents with $w > \bar{w}$ default. Since the propensity to default $\theta_{\bar{w}}$ is strictly decreasing in w, such an equilibrium must be characterized by a pivotal type w_i such that all agents with $w < w_i$ default and all agents with $w \geq w_i$ do not default. That is, the default rate is $F(w_i)$. In addition, the pivotal type must be indifferent between defaulting or not. By Lemma 1, we know that the pivotal type is indifferent when $P^e = \theta_{\bar{w}}(w_i)$. Hence, for P^e to be sustained, we must have

$$X = \theta_{\bar{w}}(w_i)F(w_i). \tag{11}$$

It turns out that the above equation admits up to two solutions. This is because, under the above assumptions $\theta_{\bar{w}}(w)F(w)$ is single-peaked (see Lemma 2 below). Thus, there are up to three equilibria in our model, two of which are inefficient. For example, consider the usual

[7] Any such equilibrium must exhibit $P < 1$, otherwise agents with $w > \bar{w}$ would not find optimal to default.

log-normal distribution and suppose $\bar{w}$ is to the left of the peak of $\theta_{\bar{w}}(w)F(w)$ denoted by $w_{\max}$, which is a reasonable assumption as otherwise default rate is very high. Figure 2 illustrates the existence of three equilibria for any enforcement capacity between $\underline{X}$ and $\overline{X}$. Proposition 1 formalize these results.

Lemma 2. *If* $\dfrac{F(w)}{wF'(w)}$ *is increasing and* $\lim_{w\downarrow 0}\dfrac{F(w)}{wF'(w)} < 1$ *then* $\theta_{\bar{w}}(w)F(w)$ *is increasing at* 0 *and single-peaked.*

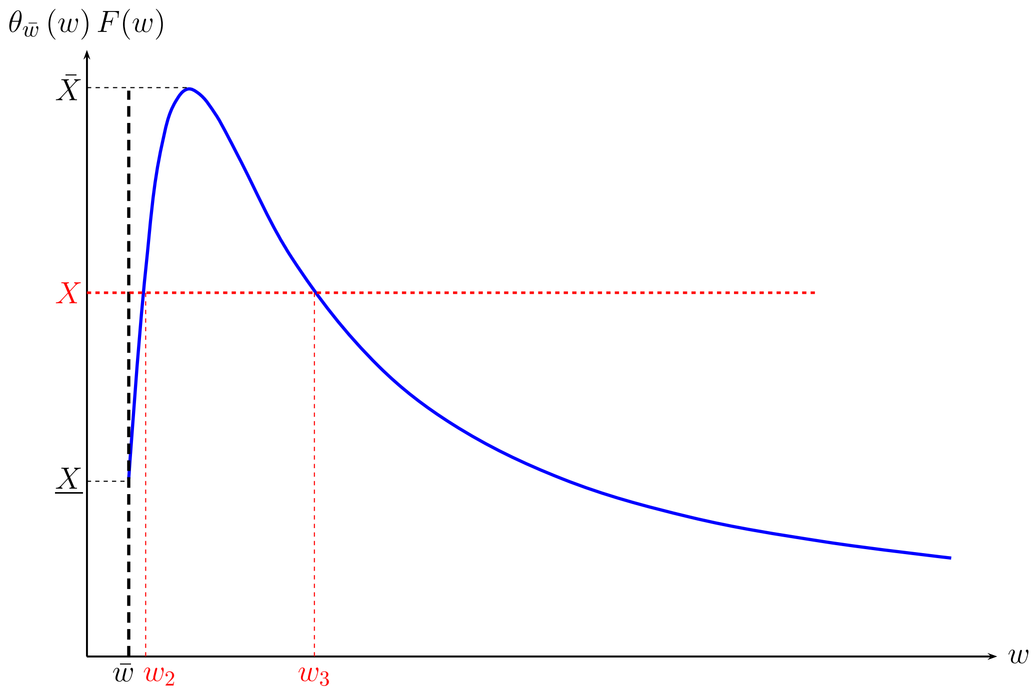

Figure 2: Multiplicity of Equilibria Under Common Knowledge.

Definition 1. *Let* $w_{\max}$=argmax$_w\,\theta_{\bar{w}}(w)F(w)$, $\underline{X} = F(\bar{w})$, *and* $\overline{X} = \theta_{\bar{w}}(w_{\max})F(w_{\max})$.

Proposition 1. *Under common knowledge of* X, *if* $\bar{w} < w_{\max}$, *equilibrium is unique iff* $X < \underline{X}$ *or* $X > \overline{X}$. *Otherwise, there are three equilibria in the case of* $X \in (\underline{X}, \overline{X})$ *and two equilibria in the case of* $X = \underline{X}$ *and* $X = \bar{X}$. *If* $\bar{w} \geq w_{\max}$ *equilibrium is always unique.*

2.3.2 Equilibrium selection

Multiplicity of equilibria in our model crucially relies on the assumption that agents are perfectly informed about enforcement capacity X. As the global games literature shows, even infinitesimal uncertainty in this respect can induce large strategic uncertainty about the equilibrium action of others, eliminating all but one equilibrium. We use this refinement to argue that multiplicity is not a robust prediction and characterize the unique equilibrium implied by the refinement.

The global game version of the enforcement game is identical to the complete information game except that each agent receives a signal $x = X + \nu\eta$, where $\nu > 0$ is a scale factor and η is an i.i.d random variable, independent of X, with continuous distribution H and full support on $[-1/2, 1/2]$. Before receiving the signals, agents' prior about X is the uniform distribution in $[0, 1]$, that is, agents do not have prior knowledge of X before receiving their signals.[8] This implies that an agent with signal x believes that $X = x - \nu\eta$, where $\eta \sim H$.

Accordingly, the expected probability of state verification P^e by each agent now depends on her signal and thus differs across agents. Specifically,

$$P^e = \min\left\{\frac{x - \nu\eta}{\psi}, 1\right\},$$

where the default rate ψ is a random variable whose distribution depends on x. Thus, by Proposition 1, the best response function of each agent depends on the realization of her private signal x, in addition to her type w.

Our goal is to characterize equilibrium as ν goes to zero.

Equilibrium uniqueness. Our first result shows that there is a unique equilibrium in threshold strategies characterized by a cutoff $k(w)$ such that an agent of type w chooses $a = 1$ when

[8] Our results do not hinge upon the uniform prior assumption. As Frankel, Morris and Pauzner (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.

her signal is above $k(w)$ and $a = 0$ when her signal is below $k(w)$. This is the only strategy profile that survives iterated elimination of strictly dominated strategies.

Proposition 2. *The game has a unique equilibrium.*[9] *Equilibrium strategies are characterized by a cutoff on signal $k(w)$, such that:*

- *If $x \geq k(w)$, agents choose to repay, $a = 1$,*
- *If $x < k(w)$, agents choose to default, $a = 0$.*

The uniqueness result is fairly standard, and our proof follows closely Frankel, Morris and Pauzner (2003). Abstracting from agent heterogeneity, the following informal argument is helpful to understand why equilibrium is unique in this type of games.

Assume there are two types, one with type lower than $\bar{w}$ and the other with w such that $\theta_{\bar{w}}(w) \in (0, 1)$, and fix noise ν at some sufficiently small level. Recall that the default rate satisfies $\psi \geq F(\bar{w})$. In this context, at extreme signals agents of type w follow a dominant strategy: they always default for low enough x, since they expect $P^e \leq X/F(\bar{w})$ to be lower than their propensity to default $\theta_{\bar{w}}(w)$, and never default for high enough x, since $P^e \geq X$ will greater than $\theta_{\bar{w}}(w)$. Denote the upper bound and the lower bound of these dominance regions by $\underline{x}$ and $\overline{x}$, respectively.

Next, consider an agent who receives a signal $x > \underline{x}$ arbitrarily close to $\underline{x}$ relative to ν. Clearly, since her signal is very close to the low dominance region, such an agent will conclude that *almost* all agents who receive a signal to the left of her signal default. Hence this agent will also find optimal to default since the default rate will be high relative to the tiny increase in expected capacity relative to signal $\underline{x}$.

After repeating this elimination procedure for all agents, and moving up along the signal range, at some point an indifference between defaulting or not defaulting must be reached

[9]Equilibrium strategy is unique up to sets of measure zero.

since expected capacity goes up with x. Note that this is inevitable due to the presence of the upper dominance region. Let $\hat{x}$ be the point of indifference and consider signals $x > \hat{x}$ such that $x - \hat{x} < \nu/2$. In such a case, the agent would prefer not to default, as her belief about default rate ψ may only go down with x within this range while her belief about X has gone up. Repeating this iteratively, we can eliminate all strategies that involve default in the region above $\hat{x}$. Hence, equilibrium is characterized by a single threshold such that the agent who receives a signal equal to the threshold is indifferent between defaulting or not.

Introducing heterogeneity changes does not change the gist of the above reasoning, except in that it introduces heterogeneity in cutoffs brought by the fact that thresholds must satisfy indifference conditions[10]

$$E(P^e|x = k(w)) = \theta_{\bar{w}}(w) \text{ for all } w \text{ with } \theta_{\bar{w}}(w) \in (0, 1). \tag{12}$$

Nonetheless, it makes solving for equilibrium thresholds $k(w)$ challenging. In the above two-type case, there is just one strategy followed by all agents with type w, implying that probability distribution of ψ involved in the expectation operator of the left hand side is uniform. This is referred to as the *Laplacian property* of the threshold type (Morris and Shin, 2003). It follows from the simple fact that the mass of agents with w that default is given by those with signals $x < k(w)$. But since signals are i.i.d., for an agent receiving the threshold signal $k(w)$ the probability that a fraction z of type-w agents receive $x < k(w)$ is the same as the probability that such fraction is z', for all $z, z' \in [0, 1]$. (For details, see Morris and Shin (2003)).

