Can Market Failure Cause Political Failure?

Madhav Aney, Maitreesh Ghatak, and Massimo Morelli *

April 28, 2009

Abstract

How can a market failure interact with the choice of institutional reform made by an electorate? We study this question in an occupational choice framework, where agents are endowed heterogeneously with wealth and talent. In our model, market failure due to unobservability of talent endogenously creates a class structure that affects voting on institutional reform. We find that the preferences of these classes are often aligned in ways that creates a tension between institutional reforms that are growth maximising and those that are politically feasible. This is in contrast to the world without market failure where the electorate unanimously vote in favour of surplus maximising institutional reform. Inefficiencies of market failure may be further amplified by political choices made by interest groups created in the inefficient market.

Keywords: occupational choice, adverse selection, property rights, asset liquidation, political failure, market failure.

JEL Classification: O12, O16, O17

1 Introduction

It is well known that market failures abound in the real world. A key insight in the institutional approach to development economics is that capital market failures prevent individuals and economies from reaching their full potential and can lead to poverty traps (see Banerjee and Newman (1993) and Galor and Zeira (1993)). In this literature institutional frictions are taken as exogenous.[1]

*We would like to thank James Peck, Tomas Sjöström and seminar participants at the Midwest Mathematical Economics and Theory Conference, North Eastern Universities Development Consortium Conference, and the EOPP Workshop for their helpful feedback.

[1] See Banerjee (2001) for a survey of this literature.

It is also well known that even fully accountable governments can fail to implement growth maximising policies when they lack sufficient instruments for compensating losers. Furthermore, the political economy approach to development has emphasized how concentration of political power in the hands of an elite, may lead to distortion of the market by the elites for maximising their own payoffs.[2] This strand within the political economy literature makes the argument that the distribution of political power may be sufficiently skewed so as to allow the elites to distort the market outcome in their favour, and this typically leads to inefficiencies.

In this paper we highlight the reverse link, namely that market failure may create a political failure even when political power is uniformly distributed. We think of political failure as the failure of the electorate to pick the surplus maximising reform.[3] In our model, in the first best world with well functioning markets, the electorate unanimously chooses institutions that maximise total surplus. However once a market imperfection in the form of unobservability of entrepreneurial talent is introduced, things change dramatically. The competitive market responds to this imperfection by screening agents based on their wealth. This leads to creation of a class structure in the economy with preferences that are aligned in ways that defeat surplus maximising reforms. In a nutshell, the motivation for this paper is to uncover the political implications of market failure.

There is an important distinction between our approach and the existing literature on political economy. Instead of taking political classes or interest groups as exogenous and studying the impact of their alignment on markets, we derive them from economic fundamentals, namely, the nature of technology, and the informational environment in the economy. In this regard, the mechanism that our paper identifies fits into a theme present in both Marxist and Neo-Classical theories of institutions that use economic forces as the base over which the political superstructure is built.[4]

We argue that in addition to the well known impact of market failure articulated in the literature on poverty traps, there may also be a political impact. The latter problem could turn out to be more persistent since unlike the solutions to

[2] This is most obvious when elites lobby for barriers to entry (Djankov et al. (2002)). Acemoglu (2003) makes the argument that concentration of political power may lead to distortion of the market through manipulation of factor prices in ways that benefit the political elites.

[3] For a discussion on somewhat different notions of political failure see Besley (2006).

[4] See chapter 1 in Bardhan (1989) for a review of the common themes in these literatures concerning the theory of institutions.

poverty traps that are easier to characterise[5], the solutions to political failure that are politically feasible may not exist. A more general message emerging from our model is that market and political failures complement each other in terms of generating economic inefficiencies.

