The Costs of Credit Booms

Guido Lorenzoni*

November 19, 2002

Abstract

In this paper we study a simple model of a credit boom driven by an expected increase in productivity in the entrepreneurial sector. We study how the presence of collateral constraints affects the welfare properties of the equilibrium. In particular we show under what circumstances entrepreneurs tend to under-insure their net worth against negative aggregate shocks and to over-invest during the boom.

Despite the presence of over-investment, we show that a monetary contraction during the credit boom is a blunt instrument to attack the inefficiency described, and it can be counterproductive. Capital requirements and a monetary policy aimed at output stabilization *ex post* are more effective tools.

Over the last two decades both industrialized and emerging economies have experienced credit booms associated to periods of high investment, high asset prices and fast growth. In some cases these have been followed by a bust associated with low investment, bankruptcies, a credit contraction and a contraction in output. An extensive literature has showed how cyclical movements can be driven and amplified by the presence of credit constraints. In this paper we use a simple model to study the welfare properties of a credit boom in presence of financial constraints. Our main objective is to show in what circumstances *over-investment* arises during a credit boom and how it is associated to excessive *fragility*, i.e. to a financial structure that is excessively sensitive to aggregate shocks.

*Princeton University, Department of Economics and Woodrow Wilson School of Public and International Affairs

The models features a fundamental investment boom driven by high expected productivity. The boom is followed, with some probability, by a bust associated to a low realization of the productivity shock. The model sheds some light on the role of financial frictions in amplifying a fundamental shock, on the difference between private and social incentives to stabilize entrepreneurial wealth and thus on the inefficiencies associated with the boom and bust cycle. In particular, the model shows that during a credit boom entrepreneurs may over-invest because they under-estimate the social damage coming from a negative shock to their net worth. The inefficiency arises solely from the presence of a credit market friction and it is not due to irrational or speculative pricing of financial assets nor to the lack of state-contingent clauses in financial contracts. *Ex post* the economy is faced with a debt-overhang problem. Even though entrepreneurs correctly forecast the probability of the debt-overhang and have access to contingent contracts that could protect their balance sheets, they tend to use contingent contracts less than optimally.

In models with collateral constraints little attention has been devoted to the costs associated to the boom side of a credit cycle. From a first best perspective models of credit constraints always display *under-investment.* Policy makers may be concerned with credit crunches but it is not clear why they should be concerned with credit booms that boost the net worth of firms, reduce the outside-finance premium and increase investment in high return projects. In a monetary economy an investment boom may generate a demand push, but as long as monetary policy adjusts the interest rate to its natural level, and keep inflationary pressures are under control, there seem to be no additional reason to increase interest rates to quench a credit boom.

This paper shows that if one moves to a second best perspective the costs of a credit boom can be analyzed in a relatively standard framework, and this can help us determine the circumstances under which a boom may be socially costly and study the effect of different policies. In particular, we show that the cost of the boom depends on the uncertainty regarding future productivity growth, on the relative importance of temporary and permanent productivity shocks and on the level of entrepreneurial wealth in the booming sector. Moreover, we show that policies oriented at reducing fragility tend to be more effective than policies that try to reduce the level of investment during the boom.

Models that incorporate financial market frictions have been used to study the propagation and amplification of real shocks. In particular models that highlight the role of real assets as collateral, starting with Kiyotaki and Moore (1997), have shown that a negative productivity shock that reduces asset prices can amplify its effects by reducing the net worth of the entrepreneurial sector. A large part of the literature[1] imposes restrictions on the liabilities used by the entrepreneurs to finance their projects, and rules out state contingent liabilities that depend on *aggregate shocks*. On the other hand the corporate finance arguments that have been used to explain the use of rigid liabilities (debt) cannot be invoked to justify this type of rigidity, given that economy-wide aggregate shocks are usually outside the control of the single manager or entrepreneur and are easily observable.

A second contribution of this paper is to show that even in presence of state contingent liabilities we observe investment and output fluctuations that are larger and more persistent, both if compared to a frictionless first best and compared to a second best benchmark. Moreover equilibrium financial contracts tend to be too rigid with respect to aggregate shocks. In section **??** we discuss the introduction of contracting costs and we show that our model can be used to show that private contracts may fail to account for some realizations of the aggregate shock that have first-order effects on *ex ante* welfare.

Our results rely crucially on the presence of pecuniary externalities in a model with collateral constraints. Collateral constraints limit both the amount of borrowing and the amount of insurance that can be supported in equilibrium. A firm can buy more insurance against a bad productivity shock by reducing the contingent payments it has to make if that shock materializes. This protects the firm net worth and allows the firm to invest more. This has a positive effect on wages that firms do not take into account. Essentially, workers would be willing to pay the firms ex ante to induce them to reduce the volatility of entrepreneurial net worth. If we compare the competitive equilibrium with a constrained optimum we have higher investment during the boom *ex ante* and higher dispersion in output *ex post*, therefore the inefficiency magnifies macroeconomic fluctuations both across states and across time.

This paper is related to the literature on the effects of aggregate liquidity shocks. In particular Holmstrom and Tirole (1998) have looked directly at

[1]Some of the papers that study liquidity crises assuming that private liabilities are not contingent on the aggregate shock are Diamond and Rajan (2002) and Allen and Gale (2001).

the second best, therefore the issues related to private vs. social gains from net-worth insurance do not arise in their model. In an international context Caballero and Krishnamurty (2001 and 2002) have studied economies in which excess investment can arise *ex ante* and entrepreneurs tend to invest too little in state contingent (e.g. peso denominated) liabilities. We share their emphasis on general equilibrium effects and on pecuniary externalities, but the mechanism at work is different, as we don't have a dual model of liquidity. They concentrate on the effects that equilibrium prices have on the reallocation of wealth among entrepreneurs, while here we concentrate on effect they have in reallocating wealth between entrepreneurs and outside investors.

