Enforcement, Incomplete Contracts, and Recessions

Cristina Arellano*, Yan Bai† and Jing Zhang**
University of Minnesota and Federal Reserve Bank of Minneapolis, Arizona State University and University of Michigan

PRELIMINARY AND INCOMPLETE
June 2006

Abstract
This paper studies the link between enforcement in financial contracts and output volatility. In particular we study how cycles and importantly recessions can be amplified in economies with weak contract enforcement. The idea of the paper is that firms in economies with enforcement problems will not be able to have the optimal scale of production because investors cannot enforce the collection of the large loans needed to operate at efficient levels. Moreover the value of a firm decreases over time after a sequence of bad shocks, because the firm needs to roll over an increasing larger amount of debt as output is insufficient to cover the uncontin- gent debt payments in low shocks. Thus as low aggregate shocks hit the economy an increasing fraction of firms experience a reduction in their value exacerbating enforcement frictions and the inefficiency in scale which further amplifies the recession.

* arellano@econ.umn.edu
† yan.bai@asu.edu
** jzhang@umich.edu
1. Introduction

Emerging markets have weaker contract enforcement and more volatile aggregate output than developed economies. For example Peru's GDP growth rate volatility is 6.5%, and its enforcement index is very low and equal to 1.7. On the contrary Canada’s GDP growth rate volatility is 2.3 while having a stronger degree of contract enforcement, with an index of 3.3.\(^1\) Moreover the effects from lack of enforcement in contracts are more important for sectors that rely on external finance and especially harsh when countries are in recessions. Rajan and Zingales (1998) find that industrial sectors in relatively more need of external finance grow faster in countries with better contract enforceability. Braun and Larrein (2004) document that this effect has an important cyclical variation in that these sectors are more strongly affected in recession in countries with weak contract enforcement.

This paper studies the link between enforcement in financial contracts and output volatility. In particular we study how cycles and importantly recessions can be amplified in economies with weak contract enforcement. The idea of the paper is that firms in economies with enforcement problems will not be able to have the optimal scale of production because investors cannot enforce the collection of the large loans needed to operate at efficient levels. Moreover the value of a firm decreases over time after a sequence of bad shocks, because the firm needs to roll over an increasing larger amount of debt as output is insufficient to cover the uncontingent debt payments in low shocks. Thus as low aggregate shocks hit the economy an increasing fraction of firms experience a reduction in their value exacerbating enforcement frictions and the inefficiency in scale which further amplifies the recession.

The framework is a dynamic model of heterogeneous firms similar to Albuquerque and Hopenhayn (2004) but under incomplete markets. Firms in the model borrow from foreign investors to finance the working capital and set up costs needed for production. Firms sign contracts with investors before their idiosyncratic productivity is known and we assume that contracts cannot be contingent on the realization of the shock. After the capital is in place,

---

\(^1\) The GDP volatility is computed using WDI statistics for the period from 1980 to 2001 in annual frequency. The enforcement index is compiled by Business Environmental Risk Intelligence and measures the relative degree to which contractual agreements are honored. It takes values from 0 to 4 with higher values indicating better enforcement.
each period the firm faces a shock to their productivity. Firms can choose to repay the capital borrowed and the long debt due and remain in operation for the next period, or default, exit, and get part of their current output. Incentives to repay debt depend crucially on the value of keeping the firm in operation at every period.

In our model the value of the firm depends inversely on the amount of debt that it owes to creditors. We show that firms with low values and high debt are inefficiently small because they under-invest relative to an unconstrained first best level. Enforcement problems are severe for low value firms, limiting the amount of risk-free funds. If firms instead borrow risky and default in some states, the high interest rates charged on loans also distort downward the optimal investment and the size of firms. Thus low value firms uniformly underinvest and produce at inefficient scales.

Incomplete contracts introduce rich dynamics in the debt firms owe and the value of firms. In our model firms with a history of bad shocks accumulate debt, while reducing its value because their low output is not enough to cover interest payments on outstanding debt. Thus our model generates ‘positive persistence’ for low productivity realizations in firms’ output because as debt grows over time the inefficiency in scale worsens and output further contracts. Incompleteness of contracts is key for generating this result because it forces debt holdings of firms to rise over time after bad shocks.

We find that enforcement problems at the firm level have important aggregate implications especially in economies with low aggregate productivity. Aggregate output in our economy is always smaller than that in an unconstrained benchmark. Moreover, the inefficiency is more accentuated when aggregate productivity is low. The reason is that values of firms are lower in low aggregate productivity economies and thus the contracts are more limited due to exacerbated enforcement frictions. Therefore, firms face higher interest rates and more limited capital inflows. This feature matches the data in that firms relying more on external finance in countries with weaker enforcement protection suffer more.

In terms of aggregate dynamics, we find that enforcement frictions and incomplete markets amplify both expansions and recessions, generating a higher volatility of aggregate output. In our model, when the economy is hit by a low aggregate shock, the old uncontingent debt for each firm becomes larger relative to their future expected lower output. Thus the ef-
The effective degree of contract enforcement decreases because the value of all firms goes down. Consequently, a larger fraction of firms go bankrupt and the financing possibilities for an efficient size are further limited. Recessions are then amplified due to more costly defaults and further underinvestment of non-defaulting firms. Incomplete contracts are essential for amplification of downturns. The model matches the data in that recessions are more severe in countries with weaker enforcement of contracts.

The amplification of expansions works in a similar fashion. When the economy has a boom, the value of all firms goes up, a smaller fraction of them go bankrupt and they can have a more efficient scale. Thus our model is in line with the cross country implication that countries with weak enforceability of contracts have larger output volatility. However, our insight is that with incomplete markets the extent of contract enforcement is endogenous to aggregate fluctuations because it is linked to the aggregate debt relative to the aggregate value of firms which varies with cycles.