[10]The proof uses the fact that games with strategic complementarities feature a smallest and largest Nash equilibrium, both in monotone cutoff strategies (Milgrom and Roberts, 1990), and the fact that the signal distribution is translation invariant. The latter means that, if an agent receives $x' = x + \Delta$ the distribution of others' signals conditional on x' is identical to the distribution conditional on x shifted by Δ. Thus, as we shift up cutoffs $k(\cdot)$ by Δ the distribution of ψ conditional on $k(w) + \Delta$ is identical to its distribution conditional on $k(w)$ before the shift in cutoffs. However, the expected capacity has gone up implying an increase in P^e. Hence, as we move from the smallest to the largest equilibrium, expected monitoring probabilities go up, implying that there is a unique profile of cutoffs at which $E(P^e|x = k(w)) = \theta_{\bar{w}}(w)$ for all agents with $\theta_{\bar{w}}(w) \in (0, 1)$.

Unfortunately, in the presence of agent heterogeneity the Laplacian property of the threshold type is lost, as it is possible that other agent types may choose a strategy that is different yet lies in the neighborhood of the other types' strategy (i.e., on its signal support), in which case the distribution of default rates are no longer uniform since some agent types will exhibit higher default rates than others.[11] In principle, solving for (12) seems hopeless: we do not know when such a clustering may occur without solving the conditions that define the equilibrium, and to solve the conditions we must know the probability distribution that underlies the expectation operator. Nonetheless, a key feature of beliefs is that the Laplacian property applies *on average*, as we explain below. We use this fact to circumvent the need to pin down individual beliefs and fully characterize the solution by averaging indifference conditions (12) of types in any potential cluster and using them to identify which cluster arises in equilibrium. Specifically, we use the joint average condition $\int_{w\in W'} E(P^e|x = k(w))f(w)dw = \int_{w\in W'} \theta_{\bar{w}}(w)f(w)dw$, where W' is a subset of types that may cluster on the same threshold $k(w)$ in the limit $\nu \downarrow 0$ and replace individual beliefs by their average. Since the threshold is identical in the limit—thresholds in a cluster must lie in a neighborhood of size 2ν, this is all we need to solve the game. The next section provides the details and characterizes the equilibrium using this approach.

Characterization of equilibrium. To characterize the solution of our game, we derive conditions for clustering of heterogeneous agent types and solve the average indifference condition of types in a cluster. To do so, we use the insight from Sakovics and Steiner (2012) to characterize the average belief of all threshold types (agents who receive threshold signal $x = k(w)$) and show that is the uniform distribution. We apply such insight to pin down the average belief of any subset W' of types in the game. In other words, if we randomly select a sample of agents with $w \in W'$, and ask them for their beliefs about the distribution of ψ when they receive their

[11] The behavior of types with threshold outside the signal support is known to an agent receiving her threshold signal $k(w)$ so she only faces uncertainty regarding the behavior of agents with thresholds in the signal support associated to $x = k(w)$.

threshold signal, the average answer is the uniform distribution.

Lemma 3 (belief constraint)**.** *Let $\psi(W')$ be the proportion of agents in a measurable set $W' \subseteq W$ choosing $a = 0$ when capacity is X, i.e.*

$$\psi(W', X) := \frac{1}{f(W')} \int_{W'} H\left(\frac{k(w) - X}{\nu}\right) f(w)dw.$$

Then, for any $z \in [0, 1]$,

$$\frac{1}{f(W')} \int_{W'} \mathbb{P}_w \left(\psi(W', X) \leq z \middle| x = k(w)\right) f(w)dw = z, \tag{13}$$

where $\mathbb{P}_w \left(\cdot \middle| x = k(w)\right)$ is the probability assessment of $\psi(W')$ by an agent receiving $x = k(w)$.

The belief constraint is instrumental to fully characterize equilibrium thresholds in our model as ν goes to zero. As highlighted by Remark 1 below, it allows us to express equation average indifference conditions in a closed form.

Proposition 3. *In the limit, as $\nu \to 0$, there exists $w^* \geq \bar{w}$ such that*

$$k(w) = \begin{cases} \theta_{\bar{w}}(w)F(w) & \text{for all } w > w^* \\ \theta_{\bar{w}}(w^*)F(w^*) & \text{for all } w \leq w^* \end{cases}$$

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

$$\int_{\bar{w}}^{w^*} \theta_{\bar{w}}(w)f(w)dw = \theta_{\bar{w}}(w^*)F(w^*) \left(1 - \log \theta_{\bar{w}}(w^*)\right) - F(\bar{w}). \tag{14}$$

Furthermore, $\theta_{\bar{w}}(w^*)F(w^*)$ is increasing in $\bar{w}$. (All agents with $w < \bar{w}$ default, regardless of the action of other agents in the economy.)

Remark 1. *Equation* (14) *follows from the fact that all agents in the cluster* $[\bar{w}, w^*]$ *must be individually indifferent between defaulting and not defaulting, implying that on average:*

$$\int_{w\in[\bar{w},w^*]} \theta_{\bar{w}}(w) f(w) dw = \int_{w\in[\bar{w},w^*]} \mathbb{E}\left(P^e(x)\middle| x = k(w)\right) f(w) dw. \tag{15}$$

The belief constraint (13) *allows to obtain equation* (14) *from equation* (15).

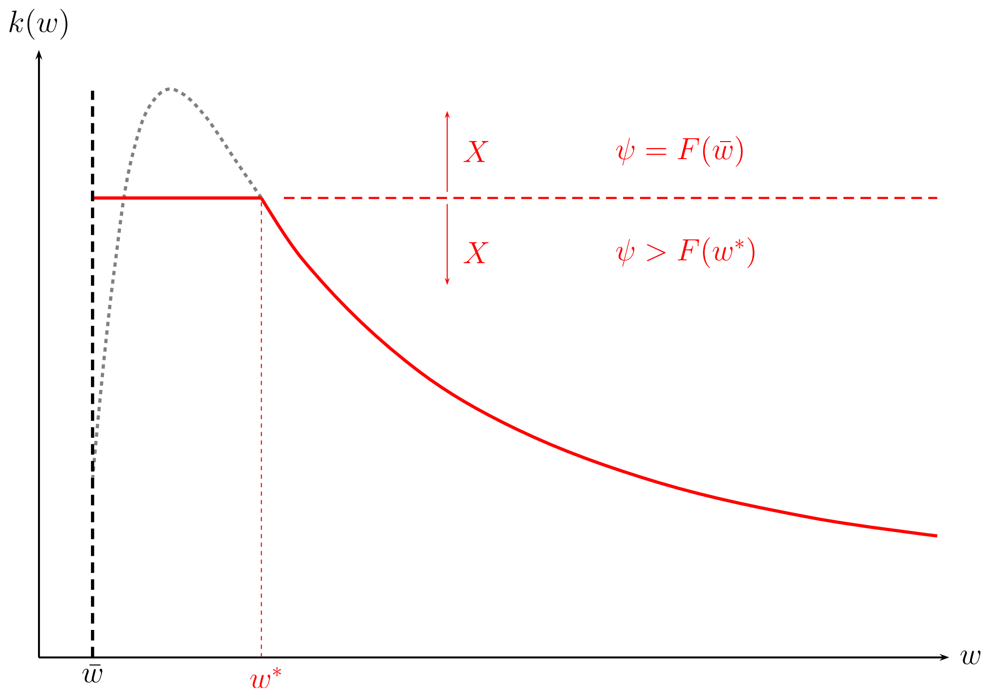

Figure 3: Equilibrium Strategy in the Global Game.

Figure 3 illustrates the equilibrium strategy derived in Proposition 3. It plots $k(w)$ in relation to $\theta_{\bar{w}}(w)F(w)$ studied earlier. Recall that the latter expression is the level of enforcement capacity X that makes agents of type w exactly indifferent between defaulting or not. As is clear from the figure, $k(w)$ is equal to it for $w > w^*$. However, this is not the case on the interval $w \in [\bar{w}, w^*]$, where agents from the interval 'cluster' on the same threshold strategy $k(w) = k(w^*)$. This is the main qualitative result of our paper. Specifically, this implies that when X is above $k(w^*)$ the efficient equilibrium is selected, as only agents below the prescribed cutoff $\bar{w}$ default. However, if X falls just below $k(w^*)$, all agents from the interval $[\bar{w}, w^*]$ choose to default and the most inefficient equilibrium is played.

Clustering of default decisions has important implications because, unless the opportunity costs of building capacity are low, it makes the efficient contract that the principal wants to implement fragile to shocks. As we show in the next section, the principal will typically tighten credit in response to shocks rather than allow for a wave of inefficient defaults. In other words, the principal will generally do "whatever it takes" to avoid widespread default and, once X is fixed, a credit contraction is the only tool she can use to lower agents' propensity to default. In addition, it endogenously introduces 'lumpiness' in capacity investment by the principal, which can lead to development traps when when distortions are high and resources are limited.

The reason why clustering arises in our model can be explained using a simple example with only three types, one below $\bar{w}$ and the other two above $\bar{w}$, namely, a low w type of mass f_l, denoted by l, and a high w type of mass f_h and denoted by h. Let her propensities to default be θ_l and θ_h, respectively, where $1 > \theta_l > \theta_h > 0$. Furthermore, let their respective threshold strategies are given by k_h and k_l. In equilibrium it must be that $\mathbb{E}_l(P|x = k_l) = \theta_l$ and $\mathbb{E}_h(P|x = k_h) = \theta_h$, where E_i denotes expectation of type $i = l, h$.