On the other hand this paper is related to the burgeoning literature on asset prices and monetary policy. A number of recent contributions to the debate on asset prices and monetary policy have assumed that an irrational bubble component may be driving the asset price boom. In particular Bernanke and Gertler (2001a, 2001b) have studied the effects of irrational mis-pricing of assets in a model with financial frictions and have reached the conclusion that monetary policy need not respond to asset price fluctuations. In their model, however, they assume that the bubble affects investment only through an effect on the borrowers balance sheet. Essentially in presence of a bubble entrepreneurs see their net worth increased and invest more because the outside finance premium declines, but are not led to issue additional equity just to take advantage of the asset price boom. A mis-pricing in their model has only the effect of temporarily boosting or reducing the net worth of entrepreneurs. Dupor (2002) reaches opposite conclusions on optimal monetary policy on the basis of a model with no financial frictions where the mis-pricing has a direct effects on investment decisions. In his model an increase in q raises investment above its efficient level, and monetary policy has to trade-off the cost of inefficiently high investment against the costs of a deflationary policy.

The presence of irrational mis-pricing in these models is entirely exogenous so that welfare assessments depend heavily on assumptions about the central bank having superior information (or superior rationality) with respect to the private sector. Here instead, the mis-pricing depends simply on the presence of the externality generated by the financial friction, and attempts to correct that externality can be easily interpreted. Our model is essentially a simplified version of the model in Bernanke-Gertler, which makes the welfare analysis very transparent. In this way we are able to show that, even without invoking irrational bubbles, a form of inefficient pricing may be present during an asset price boom, and authorities may be rightly concerned about an excessive level of investment. However, we also show that monetary policy is a blunt instrument to correct this type of mis-pricing, and that a contraction that reduces investment during an asset price boom is not effective in tackling the under-insurance problem which is at the root of the inefficiency.

1 A simple model of the financial accelerator

Consider an economy lasting three periods. There is one commodity that can be used for consumption or investment, and there are two groups of agents: consumers/workers and entrepreneurs. There is a large number of agents of each type, we normalize the population of each type at 1.

Consumers are risk neutral with preferences on consumption and labor represented by the utility function

$$\sum_{t=0}^{2} (c_t - v(l_t))$$

where c_t is consumption and l_t is labor. Consumers own the firms producing the only consumption good. We assume that in equilibrium c_t is positive, so that in both periods the (gross) real interest rate in this economy is 1.

Entrepreneurs have preferences represented by the utility function

$$\sum_{t=0}^{2} c_t^E$$

and they are endowed with N_0 units of the consumption good at date zero. Entrepreneurs have access to a technology that allows them to transform 1 unit of new capital at date t into one unit of capital ready for production at date $t+1$. They rent the capital stock each period to the firms producing consumption goods at the rental rate r_t.[2]

[2] Assuming that the entrepreneurs produce consumption goods directly by hiring labor at the competitive wage w_t will give identical results. The assumption made here will be useful when we introduce monopolistic competition in the production of consumption goods in order to study monetary policy with sticky prices.

The technology is described by the constant returns to scale production functions $A_t F_t(K_t, L_t)$. Used capital depreciates at rate δ.

The markets for consumption goods, capital and labor are competitive. The wage rate is w_t and the rental rate of capital is r_t. The level of total factor productivity in period 1 A_1 is the only source of uncertainty in the model, and all uncertainty is resolved in period 1. There is a discrete set S of states of the world. At date 1 the state of the world $s \in S$ is realized with probability π_s and productivity takes the value A_{1s}.

The only financial market friction we introduce is a collateral constraint on the entrepreneurs. In particular we assume that the only collateral in the economy is a fraction θ of existing old capital,

$$\theta K_t$$

with $0 < \theta \leq 1 - \delta$. Any promise of payment made by any agent in the economy must be backed by collateral. In particular, entrepreneurs will be able to borrow at date 0 only by promising repayments fully backed by collateral.

Firms behavior is standard. The behavior of firms producing consumption goods is characterized by the first order conditions

$$\begin{aligned} A_t F_{tK}(K_t, L_t) &= r_t \\ A_t F_{tL}(K_t, L_t) &= w_t. \end{aligned}$$

1.1 Equilibrium financial contracts

At date 0 entrepreneurs' net worth is equal to their intial wealth N_0. Each period entrepreneurs raise B_t in outside finance in exchange for state contingent promises of repayment at date $t+1$. In particular at date 1 entrepreneurs promise to repay $p_s K_1$ in state s. Given that consumers are risk neutral the participation constraint for outside investors is

$$B_t \leq E_t [p_{t+1}] K_1 \tag{1}$$

The payments $\{p_s\}_{s=1}^S$ are state contingent. However, they have to satisfy the collateral constraint in each state of the world:

$$p_s K_1 \leq \theta K_1 \tag{2}$$

Finally, the budget constraint for an entrepreneur at date 0 is

$$K_1 \leq N_0 + B_0 \tag{3}$$

At date 1 entrepreneurial wealth is

$$N_{1s} = r_{1s} K_1 + (1 - \delta) K_1 - p_s K_1 \tag{4}$$

Again, the amount of external funds they can raise at date 1 is limited by the collateral constraint. Given the absence of uncertainty we can summarize the effects of the financial friction at date 1 imposing the maximum leverage constraint

$$K_{2s} \leq \frac{1}{1-\theta} N_{1s}$$

In equilibrium the two conditions

$$\begin{aligned} E\left[r_1 + 1 - \delta_1\right] &\geq 1 \\ r_{2s} + 1 - \delta &\geq 1 \end{aligned}$$

will be satisfied and under these conditions the entrepreneurs' problem is well defined.

Under these condition the entrepreneur problem can be stated as follows: choose K_1 and a financial contract $\{p\}$ at date 0 that solve:

$$\begin{aligned} \max_{K_1, \{p\}} \quad & E\left[Z\left(r_1 - p + 1 - \delta\right) K_1\right)] \\ s.t. \quad & K_1 \leq N_0 + E\left[p\right] K_1 \\ & p_s \leq \theta \end{aligned}$$

where the gross rate of return on entrepreneurial wealth at date 1 is:

$$Z = \max\{\frac{r_2 + 1 - \delta - \theta}{1 - \theta}, 1\}$$

Given that the interest rate is 0 $Z - 1$ is the outside finance premium at time 1.