Our work is closely related to the literature that studies the aggregate implications of financial imperfections. Cooley, Quadrini and Marimon (2004) find that enforcement frictions under a complete set of assets can amplify expansions. In their model entrepreneurs can search for better productivity projects after default and thus when the economy experiences a boom, the enforcement constraint becomes tighter because incentives to default are higher. The optimal contract then encompasses higher capital transfers to the entrepreneur to provide him incentives to stay in the credit relation. Our model shares the amplification of expansions, but our mechanism is different from theirs and has to do with the dynamics of aggregate debt. More importantly, our model also delivers amplification of recessions, while in their paper recessions might be dampened precisely because the outside option is lower and capital constraints are relaxed.

Another important paper that is related to our work is Kiyotaki and Moore (1997), in which firms use their physical assets to collateralize loans. Their paper generates the amplification of recessions, because in downturns the value of the capital stock is lower and thus collateral constraints become tighter which further limits the investment. While collateral financing may be a common type of contracts for domestic transactions, in the international context collateral contracts have limited applicability. Confiscation of foreign capital in case of de-
fault is rarely observed and only firms with highly tradable goods (like oil) can use their output as collateral for foreign financing. In fact most of the firms in emerging countries obtain foreign financing from transactions through banks and the main channel by which foreign investors insure against bankruptcy is through higher interest rates. Thus incentive considerations are at the core of the limited international capital flows. Our model with incomplete markets and enforcement captures the incentive consideration and also generates the empirical implications regarding levels and dynamics of interest rates across countries and over time.

2. Model Economy

The economy consists of a small country populated by a continuum of infinitively lived entrepreneurs who have access to risky projects and finance those projects by borrowing from foreign lenders. Financial contracts are not enforceable and entrepreneurs can default on the debt they owe to creditors. Contracts in this economy are also incomplete and thus default occurs along the equilibrium.

2.1 Firms

Entrepreneurs in the economy have access to a mass of $N$ project opportunities to produce consumption goods. We define each project opportunity as a firm. Every period a fraction $\xi$ of the projects are available to new entrepreneurs. Entrepreneurs are risk neutral, infinitively lived and decide on entry, exit and production and financing plans to maximize the lifetime value of dividends from the project

$$E_0 \sum_{t=0}^{\infty} \left[ \beta (1 - \delta) \right]^t d_t .$$

(1)

Entrepreneurs discount time at rate $\beta < 1$ and face an exogenous probability of dying $\delta < 1$ each period. New entrepreneurs have no initial wealth and their outside option is equal to 0. Thus, it is always worth it for an entrepreneur to enter and operate the firm if he draws a pro-
ject that gives him positive expected present value of dividends. Each entrepreneur owns at most one firm.

Firms are funded by foreign lenders who offer contracts to cover two distinctive needs: initial set up costs and working capital requirements. Each new firm needs to pay an initial set-up cost $I$ to start its operation. Since new firms have no initial wealth, they borrow from foreign long-lenders this initial set up cost. The set up cost is the initial loan for every firm $L_0 = I$. This initial loan is associated with a starting long-debt position for the firm $B_1$ that is due the next period. Every subsequent period the firm can choose a new loan and long-debt position $(L, B')$ from the set of long term contracts offered. By choosing a different long-term contract corresponding to a smaller long debt position, $B < B_1$, the firm effectively repays off part of its initial set up cost.

Every period each firm that is in operation produces output $y$ with a stochastic production technology that transforms capital goods into consumption goods with a decreasing returns technology. For a given level of capital input $K$ invested at the beginning of the period, the firm produces output $y$ at the end of the period such that

$$y = zK^\alpha,$$

where $0 < \alpha < 1$ and $z$ is the realization of the shock. The random variable $z$ is assumed to be i.i.d. and distributed according to $G$ with density $g$ on $[\underline{z}, \overline{z}]$. Capital depreciates completely every period.

The timing of decisions within the period is as follows. At the beginning of the period, each entrepreneur who owns a project with long-debt $B$, decides how much capital $K$ to invest in the project depending on expectations of productivity. Firms borrow the working capital desired for production from foreign short-lenders. At the end of the period the productivity shock $z$ is realized and the firm decides whether to repay or default on all debt due. If the firm repays the working capital and the long debt, then it chooses a long-debt contract from the ones offered and continues operations for the next period. If the firm defaults on all the debt due, both short and long term, then it gets part of their output, and exits. Every firm that exits frees up an additional project opportunity for a new entrepreneur the following period.
2.2 Contracts

The set of contracts between the firm and lenders are of two types: short-term contracts and long-term contracts. Short-term contracts are for capital requirements. These are traded in the beginning of the period before the productivity shock is known and consist of two numbers \((K, P)\). \(K\) denotes the payment in the beginning of the period from the lender to the firm which is used as capital input for producing output. \(P\) is the payment the firm promises the lender at the end of the period conditional on not defaulting. The set of short-term contracts are a function only of the firms’ long-debt position in the beginning of the period \(B\).

Long-term contracts are traded at the end of the period after the productivity shock is known and consist also of two numbers \((L, B')\). \(L\) is the immediate payment from the lender to the firm that is used for interest payments of the original set up costs and to increase period dividends. \(B'\) is payment the firm promises the lender next period conditional on not defaulting. \(B'\) is the new long debt position of the firm next period. The set of long-term contracts are independent of the firm’s long debt in the beginning of the period \(B\).\(^2\) Both contracts are independent of the history of shocks due to our \(i.i.d\). assumption. However, short-term and long-term are exogenously incomplete in that they are not a function of the output realization when the payments are due.

2.3 Incentives, Dividends and Production

The dividend each entrepreneur receives at the end of the period depends on whether he decides to repay his debts and remain in operation, or default and exit. After observing the productivity shock \(z\) at the end of the period, if repayment is chosen, then the entrepreneur of each firm receives as dividend

\[
d = zK'' - P - B + L.
\]  

\(^2\) We have written the long-term contract as a one period debt contract for simplicity. It can be easily shown that these contracts can be written as true long-debt contracts where firms change their stock of long debt every period.
The dividend conditional on repayment equals the output produced with capital $K$ in the period minus payment of the short-term working capital debt $P$ and long debt due $B$ plus new long debt issuances $L$. Also define as ‘profit’ the period return from the project excluding long debt operations: $zK^\alpha - P$.