We argue that for some values of θ and f $k_l > k_h$ cannot arise ($k_l < k_h$ cannot happen in equilibrium since types l have a higher propensity to default). To see why, suppose $k_l > k_h$ and consider the belief of a threshold type h about the expected default rate in the economy—for simplicity we provide intuition using expectations rather than integrating over all possible values of ψ. Clearly, for ν sufficiently low type h will think that, on average, half of the agents of her type have a signal below k_h and half have the signal above this level. Thus, such an agent will conclude that, in expectation, half of the agents of her own type default. At the same time, she will believe that all agents of type l must default since signals will be very close to k_h. Hence, the belief of a threshold type h about the default rate in the economy is in this case $\mathbb{E}_h(\psi) = F(\bar{w}) + f_l + f_h/2$. We can similarly find that the belief of a threshold type l is $\mathbb{E}_l(\psi) = F(\bar{w}) + f_l/2$, implying $\mathbb{E}_h(\psi) - \mathbb{E}_l(\psi) = f_l/2 + f_h/2$. Given this, if $E_l(P|x = k_l) = \theta_l$ we can find a range of θ_h such that $\mathbb{E}_h(P|x = k_h) < \theta_h$. This is because $P = \frac{X}{\psi}$, and both the expected X and ψ are lower for type h.

In this context, the only possibility is that $k_l \downarrow k_h$ as $\nu \downarrow 0$ so as to make the two types indifferent. That is, k_l must fall within noise range of k_h, breaking the above reasoning. It is easy to see that both indifference conditions can be satisfied in such a case by positioning the thresholds close to each other to exactly offset the difference in the propensity to default.

2.4 Comparative Statics

What is the effect of X on credit provision in our model? It turns out that, because of the inefficiency of clustering due to the dead-weight loss associated to project liquidation, the planer typically adjusts the loan size to ensure that the efficient enforcement equilibrium is sustained. As a result, shocks lead to sizable credit crunches, rather than resulting in a wave defaults. In addition, political economy frictions in our economy can have a highly non-linear effect. We demonstrate these properties of the model by considering concrete numerical examples. But first we show that indeed adjusting credit is the way to guarantee the equilibrium with default rate $F(\bar{w})$.

Proposition 4. *The solution to the principal's problem* (8)-(9) *involves finite* $b(s)$ *for all* s*. If the solution involves* $X(s) = \theta_{\bar{w}(s)}(w^*)F(w^*)$ *for some* s *then* $b(s)$ *is increasing in* X*.*

The result follows from the fact that higher borrowing levels are associated to a higher default rule $\bar{w}$ since they require more revenue from repayment to satisfy zero profit, which in turn leads to a higher pooling threshold by the last part of Proposition 3.

3 Applications

This section highlights implications of our model that are of substantive interest. The first application concerns propagation of shocks by the financial sector. The second application concerns the impact of political economy distortions on a developing economy. In this economy enforcement constraint binds at all times, but there are no shocks. Instead, the economy accumulates enforcement capacity in the process of building its state infrastructure so as to enhance the depth of its financial markets and this way spur entrepreneurship and growth. We study how the presence of the constraint affects deterministic dynamics. In both applications we consider the full equilibrium of our model. That is, the solution involves the ex ante endogenous choice of enforcement capacity X and credit contracts, while factoring in the expected outcome of the ex post enforcement game.

Table 1: Paramaters and Aggregate Statistics

Parameter	Value	Source
y	1	
Ew	1.03	BGG[a], McGrattan and Prescott (2005)[b]
F	Lognormal ($\sigma^2 = 0.3$)	BGG
μ	0.12	BGG
$C(X)$	$0.055X$	
Statistic	**Value**	**Target**
Leverage ratio ($\frac{y+b}{y}$)	1.61	$[1.5, 2]$[c]
Default rate (ψ)	4.1%	3%[d]
Return on Equity (ROE)	4.6%	
Capacity Costs/ROE	20%	

[a]BGG refers to Bernanke et al. (1999).
[b]The Aggregate project return is 1.02 in BGG and 1.04 in McGrattan and Prescott (2005).
[c]Leverage ratio is 1.5 in Christiano et al. (2014) and 2 in BGG.
[d]Business failure rate reported in BGG.

3.1 Propagation of Aggregate Shocks

In our first application the problem of the principal describes the allocation implied by a competitive market in which Bertrand lenders compete to provide funding to a population of entrepreneurs. The economy is in a stationary equilibrium in the sense that R is high enough so that the ex ante non-negativity constraint $g \geq 0$ does not bind. In this case $\alpha = 1$, implying that the principal values ex ante resources on par with ex post utility of entrepreneurs.[12]

The question we ask is how a single shock s that occurs with probability $p = 5\%$ may affect credit provision. To answer this question, we plot how much b declines due to the shock as a function of shock size s, while solving the model for each case separately so that shock size is perfectly anticipated. Other parameters are set in line with the values typically used in the literature on financial frictions. The calibrated value of the parameters, as well as the targets we set, are all listed in Table 1.[13]

The results are presented in Figure 4. As is clear from the figure, the supply of credit b (vertical axis) is lower when the shock hits, triggering a credit crunch when the shock is relatively large—e.g., for a shock of $s = 0.05$ credit is 30% lower. The model implies a significant contraction of credit even in response to small shocks.

The conceptual mechanism behind the observed effect is illustrated in Figure 5. In the absence of the shock the planner brings b up as much as possible without triggering default by types in the cluster. However, when the shock hits, capacity goes down to $X - s$, implying that the initial contract would lead to a default rate $F(\hat{w}(s)) >> F(\bar{w})$. In light of this, the principal chooses to reduce b so as to prevent the cluster from defaulting.

This example illustrates a novel mechanism that leads the abrupt tightening of credit supply

[12] If the principal distributes resources back to entrepreneurs the optimization problem consists on maximizing aggregate utility subject to a zero profit condition in the competitive credit market.

[13] They are set for a benchmark economy that faces no shocks.

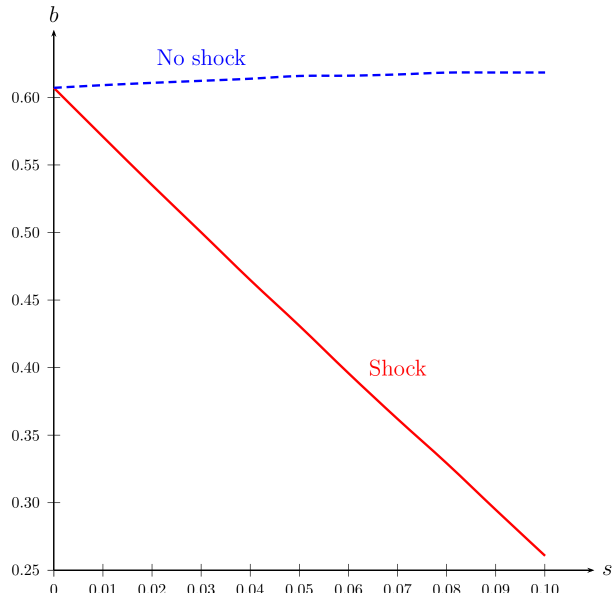

Figure 4: Enforcement Capacity and Shock Propagation.

in response to adverse economic conditions. Our mechanism crucially differs from the usual shock to initial net worth y. From an empirical point of view a lot more work is required to sort out how contraction of credit is be linked to different factors such as net worth of a firm. Interestingly, in a recent working paper Gopinath et al. (2014) look at micro-level regressions and find that the link between net worth and financial frictions alone cannot account for micro-level variations in credit supply.

3.2 Political Economy and Economic Development

Our second application describes an economy-wide planning problem that involves accumulation of state capacity X. In contrast to the first application, R is such that for our benchmark value of $\alpha = 1$ the constraint $g \geq 0$ is binding. That is, the principal would like to borrow against future provision of public goods to relax the enforcement constraint that severely restricts the economy's potential. In contrast to our first application, there are no shocks.

The key question that we ask in this environment is how accounting for the presence of

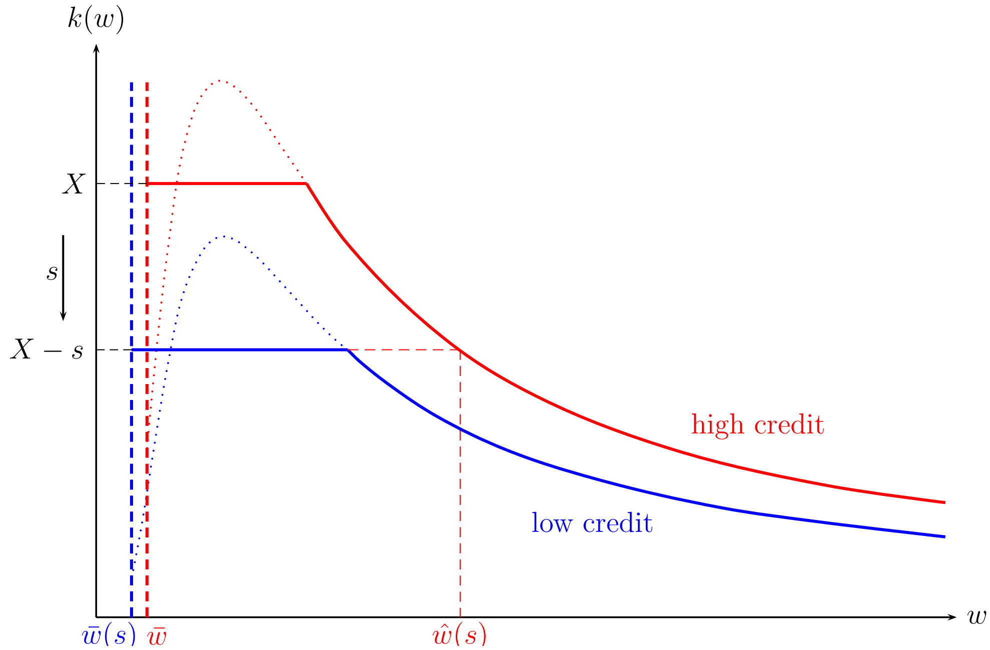

Figure 5: Equilibrium Strategy in the Global Game.

the constraint affects deterministic allocation in the model and the effect of political economy distortion captured by α. For the most part, parameter values are similar to the previous setup.[14]

The results of this exercise are illustrated in Figure 6. We compare two cases as we vary α: the first case is our benchmark economy, which is labeled as *externalities*; while the second case, labeled *no externalities*, is a counterfactual economy in which all constraints are identical except that the agents are unaware of the enforcement capacity constraint (i.e. they do not act strategically and only default when $w < \bar{w}$).