Our paper is related to the growing literature on micro political economy. This literature looks at failure of alternative institutions and asks two questions:

1. Which institutions make an economy more productive?

2. Which institutions are more likely to be chosen given a certain distribution of political power?

We now present a review of papers that ask similar questions. Boyer and Laffont (1999) present a model where a monopolist produces a socially valuable good and some amount of pollution as a byproduct. The regulator has a choice of several instruments that can be used to make transfers to the monopolist. The incentives of the electorate may not be aligned with those of a total surplus maximising regulator since the electorate is composed of voters of whom a certain proportion are also shareholders in the monopoly. This can lead to non surplus maximising policies being chosen, regardless of information asymmetries.

Perotti and Volpin (2004) have a model where agents are endowed with wealth and are either consumers or entrepreneurs. There is a non convexity in the production function and entrepreneurs with wealth lower than a certain threshold are financed by equity. Project returns are subject to the ex post moral hazard problem and investor protection, the institution that they study, can mitigate the problem. Elites that have wealth over the threshold required to start an enterprise, lobby for lower investor protection so as to face a lower competition in the product market. The political economy process is modelled as a social planner that maxmises the weighted sum of the total surplus and bribes from lobbies. As the weight on the bribes increases, investor protection goes down.

Rajan and Zingales (2006) study a model evaluating the incentives of the educated and non-educated class to pass educational and pro market reform. Educational reforms allow the uneducated to become educated and increase their wages through an increase in their productivity. Pro market reforms allow educated workers to setup their own firms. An agent's preference for any reform is driven by which group the agent belongs to.

[5]Micro-lending has been a big theme in this literature. See for example Ghatak and Guinnane (1999).

Biais and Mariotti (2003) address the question of optimal bankruptcy laws. They have a model of occupational choice where agents can be entrepreneurs or workers. Credit market is imperfect because entrepreneurial effort is unobservable. The mechanism through which bankruptcy law affects total surplus is the following: a tough bankruptcy law implies a strong threat of liquidation ex-ante. This induces high effort which increases surplus. However liquidation is ex-post inefficient since some surplus is lost when a company is harvested for its assets at liquidation. In terms of the political economy aspects, the rich want soft laws to induce lower wages. The poor want the opposite. The agents with intermediate wealth align with rich if they are entrepreneurs and align with poor otherwise. This paper is similar to ours in the sense that here too a market failure generates the need for institutions. The paper differs from ours in terms of the result they find on the choices an electorate make. In their model soft laws which are often chosen by the electorate are often efficient due to inefficiency of liquidation ex post. In contrast, our results indicate that their exists an inherent tension between politically feasible and surplus maximising reforms.

Another paper that is related to ours is Caselli and Gennaioli (2008). In their model agents differ in two discrete dimensions; talent and license. There is an exogenous mismatch between talent to run an enterprise and the endowment of license that is required to run an enterprise. They model how this exogenously conferred incumbency and talent interact to create preferences for deregulation and legal reform. Deregulation lowers the cost of acquiring a new license whereas legal reform makes the trade of licenses between agents easier.

In these models, markets can be complete and perfectly competitive if the best possible institutions are chosen. In absence of such institutions, frictions are created that take the economy away from the growth maximising outcome. The source of problems in these models is purely the exogenous presence of political alignments that undermine the support for best possible institutions. In our model on the other hand these political alignments are endogenised and the fundamental source of inefficiency will be the adverse selection problem created by the unobservability of entrepreneurial talent. Institutions, depending on their quality, would mitigate or worsen this problem.

In our model, even with fully benevolent government and perfectly competitive markets, there are market frictions arising from informational (i.e., adverse selection) and transactional constraints (limited liability). As in the standard neoclassical model, preference and technology differences might have seemingly similar implications: e.g., in the Solow model, low steady state output could result from lower saving propensity or use of less efficient technology. However, the policy implications are dramatically different: preference differences are more intractable than technology differences and this is especially so if we recognize the potential mutual interaction of preferences and technology adoption which, for example, reflects some underlying market failure. Analogously, we argue that with government frictions the policy implications are to be found in the political domain and are relatively easy to characterize which is not to say they are easy to implement: improve political institutions to improve the quality of candidates, improve incentives for incumbents so that inefficient rent-extracting policies are removed. In contrast, with market frictions the policies are far less easy to characterize, and this is especially so if they interact with an otherwise frictionless political system where the distribution of political power is uniform.