It is also useful to define the expected gross rate of return on entrepreneurial wealth at date 0 which is equal to

$$\lambda = \max\left\{\frac{E\left[Z\left(r_1 + 1 - \delta - p\right)\right]}{1 - E\left[p\right]}, 1\right\}$$

and corresponds to the lagrange multiplier on the first constraint in entrepreneurs' problem. With this notation optimal investment and financial structure are simply characterized by the first order conditions

$$\lambda \geq Z_s \tag{5}$$

$$\theta \geq p_s \tag{6}$$

where for each state s one of the two conditions has to hold as a strict equality.

These conditions characterize the optimal financial structure $\{p\}$. To have an economic interpretation of these conditions consider the following marginal choice between two financial strategies. The entrepreneur can employ a dollar of inside funds by buying capital at date 0 or by saving funds at the riskless rate 1 and using the receipts to buy capital at date 1. Investing in physical capital can be leveraged so investing 1 dollar of inside funds at date 0 he can purchase $\frac{1}{1-E[p]}$ units of new capital and earn the shadow rate of return λ. Alternativley he can save a dollar at date 0 and exchange it for a certain repayment in state s. He would get $\frac{1}{\pi_s}$ dollars in state s and invest them in physical capital at date 1 with an ex post shadow return of Z_s. The expected return of this investment will be $\pi_s \frac{1}{\pi_s} Z_s$, which gives the right hand side of (5). If the return on investment at date 0 is higher than the return on the second strategy then the entrepreneur finds it optimal to increase his debt in state s to its maximum amount and the collateral constraint (6) will hold as an equality. However, if the return on investment at date s is high enough then it is optimal to save collateral in state s. That is, it is optimal to protect entrepreneurial net worth from the adverse shock by choosing a level of state contingent lower than the maximum. In this case (5) holds as an equality and the collateral constraint (6) is slack.

The shadow rate of return Z governs the incentives to save entrepreneurial net worth. Entrepreneurs are willing to reduce investment at date 0 to buy insurance that pays off in states where the rate of return Z is high enough. Therefore the variability of Z induces a demand for insurance. We will return on the interpretation of condition (5) as reflecting a demand for insurance after having considered the equilibrium determinants of Z.

1.2 Equilibrium

The features of a competitive equilibrium depend on the parameters of the model, and in particular on entrepreneurs' net worth N_0 and on the fraction of collateralizable assets θ. When N_0 is large enough the unique competitive equilibrium corresponds to the first best allocation and financial constraints are not binding. The next proposition gives a characterization of the equilibria that arise in this economy.

Proposition 1 *A competitive equilibrium exists and is unique. The equilibrium is characterized by two cutoff levels A_1^I and A_1^{II}, with $A_1^I \leq A_1^{II}$, such* *that:*

1. *If* $A_{1s} > A_1^{II}$ *then* Z_s *is constant and*

$$r_{2s} + 1 - \delta = 1$$

2. *If* $A_1^I < A_{1s} < A_{1t} \leq A_1^{II}$ *then* $Z_t < Z_s < \lambda$ *and collateral is exhausted in all these states.*

3. *If* $A_{1s} \leq A_1^I$ *then* $Z_s = \lambda$ *and the collateral constraint is slack in state* s.

The characterization above is simply driven by the equilibrium relation between the level of entrepreneurial net worth and the equilibrium rate of return Z. When productivity at date 1 is high entrepreneurial net-worth is large. Then entrepreneurs can finance the first best level of investment at date 1 and the rate of return Z is equal to the interest rate. If productivity is in an intermediate range entrepreneurs are credit constrained at date 1, capital is scarcer and has higher marginal productivity, and the return Z is higher. As productivity at date 1 and profits are lower investment in the entrepreneurial sector is reduced. As entrepreneurial capital gets scarcer the rate of return r_2 and Z increase. However as long as Z is smaller than λ entrepreneurs maximize their leverage and save no collateral. When productivity falls below the level A_1^I entrepreneurial capital become so scarce and Z so large that entrepreneurs prefer to save collateral by borrowing less than the maximum amount.

Everyone in this economy is risk-neutral, however the presence of decreasing returns to capital and the presence of financial frictions induce a motive for insuring the net-worth of entrepreneurs. In the absence of financial frictions investment would be independent of entrepreneurial net-worth and the return on entrepreneurial net worth would be equal to 1 for any net worth level. When financial frictions are present, though, investment depends on entrepreneurial net worth and thus the rate return on capital is negatively related to the total net worth of the entrepreneurial sector. This generates a motive for stabilizing entrepreneurial net-worth. Entrepreneurs face a trade-off between net-worth insurance *ex post* and investment *ex ante* (at date 0). Given the limited collateral available in states where realized profits are high, in order to raise additional funds at date 0 entrepreneurs must promise part of the collateral available when profits are low. By doing so, they reduce their net worth in the low states and this increases the return Z. The degree of stabilization obtained in a competitive equilibrium depends on the slope of the relation between Z and entrepreneurial net worth. We can write this equilibrium relation as $Z = h(N_1)$. If the function h is steeper this will induce a more prudent behavior on the part of the entrepreneurs, and will result in a higher cutoff A_1^I. The slope of the function h effectively reflects the risk-aversion of the entrepreneurial sector and determines the equilibrium level of investment and insurance.

In this simple framework we assume that entrepreneurs use state contingent claims to insure their net worth against aggregate shocks. In a more realistic framework the type of contracts that can be used to stabilize the wealth of entrepreneurs can involve a variety of financial contracts: firms can buy financial assets with countercyclical payoffs, can accumulate and decumulate liquid assets, use credit lines or use informal credit arrengements with their suppliers. In this paper we abstract from the the specific instruments used to stabilize the level of inside funds and we focus on state contingent debt that most efficient form of insurance. In the next section we will analyze in detail the difference between the private and the social margins affecting the motive for net-worth insurance.