If the entrepreneur chooses to default on his debt then the firm exits, the entrepreneur loses the project and receives a portion $\eta$ of the output

$$d = \eta zK^\alpha.$$  

(4)

We assume that default entails some direct bankruptcy costs such that $0 \leq \eta \leq 1$. The parameter $\eta$ controls the degree of enforceability of financial contracts in the country. If $\eta$ is very low the default is very unattractive for entrepreneurs and we interpret this as the case with higher enforceability of financial contracts. If $\eta$ is high then default becomes more attractive as contracts are less enforceable.

Given that the entrepreneur’s only financial relations are with the short-term lender and the long-term lender we also impose a limited liability condition such that

$$d \geq 0.$$  

(5)

Note that our limited liability condition is always feasible for all set of long-term contracts $(L, B')$ because default is always an option to the entrepreneur and $\eta \geq 0$.

The decision to default or repay debts depends on the current dividend the entrepreneur receives and on the entire future stream of dividends the project will generate. The firm will experience two types of default events: voluntary and involuntary. We first discuss the voluntary default decision.

The voluntary decision to default $d$ or repay $r$ for a particular firm that has long debt $B$ and short contract $(K, P)$ after observing its productivity shock $z$ is given by

$$W^o(B, K, P, z) = \max_{r,d} \{W^r(B, K, P, z), \eta zK^\alpha\}.$$  

(6)

$W^o(B, K, P, z)$ is the present value of a firm with $B$ debt, short term contract $(K, P)$, and $z$ productivity that has the option to default on all debt. $W^r(B, K, P, z)$ is the present value of such firm conditional on repaying debt today, but that tomorrow has the option to default.
We assume that if the firm defaults, then it exits and the entrepreneur gets part of the current output but loses the project forever. After a firm exists an additional project opportunity is available for a new entrepreneur the following period such that the mass of projects operating in every period is $N$.

If the entrepreneur decides to repay his debt, then he chooses a long-term contract $(L, B')$ to maximize the value of staying in the contract

$$W^c(B, K, P, z) = \max_{L, B'} \left( zK'' - P - B + L + \beta(1 - \delta)W(B') \right). \quad (7)$$

$W(B')$ is the present value of a firm with loan $B'$ in the beginning of the period before the productivity shock is realized.

By paying the short-term debt and choosing the new long-term contract the entrepreneur can remain in operation for one more period. The entrepreneur understands that the decision to repay debts is a period-by-period decision and that tomorrow he will again have the option to default.

The second type of default firms will experience is involuntary. Under this class of default event, the firm cannot satisfy its non-negative dividend condition with the set of long contracts available and it is forced to default although the lifetime value of remaining in the contract is larger. Formally, a firm that has long debt $B$, short contract $(K, P)$ and productivity shock $z$ is forced to default if

$$d = zK'' - P - B + L < 0 \text{ for all } (L, B').$$

The default policy of each firm can be summarized by repayment sets and default sets. For a given initial level of debt $B$ and short contract is $(K, P)$, denote the involuntary default set $D_{in}(B, K, P)$ as the $z$ set for which the firm cannot have positive dividends with any of the long contracts available

$$D_{in}(B, K, P) = \{z : zK'' - P - D + L < 0 \ \forall \ (L, B')\}. \quad (8)$$

Voluntary default sets $D_{vol}(B, K, P)$ for a firm with debt $B$ and short contract $(K, P)$ are the set of $z$ for which the value under the contract is less than the value under default.
\[ D_{\text{vol}}(B,K,P) = \{ z \in Z / D_{\text{inv}}(B,K,P) : W^z(B,K,P,z) < \eta z K^\alpha \} . \]  

The default set is then the union of the voluntary default set and the involuntary default set
\[ D(B,K,P) = D_{\text{vol}}(B,K,P) \cup D_{\text{inv}}(B,K,P). \]

The repayment set is defined as the complement of the default set:
\[ A(B,K,P) = [D(B,K,P)]^c. \]

Besides debt and default choices, the entrepreneur makes investment decisions every period. In the beginning of the period before the productivity shock is observed the entrepreneur decides on capital and production plans. Given an initial level of debt \( B \) the entrepreneur chooses the short-term contract \((K,P)\) to finance investment such that it maximizes his value
\[ W(B) = \max_{K,P} \int_0^1 W^z(B,K,P,z) dG(z). \]

The decision of how much capital to invest depends on the expectations of average productivity and on the set of contracts available. For example, if productivity is expected to be high, the optimal size of the project is larger and the entrepreneur has an incentive to invest a large level of capital. In fact in a world without enforcement frictions, the entrepreneur will choose the size of the project proportionally to productivity expectations. However with enforcement frictions the optimal size of the project can be distorted by the set of short-term and long-term contracts available. In particular if the firm already starts the period with a large level of long term debt \( B \) then default would be more likely and thus the set of short term and long term contracts will very limited. The entrepreneur might then not be able to run his project with an optimal size and the project will be inefficiently small. We show in the next session that the link between the level of long debt \( B \) and the scale of project imply that firms with high debt under-invest because enforcement problems are very stringent.

2.4 Lenders

Lenders in the model are assumed to be competitive and discount time at the rate of the international risk free interest rate \( r \). They behave passively and are willing to finance the firm’s
initial set up costs and working capital needs as long as they are compensated for the expected loss in case of default. In particular a short-term lender offers contracts \((K, P)\) to a firm with long debt \(B\) such that

\[
K = \frac{P}{(1+r)} \left(1 - \int_{D(B,K,P)} dG(z)\right). \tag{12}
\]

With every short term contract \((K, P)\) the lender breaks even in expected value.

A long-term lender offers contracts \((L, B')\) such that with every contract the lender receives in expectation the risk free interest rate

\[
L = \frac{B'}{1+r} \left(1 - \int_{D(B',K'(B'),P'(B'))} dG(z)\right). \tag{13}
\]

Every long-term contract offered \((L, B')\) forecasts the firm’s choice of the short-term contract next period \((K(B'), L(B'))\). This is because the firm’s decision to default or repay is carried out after the capital stock is in place and the productivity is realized.