As is clear from the figure, first of all, the supply of credit b (vertical axis) is uniformly lower in the economy with enforcement externality. This should be clear from Proposition 3. Clearly, in order to prevent widespread default in the presence of the externality, X must be at least $\theta_{\bar{w}}(w^*)F(w^*)$. In the absence of the externality, the needed enforcement capacity

[14]We set $R = 0.005$, where the non-negatively constraint $g \geq 0$ is relaxed at about $R = 0.012$, which is the setting in the other example. The probability of the shock is $p = 0$. Other parameter values are identical.

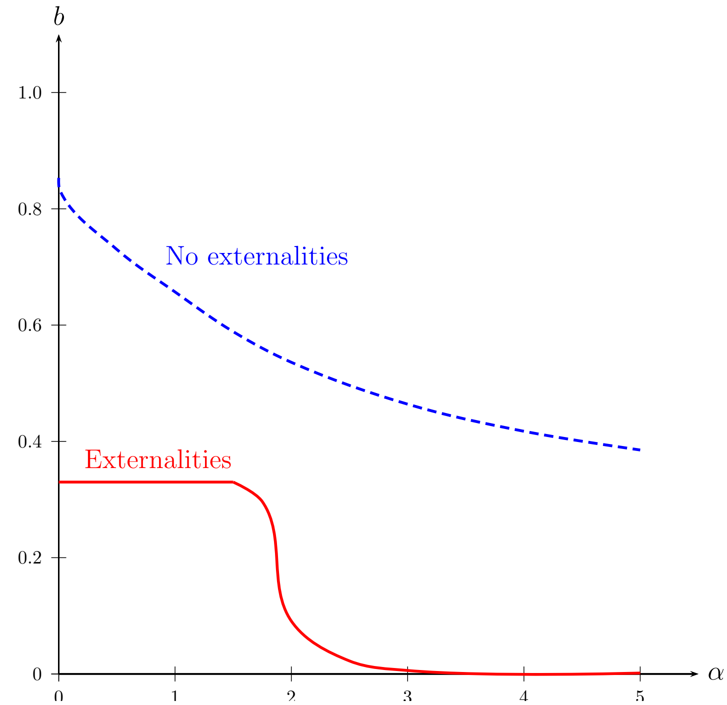

Figure 6: Political Economy Distortions and Provision of Credit. (Red line denote benchmark model; blue line (dashed) line denotes a counterfactual commitment case in which agents are unaware of the constraint.)

is just $F(\bar{w})$, allowing the principal to choose a higher b. Most importantly, however, small increases in α can lead to a regime change, going from growing the economy at the fullest extent possible by investing all of R into building X ($\alpha < 1.75$) to a severe underdevelopment of credit markets ($\alpha \geq 2$). In such a case, very quickly the economy collapses ending up in a financial development trap: at a high levels of α, virtually no projects are funded beyond the initial wealth of the entrepreneurs. In contrast, in the absence of enforcement externality the effect of political economy distortions is very gradual.

4 Conclusions

In this paper we have relaxed one of the central assumptions of the costly state verification framework, namely the fact that the principal can commit to a particular verification strategy regardless of what happens in the economy. We have shown how the short-run rigidity of enforcement resources can lead to strategic complementarities that crucially affect equilibrium outcomes. We have illustrated its economic relevance in the context economic development and financial fragility.

A Appendix

In order to prove the results we proceed as follows. First we present equilibrium existence, selection and characterization results for the model with general discrete distributions of returns. We then provide the proofs of the results in the paper by expressing equilibrium conditions for continuous F as the limit of the discrete-type equilibrium conditions.

A.1 Enforcement Game with Discrete Returns

In this economy $\mathcal{W} \subset [0, \infty)$ is a finite set of possible returns, each with a positive mass, which are distributed according to the commonly known discrete distribution F, with probability mass function f. Since the contract is fixed at the enforcement stage, we drop the subscript from $\theta_{\bar{w}}$ to lighten notation.

Given contract $(\bar{w}, b)$, we make the following assumption about agent payoffs.

Assumption 1. *For all $w \in \mathcal{W}$*

(i) $U(0, w, 0) \neq U(1, w, 0)$; and

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

Condition (i) implies that no agent is indifferent between paying back the loan and defaulting when $P = 0$, i.e., there is no agent with $\theta(w) = 0$. Similarly, (ii) means that that no agent is indifferent at $P = 1$, that is, there is no agent with $\theta(w) = 1$. This technical assumption simplifies the proof of uniqueness since it leads to the existence of dominance regions: for X sufficiently close to zero or one, all agents have strictly dominant strategies.

Note that agents with $w < \bar{w}$ ($\theta(w) > 1$) and those with $\theta(w) < 0$ behave in a non-strategic fashion: the former always choose $a = 0$ and the latter $a = 1$, regardless of P. Hence, our focus is on pinning down the behavior of types in the set $\mathcal{W}^* := \{w \in \mathcal{W} : \theta(w) \in (0, 1)\}$, with its lowest and highest elements respectively denoted w_l and w_h.

We assume that the noise scale factor satisfies $0 < \nu < \bar{\nu} := \min \{\theta(w_h)F(\bar{w}), 1 - \theta(w_l)\}$. [15]

We first establish that there exists a unique equilibrium of the game with finite types, featuring cutoff strategies.

Theorem 1. *The game has an essentially unique equilibrium.*[16] *Equilibrium strategies are characterized by cutoffs $k(w)$ on signal x, such that all agents of type $w \in \mathcal{W}^*$ choose action $a = 1$ if $x \geq k(w)$ and $a = 0$ otherwise.*

Proof. 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, i.e. cutoff strategies. Second, we show that there is at most one equilibrium in monotone strategies (up to differences in behavior at cutoff signals). But this implies that the equilibrium is essentially unique.

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 $\mathbf{x}$ 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, $\underline{\mathbf{a}}(\mathbf{x})$ and $\bar{\mathbf{a}}(\mathbf{x})$ such that any equilibrium profile $\mathbf{a}(\mathbf{x})$ satisfies $\underline{\mathbf{a}}(\mathbf{x}) \leq \mathbf{a}(\mathbf{x}) \leq \bar{\mathbf{a}}(\mathbf{x})$. Moreover, if we fix the action profile of all agents, the difference in expected payoff from choosing $a = 0$ versus $a = 1$ for any given agent is increasing in $\mathbf{x}$, since default rates are the same across signal profiles while X is higher in expectation the higher the signal profile is, thus implying a higher expected monitoring probability. That is, expected payoffs exhibit increasing differences w.r.t. $\mathbf{x}$, and Theorem 6 in Milgrom and Roberts (1990) applies: $\underline{\mathbf{a}}(\mathbf{x})$ and $\bar{\mathbf{a}}(\mathbf{x})$ are non-decreasing functions of $\mathbf{x}$. But since an agent's strategy can only depend on her own signal, this means that $a(x, w)$ is a cutoff strategy for all $w \in \mathcal{W}^*$.

To show that there is at most one equilibrium in monotone strategies we make use of the following two lemmas. The first one shows that equilibrium cutoffs are bounded away from zero and one. The second lemma uses these bounds to to establish the following translation result: when all cutoffs are shifted by the same amount Δ expected monitoring probabilities go up. Equipped with such result we will show how as we move from the smallest to the largest equilibria monitoring probabilities go up, implying that there must be a unique profile of cutoffs at which indifference conditions (16) are satisfied.

Let $\mathbf{k} + \mathbf{\Delta} = (k(w) + \Delta)_{w \in \mathcal{W}^*}$, while $\underline{\mathbf{k}}$ and $\bar{\mathbf{k}}$ represent the profile of cutoffs associated to the smallest and largest equilibrium, respectively. Abusing notation, let $\mathbb{E}[P|\mathbf{k}; x]$ represent the expected monitoring probability of an agent receiving signal x when agents use strategy profile $\mathbf{k}$.

Lemma 4. *If $\mathbf{k}$ is a profile of equilibrium strategies then $k(w) \in [(\theta(w) - \nu/2)F(\bar{w}), \theta(w) + \nu/2]$ for all $w \in \mathcal{W}^*$.*

[15] The upper bound on ν 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.

[16] In the sense that equilibrium strategies may differ in zero probability events.