2 Model

The basic setup extends the model presented in Ghatak et al. (2007).

2.1 Technology

There are two technologies in the economy: a subsistence technology that yields $\underline{w}$ with certainty for one unit of labour and a more productive technology y that yields a return R in case of success and 0 in case of failure and requires n workers and 1 entrepreneur to run it.

2.2 Preferences

All agents are assumed to be risk neutral with a utility function that is additively separable in effort and money. The net disutility of labour effort relative to entrepreneurial effort is normalised to M. This can also include any perks that entrepreneurs enjoy relative to workers such as the comfort of sitting in an air conditioned office, or the psychological payoff from not having a boss.

2.3 Endowments

Agents are endowed with one unit of labour, entrepreneurial talent and illiquid wealth. Talent θ of an agent is the probability of success of the more productive technology if she becomes an entrepreneur. θ is distributed with a cdf $F(\theta)$.

Agents are endowed with illiquid wealth a with a distribution $G(a)$. We assume that the distributions of wealth and talent are independent.

2.4 Informational and Institutional Frictions

The entrepreneurial ability θ can be either observable or unobservable. In the first best world θ is observable and the first welfare theorem operates ensuring that the competitive equilibrium is Pareto efficient. In contrast when θ is unobservable, a market failure arises. The illiquid wealth a, and output y, are verifiable. M is also verifiable but is not appropriable since it is the psychological net benefit of being an entrepreneur.

The 2 institutional parameters in the model are ϕ and τ. ϕ is the proportion of collateral that is recovered from a borrower when she defaults. This can be thought of as the strength of judicial enforcement of contracts. τ is the probability with which the wealth a is expropriated. The efficiency of both these institutions affect the credit contract that an agent is offered in the second best world as the credit market takes into account the efficiency of the judiciary and the risk of expropriation when accepting the agent's wealth as collateral. We discuss this in greater detail in section 5.

In addition to these institutional variables, a limited liability constraint also operates in the economy. This implies that in the event an entrepreneurial project fails, the agent can only be liable upto the illiquid asset a. In other words agents are guaranteed a non negative payoff in all states of the world.

2.5 Occupational Choice

Agents choose their occupation. They can either choose to work in the subsistence sector, become workers, or become entrepreneurs. They are paid a wage w at the end of the period if they choose to work for a wage. If they choose entrepreneurship, their payoff is stochastic. The project succeeds with a probability θ which is the unobservable talent of the agent. To set up a firm an entrepreneur needs to hire n workers and pay them a wage w up front. Where $w \geq \underline{w}$ since working with the subsistence technology is an outside option that all agents have.

Our assumption that the productive technology requires n workers and 1 entrepreneur implies that workers and the entrepreneur are perfect complements in the production function. This assumption greatly simplifies our analysis and allows us to get sharp political economy results, though is not central to our analysis.

2.6 Markets

We will present a general equilibrium model with two markets; the labour and credit market. The need for credit arises as workers need to be paid up front when an entrepreneurial project is set up and the wealth of agents is illiquid. Both the markets are assumed to be perfectly competitive. The risk free interest rate is assumed to be zero.

3 Credit Contracts

Since the wealth of an agent is illiquid, agents need to borrow from the credit market to become entrepreneurs. The credit market is assumed to be perfectly competitive. The supply of credit is assumed to be perfectly elastic at interest rate equal to 1.