When a new firm starts its operation, its initial long-term contract \((L_0, B_1)\) is such that the firm receives as first payment the initial set-up costs needed \(L_0 = I\) with

\[
I = \frac{B_1}{1+r} \left(1 - \int_{D(B_1,K_1(B_1),P_1(B_1))} dG(z)\right). \tag{14}
\]

If initial set up costs are so high then there might not exist a long-term contract that can finance the entry of the firm because probabilities of default are too high.

### 2.5 Equilibrium

We now define the equilibrium:

**Definition.**

The recursive equilibrium for this economy is defined as: (i) the policy functions for short-term contracts \((K(B), P(B))\), long-term contracts \((L(B,K,P,z), B'(B,K,P,z))\), dividends
Taking as given the menu of short-term contracts \((K,P)\), and long-term contracts \((L,B')\) offered, the policy functions \((K(B),P(B))\), \((L(B,K,P,z),B'(B,K,P,z))\), \(d(B,K,P,z)\), repayment sets \(A(B,K,P)\) and default sets \(D(B,K,P)\) satisfy the firm’s optimization problem.

2. Short-term contracts and long-term contracts available to each firm reflect the firm’s default probabilities such that lenders break even in expected value.

3. The distribution of firms \(\Upsilon(B,z)\) is consistent with individual decisions and shocks.

4. The mass of new entrants \(\xi(\Upsilon(B,z))\) is equal to the measure of all the firms that default and die in the limiting distribution \(\Upsilon(B,z)\).

2.6 First Best Benchmark

We will compare our model to the benchmark where contracts are perfectly enforceable but incomplete. We denote this benchmark model as the first best. The capital stock of each firm in the first best is such that the expected marginal product of capital equals the interest rate. Thus all firms have the same size, and the distribution of firms is degenerate at the optimal capital stock. The firm pays each period the perpetuity value of the set-up cost and receives as dividend the residual. In the first best the capital stock is determined by

\[
E(z)\alpha K_{fb}^{-\alpha-1} = (1 + r). 
\]

Thus, the first best short contract is given by

\[
\{K_{fb}, P_{fb}\} = \left\{ \left( \frac{E(z)\alpha}{1 + r} \right)^{\frac{1}{1-\alpha}}, \left(1 + r\right) \left( \frac{E(z)\alpha}{1 + r} \right)^{\frac{1}{\alpha}} \right\}.
\]
The value of the firm under the first best for initial set up costs $I$ is

$$W(I) = \frac{(E(z)K_{fb} - P_{fb} - rI)(1 - \beta \delta)}{\beta \delta}.$$ 

3. Firms Dynamics

This section studies the cross section implications from lack of enforcement in contracts and incomplete markets in terms of firm size as well as the dynamics for output and investment for individual firms. We focus on the case with two idiosyncratic shocks. Our two main results are that in our model firms with high levels of long debt are inefficiently small, and this inefficiency is exacerbated over time if firms receive a sequence of bad productivity realizations. Our model also delivers a non-degenerate distribution of firms because the ones that have a sufficient level of long debt choose to under-invest relative to the first best benchmark. Moreover we show that the size of a particular firm that receive a series of bad productivity realizations become smaller and smaller over time as long debt grows. We show how lack of enforcement plus incomplete markets are both key in generating the amplification in output of a particular firm after low productivity realizations.

3.1 Underinvestment

This section characterizes default and investment decisions for firms. The productivity shocks are denoted by $z = \{z_L, z_H\}$. The probability of the low shock is $\pi_L$ and the probability for the high shock is $\pi_H = 1 - \pi_L$. We study how the firms’ decisions vary as a function of the level of long debt it owes. With only idiosyncratic uncertainty that is i.i.d. over time, the set of long debt contracts $(L, B')$ offered to firms is constant over time and across firms. In particular the available set is independent of the level of level of long debt $B$, productivity realization $z$ and short-term contract $(K, P)$.

Lemma 1. Default sets are increasing in long debt.
If \( z \in D(B^1, K, P) \), then \( z \in D(B^2, K, P) \) \( \forall B^2 > B^1 \).

Firms default when they hold high long debt. This is because both current dividend \( d(B, K, P, z) \) and the contract value \( W^c(B, K, P, z) \) are decreasing in long debt while the default value is independent of long debt.

**Lemma 2.** Long debt contracts are bounded. There exists an \( L > 0 \) such that \( L \leq L \) for all \((L, B')\) contracts.

The set of long debt contracts are bounded by the maximum set up costs that a project can have while remaining profitable. In an environment without perfect enforcement, the entrepreneur will only participate in the project if the expected value delivers a positive value. The project at most will generate the expected lifetime discounted profit under the optimal investment \( K_{fb} \), \( M = \left[ E(z)K_{fb}^{\infty} - P_{fb} \right]/r \). Thus projects with larger set up costs will for sure not be financed because they can not deliver expected positive value to both the lender and the entrepreneur. Hence, the set of long contracts every period offered to firms that remain in operations are bounded such that \( L \leq M \).

Moreover, the value of each project in the economy with imperfect enforcement will be lower than in an economy with perfect enforcement due to the added constraints. We show below that in our model capital depends on the level of long debt and firms will be inefficiently small. This implies that the maximum set up cost for which projects will be financed are strictly smaller in our economy relative to the first best \( \bar{L} < M \). In addition, if with this maximum loan of \( L \) default sets are non-empty the corresponding long contract \((L, B)\) will imply that \( \bar{B} \geq \bar{L}(1 + r) \).

**Definition.** Debt Overhang. A firm with long debt \( B \) faces ‘debt overhang’ if when it invests the first best capital stock, it cannot satisfy the non-negative dividend conditions for all shock realizations.
**Proposition 1.** The debt overhang exists. There exists a $B^*$ such that for all $B > B^*$, 
\[ z_L K^*_{fh} - P_{fh} - B + L < 0 \forall (L, B^*). \]

Define $B^* = L + z_L K^*_{fh} - P_{fh}$. For any $B > B^*$, the non-negative dividend condition for $z_L$ cannot be satisfied. The first best capital stock is chosen based on average productivity. However the non-negative dividend condition must be satisfied for every shock realization. Thus if ex-post the low shock is realized for some firm the first best capital stock is too big because the short-term loan is on an inefficiently big capital stock. Given that enforcement problems limit the resources long debt contracts provide, a firm with large long debt that invests the first best capital and receives the low productivity shock will have negative dividends.