Proof. Note that k is an equilibrium if it solves (16). 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 (10) we have that, if $k(w) \in [\nu/2, 1-\nu/2]$,

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

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

$$\mathbb{E}[P|\mathbf{k}; k(w)] \leq \mathbb{E}\left[\frac{X}{F(\bar{w})}\Big|\mathbf{k}; k(w)\right] \leq \frac{k(w) + \nu/2}{F(\bar{w})},$$

which, using a symmetric argument, yields the above lower bound on $k(w)$. □

Lemma 5. *If* $\mathbf{k}$ *is an equilibrium profile then* $\mathbb{E}[P|\mathbf{k}; k(w)] < \mathbb{E}[P|\mathbf{k} + \boldsymbol{\Delta}; k(w) + \Delta]$ *for all* $\Delta > 0$ *and all* $w \in \mathcal{W}^*$ *such that* $k(w) + \Delta \leq \bar{k}(w)$.

Proof. First note that the density of X conditional on an agent receiving signal $x \in [\nu/2, 1-\nu/2]$ is given by $h\left(\frac{x-X}{\nu}\right)$. Also notice that an agent of type w does not comply if she receives a signal $x < k(w)$ and thus, the fraction of type-w agents not complying when capacity is X is given by $H\left(\frac{k(w')-X}{\nu}\right)$. Since $\nu \leq \theta(w_h)F(\bar{w})$ and, by Lemma 4, $k(w) \geq (\theta - \nu/2)F(\bar{w})$ we have that $k(w) \geq \nu/2$. Likewise, $k(w) + \Delta \leq \bar{k}(w) \leq 1 - \nu/2$ by Lemma 4 and the fact that $\nu \leq 1 - \bar{\theta}$. Hence, we can obtain the following inequality by a well-defined change of variables:

$$\begin{aligned}
&\mathbb{E}[P(X)|\mathbf{k}; k(w)] = \\
&\qquad \int_{-1/2}^{1/2} \min\left\{\frac{X}{F(\bar{w}) + \sum_{w'} H\left(\frac{k(w')-X}{\nu}\right) f(w')}, 1\right\} h\left(\frac{k(w) - X}{\nu}\right) dX \\
&\qquad < \int_{-1/2}^{1/2} \min\left\{\frac{X + \Delta}{F(\bar{w}) + \sum_{w'} H\left(\frac{k(w')-X}{\nu}\right) f(w')}, 1\right\} h\left(\frac{k(w) - X}{\nu}\right) dX \\
&\qquad = \int_{-1/2}^{1/2} \min\left\{\frac{X'}{F(\bar{w}) + \sum_{w'} H\left(\frac{k(w')+\Delta-X'}{\nu}\right) f(w')}, 1\right\} h\left(\frac{k(w) + \Delta - X'}{\nu}\right) dX' \\
&\qquad = \mathbb{E}[P|\mathbf{k} + \boldsymbol{\Delta}; k(w) + \Delta].
\end{aligned}$$

The inequality is strict since $\mathbf{k}$ being an equilibrium profile means that $\mathbb{E}[P|\mathbf{k}; k(w)] = \theta(w) < 1$ for all $w \in \mathcal{W}^*$. Accordingly, monitoring probabilities, conditional on $x = k(w)$, 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 Δ. □

Equipped with Lemma 5 we know argue that $\underline{\mathbf{k}} = \bar{\mathbf{k}}$. Assume, by way of contradiction, that $\underline{k}(w) < \bar{k}(w)$ for some $w \in \mathcal{W}^*$. Denote $\hat{w} = \arg\max_{w \in \mathcal{W}^*}(\bar{k}(w) - \underline{k}(w))$ and $\hat{\Delta} = \bar{k}(\hat{w}) - \underline{k}(\hat{w})$. By Lemma 5, we have that

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

where the last inequality comes from the fact that default rates at $\bar{\mathbf{k}}$ are lower than at $\underline{\mathbf{k}} + \hat{\mathbf{\Delta}} \geq \bar{\mathbf{k}}$, and thus so are monitoring probabilities conditional on $x = \bar{k}(\hat{w})$. □

Next, we proceed to characterize the limit equilibrium as $\nu \to 0$. First, note that for threshold profile $k(\cdot)$ to be an equilibrium profile it must satisfy the following set of indifference conditions:

$$\mathbb{E}_\theta[P|x = k(w)] = \theta(w) \qquad \forall w \in \mathcal{W}^*. \tag{16}$$

In order to solve this system of equation we need to pin down agents beliefs when they receive their threshold signals. In order to do so we make use of the full force of the belief constraint (Sakovics and Steiner, 2012): on average, conditional on $x = k(w)$, agents with types in a subset $W' \subseteq \mathcal{W}^*$ believe that the default rate of agents in W' is uniformly distributed in $[0,1]$. The default rate in W' when capacity is X is given by

$$\psi(X, W') = \frac{1}{\sum_{W'} f(w)} \sum_{W'} H\left(\frac{k(w) - X}{\nu}\right) f(w). \tag{17}$$

Lemma 6 (Belief Constraint)**.** *For any subset $W' \subseteq \mathcal{W}^*$ and any $z \in [0,1]$,*

$$\frac{1}{\sum_{W'} f(w)} \sum_{W'} \mathbb{P}\left(\psi(X, W') \leq z \middle| x = k(w)\right) f(w) = z. \tag{18}$$

Proof. The result follows directly from the proof of Lemma 1 in Sakovics and Steiner (2012). To see why note first that Lemma 4 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 $w \in \mathcal{W}'$ and focus on the aggregate action of agents with types in W', rather than the aggregate action in the population.[17] □

The above result is instrumental to characterize equilibrium thresholds as ν goes to zero. In particular, it allows to derive closed-form solutions for the above indifference conditions from which we can obtain $\mathbf{k}$. We state this remarkable result in its full generality below. In stating this result we refer to a partition $\Phi = \{W_1, \cdots, W_I\}$ of $\mathcal{W}^*$ as being monotone if $\max W_i < \min W_{i+1}$, $i = 1, \cdots, I-1$, and denote the lowest and highest elements of W_i by $\underline{w}_i$ and $\bar{w}_i$, respectively.

[17] When thresholds do not fall within ν of each other, the distribution $\tilde{F}$ of virtual errors $\tilde{\eta}$ need not be strictly increasing and thus its inverse may not be well-defined. Defining $\tilde{F}^{-1}(u) = \inf\{\tilde{\eta} : \tilde{F}(\tilde{\eta}) \geq u\}$ takes care of this issue and ensures that the proof of Lemma 2 in Sakovics and Steiner (2012) applies to the general case.

Theorem 2. *In the limit, as $\nu \to 0$, the equilibrium cutoff strategies are given by a unique monotone partition $\Phi = \{W_1, \cdots, W_I\}$ and a unique vector $(k_1, \cdots, k_I)$ satisfying the following conditions:*

(i) $k(w) = k(w') = k_i$ *for all* $w, w' \in W_i$.

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

(iii) $\theta(\underline{w}_i)F^-(\underline{w}_i) \leq k_i \leq \theta(\bar{w}_i)F(\bar{w}_i)$ *for all* $i = 1, \cdots, I$.

(iv) $\int_{F^-(\underline{w}_i)}^{F(\bar{w}_i)} \min\left\{\frac{k_i}{z}, 1\right\} dz = \sum_{W_i} \theta(w)f(w)$ *for all* $i = 1, \cdots, I$.

Proof. From Theorem 1 we know that for each $\nu > 0$ there exists essentially a unique equilibrium, which is in monotone strategies. Let $k^\nu(w)$ represent the equilibrium threshold of type-w agents associated to $\nu > 0$, with $\mathbf{k}^\nu$ denoting the equilibrium cutoff profile. The first step of the proof is to show that $\mathbf{k}^\nu$ uniformly converge as $\nu \to 0$ and identify the set of indifference conditions that pin down the limit equilibrium. Let

$$A_w(z|\mathbf{k}^\nu, W') := \mathbb{P}\left(\psi(X, W') \leq z \middle| x = k^\nu(w)\right)$$

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

Lemma 7. *There exists a unique partition $\{W_1, \cdots, W_I\}$ and a set of thresholds $k_1 > k_2 > \cdots > k_I$ such that, as $\nu \to 0$, for all $w \in W_i$, $i = 1, \cdots, I$, $k^\nu(w)$ uniformly converges to k_i. Moreover, thresholds $\mathbf{k} = (k_1, \cdots, k_I)$ are the solution to the set of limit indifference conditions*

$$\int_0^1 \min\left\{\frac{k_i}{F(\bar{w}) + \sum\limits_{\cup_{j<i}W_j} f(w') + z\sum\limits_{W_i} f(w')}, 1\right\} dA_w(z|\mathbf{k}, W_i) = \theta(w), \quad \forall w \in W_i,\ \forall i, \tag{19}$$

where $A_w(z|\mathbf{k}, W_i)$ represents the strategic beliefs of type-w agents in the limit and satisfies the belief constraint (18).