2 The credit boom: under-insurance and over-investment

We now turn to the efficiency properties of the competitive equilibrium. First of all we can compare the equilibrium to the allocation arising in an economy with no collateral constraints. This is our first best benchmark. The first best equilibrium is characterized by prices and quantities such that r_{2s}^* and K_{2s}^* are constant across states and satisfy the relations:

$$\begin{aligned} r_{t,s}^* + 1 - \delta &= 1 \text{ for all } (t,s) \\ E\left[r_1^* + 1 - \delta_s\right] &= 1 \end{aligned}$$

Neither investment nor the capital stock depend on the productivity shock at time 1, given that this shock carries no information on future profitability. This means that the dependance of K_2 on the productivity shock in this model is *only* due to the presence of financial frictions. The following proposition summarizes the comparison between the first best and the competitive economy.

Proposition 2 *Compared to the frictionless first best a competitive equilibrium displays: (1)* $K_1 \leq K_1^*$ *(underinvestment at date 0); (2)* $K_{2s} < K_2^*$ *if* $A_{1s} < A_1^{II}$ *(underinvestment at date 1); (3) persistent effects of temporary productivity shock.*

Note that in equilibrium it is always the case that

$$\begin{aligned} Z &\geq 1 \\ \lambda &\geq 1 \end{aligned}$$

so $\lambda = 1$ implies $Z_s = 1$ and $E[r_1 + 1 - \delta_s] = 1$, that is first best investment both at time 0 and 1. Clearly, our interest will focus on economies with scarce entrepreneurial collateral, where $\lambda > 1$ and where $Z_s > 1$ in some states s.

Proposition 2 makes it clear that from a first best point of view the model can only display underinvestment. If the financial constraint is binding the amount of outside funds that can be raised by entrepreneurs is limited and the rate of return on investment is higher than the interest rate 1. If the financial constraint is not binding the rate of return on investment is 1, the outside-finance premium is zero and we achieve the first best level of investment.

Let us now turn to a second best analysis. In particular we want to study wether, taking as given the financial frictions assumed in the model, entrepreneurs' private choice of investment and financing (K_1 and $\{p_s\}$) at date 0 can be modified so as to induce a Pareto improvement. We allow for compensating transfers between entrepreneurs and the consumers at date 0. The transfers are made to reallocate the welfare gains *ex ante*. However a transfer from entrepreneurs to consumers reduces entrepreneurial wealth, and reduces the amount of external finance entrepreneurs can borrow. So the financial constraints are not relaxed by allowing compensating transfers at date zero. The definition of constrained efficiency we use is analogous to the definition used in the literature on general equilibrium with incomplete markets.[3] In numerical examples we will eliminate side transfers and ask wether a unilateral change in the financial contract by entrepreneurs can induce a Pareto improvement. This second approach may be more informative if one thinks about the effect of regulatory requirements on financial variables.

[3]See Geanakoplos and Polemarchakis (1986) , and Allen and Gale (2001) for a recent application of constrained efficiency to a setting with financial frictions.

In order to define a constrained optimal allocation let us introduce the compensated welfare of entrepreneurs as a function of the promised payments $\{p\}$:

$$\begin{aligned} W(p) &= E\left[ZN_1\right] \\ N_{1s} &= \frac{r_{1s}+1-\delta-p_s}{1-E\left[p\right]}\left(N_0-z\right) \end{aligned}$$

where the prices satisfy the equilibrium relation:

$$(\mathbf{r},\mathbf{w})=\phi(p) \tag{7}$$

and z is the transfer required to make the consumers as well off as in the competitive allocation, and is a function of equilibrium prices according to:

$$z=\psi(\mathbf{r},\mathbf{w}) \tag{8}$$

Notice that the transfer z reduces the resources available to entrepreneurs at date 0 and must be financed with inside funds N_0. As we noticed above, the problem is set up so that the presence of the transfer does not change the nature of the financial constraint. The map ϕ gives the equilibrium prices as a function of the entrepreneur choices at date 0, the map ψ gives the expected utility of consumers at the new prices minus the same expected utility at the competitive equilibrium

$$z=U_C^{CE}-E\left[\sum_{t=0}^{2}\left(c_t-v(l_t)\right)\right].$$

A financial contract $\{p\}$ is defined to be *constrained efficient* if it is a solution to the problem

$$\begin{aligned} \max_{p} \quad & W(p) \\ s.t. \quad & p_s\leq\theta \\ & (7),(8) \end{aligned} \tag{9}$$

It is worth noticing that the notion of constrained efficiency is different from a criterion simply based on a measure of the total surplus $\sum^t\left[AF(K_t,L_t)-v(L_t)\right]$. The model is inherently characterized by heterogeneity, therefore a welfare criterion based only on total surplus can be misleading. In models characterized by borrowing constraints an increase in the net worth of the entrepreneurs has a positive effect on total surplus, but at the same time it has distributional side effects. Here we consider proper Pareto improvements and we require all transfers between consumers and entrepreneurs at date 0 to satisfy the given set of financial constraints: if entrepreneurs have to compensate consumers for changes in future utility they have to use their scarce internal funds. In particular the planner cannot tax the non-pledgeable part of asset returns to repay consumers in date 1 or 2. In this sense the planner in this model is not allowed to create "public liquidity" in the sense of Holmstrom and Tirole.

The following lemma provides the main step for constrained efficiency analysis.

Lemma 3 *At a competitive equilibrium the marginal effect of p_s on entrepreneurs' compensated welfare is given by the expression*

$$\frac{\partial W}{\partial p_s} = \pi_s \left(\lambda - Z_s\right) K_1 + E\left[\left(\lambda - Z\right) L_1 \frac{\partial w_1}{\partial p_s}\right] + E\left[\left(\lambda - 1\right) L_2 \frac{\partial w_2}{\partial p_s}\right] \tag{10}$$

where the derivatives $\frac{\partial w_u}{\partial p_s}$ are obtained differentiating the equilibrium map ϕ.