Note that in our model firms can have debt overhang for $B < B^*$ if $L + z_L K^*_{fh} - P_{fh} < B$. Given that $B \geq L(1 + r)$ the necessary condition is that $z_L K^*_{fh} - P_{fh} < rL$. We can then choose a particular distribution of shocks to guarantee this condition for any $rL \geq 0$.

**Proposition 2.** Default sets are the lower set.

If $z \in D(B, K, P)$, then $z_i \in D(B, K, P)$ for any $z_i < z$.

*Proof.* See Appendix.

Firms default when productivity is low. If default is involuntary then it is clear that default happens for low shocks because current dividends $d(B, K, P, z)$ are decreasing in $z$. If default is voluntary, then firms default for low productivity because the contract value $W^c(B, K, P, z)$ is increasing in productivity at a faster rate than the default value. The reason is that the set of long debt contracts the firm can choose under the low shock, such that the non-negative dividend condition is satisfied, are also feasible under the high shock. It is then clear that given that $\eta < 1$ the contract value increases in productivity at a faster rate than default value and we get the result.

**Proposition 3.** A firm with debt overhangs under-invests relative to the first best.
Proof. See Appendix.

A firm that faces ‘debt overhang’ has two choices: (1) the firm chooses not to default for both shocks, but adjusts the investment such that the non-negative dividend condition is satisfied for $z_L$; (2) the firm chooses to default for some shock and adjusts the investment decision according to default choices. The proposition shows both cases lead to underinvestment relative to the first best.

A firm that faces debt overhang and decides to repay debt, cannot have the first best capital because this would lead to having negative dividends. Thus the firm decreases its investment such that profits under $z_L$ are high enough to deliver a non-negative dividend. This necessarily leads to underinvestment relative to the first best because smaller sizes of firms increase ex-post profits for $z_L$.

The firm with debt overhang could also decide to default for some shock realizations. If the firm chooses to default, this will happen only for the case of low productivity $z_L$. This is because default sets are the lower set due to proposition 2, and because they are less than the whole set when projects produce positive output.

Capital choices are based on maximizing expected profits and relaxing the non-negative dividend condition ex-post state by state in the repayment states. Regarding the maximization of profits firms choose capital such that the marginal product equals the marginal cost. When firms choose to default under some shocks, the expected marginal product of capital is lower than if firms do not default because default states involve direct costs as the firm loses $(1-\eta)$ of output. However the expected marginal cost of capital is the same regardless of the default decisions because short-term lenders break even. These two features push the optimal capital stock to be smaller than the first best when default sets are non-empty. Regarding relaxing the non-negative dividend condition, this is only relevant for the choice of capital if it is binding under $z_L$. However, under $z_H$ the non-negative dividend condition is relaxed by also decreasing the capital stock. The reason is that given that the short-term contract carries a de-

---

3 The non-negative dividend condition under $z_L$ is irrelevant if default is chosen here.
fault premium and the capital that maximizes \( z_h K^\alpha - (1 + r) / \pi_h K \) is smaller than \( K_{fb} \). Thus if the firm wants to use capital to relax the non-negative dividend condition in the high shock, this also implies that it must under-invest relative to the first best. It is key that the short-term contract adjusts according to default choices. If the short-term contract would not be adjusted to default choices a defaulter firm could find it optimal to over-invest relative to the first best.

Thus firms with high levels of long-term debt will for sure under-invest regardless of their default choices. The next session explores quantitatively this magnitude of the inefficient scale in production.

### 3.2 Numerical Examples

This section presents some numerical examples to quantify the underinvestment feature our model has. Table 1 presents the parameters used in these examples. The capital share \( \alpha \), the discount factor \( \beta \) and the risk free interest rate \( r \) are set at the standard literature value. The death rate of entrepreneurs \( \delta \) is set to 5%, which is consistent with the numbers reported in industry dynamics studies (Evan 1987). The enforcement parameter \( \eta \) is set to 0.5, which means that the bankruptcy cost is 50% of output, or that entrepreneurs are able to steal 50% of output when defaulting. The startup cost parameter is set to be 0.58. In flow terms, \( rI_0 \) is about 8% of aggregate output in the model economy.

<table>
<thead>
<tr>
<th>Table 1: Parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital share ( \alpha )</td>
<td>0.33</td>
</tr>
<tr>
<td>Discount factor ( \beta )</td>
<td>0.96</td>
</tr>
<tr>
<td>Risk free rate ( r )</td>
<td>0.04</td>
</tr>
<tr>
<td>Death rate ( \delta )</td>
<td>0.05</td>
</tr>
<tr>
<td>Enforcement friction ( \eta )</td>
<td>0.50</td>
</tr>
<tr>
<td>Initial set up cost ( I_0 )</td>
<td>0.58</td>
</tr>
</tbody>
</table>
Given the above parameterization, we first analyze the firm’s decision rules as a function of long debt levels, and then present the limiting distribution of firms. Based on understanding of decision rules and the limiting distribution, we analyze the time series dynamics of firms using simulated examples.

Figure 1 plots investment, output and dividend of a firm as a function of its long debt normalized by GDP. The upper panel of Figure 1 presents investment decisions which is independent of shock realizations given the nature of incomplete markets. The middle panel and the lower panel plot output and dividend conditional on receiving a high or a low productivity shock. We could classify 4 distinct regions across long debt over GDP ratios according to investment decisions. In the first region long debt is small enough such investment is equal to the first best. The second region corresponds to the case where the firm faces a binding non-negative dividend condition when the low shock is realized. Here the firm under-invests as prescribed in Proposition 3 to avoid default and be able to increase profits in the low state such that the non-negative dividend condition is satisfied. As the figure shows in region 2 the firm is willing to become very small and inefficient in order to avoid default. This is because the continuation value of keeping the project $W(B')$ is very large once the initial setup costs have been already paid.