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 $w > \hat{w}$ such that $w \in \mathcal{W}_i$ and $\hat{w} \in \mathcal{W}_m$ with $m > i$. First note that the LHS in (19) is bounded below by $\frac{k_i}{F(\bar{w}) + \sum_{\cup_{j\leq i}W_j} f(w')}$ and bounded above by $\min\left\{\frac{k_i}{F(\bar{w}) + \sum_{\cup_{j<i}W_j} f(w')}, 1\right\}$. Given this, since $\theta(\hat{w}) < 1$ the monitoring probability when all agents with types in W_m default is strictly less than one, i.e., $\frac{k_m}{F(\bar{w}) + \sum_{\cup_{j\leq m}W_j} f(w')} < 1$. Otherwise (19) would be violated. In addition,

$m > i$ implies that $k_m < k_i$ by the above lemma and that $\sum_{\cup_{j<i} W_j} f(w') < \sum_{\cup_{j\le m} W_j} f(w')$. Combining all this we arrive to the following contradiction

$$\theta(w) \geq \min\left\{\frac{k_i}{F(\bar{w}) + \sum\limits_{\cup_{j<i} W_j} f(w')}, 1\right\} > \frac{k_m}{F(\bar{w}) + \sum\limits_{\cup_{j\le m} W_j} f(w')} \geq \theta(\hat{w}).$$

The monotonicity of the type partition implies that $F(\bar{w}) + \sum_{\cup_{j\le i} W_j} f(w') = F(\overline{w}_i)$ and that $F(\bar{w}) + \sum_{\cup_{j<i} W_j} f(w') = F^{-}(\underline{w}_i)$. Given this, it is straightforward to check that the above bounds on the LHS of (19) lead to condition (iii) in the theorem.

Finally, in order to obtain condition (iv) from (19) we make use of the belief constraint in the limit, which can be written as

$$\frac{1}{\sum\limits_{W_i} f(w)} \sum_{W_i} A_w(z|\mathbf{k}, W_i) f(w) = z. \tag{20}$$

Multiplying both sides of (19) by $\frac{f(w)}{\sum_{W_i} f(w)}$ and summing over $w \in W_i$ we get

$$\int\limits_0^1 \min\left\{\frac{k_i}{F^{-}(\underline{w}_i) + z\sum\limits_{W_i} f(w')}, 1\right\} d\left(\frac{1}{\sum\limits_{W_i} f(w)} \sum_{W_i} A_w(z|\mathbf{k}, W_i) f(w)\right) = \frac{\sum_{W_i} \theta(w) f(w)}{\sum_{W_i} f(w)}.$$

Finally, using the belief constraint (20) to substitute for the last term in the LHS we obtain

$$\int_0^1 \min\left\{\frac{k_i}{F^{-}(\underline{w}_i) + z\sum\limits_{W_i} f(w')}, 1\right\} dz = \frac{\sum_{W_i} \theta(w) f(w)}{\sum_{W_i} f(w)}. \tag{21}$$

Note that $F^{-}(\underline{w}_i) + z\sum_{W_i} f(w') \sim U[F^{-}(\underline{w}_i), F(\bar{w}_i)]$ with density $\frac{1}{\sum_{W_i} f(w)}$ since $z \sim U[0,1]$. Hence, we can rewrite (21) as

$$\frac{1}{\sum\limits_{W_i} f(w)} \int_{F^{-}(\underline{w}_i)}^{F(\bar{w}_i)} \min\left\{\frac{k_i}{z}, 1\right\} dz = \frac{\sum_{W_i} \theta(w) f(w)}{\sum_{W_i} f(w)},$$

yielding condition (iv). □

Proof of Lemma 7. To prove convergence we first partition the set of types into subsets W_i of types for sufficiently small ν as follows: (i) if we order the signal thresholds of all types, any adjacent thresholds that are within ν of each other belong to the same subset; and (ii) $j > i$ implies that the thresholds associated to types in W_j are lower than those associated to W_i—by at least ν. Also, let $Q^\nu_w(\chi|\mathbf{k}^\nu, z) := \mathbb{P}\left(X \leq \chi \middle| x = k^\nu(w), \psi(X, W_i) = z\right)$ represent the beliefs about capacity of an agent of type $w \in W_i$ conditional on receiving her threshold signal $k^\nu(w)$ and on the event that the default rate in W_i is equal to z.

Note that a type-w agent receiving signal $x = k^\nu(w)$ knows that all agents with types in W_j are defaulting if with $j < i$, and repaying if $j > i$. Also, the support of $Q^\nu_w(\cdot|\mathbf{k}^\nu, z)$ must lie within $[k^\nu(w) - \nu/2, k^\nu(w) + \nu/2]$. Given this, by the law of iterated expectations, her expected monitoring probability conditional on $x = k^\nu(w)$ can be written in terms of her strategic belief as follows:

$$\mathbb{E}(P|\mathbf{k}^\nu; k^\nu(w)) =$$
$$\int_0^1 \int_{k^\nu(w)-\nu/2}^{k^\nu(w)+\nu/2} \min\left\{\frac{\chi}{F(\bar{w}) + \sum\limits_{\cup_{j<i}W_j} f(w') + z\sum\limits_{W_i} f(w')}, 1\right\} dQ^\nu_w(\chi|\mathbf{k}^\nu, z) dA_w(z|\mathbf{k}^\nu, W_i). \tag{22}$$

In addition, notice that we can always express $\mathbb{E}(P|\mathbf{k}^\nu; k^\nu(w))$ in terms of the threshold signal $k^\nu(w)$ and relative threshold differences $\Delta_{w'} = (k^\nu(w') - k^\nu(w))/\nu$. Importantly, as Sakovics and Steiner (2012) emphasize, strategic beliefs depend on the relative distance between thresholds $\Delta_{W_i} = \{\Delta_{w'}\}_{w' \in W_i}$. , rather than on their absolute distance. That is, keeping Δ_{W_i} fixed, $A_w(z|\mathbf{k}^\nu, W_i)$ does not change with ν. [18] This directly implies that strategic beliefs satisfy the belief constraint when $\nu = 0$.

Fix $k^\nu(w) = k_i$ for some $w \in W_i$ and also fix Δ_{W_i}, for all $i = 1, \cdots, I$ and all ν sufficiently small. By fixing relative differences, the partition $\{W_i\}_{i=1}^I$ still satisfies the above definition, and thus does not change as $\nu \to 0$. We are going to show that indifference condition $\mathbb{E}(P|\mathbf{k}^\nu; k^\nu(w)) = \theta(w)$ is approximated by the limit condition in the lemma for ν sufficiently small.

Note that the inner integral in (22) is bounded below by $\min\left\{\frac{k^\nu(w)-\nu/2}{F(\bar{w})+\sum\limits_{\cup_{j<i}W_j} f(w')+z\sum\limits_{W_i} f(w')}, 1\right\}$ and above by $\min\left\{\frac{k^\nu(w)+\nu/2}{F(\bar{w})+\sum\limits_{\cup_{j<i}W_j} f(w')+z\sum\limits_{W_i} f(w')}, 1\right\}$. Hence,

$$\int_0^1 \min\left\{\frac{k_i - \nu/2}{F(\bar{w}) + \sum\limits_{\cup_{j<i}W_j} f(w') + z\sum\limits_{W_i} f(w')}, 1\right\} dA_w(z|\mathbf{k}^\nu, W_i) \leq \mathbb{E}(P|\mathbf{k}^\nu; k^\nu(w))$$
$$\leq \int_0^1 \min\left\{\frac{k_i + \nu/2}{F(\bar{w}) + \sum\limits_{\cup_{j<i}W_j} f(w') + z\sum\limits_{W_i} f(w')}, 1\right\} dA_w(z|\mathbf{k}^\nu, W_i). \tag{23}$$

The first term in these integrals is Lipschitz continuous. In addition, the next lemma shows

[18] This is straightforward to check. First, if we substitute $X = k^\nu(w) - \nu\eta$ (since agents with type w get her threshold signal) and $k(w') = \nu\Delta_{w'} + k^\nu(w)$ into (10), we find that $\psi(X, W_i)$ only depends on Δ_{W_i} and $k^\nu(w)$. But this means that $A_w(z|\mathbf{k}^\nu, W_i)$ only depends on Δ_{W_i} and $k^\nu(w)$ since h is independent of ν.

that $dA_w(z|\mathbf{k}^\nu, k^\nu(w))$ is bounded for all ν.

Lemma 8. $0 \le \dfrac{\partial A_w(z|\mathbf{k}^\nu, k^\nu(w))}{\partial z} \le \dfrac{\sum_{W_i} f(w')}{f(w)}$ *for all $w \in W_i$ and all z in the support of $A_w(\cdot|\mathbf{k}^\nu, k^\nu(w))$.*

See proof below.

Hence, the LHS and the RHS of (23) uniformly converge to each other as $\nu \to 0$, leading to limit indifference conditions (19). Note also that $k^\nu(w) \in [-\bar{\nu}/2, 1+\bar{\nu}/2]$ and, keeping $\{W_i\}_{i=1}^I$ fixed, $\Delta_{w'} \in [-1,1]$ for all $w' \in W_i$. That is, the solution to the system of indifference conditions $\mathbb{E}(P|\mathbf{k}^\nu; k^\nu(w)) = \theta(w)$ lies in a compact set.[19] Accordingly, we can find $\hat{\nu}$ so that indifference conditions are within ε of the limit condition for all $\nu < \hat{\nu}$, leading to their solutions being in a neighborhood of the solution $\mathbf{k}$ of limit indifference conditions (19). □

Proof of Lemma 8. Let $\psi^{-1}(z, W_i)$ be the inverse function of $\psi(X, W_i)$ w.r.t. X. The latter function is decreasing in X as long as $0 < \psi(X, W_i) < 1$, implying that ψ^{-1} is well defined and decreasing in such a range of capacities. Since the signal of an agent of type w satisfies $x = X + \nu\eta$ we can express her strategic belief as

$$A_w(z|\mathbf{k}^\nu, W_i) = \mathbb{P}\left(\psi^{-1}(z, W_i) \le k^\nu(w) - \nu\eta\right) = H\left(\frac{k^\nu(w) - \psi^{-1}(z, \mathbf{k}^\nu, W_i)}{\nu}\right).$$

Differentiating w.r.t. z yields

$$\begin{aligned}\frac{\partial A_w(z|\mathbf{k}^\nu, W_i)}{\partial z} &= \frac{1}{\nu} h\left(\frac{k^\nu(w) - \psi^{-1}(z, W_i)}{\nu}\right)\left(-\frac{\partial \psi^{-1}(z, W_i)}{\partial z}\right) \\ &= \frac{h\left(\frac{k^\nu(w) - \psi^{-1}(z, W_i)}{\nu}\right)}{\frac{1}{\sum_{W_i} f(w')} \sum_{W_i} H\left(\frac{k^\nu(w') - \psi^{-1}(z, W_i)}{\nu}\right) f(w')}.\end{aligned}$$

For all $z \in (0,1)$ we must have $h\left(\frac{k^\nu(w) - \psi^{-1}(z, W_i)}{\nu}\right) > 0$ since h is bounded away from zero in its support. Hence, the last term is positive and weakly lower than $\dfrac{\sum_{W_i} f(w')}{f(w)}$. □

A.2 Proofs

Proof of Lemma 1. The propensity to default θ is found by equating the expected payoff from repaying to that of defaulting. From (6), the payoff from paying back the loan is $u(w, 1, 0) =$

[19] If $\{W_i\}_{i=1}^I$ is not kept fixed then when ν is very small $\mathbb{E}(P|\mathbf{k}^\nu; k^\nu(w))$ would be discontinuous at some ν, implying a violation of the indifference condition for some $w \in \mathcal{W}^*$.