3.1 First Best

If talent was observable, then credit contract would not be based on collateral due to the presence of contractual friction ϕ that arises when collateral is used. Hence an agent with talent θ would be offered a contract with an interest rate $\frac{1}{\theta}$. Since the mass of entrepreneurs in this economy cannot exceed $\frac{1}{n+1}$, in equilibrium the wage would ensure that agents with talent less than θ^* become workers where

$$\theta^* : \int_{\theta^*}^{1} f(\theta) d\theta = \frac{1}{n+1}.$$

For the labour market to be in equilibrium, an agent with talent lower than θ^* should prefer working for a wage and agents with greater talent should prefer entrepreneurship. This implies that in equilibrium the agent with talent θ^*, who is indifferent between working for a wage and becoming an entrepreneur, has the following occupational choice condition:

$$\theta^* \left(R - \frac{n\overline{w}}{\theta^*} \right) + M + (1-\tau)a = \overline{w} + (1-\tau)a.$$

We can rearrange this condition to back out the equilibrium wage $\overline{w}$ at which the labour market clears:

$$\overline{w} = \frac{\theta^* R + M}{n+1}. \tag{1}$$

It follows that in the first best world the value of ϕ will not matter since wealth will not be used in the credit contract. In contrast, the value of τ would matter since an increase in τ would lower the expected payoff of agents due to the increased risk of expropriation.

We assume that

$$\overline{w} > M > \underline{w}.$$

The first part of the assumption ensures that the returns from the project when it succeeds are large enough to make interest payments.[6]

The second part of the assumption, $M > \underline{w}$ is necessary for the existence of a credit constraint in this economy.[7]

3.2 Second Best

The second best world is characterised by the unobservability of entrepreneurial talent. In all other respects it is identical to the first best world. Since talent is unobservable, the credit market can no longer offer contracts that are indexed by the agent's talent. However agents are endowed with wealth which they can use as collateral to access credit. Hence the credit contract will be defined by a pair (r, a) that is, interest rate and collateral.

We now discuss the possible credit contracts that can be offered to entrepreneurs and we characterise the equilibrium in the credit and labour market. The reader interested in the choice of institutions by the electorate in the first and second best world can see the figure in section 4 that captures the characterisation of the equilibrium and skip directly to section 5.

[6] Note that he interest rate offered to entrepreneur with talent θ^* is $\frac{1}{\theta^*}$. Backing out the value of θ^* from equation (1), we can check that

$$R - \frac{n\overline{w}}{\theta^*} > 0$$

is satisfied when $\overline{w} > M$.

[7] Consider an agent with zero wealth and talent. He would be attracted to entrepreneurship only if $M > \underline{w}$. Hence if this condition is not satisfied, his occupational choice condition in the second best world would be such that he would prefer working for a wage when $R - rnw < 0$ and consequently there may not be a credit constraint in the economy. The existence of a credit constraint introduces interesting results. We discuss this in greater detail in section 5.

3.2.1 Separating Contract

Let us first consider the separating contracts that can be offered to the agents. A separating contract exists if the contract is such that agents have an incentive to reveal their types. Since the probability of success is increasing in type, agents with higher entrepreneurial ability are offered contracts with lower interest rates. This feature of the credit contract creates an incentive to lie for low ability agents. Hence for such contracts to be incentive compatible, agents need to have sufficient wealth that the credit market can use as a screen. The separating contract is defined by the incentive compatible pair $(r_s(\theta, a(\theta)), a(\theta))$ which is the interest rate and the collateral that is offered to an agent with talent θ. The wealth level below which a separating contract is not feasible is determined by the constraint

$$R - r_s nw \geq 0$$

holding with an equality. This is shown by the following lemma.

Lemma 1. *No separating contract (r_s, a) can exist if $R < r_s nw$*

Proof. In the appendix. □

The intuition for this result is the following. When $R < r_s nw$, the entire return from the project has to be handed over to the bank when the project succeeds. In addition to R, agents also need to hand over a proportion of their wealth when the project succeeds. This additional requirement makes separation impossible. This happens because the separating contracts that are offered to high types are ones that return a large proportion $\gamma(a)$ of collateral in the success state. However these contracts are attractive to all agents that choose entrepreneurship regardless of their type.