The first term in the expression above represent the direct effect on entrepreneurs' utility as it appears in entrepreneurs' first order condition (5). The next two terms are due to the general equilibrium effects of entrepreneurial wealth on equilibrium prices. A change in p_s affects the capital stock at times 1 and 2. These changes in turn affect equilibrium wages w_t. The distributional effect of a wage change is a transfer of resources from entrepreneurs to consumers equal to $L_t dw_t$. If these transfers were internalized in individual financial contracts consumers would be willing to pay $E\left[L_t dw_t\right]$ for them at date 0, and this would affect available resources for investment at date 0. The shadow marginal benefit of these funds to entrepreneurs is given by λ. The shadow marginal cost of this transfer depends on the return on inside funds which is Z_u at date 1 in state u and it is equal to 1 at date 2.

Notice, first of all, that if $\lambda - Z_s > 0$ and $\frac{\partial W}{\partial p_s} > 0$ in all states then social and private incentives are aligned. It is optimal, both from a social and from a private point of view to fully exhaust collateral in all states of the world. Essentially, the average returns in period 1 are so large that it is optimal to borrow up to the limit in all possible circumstances in order to maximize investment at date zero.

Excessive borrowing arises when in some states of the world $\lambda - Z_s \geq 0$ and $\frac{\partial W}{\partial p_s} < 0$. In this case it would be desirable from a social point of view to save collateral in state s but private incentives to save collateral are too small. The analysis of the expression $\frac{\partial W}{\partial p_s}$ is in general complicated by the presence of the endogenous transfer z and of the associated wealth effects on equilibrium prices. However, assuming a linear production function at date 1 we can get an expression for $\frac{\partial W}{\partial p_s} < 0$ that is easy to analyze. In this case we obtain a simple characterization of economies with excess fragility, with a straightforward economic interpretation.

Assumption I The production function at date 1 is linear $F_1(K, L) = K + bL$

With this assumption the second term in (10) disappears and Lemma **??** in the appendix shows that (10) can be rewritten as

$$\frac{\partial W}{\partial p_s} = \pi_s \left(\lambda - Z_s\right) K_1 + \pi_s \frac{\lambda - 1}{1 - \tilde{\lambda}} \left(\tilde{\lambda} - \tilde{Z}_s\right) K_1 \tag{11}$$

where the variables $\tilde{Z}$ and $\tilde{\lambda}$ are

$$\begin{aligned} \tilde{Z}_s &= L_{2s} \frac{dw_{2s}}{dK_{2s}} \frac{1}{1-\theta} \\ \tilde{\lambda} &= \frac{E\left[\tilde{Z}\left(r_1 + 1 - \delta - p\right)\right]}{1 - E\left[p\right]} \end{aligned}$$

and where the derivative $\frac{dw_{2s}}{dK_{2s}}$ is obtained from the equilibrium relation between the capital stock and the wage rate in the spot market at date 2.

Notice the symmetry between the two terms in $\frac{\partial W}{\partial p_s}$. We can interpret the second term as expressing an additional trade-off between insurance and investment. However, the marginal return on entrepreneurial wealth in different states of the world is not determined by the rate of return on inside funds Z but by $\tilde{Z}$. The difference between Z and $\tilde{Z}$ is that the first captures the private return on an extra dollar of net worth N_1, while the second captures the effect of an extra dollar of net worth on wages. An increase in entrepreneurial net worth affects total capital and total production, however entrepreneurs only capture the private portion of this effect and this can generate a gap between social and private incentives to save on collateral.

Let us introduce a function $\tilde{h}$, analogous to the function h introduced in section 1.2, that associates to the net worth N_1 the corresponding level of the external effect $\tilde{Z}$. The slope of h captures the risk aversion of the private sector, while an average of h and $\tilde{h}$ captures risk aversion according to a social welfare function. The possibility of overinvestment and underinsurance is linked to the relative slopes of these two functions. In particular when $\tilde{h}$ is steeper than h the entrepreneurs underestimate the cost of a negative shock to net worth. The relative slope of h and $\tilde{h}$ depends on the form of the production function F_2, and therefore the distortion can go in both directions.

A simple case to analyze is the case of a Cobb-Douglas production function at date 2,

Assumption II The production function at date 2 is Cobb-Douglas $F_2(K, L) = K^{\alpha} L^{1-\alpha}$

Under Assumptions I and II we can prove the following proposition.

Proposition 4 *Consider an economy with F_1 linear, F_2 Cobb-Douglas and $\delta > 0$. Suppose that in a competitive equilibrium $\lambda = Z_s$ for some state s. Then the competitive equilibrium is constrained inefficient. There is a $\tilde{A} > A_1^I$ such that*

1. *If $A_{1s} < \tilde{A}$ then $\frac{\partial W}{\partial p_s} < 0$*
2. *If $A_{1s} \geq \tilde{A}$ then $\frac{\partial W}{\partial p_s} \geq 0$*

In this case the competitive equilibrium displays the following inefficiencies: the cutoff A_1^I is too low, the promised repayments p_s in the low productivity states are too large, investment at date 0 is too high.

It is useful to highlight the circumstances under which the inefficiency arises. Two conditions tend to generate overinvestment according to Proposition 4. We can rewrite λ, as

$$\lambda = E\left[Z \frac{r_1 + 1 - \delta - p}{E\left[r_1 + 1 - \delta - p\right]}\right] \left(\frac{E\left[r_1 + 1 - \delta - p\right]}{1 - E\left[p\right]}\right)$$

The first term of this expression is a weighted average of the Z's, the second term is a measure of the return on inside funds at date 0. If the second term of this expression is equal to 1 then we will have Z constant and $\lambda = Z$ in every state, and Proposition 4 applies. If the second term is greater than 1, if the variability of Z is large enough we will still have states of the world in which $\lambda = Z$ holds and Proposition 4 applies. These two conditions can be interpreted as follows: inefficiency arises when investment at date 0 is close to first best and when there is enough variability of investment at date 1. These investment levels are endogenous and are driven by the productivity levels $E[A_1]$ and A_2, and by the variability of A_1.