When long debt is big enough, the firm finds it optimal to default if the low shock is realized; this is region 3. In anticipation of such an event the firm under-invests relative to the first best as in Proposition 3, but increases investment relative to region 2. The reason for the increase in investment is that given that in the low shock the short loan will not be repaid, the firm finds it optimal to adjust its size for a more appropriate scale in the high shock. However given the more expensive short contracts, this increase in investment is never big enough such that capital is higher than in the first best. For the highest debt levels in region 3, the firm can further under-invests because the non-negative dividend condition binds for the high shock, and decreasing investment relaxes this constraint.
Finally in region 4, long debt is so high, that for any positive output the firm will find it optimal to default. Thus in equilibrium short lenders are unwilling to provide the firm with any contracts that would give positive output. An alternative way to think about region 4 is that projects with initial set up costs corresponding to region 4 will not be financed in equilibrium. Note though that under perfect enforcement some of these projects would be financed because the value of the firm under perfect enforcement is uniformly higher than under imperfect enforcement for all initial set up costs. This is because the value of the firm is lower due to underinvestment when debt is high.
Figure 1. Firm’s investment, output and dividend as a function of long debt holdings

Given the parameters of the model and the above decision rules, the model generates a limiting distribution of firms, shown in Figure 2 where we report the capital stock relative to the first best. In the limiting distribution, substantial portions of firms hold large long debt positions, and thus the firm size distribution contains many inefficiently small firms. About 73% of firms have a capital stock almost as big as the first best with a ratio of 0.9. Some firms have very small size, or about 10 times smaller than in a frictionless world. The cross sectional variance in output across firms is 0.2 in the limiting distribution. Without aggregate uncertainty the limiting distribution of firms is invariant and aggregate output is constant. However, each firm has volatile output as they experience high or low productivity and move through the distribution of debt. Clearly aggregate investment and output in our economy with enforcement problems are smaller than in an economy with perfect enforcement because of the link between debt and optimal size.
The result that enforcement problems induce firms to be inefficiently small is also present in models of enforcement frictions with a complete set of assets (Quintin (2000), Cooley, Marron and Quadrini (2004), and Clementi and Hopenhayn (2006)). Thus our added feature of incomplete markets is unnecessary for qualitatively getting this result. However, our key difference is that incomplete markets amplify the inefficiency in scale for some firms because the value of the firm can decrease over time due to higher long debt holdings. Enforcement problems are more severe when the value of the firm is low, thus as firms increase debt holdings the enforcement friction limits further their ability to produce efficiently. For a given initial start up cost, lack of enforcement plus incomplete markets generates a larger portion of inefficiently small firms in the invariant distribution.

Equipped with the investment decision rules for firms, we now present some time series examples. The time series dynamics of a firm’s investment and output after a series of shocks is driven by the dynamics of long debt, as short-term investment contracts do not carry any persistent effects in our model.⁴ Figure 3 presents the time series dynamics for two firms: the upper panel plots output and long-term bond dynamics for a firm that experiences a sequence of only low productivity shocks; the lower panel plots the same statistics for a firm that ex-

---

⁴ This is because we have assumed that depreciation of capital equals 1. If depreciation was to be lower, capital dynamics would interact with long debt dynamics.
periences a sequence of only high productivity shocks. The initial debt position was chosen such that firms start in region 2.

Let’s first consider the case of a firm that faces low productivity shocks. This firm produces less and less output while it increases its long debt holdings over time. The firm chooses to remain in operation because of better future prospects, but it produces very little contemporaneously. After a long enough sequence of low shocks, 6 periods in this simulation, long debt grows so much, such that the firm finds it optimal to go bankrupt and exit. Anticipating a default event in period 6, the capital invested increases because the firm does not repay back the short loan in the low shock and thus dividends are strictly positive.

Incomplete markets are essential for enforcement frictions to amplify bad productivity shocks over time. Our interpretation is that enforcement problems are exacerbated for firms that have a history of low productivity. A sequence of low shocks lowers the value of the firm because uncontingent long debt has to grow over time to cover interest payments of an increasingly higher stock of debt. This feature differentiates our model from existing models that look at the implications of enforcement problems under a complete set of assets. In those models the value of the firm cannot go down ever because state contingent long debt allows the firm to efficiently pay off the set up costs only when high shocks are realized. By introducing incomplete markets our model is able to maintain the underinvestment feature of those models, while producing the additional feature of amplification for low shocks.
Let’s now consider the case of a firm who faces a sequence of good productivity shocks as in the lower panel of Figure 3. A firm who initially has a high initial long debt, starts operating at an inefficient scale. However if the firm receives a high productivity shock then it is able to repay enough of its long debt, such that in the next period it becomes unconstrained. This shows that our model also generates amplification of good shocks because after a sequence of high shocks the value of the firm increases enough such that it becomes unconstrained. Our model shares the ‘positive persistence’ feature of models with enforcement frictions and a complete set of assets have.

4. Aggregate Implications

This section compares levels and dynamics of aggregate output across two economies with different aggregate productivities. We parameterize the level of aggregate productivity by choosing two distinct transition matrices for shocks. Table 2 presents the transition matrices used in the two experiments of this section: high regime and low regime. In the high regime, the fraction of firms has the high shock $\pi_H$ larger than that in the low regime. All other parameters across economies are equal to those in Table 1.
Table 2. Two Economies

<table>
<thead>
<tr>
<th></th>
<th>Transition Matrix</th>
<th>Shock Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>High regime</td>
<td>(\pi_L = 0.30), (\pi_H = 0.70)</td>
<td>(z_L = 0.1), (z_H = 0.9)</td>
</tr>
<tr>
<td>Low regime</td>
<td>(\pi_L = 0.35), (\pi_H = 0.65)</td>
<td>(z_L = 0.1), (z_H = 0.9)</td>
</tr>
</tbody>
</table>

As shown in the previous section aggregate output in our model with enforcement and incomplete markets is lower than that in the first best benchmark because firms are inefficiently small. However, this inefficiency is more severe in economies with lower aggregate productivity. Table 3 presents aggregate statistics in the limiting distribution for the model economy and the first best benchmark across both regimes. Aggregate investment and output are lower in both regimes relative to the respective first best benchmarks, but they are relatively smaller in the regime with the low aggregate productivity. In the high regime, aggregate output is about 95% of the first best level, while in the low regime it is 93% of the first best. In the limiting distributions, the long debt over GDP ratio in the low regime is lower than that in the high regime because less debt can be sustained under the low aggregate productivity.