Given Lemma 1, we can restrict our attention to the region where $R > r_s nw$. In this region, the zero profit condition for the bank is

$$\tilde{\theta} r_s(a) nw + (1 - \tilde{\theta})(1 - \tau)\phi a = nw \tag{2}$$

when lending to an agent of type $\tilde{\theta}$. Similarly, the feasibility condition for the loan is

$$R - r_s(a) nw \geq 0.$$

At the point where this feasibility constraint binds, we can find the talent of the least talent agent that becomes an entrepreneur by plugging in the zero profit condition, the feasibility condition for the loan, and the agent's occupational choice condition to find the lowest level of talent and collateral that is consistent with a separating contract. The occupational choice constraint of an agent indifferent between entrepreneurship and working for a wage is

$$\theta(R - r_s(a)nw) + M - (1-\tau)(1-\theta)a = w.$$

When the feasibility constraint of the loan binds, we have $R = r_s(a)nw$. We can now back out the talent of the least talented agent who could become an entrepreneur. This is $\underline{\theta}_s$ such that

$$M - (1-\tau)(1-\underline{\theta}_s)a = w. \tag{3}$$

Using equations (2) and (3) along with the feasibility condition for the loan, we can substitute out the equilibrium interest rate to find the expressions for $\underline{\theta}_s$ and $\underline{a}_s$, the lowest level of talent and wealth that are consistent with the existence of a separating contract. These are

$$\underline{\theta}_s = \frac{nw - \phi(M-w)}{R} \tag{4}$$

and

$$\underline{a}_s = \frac{(M-w)R}{(1-\tau)(R - nw + \phi(M-w))}. \tag{5}$$

$\underline{a}_s$ is a threshold wealth below which a separating contract is not feasible.

The strategy for deriving the separating contract schedule is the following. Equation (2) gives us the expression for the interest rate that is charged to an agent with type $\tilde{\theta}$. An agent with type θ has an incentive to declare his true type if a truthful declaration maximises his payoff from entrepreneurship. Hence if a separating contract can be designed such that a truthful declaration by the agent globally maximises her payoff from entrepreneurship, then we can say that such a separating contract is incentive compatible.

The existence of the separating contract depends on the existence of a type dependent collateral schedule that is implementable. In other words, letting $\tilde{\theta}$ be the type that an agent declares in a direct mechanism, if we can find a schedule of collateral $a(\tilde{\theta})$ such that agents find it optimal to declare their true types ($\tilde{\theta} = \theta$), then $(r_s(\tilde{\theta}, a(\tilde{\theta})), a(\tilde{\theta}))$ is a separating contract. It is optimal for an agent of type θ with wealth a to declare her type truthfully if:

$$\operatorname*{argmax}_{\tilde{\theta}} v_\theta(\tilde{\theta}) = \theta \tag{6}$$

where

$$v_\theta(\tilde{\theta}) = \theta\left(R - r(\tilde{\theta}, a(\tilde{\theta}))nw + (1-\tau)a(\tilde{\theta})\right) - (1-\tau)a(\tilde{\theta}) + M. \tag{7}$$

The first order condition of this problem yields a differential equation that we can use to solve for the collateral schedule $a(\tilde{\theta})$ such that agents have an incentive to reveal their types truthfully. This is

$$a(\tilde{\theta}) = \frac{nw}{\phi(1-\tau)}\left(1 - \frac{(R-nw)}{nw}\left(\frac{\underline{\theta}_s}{1-\underline{\theta}_s}\right)^{\frac{1}{1-\phi}}\left(\frac{1-\tilde{\theta}}{\tilde{\theta}}\right)^{\frac{\phi}{1-\phi}}\right). \tag{8}$$

The steps for the derivation of $a(\tilde{\theta})$ and the expression showing the concavity of the objective function are in the appendix.

The uniqueness of the solution to the differential equation tells us that there exists a unique collateral schedule such that agents find it optimal to declare their types truthfully. Since the zero profit condition for the banks is embedded into the expression for the interest rate in the objective function of the agents. We can recover the interest rate by plugging in the collateral schedule $a(\tilde{\theta})$ into the interest rate $r_s(\tilde{\theta}, a(\tilde{\theta}))$. It is possible to check that $a(\tilde{\theta})$ is monotonically increasing in $\tilde{\theta}$. High types are willing to post higher collateral since the value that an entrepreneur places on the reduction in the interest rate relative to the increase in collateral is increasing in her type. When the type of the agent is the highest possible, that is, $\theta = 1$, the corresponding collateral is $\overline{a}$ and the interest rate charged is 1.