3 Some policy implications

The inefficiency highlighted in the previous section can be attacked with a number of possible instruments. Here, we will focus on the activity of the central bank and discuss implications of the model for prudential regulation and for monetary policy.

It is immediate to interpret the model above as a rationale for capital requirements. Regulatory interventions that impose minimum capitalization to financial firms are widespread in industrialized economies, and often their introduction is justified on the basis of the idea that competition in the financial sector may bring about excessive fragility. The model presented gives a simple framework that rationalizes this idea.

Consider a capital requirement at date 0 of the type

$$\frac{N_0}{K_1} \geq \nu$$

that imposes a limit on the leverage of firms at time 0. The presence of this constraints effectively reduces the rate of return on investment at date 0, λ, by increasing the cost of capital. This tilts the trade-off in favor of insurance, expands the set of states of the world against which entrepreneurs purchase net-worth insurance and increases net worth in these states. The next proposition shows that an appropriate capital requirement implements a constrained efficient allocation.

Proposition 5 *Suppose the financial contract $\{p\}$ and the transfer z solve the problem (9) and consider the associated, constrained efficient, allocation. Let*

$$\nu^* = \frac{N_0 - z}{K_1^*}$$

Then a competitive equilibrium with the transfer z, subject to the capital requirement ν^*, supports the same constrained efficient allocation.

In reality, capital requirements are imposed on a specific class of firms, namely on financial firms and banks in particular. To have a fully fledged theory of capital requirements the model above needs to be integrated into a model of financial intermediation. If financial intermediaries specialize in the provision of contingent credit lines and other forms of net-worth insurance we can expect that the stabilization of their net-worth is instrumental in providing net-worth insurance to the non-financial corporate sector. Also, if entrepreneurial firms which rely more on outside funding are more dependent on bank credit, capital requirements on banks could indirectly help to stabilize the balance sheet of these firms.

A second type of intervention that can be studied in our framework is a direct intervention that transfers resources to distressed firms at date 1. This type of intervention is essentially useless in our model. If the private incentives to save collateral are unchanged private contracts would exactly undo whatever government transfers do. More precisely suppose the government can induce a non-distortionary transfer of resources $t_s K_1$ from investors to entrepreneurs at date 1, and suppose that these transfers are compensated by an equivalent transfer $E[t]K_1$ from entrepreneurs to investors at date 0. Proposition 6 shows that the competitive equilibrium is unchanged.

Proposition 6 *Suppose the government implements a scheme of state contingent transfers from investors to entrepreneurs $\{t_s K_1\}_s$ associated to an ex ante compensating transfer $-E[t]K_1$. Then the equilibrium prices and quantities are unchanged and the equilibrium financial contract at date 0 is*

$$\hat{p} = p + t$$

Finally, one can consider the effects of monetary policy. Monetary policy can be introduced in a number of different ways, here I will concentrate on two possible effects that arise in a simple sticky price version of the model above. The sticky price model is sketched in Appendix II. First, one could take our model of overinvestment to justify a contractionary monetary policy that attempts to control the credit boom at date 0. In a sticky price environment a contractionary monetary policy at date 0 would affect investment through a simple cash-flow channel.[4] So far we have assumed that N_0 was given, it is easy to add an initial condition for the capital stock K_0 and for corporate debt p_0 and have a relation that determines the internal funds at date 0

$$N_0 = r_0 K_0 + (1 - \delta) K_0 - p_0 K_0.$$

A monetary contraction would have a negative effect on r_0 and thus on N_0 and reduce investment through this channel. This effect however is completely useless from the point of view of restoring constrained efficiency. Even though this policy achieves a reduction in investment this policy is not attacking the source of the inefficiency, that is in the composition of investment finance and is reflected in the excess fragility.

Looking at the numerical example presented in the previous section we can clearly see that a reduction in N_0 is associated to a reduction of N_1 in all states of the world, and not to a stabilization of N_1 across states.

A very different type of monetary intervention that could have more desirable effects in the model above is a state contingent monetary policy at date 2. In particular an expansionary policy in bad states of the world that has a positive effect on r_2 would boost the profits of those firms that have maintained a sufficient level of capitalization. This policy would increase Z in the bad states and have favorable effects on ex ante incentives for net-worth stabilization. We can think of this policy as a state contingent subsidy to the rental rate on capital r_2. Suppose that the constrained-efficient problem is well behaved and the first order conditions

$$\begin{aligned} \frac{\partial W}{\partial p_s} &\geq 0 \\ p_s &\leq \theta \end{aligned}$$

characterize a second best allocation.[5] In this case an appropriate system of subsidies to capital income in bad states of the world, together with appropriate transfers at date 0, can implement the second best efficient allocation.

A monetary policy that increases output above its "natural" level in bad states at date 2, has also distortionary effects on output and on relative prices,

[4] Given our simple environment with linear utility, monetary policy would have no effects on the real interest rate.

[5] So far we have only looked at inefficiency results, and we have not characterized second best efficient outcomes, that is why we needed no assumptions about the concavity of $W(p)$.

so the positive incentives effects just discussed would have to be balanced against these distortionary effects. Notice also that the notion of "natural" output is not obvious in our setup. If one thinks of natural output as the equilibrium level of output associated to "healthy" balance sheets of the entrepreneurial sector, then the monetary policy just described takes the form of a standard form of output stabilization.

The general message one gets from our model is that a monetary policy that tries to quench the credit boom by stifling investment ex ante is very ineffective at restoring efficiency, while a monetary policy that tends to stabilize output during credit crunches ex post may also have favorable incentive effects ex ante.