$(y+b)(w-\bar{w})$, while the expected payoff from defaulting when verification probability is P is given by $(1-P)\gamma\mu(y+b)w$. Thus, θ solves

$$(y+b)(w-\bar{w}) = (1-\theta)\gamma\mu(y+b)w,$$

which leads to $\theta = 1 - \frac{1}{\mu\gamma}\left(1 - \frac{\bar{w}}{w}\right)$. □

Proof of Lemma 2. The derivative of $\theta(w)F(w)$ is given by

$$\left(1 - \frac{1}{\mu\gamma}\left(1 - \frac{\bar{w}}{w}\right)\right) f(w) - \frac{\bar{w}}{w^2}\frac{1}{\mu\gamma}F(w),$$

which has the same sign as

$$1 - \frac{w}{\bar{w}}(1-\mu\gamma) - \frac{F(w)}{wf(w)}.$$

If $\dfrac{F(w)}{wf(w)}$ is increasing the expression is strictly decreasing. Now, since the second term is zero at $w=0$ and $\dfrac{F(w)}{wf(w)}$ is increasing and $\lim_{w\downarrow 0}\dfrac{F(w)}{wf(w)} < 1$ the expression—and hence the slope of $\theta(w)F(w)$—is initially positive and eventually negative for high enough w. That is, $\theta(w)F(w)$ is single-peaked. □

Proof of Proposition 1. First consider the case of $X < \underline{X}$. Since $\bar{w} < w_{max}$ it must be that (14) has only solutions, $w_1 < \bar{w}$ and $w_2 > w_max$. The former cannot be an equilibrium since it would require $\psi = F(w_1) < F(\bar{w})$, i.e., that some agents that strictly prefer to default choose to repay, and hence equilibrium is unique. The same argument applies when $\bar{w} \geq w_{max}$. When $X = \underline{X}$ then $w_1 = \bar{w}$ is also an equilibrium.

Second, let $X > \bar{X}$. In this case, $\theta_{\bar{w}}(w)F(w)$ lies below X, implying that for any given P such that all agents with default propensity less than P default there is enough capacity so that the monitoring probability is higher than P. Thus equilibrium is unique and involves $\psi = F(\bar{w})$. If $X = \bar{X}$ then w_{max} is a solution of (14), representing a second equilibrium.

Finally, if $X \in (\underline{X}, \bar{X})$ there are three equilibria given by the two solutions in $[\bar{w}, \infty)$ to (14) and another equilibrium with $\psi = F(\bar{w})$ since $F(\bar{w}) < X$ and thus the principal can credibly sustain $P = 1$ in equilibrium. □

Proof of Proposition 2. As stated in text, uniqueness should be interpreted as the existence of a unique equilibrium in nearby discrete-return economies (Theorem 1). □

Proof of Lemma 3. The statement of the lemma is just the continuous-returns version of Lemma 6. □

Proof of Proposition 3. In order to obtain the characterization of equilibrium threshold in our model we proceed as follows. First, we express Theorem 2 in terms of continuous type distributions. Second, we argue that the single peakedness of $\theta(w)F(w)$ implies the existence of a unique interval of types $(\bar{w}, w^*)$ such that $k(w) = \theta(w^*)F(w^*)$ for types in the interval and $k(w) = \theta(w)F(w)$ for $w > w^*$. Finally, we use the conditions in the theorem to pin down w^*. The last part of the proof simply shows that $\theta(w^*)F(w^*)$ is increasing in $\bar{w}$.

The version of Theorem 2 for continuous F implies the existence of a unique partition of types with propensity to default between 0 and 1 into intervals $\{(\underline{w}_j, \bar{w}_j]\}_{j=1}^{J}$ such that:

(a) if $k(w)$ is strictly decreasing in an interval i then it is constant in intervals $j-1$ and $j+1$ and vice versa;

(b) if $k(w)$ is strictly decreasing in interval j then $k(w) = \theta(w)F(w)$ for all $w \in (\underline{w}_j, \bar{w}_j]$;

(c) if $\theta(w)F(w)$ is not strictly decreasing in $(\underline{w}_j, \bar{w}_j]$ then $k(w) = k_j$ for all $w \in (\underline{w}_j, \bar{w}_j]$ with k_j satisfying $k_j = \theta(\bar{w}_j)F(\bar{w}_j) \geq \theta(\underline{w}_j)F(\underline{w}_j)$ (with equality if $\underline{w}_j > \bar{w}$) and

$$\int_{F(\underline{w}_1)}^{F(\bar{w}_1)} \min\left\{\frac{k_i}{z}, 1\right\} dz = \int_{\underline{w}_j}^{\bar{w}_j} \theta(w)f(w)dw. \tag{24}$$

Part (a) follows from conditions (i)-(ii) in Theorem 2, which mean that k is decreasing, so we can partition the space of types into a collection of successive intervals in which k alternates between being strictly decreasing and constant. Part (b) follows from (ii)-(iii): a strictly decreasing k in a given interval of types is approximated by a (growing) collection of consecutive, singleton W_i in the discrete economy. But then, as the mass associated to each of these singletons goes to zero, F^- approximates F and condition (iii) implies that k converges to $\theta(w)F(w)$.

Part (c) follows from parts (a)-(b) and conditions (iii)-(iv). Since $\theta(w)F(w)$ is continuous parts (a) and (b) imply that $k(\bar{w}) = \theta(w)F(w)$ at the boundaries of an interval in which k is constant, except possibly when $\underline{w}_i = \bar{w}$, in which case condition (iii) requires that $k_j \geq \theta(\underline{w}_j)F(\underline{w}_j)$. Expression 24 is the continuous counterpart of (iv).

We now argue that the single-peakedness of $\theta(w)F(w)$ leads to a partition consisting of two intervals, the fist one where k is constant and the second one in which it is strictly decreasing.

First notice that k being decreasing implies that there must be at least one pooling threshold since $\theta(w)F(w)$ is initially increasing. To show why there is only one we use the fact that condition (c) requires that $k_1 = \theta(\bar{w}_1)F(\bar{w}_1) \geq \theta(\underline{w}_1)F(\underline{w}_1)$. Given the single-peakedness of $\theta(w)F(w)$ and $k(w)$ being decreasing, we must have that $\theta(w)F(w)$ is increasing at $\underline{w}_1$ and decreasing at $\bar{w}_1$. Otherwise, either $\theta(w)F(w)$ is decreasing at $\underline{w}_1$ or $\theta(w)F(w)$ is increasing in $(\underline{w}_1, \bar{w}_1)$. The former case implies that $\theta(\bar{w}_1)F(\bar{w}_1) < \theta(\underline{w}_1)F(\underline{w}_1)$, violating (c). The latter case implies that $k_1 = \max_{w \in [\underline{w}_1, \bar{w}_1]} \theta(w)F(w)$, which implies that the LHS of (24) is greater than the RHS.[20]

[20]Since $k_1 = \max_{w \in [\underline{w}_1, \bar{w}_1]} \theta(w)F(w)$ we have that

$$\int_{F(\underline{w}_1)}^{F(\bar{w}_1)} \min\left\{\frac{k_i}{z}, 1\right\} dz > \int_{F(\underline{w}_1)}^{F(\bar{w}_1)} \min\left\{\frac{\theta(F^{-1}(z))z}{z}, 1\right\} dz = \int_{\underline{w}_j}^{\bar{w}_j} \theta(w)f(w)dw,$$

Accordingly, by single-peakedness, if $\theta(w)F(w)$ is increasing at $\underline{w}_1$ and decreasing at $\bar{w}_1$ we cannot find another interval satisfying the same condition that does not intersect with $[\underline{w}_i, \overline{w}_i]$. Thus, there must be a unique interval of returns at which k is constant. Finally, since $\theta(w)F(w)$ is increasing in $[\bar{w}, \underline{w}_1]$ the monotonicity of $k(w)$ requires that $\underline{w}_1 = \bar{w}$.

We finish the characterization of equilibrium thresholds by showing that $\bar{w}_1 = w^*$, where w^* is the unique solution to (14) in $(\bar{w}, \infty)$ when $\bar{w} < w_{max}$.