Table 3. Aggregate Statistics for Model Economics in Limiting Distribution

<table>
<thead>
<tr>
<th></th>
<th>Output</th>
<th>Capital</th>
<th>Interest Rate</th>
<th>Long Debt / Output</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model Economy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Regime</td>
<td>0.29</td>
<td>0.084</td>
<td>7.1%</td>
<td>1.35</td>
</tr>
<tr>
<td>Low Regime</td>
<td>0.26</td>
<td>0.073</td>
<td>8.2%</td>
<td>1.16</td>
</tr>
<tr>
<td>High/Low Ratio</td>
<td>1.12</td>
<td>1.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>First Best Benchmark</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Regime</td>
<td>0.31</td>
<td>0.097</td>
<td>4%</td>
<td>---</td>
</tr>
<tr>
<td>Low Regime</td>
<td>0.28</td>
<td>0.088</td>
<td>4%</td>
<td>---</td>
</tr>
<tr>
<td>High/Low Ratio</td>
<td>1.10</td>
<td>1.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Now we focus on amplifications of aggregate output from the high regime to the low regime in the limiting distributions. In our model economy, the ratio of aggregate output across two regimes is 1.12, higher than 1.10 in the first best benchmark. The difference in steady state aggregate output across these two regimes is driven by two factors: changes in the investment decision of individual firms and changes in the distribution of firms.

First, we find the investment decision for individual firms is the main driver for the difference in amplification across economies. In the first best benchmark, firms’ decisions on investment only depend on changes in regimes, not on their debt levels. In our model, changes in investment across regimes also depend on firms’ different debt levels. For example, in our model a firm with a ratio of debt to GDP of 150% chooses a capital stock of 93% of the corresponding first best capital in the high regime. However, in the low regime a firm with the same debt to GDP ratio, chooses a much lower capital stock of 78% of the corresponding first best capital. Thus, the impact from lack of enforcement for individual firms is most severe in economies with a low aggregate productivity, which leads to even lower investment and output relative to those in the first best.

Second, in terms of changes in the limiting distributions, we first look at defaults rates across the two regimes: 3% in the low regime and 2% in the high regime. The reason for higher default rates is that a larger portion of firms have lower productivities and are pushed to borrow risky. Higher default rate translates into lower aggregate output because of the direct bankruptcy costs. Additionally, higher default rates translate into more expensive loans and higher interest rates which distort downward the investment. The average interest rate in the low regime is 8.2 percent versus 7.1 percent in the high regime. Thus higher default rates decrease aggregate output in the low productivity economy because of higher marginal cost of capital for non-defaulting firms and a larger number of costly defaults.

However, overall the difference across steady states output from the two regimes in the limiting distribution is not that strong. This is because the distribution adjusts such that a lower percentage of firms end up with such high levels of debt in the new limiting distribution. To see this, we compare fractions of firms with a debt to GDP ratio higher than 1.4 across two regimes. In the high regimes, the fraction is 99% whereas it is only 19% in the low regime. We could also look at fractions of firms in the constraint region (the second region).
The fraction is 93% in the high regime, but 92% in the low regime.

It is worth noting that the difference between aggregate outputs will be amplified if the low regime is associated with higher start-up costs. Empirical evidence favors the view that the process to set up firms is more costly in poor countries. For example, Djankov et al. (2002) document that the number of days it takes to set up a firm in Peru is 83 days, whereas in Canada is 2 days. In monetary terms, the average cost to set up a firm is $435 (2.25% of GDP per capita) in Canada, whereas in Peru it is $1268 (53.06% of GDP per capita).

We next explore transitions of aggregate output, capital and long debt when the economy starts in the high (low) regime and changes unexpectedly to the low (high) regime. Let’s first look at transition dynamics when the economy moves from a high productivity regime to a low productivity regime as in Figure 4a. All variables are normalized by the respective statistics in the first period. We assume all firms start in period 1 with the high-regime limiting distribution, and make investment and long debt decisions based on the high regime. In period 2 all firms are unexpectedly hit by a permanent low aggregate productivity regime. Firms understand that from now onward the new regime encompasses low aggregate productivity and they make decisions accordingly.

As the figure shows aggregate output will decrease in both economies: the first best benchmark and our model economy. However, recessions are amplified in our model relative to the first best benchmark, because after a low aggregate shock a larger portion of firms default and exit and a larger portion of firms under-invest. The long debt over GDP ratio is larger in period 2 than in period 1, which makes the economy more vulnerable to underinvestment and default. GDP in period 2 is 4% lower than in the first best benchmark after the negative aggregate shock. Aggregate capital is much lower in period 2, 8% lower than in the first best. The economy converges to a new lower steady state level of output in period 6.

Let’s now look at the transition dynamics from the low to high regime. We do a similar experiment as for the case of the recession. Figure 4b shows aggregate output, capital and long debt. The model also amplifies expansions, as aggregate output and capital in our economy are 5% and 12% larger than in the first best. Booms are amplified in our model because the fraction of firms that default is smaller when the economy experiences booms. In addition, under-investment is less severe because aggregate long debt holdings are smaller in period 2.
than in period 1. The economy converges to a permanently higher level in period 6 which is relatively higher than under the first best.\footnote{The new steady state level is of aggregate output is lower in the economy with enforcement frictions, however relatively to the period 1 level of output this output is higher as Figure 4b shows.}

Figure 4: Impulse Responses

From the above two experiments, it is clear that our model economy amplifies responses to both booms and to recessions. The main mechanism we want to stress is that we can generate such amplification because with incomplete markets and lack of commitment, the effec-
tive degree of enforceability in contracts varies endogenously with cycles. This is because the effective degree of contract enforcement is linked to the aggregate uncontingent debt relative to the aggregate value of firms and this varies with aggregate fluctuations.
References


Appendix

**Proposition 2.** Default sets are the lower set.
If \( z_2 \in D(B, K, P) \), then \( z_1 \in D(B, K, P) \) for any \( z_1 < z_2 \).