Even though our model displays overinvestment, overinvestment is just a symptom of an overly fragile financial structure, that is, of excess liabilities in bad states of the world. A policy that reduces investment ex ante has no bite on the composition of investment finance, it reduces the level of entrepreneurial net-worth in all future states and has no stabilizing effects. In order to attack the inefficiency in our model the policy maker has to resort either to regulatory interventions or to well announced state contingent interventions ex post. The latter should not be oriented to restore the capitalization of troubled firms by injecting funds into them, rather they should favor those firms that have protected their balance sheets by increasing the ex post return on capital.

4 Conclusions

Asset prices

Investment composition

Persistent shocks to A_2

Changing θ (financial liberalizations)

References

[1] F. Allen and D. Gale

[2] P. Gourinchas, R. Valdes and O. Landerretche, 2001, "Lending Booms: Latin America and the World," NBER Working Paper 8249.

[3] N. Kiyotaki and J. Moore

[4] P. Aghion, A. Banerjee and T. Piketty (1999) Dualism and Macroeconomic Volatility, *Quarterly Journal of Economics*.

5 Appendix I

Proof of Proposition 1

Suppose $\mu_h = \mu_l = \mu$ in all two states $l < h$. Then we have $r_{2l} = r_{2h} = \lambda$ which means that K_{2s} is the same in both states and so is q_{2s}. From the entrepreneur budget constraint at time 1 we derive that $c_{1h} > 0$ which implies first best asset prices in the high state

$$\frac{r_{2h}}{q_{1h}} = 1$$

which in turns implies first best investment and first best asset pricing in the state l. If for all states $\mu_s = 0$ this also implies $\lambda = 1^2$. From the latter we can rewrite the first order condition **??** as

$$E\left[r_1 - p + (1-\delta)\, q_1\right] = 1 \left(q_0 - \frac{E\left[p\right]}{1} \right)$$

which immediately gives the first best pricing equation at time 0

$$\frac{E\left[r_1 + (1-\delta)\, q_1\right]}{q_0} = 1.$$

The remaining characterization of the equilibria is immediate, except to notice that $\mu_l > \mu_h$ cannot arise in equilibrium if $l < h$. To show this just observe that as $\mu_h \geq 0$, if we had $\mu_l > \mu_h \geq 0$ we would have, from **??**, $\frac{r_{2l}}{q_{2l}} < \frac{r_{2h}}{q_{2h}}$ which implies $K_{2l} > K_{2h}$ and $q_{1l} > q_{1h}$. But the binding constraint in state l, together with $A_{1l} < A_{1h}$ would imply the chain of inequalities

$$K_{2l} \leq \frac{r_{1l}}{q_{1l}} + (1-\theta)(1-\delta) \frac{\hat{q}_{1l}}{q_{1l}} < \frac{r_{1h}}{q_{1h}} + (1-\theta)(1-\delta) \frac{\hat{q}_{1h}}{q_{1h}} = K_{2h}$$

which gives a contradiction.

Derivation of condition (5)

$$\begin{aligned}\frac{dN_{1u}}{dp_s} &= \pi_s \frac{r_{1u} + 1 - \delta - p_u}{1 - Ep} K_1 \text{ for } u \neq s \\ \frac{dN_{1s}}{dp_s} &= \pi_s \frac{r_{1s} + 1 - \delta - p_s}{1 - Ep} K_1 - K_1\end{aligned}$$

$$E\left[Z \frac{dN_1}{dp_s}\right] = \pi_s E\left[Z \frac{r_1 + 1 - \delta - p}{1 - Ep}\right] K_1 - \pi_s Z_s K_1$$

Proof of Lemma 3. Because of constant returns to scale the equilibrium price changes dr_t and dw_t are related by

$$L_t dw_t + K_t dr_t = 0$$

the effect on z is

$$dz = E\left[L_1 dw_1\right] + E\left[L_2 dw_2\right]$$

and given that the lagrange multiplier on the constraint is λ the marginal effect through z is

$$\lambda E\left[L_1 dw_1\right] + E\left[L_2 dw_2\right]$$

The effects

$$\begin{aligned}dN_1 &= K_1 dr_1 = -L_1 dw_1 \\ dc_2 &= K_2 dr_2 = -L_2 dw_2\end{aligned}$$

have an effect equal to

$$-E\left[Z L_1 dw_1\right] - E\left[L_2 dw_2\right]$$

summing up the effects above gives us the second and third term in (10).

Proof of Lemma ??. The promised repayment p_s affects wages through two channels: it directly affects wages in state s by reducing investment in state s, it indirectly affects wages in all states u by affecting the leverage at date 0 which determines K_1. We can write

$$L_{2u} \frac{\partial w_{2u}}{\partial p_s} = L_{2u} \frac{dw_{2u}}{dK_{2u}} \frac{dK_{2u}}{dN_{1u}} \frac{\partial N_{1u}}{\partial p_s}$$

and derive

$$
\begin{aligned}
\frac{\partial N_{1u}}{\partial p_s} &= \pi_s \frac{r_{1u} + \frac{dr_1}{dK_1} + 1 - \delta - p_u}{1 - E\left[p\right]} K_1 + \frac{r_{1u} + \frac{dr_1}{dK_1} + 1 - \delta - p_u}{1 - E\left[p\right]} \frac{\partial z}{\partial p_s} \text{if } u \neq s \\
\frac{\partial N_{1u}}{\partial p_s} &= \pi_s \frac{r_{1u} + \frac{dr_1}{dK_1} + 1 - \delta - p_u}{1 - E\left[p\right]} K_1 - K_1 + \frac{r_{1u} + \frac{dr_1}{dK_1} + 1 - \delta - p_u}{1 - E\left[p\right]} \frac{\partial z}{\partial p_s} \text{ if } u = s
\end{aligned}
$$

taking expectations we obtain

$$E\left[L_{2u}\frac{\partial w_{2u}}{\partial p_s}\right] = \pi_s\left(\tilde{\lambda} - \tilde{Z}\right)K_1 + \tilde{\lambda}\frac{\partial z}{\partial p_s}$$

$$L_{1u}\frac{\partial w_{1u}}{\partial p_s} = L_{1u}\frac{dw_{1u}}{dK_1}\frac{\partial K_1}{\partial p_s}$$

$$\frac{\partial K_1}{\partial p_s} = \pi_s \frac{1}{1 - E\left[p\right]} K_1 + \frac{1}{1 - E\left[p\right]} \frac{\partial z}{\partial p_s}$$

The effect on the transfer z satisfies

$$\frac{\partial z}{\partial p_s} = E\left[L_{2u}\frac{\partial w_{2u}}{\partial p_s}\right] = \pi_s\left(\tilde{\lambda} - \tilde{Z}\right)K_1 + \tilde{\lambda}\frac{\partial z}{\partial p_s}$$

From this we immediately obtain expression ....