Condition (c) implies that $k_1 \geq F(\bar{w})$. Hence, solving the integral and substituting for $k_1 = \theta(w^*)F(w^*)$ and $\underline{w}_1 = \bar{w}$ we can express the LHS of (24) as

$$\int_{k_i}^{F(w^*)} \frac{k_i}{z} dz + \int_{F(\bar{w})}^{k_i} dz = k_i \log\left(\frac{F(w^*)}{k_i}\right) + k_i - F(\bar{w}) = \theta(w^*)F(w^*)(1 - \log\theta(w^*)) - F(\bar{w}).$$

Equating the RHS of the last expression to the RHS of (24) yields (14). To show that it has a unique solution in $(\bar{w}, \infty)$ we express it as

$$\theta(w^*)F(w^*)(1 - \log\theta(w^*)) - \int_{\bar{w}}^{w^*} \theta(w)f(w)dw = F(\bar{w}), \tag{25}$$

and differentiate the LHS w.r.t. w^*, which yields $(-\log\theta(w^*))\left(\theta'(w^*)F(w^*) + \theta(w^*)f(w^*)\right)$. The first term in this expression is positive while the second term is the slope of $\theta(w^*)F(w^*)$, which is first positive then negative in $[\bar{w}, \infty)$ when $\bar{w} < w_{max}$. That is, the LHS is first increasing and then decreasing in $\bar{w}, \infty)$. Hence, since the RHS is constant, the above expression has at most two solutions in $[\bar{w}, \infty)$. But notice that $\bar{w}$ is always a solution and that the LHS is increasing at $\bar{w}$ if $\bar{w} < w_{max}$. This, combined with the fact that the LHS approaches zero as w grows while the RHS is strictly positive implies that there exists a unique solution in $(\bar{w}, \infty)$.

Obviously, if $\bar{w} \geq w_{max}$ then $\theta(w)F(w)$ is strictly decreasing for $w \geq \bar{w}$ and conditions (a)-(c) lead to $k(w) = \theta(w)F(w)$, i.e., to $w^* = \bar{w}$.

Finally, we need to show that $\theta(w^*)F(w^*)$ is increasing in $\bar{w}$. Note that the propensity to default θ goes up with $\bar{w}$ and that $\theta(w)F(w)$ is decreasing at w^*. Given this, if we can show that w^* goes down after an increase in $\bar{w}$ then we would have proven that $\theta(w^*)F(w^*)$ increases with $\bar{w}$. We do so by implicitly differentiating (25):

$$\frac{\partial LHS}{\partial w^*}\frac{dw^*}{d\bar{w}} = \int_{\bar{w}}^{w^*} \frac{\partial\theta(w)}{\partial\bar{w}} f(w)dw.$$

From the above argument we know that $\frac{\partial LHS}{\partial w^*} < 0$, while the RHS of the last expression is positive since $\frac{\partial\theta(w)}{\partial\bar{w}} > 0$. Hence, it must be that $\frac{dw^*}{d\bar{w}} < 0$. □

Proof of Proposition 4. We first prove the last part and then the finiteness of $b(s)$.

To prove that $b(s)$ is increasing in $X(s)$ when the constraint $X \geq \theta_{\bar{w}(s)}(w^*)F(w^*)$ we need to show that $b(s)$ is increasing in $\bar{w}(s)$, since $\theta_{\bar{w}(s)}(w^*)F(w^*)$ is increasing in $\bar{w}(s)$ by Proposition 3. To do so, we first note that the propensity to default $\theta_{\bar{w}(s)}(\cdot)$ and hence $k(\cdot)$ do not depend on borrowing level $b(s)$ by Proposition 1. Accordingly, for any given $\bar{w}(s)$ the objective function

where the last equality comes from the change in variable $w = F^{-1}(z)$ $(dz = f(w)dw)$.

(8) is strictly increasing in $b(s)$ since $P(s)$, $\{w : a(s,w) = 0\}$ and $\{w : a(s,w) = 1\}$ are constant for all b whereas payoffs under repayment and under default are strictly increasing in $b(s)$.

In addition, note that we can express the budget constraint (9) as follows:

$$\frac{b(s)}{y + b(s)} \leq \int_{\{w:a(s,w)=1\}} \bar{w}(s)dF + \mu \left(P(s) + (1-\gamma)(1 - P(s))\right) \int_{\{w:a(s,w)=0\}} wdF. \tag{26}$$

Since the LHS is strictly increasing in $b(s)$ while the RHS is constant in $b(s)$, the principal will set $b(s)$ as high as possible, i.e., (26) must hold with equality.

Given this, to show that $b(s)$ is increasing in $\bar{w}$ under the constraint $X \geq \theta_{\bar{w}(s)}(w^*)F(w^*)$, it suffices to prove that the RHS of (26) is continuous and initially increasing in $\bar{w}(s)$. The reason is that $X \geq \theta_{\bar{w}(s)}(w^*)F(w^*)$ implies $P(s) = 1$ since only those with $w \leq \bar{w}(s)$ default and $\psi = F(\bar{w}(s)) \leq \theta_{\bar{w}(s)}(w^*)F(w^*)$ by condition (c) in the proof of Proposition 3. But when $P(s) = 1$ the expected payoff of entrepreneurs reduces to $\int_{\bar{w}(s)}^{\infty}(y + b(s))(w - \bar{w}(s))dF$, which is strictly decreasing in $\bar{w}(s)$. Accordingly, if the RHS of (26) is continuous and initially increasing, for any borrowing level satisfying (26) with equality for some $\bar{w}(s)$ at which the RHS is decreasing we can always find a lower $\bar{w}(s)$ satisfying (26) such that the RHS is decreasing. But then the principal will never choose a default rule $\bar{w}(s)$ at which the RHS is decreasing, since she can sustain the same borrowing level at a lower $\bar{w}(s)$, thereby raising entrepreneur payoffs.

To prove that the RHS is continuous and initially increasing notice that, when $P(s) = 1$ and $\{w : a(s,w) = 1\} = \{w : w \leq \bar{w}(s)\}$, it is given by

$$w(s)(1 - F(\bar{w}(s))) + \mu \int_0^{\bar{w}(s)} wf(w)dw. \tag{27}$$

Continuity follows from the continuity of F. Differentiating w.r.t. $w(s)$ we obtain

$$1 - F(\bar{w}(s)) - (1-\mu)\bar{w}(s)f(\bar{w}(s)),$$

which is strictly positive at $\bar{w}(s) = 0$.

We finish the proof by arguing that the optimal $b(s)$ is finite whenever $\mu\mathbb{E}w < 1$. To see why note that the RHS of (26) is bounded above by (27), which converges to $\mu\mathbb{E}w$ as $\bar{w}(s) \to \infty$. On the other hand $\frac{b}{y+b}$ goes to one as $b \to \infty$. Hence, $b(s)$ will be bounded as long as $\mu\mathbb{E}w < 1$. □

References

Arellano, Cristina and Narayana R Kocherlakota, "Internal debt crises and sovereign defaults," Technical Report, National Bureau of Economic Research 2008.

Bernanke, Ben S, Mark Gertler, and Simon Gilchrist, "The financial accelerator in a quantitative business cycle framework," *Handbook of macroeconomics*, 1999, *1*, 1341–1393.

Besley, Timothy and Torsten Persson, "The Origins of State Capacity: Property Rights, Taxation, and Politics," *American Economic Review*, September 2009, *99* (4), 1218–44.

_ **and** _ , "State Capacity, Conflict, and Development," *Econometrica*, 2010, *78* (1), 1–34.

_ **and** _ , *Pillars of Prosperity: The Political Economics of Development Clusters*, Princeton University Press, 2011.

Bond, Philip and Ashok S. Rai, "Borrower runs," *Journal of Development Economics*, 2009, *88* (2), 185 – 191.

Carlsson, Hans and Eric van Damme, "Global Games and Equilibrium Selection," *Econometrica*, September 1993, *61* (5), 989–1018.

Christiano, Lawrence J, Roberto Motto, and Massimo Rostagno, "Risk Shocks," *American Economic Review*, 2014, *104* (1), 27–65.

Cordell, Larry, Liang Geng, Laurie Goodman, and Lidan Yang, "The cost of delay," 2013.

Frankel, David M, Stephen Morris, and Ady Pauzner, "Equilibrium selection in global games with strategic complementarities," *Journal of Economic Theory*, 2003, *108* (1), 1–44.

Gale, Douglas and Martin Hellwig, "Incentive-compatible debt contracts: The one-period problem," *The Review of Economic Studies*, 1985, *52* (4), 647–663.

Gopinath, Gita, Loukas Karabarbounis, Sebnem Kalemli-Ozcan, and Carolina Villegas-Sanchez, "Productivity and Capital Allocation in Europe," *working paper*, 2014.

McGrattan, Ellen R and Edward C Prescott, "Taxes, Regulations, and the Value of U.S. and U.K. Corporations," *The Review of Economic Studies*, 2005, *3* (72), 767–796.

Milgrom, Paul and John Roberts, "Rationalizability, learning, and equilibrium in games with strategic complementarities," *Econometrica*, 1990, pp. 1255–1277.

Morris, Stephen and Hyun Song Shin, "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *The American Economic Review*, 1998, *88* (3), pp. 587–597.

_ and _ , *Global Games: Theory and Applications* Advances in Economics and Econometrics, (proceedings of the eighth world congress of the econometric society) ed., Cambridge University Press, 2003.

Sakovics, Jozsef and Jakub Steiner, "Who matters in coordination problems?," *The American Economic Review*, 2012, *102* (7), 3439–3461.

Townsend, Robert, "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, 1979, *21* (2), 265–293.

Vives, Xavier, "Nash equilibrium with strategic complementarities," *Journal of Mathematical Economics*, 1990, *19* (3), 305–321.