Proof: We need to show that for any \( z_1 < z_2 \), \( W^c(B, K, P, z_1) < \eta z_1 K^a \), i.e.,
\[
(1-\eta)z_1 K^a - P - B + L_1^* + \beta W^a(B_1^*) < 0,
\]
where \((L_1, B_1^*)\) is the optimal long debt contract.

From \( z_2 \in D(B, K, P) \), we have \( (1-\eta)z_2 K^a - P - B + L_2^* + \beta W^a(B_2^*) < 0 \), where \((L_2, B_2^*)\) is the optimal long debt contract. Define the feasible set of long debt contract \((L, B')\) under \( W^c(B, K, P, z) \) as \( \Gamma(K, P, B, z) = \{(L, B') \mid z K^a - P - B + L \geq 0\} \). Clearly \( \Gamma(K, P, B, z_1) \subseteq \Gamma(K, P, B, z_2) \). Thus, we have
\[
0 > (1-\eta)z_2 K^a - P - B + L_2^* + \beta W^a(B_2^*) \geq (1-\eta)z_1 K^a - P - B + L_1^* + \beta W^a(B_1^*)
\]
\[
\geq (1-\eta)z_1 K^a - P - B + L_1^* + \beta W^a(B_1^*) \).
Q.E.D.

**Proposition 3.** A firm with debt overhangs under-invests relative to the first best.

Proof: A firm with debt overhangs has two choices: either repays under both shocks or defaults under the low shock, but repays under the high shock. We show that the firm will under-invest relative to the first best under either choice.

Choice 1: To repay debt under both shocks, the firm has to adjust investment away from the first best to guarantee the non-negative dividend (NND) condition under \( z_L \). That is, the firm has to increase the profit under \( z_L \), only by lowering investment from \( K_{fb} \).

Choice 2: We first show that \( K_{fb} \) is dominated by some investment level smaller than the first best, and then show that \( K_{fb} \) dominates any investment higher than the first best.
(1) Underinvestment: the first best investment $K_{fb}$ is dominated by some $K_a$, smaller than $K_{fb}$, i.e., $K_a = K_{fb} - \varepsilon < K_{fb}$. We need to show that the expected welfare under $K_a$, denoted by $A(B, K_a, P_a)$ is higher than that under $K_{fb}$, denoted by $A(B, K_{fb}, P_{fb})$, where the expected welfare is defined as $A(B, K, P) = E\left(\max\{W^c(B, z, K, P), \eta z K^\alpha\}\right)$.

Assume that the short contracts are designed according to defaulting only for the low shock. We can find an $\varepsilon_0 > 0$ such that choosing $K_a > K_{fb} - \varepsilon_0$ increases short profits under both shocks and the expected profit relative to choosing $K_{fb}$, where $\varepsilon_0$ is defined as $\varepsilon_0 = K_{fb} - (\alpha (\pi_H z_H + \eta \pi_L z_L) / (1 + r))^{1/(1-\alpha)} > 0$. Consider two cases when the firm chooses $K_a$: defaults under $z_L$ and repays under both shocks.

(a) When choosing $K_a$, the firm defaults under $z_L$.

In this case, we know $A(B, K_a, P_a) = \pi_H W^c(B, z_H, K_a, P_a) + \eta \pi_L z_L K^\alpha_a$ and $A(B, K_{fb}, P_{fb}) = \pi_H W^c(B, z_H, K_{fb}, P_{fb}) + \eta \pi_L z_L K^\alpha_{fb}$. Assume that $(L'(K_{fb}), B'^\alpha(K_{fb}))$ is the optimal long contract of $W^c(B, z_H, K_{fb}, P_{fb})$. Since the profit under $K_a$ is higher than that under $K_{fb}$ when $z_H$ occurs, $L'(K_{fb})$ is feasible under $W^c(B, z_H, K_a, P_a)$.

To show $A(B, K_a, P_a) > A(B, K_{fb}, P_{fb})$, it is suffice to show

$$(\pi_H z_H + \pi_L \eta z_L) K^\alpha_a - (1 + r) K_a > K^\alpha_{fb} - (\pi_H z_H + \pi_L \eta z_L)(1 + r) K_{fb},$$

which is guaranteed by our choice of $K_a$.

(b) When choosing $K_a$, the firm repays for both shocks.

In this case, we have $A(B, K_a, P_a) = \pi_H W^c(B, z_H, K_a, P_a) + \pi_L W^c(B, z_L, K_a, P_a)$.

We know that $A(B, K_a, P_a) \geq \pi_H W^c(B, z_H, K_a, P_a) + \pi_L \eta z_L K^\alpha_a$ because the firm chooses to repay under the low shock. As in case (a), we have

$W^c(B, z_H, K_a, P_a) \geq z_H K^\alpha_a - (1 + r) K_a - B + L'(K_{fb}) + \beta \delta W'^\alpha(B'^\alpha(K_{fb}))$

$\geq z_H K^\alpha_a - (1 + r) K_a / \pi_H - B + L'(K_{fb}) + \beta \delta W'^\alpha(B'^\alpha(K_{fb}))$. 

Thus, to show $A(B, K_a, P_a) > A(B, K_{fb}, P_{fb})$ is equivalent to show inequality (16), which is true for our choice of $K_a$.

(2) No overinvestment: any investment greater than the first best $K_{fb}$ is dominated by $K_{fb}$.

The profits under both shocks and also the expected profit will decrease when the firm over-invests. Follows the similar argument above, we could show that any investment greater than the first best $K_{fb}$ is dominated by $K_{fb}$. Q.E.D.