6 Appendix II (monetary model)

Here is a sketch of the monetary model used to discuss the effects of monetary policy in section 3. There is a continuum of intermediate goods which give the final (consumption) good according to the CES aggregator:

$$x = \left(\int_0^1 x_j^{\frac{\sigma-1}{\sigma}} dj\right)^{\frac{\sigma}{\sigma-1}}$$

From consumer optimization we have $x = \frac{y}{p}$, where y is nominal income and $p = \left(\int p_j^{1-\sigma} dj\right)^{\frac{1}{1-\sigma}}$. There is a continuum of monopolistic competitive firms. Firms' technology is described by the production function

$$AF(K, L)$$

Assume prices fixed at date 0, then at dates 1 and 2 a fraction ρ of firms adjust their prices. Factor prices w and r are assumed to be flexible. We have inverse demand faced by firm j

$$\frac{p_j}{p} = \left(\frac{x_j}{x}\right)^{-\frac{1}{\sigma}}$$

Profits are given by

$$\max_{x_j} (1+s)\left(\frac{x_j}{x}\right)^{-\frac{1}{\sigma}} x_j - c(r,w)x_j$$

where the linear cost function is derived by standard cost minimization.

For flex price firms we have:

$$(1+s)\left(1-\frac{1}{\sigma}\right)\frac{p_j}{p} = c(r,w)$$

Because of constant returns to scale for all firms we have the same optimal factor proportion $\frac{K}{L}$ as a function of the relative price $\frac{r}{w}$.

The subsidy s is set to eliminate the distortion due to monopoly pricing

$$(1+s)\left(1-\frac{1}{\sigma}\right) = 1$$

Since $x_j = AF(K_j, L_j) = AF(K,L)\frac{L_j}{L}$ and $\frac{L_j}{L} = \frac{x_j}{x} = \left(\frac{p_j}{p}\right)^{-\sigma}$ we get total final goods output

$$\begin{aligned} x &= \left[\left[\int \left(\frac{p_j}{p}\right)^{1-\sigma} dj\right]^{\frac{1}{1-\sigma}}\right]^{-\sigma} AF(K,L) = \\ &= vAK^{\alpha}L^{1-\alpha} \end{aligned}$$

In this setup inflation has two costs:

(1) it distorts

$$c(r,w) \neq 1$$

(2) it reduces output in terms of consumption goods because of the variance term v above, which is $v < 1$ whenever p_j is not constant across sectors.

Consider the special case of an economy where ρ is close to 0. In this case only the first distortion is present.

In this case we can think of monetary policy as choosing an L which satisfies

$$w = L^{\eta}$$

and then determining the schedule for r:

$$r = \frac{\alpha}{1-\alpha}\frac{L}{K}L^{\eta}$$

the problem is that here it is no longer true that

$$Ldw + Kdr = 0$$

because there is a profit whenever $\frac{p_j}{p} \neq 1$.
profits are

$$\begin{aligned}
\Pi &= \left(\frac{\sigma}{\sigma-1} - \frac{r^{\alpha}w^{1-\alpha}}{A\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\right)x = \\
&= (1+s)\,AK^{\alpha}L^{1-\alpha} - wL - rK = \\
&= (1+s)\,AK^{\alpha}L^{1-\alpha} - L^{1+\eta} - \frac{\alpha}{1-\alpha}L^{1+\eta}
\end{aligned}$$

$$\frac{d\Pi}{dL} = (1-\alpha)\,AK^{\alpha}L^{-\alpha} - (1+\eta)\,\frac{1}{1-\alpha}L^{\eta}$$

Simple solution: assume that workers own final good firms and receive the lump-sum subsidy sY and just look at the effect of policy that sets ψ on

$$\frac{d\,(rK)}{d\psi}, \frac{d(Y)}{d\psi} - \frac{d\,(rK)}{d\psi}$$

this amounts to assume that all that matters for the economy from a distributional viewpoint is the distribution between entrepreneurs and the rest of the economy.

Choose $\psi = \langle L \rangle$ then look at effect on Y and rK.

In this framework, monetary policy is essentially a policy that distorts the labor supply margin L and in this way it distorts r and change the incentives for capital accumulation in the entrepreneurial sector.

Suppose K is fixed then mon.pol. sets L and you get $Y = AK^{\alpha}L^{1-\alpha}$ and

$$\begin{aligned}
\frac{dY}{dL} &= (1-\alpha)\, AK^{\alpha}L^{-\alpha} \\
\frac{d(rK)}{dL} &= (1+\eta)\,\frac{\alpha}{1-\alpha}L^{\eta} \\
\frac{d(wL)}{dL} &= (1+\eta)\, L^{\eta} \\
\frac{d(\Pi)}{dL} &= (1-\alpha)\, AK^{\alpha}L^{-\alpha} - (1+\eta)\,\frac{1}{1-\alpha}L^{\eta} \\
\frac{dY}{dL} &= \frac{d(rK)}{dL} + \frac{d(wL)}{dL} + \frac{d(\Pi)}{dL}
\end{aligned}$$

notice that $\frac{d(rK)}{dL} = (1+\eta)\,\frac{\alpha}{1-\alpha}L^{\eta}$ is valid also when K is endogenous, but the expression for $\frac{dY}{dL}$ needs to account for changes in K.