3.2.2 Pooling Contract

In addition to a separating contract, there may also exist pooling contracts in this economy. Unlike the separating contract that is only available when $R \geq r_s(a)nw$, a pooling contract is possible both for the region of wealth that satisfies the corresponding condition, and also for a certain interval of wealth where this constraint is violated.

Let us first consider the region of wealth such that $R \geq r_p(a)nw$. Any pooling contract that could be offered must satisfy the necessary condition of zero profit for competitive banks:

$$r_p(a)\theta_p(a)nw + (1-\theta_p(a))(1-\tau)\phi a = nw. \tag{9}$$

Like we saw in the case of the separating contract, we can use the occupational choice constraint of the agents to evaluate the talent of the least talented agent that chooses entrepreneurship. This is $\hat{\theta}$ such that

$$\hat{\theta}(R - r_p(a)nw) + M - (1-\tau)(1-\hat{\theta})a = w \tag{10}$$

Now let us consider the zero profit condition for banks when $R < r_p nw$. In this region, in addition to the project returns R, the banks also need to be pledged a proportion of collateral for them to break even. The zero profit contract is now defined by

$$\theta_p(a)(R + (1-\gamma(a))(1-\tau)\phi a) + (1-\theta_p(a))(1-\tau)\phi a = nw. \tag{11}$$

where $(1-\gamma(a))$ is the proportion of collateral that is taken over by the bank in case the project succeeds. It is important to note that entrepreneurship is attractive not just because of the appropriable return R but also for the non-appropriable return M. If the latter is large enough, agents would be willing to choose entrepreneurship in exchange for their wealth even in the case when the project succeeds. Note that this formulation implicitly assumes that the optimal contract is one where all wealth is seized when the agent defaults. On the other hand, when the project succeeds, the minimum wealth $a\phi(1-\tau)(1-\gamma(a))$ that satisfies the zero profit condition of the bank is seized. It is easy to see that the pooling contract will take this form since this is the preferred contract for agents with high talent. Agents with high talent succeed with a higher probability and hence, relative to less talented agents, prefer contracts that are tougher in the bad state and yield a high payoff in the good state. Note that $\gamma(a)$ is increasing in a since banks would have to appropriate a larger share of wealth in the good state to satisfy the zero profit condition when the agent has lower wealth.

In both these regions, $\theta_p(a)$ is the average talent in the pool at wealth level a:

$$\theta_p(a) = \frac{1}{1-F(\hat{\theta}(a))} \int_{\hat{\theta}(a)}^{1} \theta f(\theta) d\theta \tag{12}$$

and $\hat{\theta}(a)$ is the agent with the lowest talent in the pool, who must be indifferent between working for a wage and becoming an entrepreneur with the pooling contract. In the region where $R < r_p nw$ this is determined by

$$M - (1-\tau)(1-\hat{\theta}(a)\gamma(a))a = w. \tag{13}$$

Plugging (12) in (11), the system of two equations (11) and (13) simultaneously determines the $\gamma(a)$ which can be thought of as the pooling interest rate and the lower bound $\hat{\theta}(a)$ of types that could choose the pooling contract $(\gamma(a), a)$ if they have wealth a. However, there exists a lower bound of wealth below which banks are not willing to offer such a contract. Note that credit contracts can only be offered when

$$\theta_p(a)R + (1-\tau)\phi a \geq nw.$$

This condition only holds when agents have sufficient wealth. This in turn defines the wealth level $\underline{a}_p$, such that agents with wealth less than this threshold will not be offered a pooling contract. Note that at this wealth level $\gamma(a) = 0$ must hold since agents would have to forgo their entire wealth in order to secure the credit contract.

$$\underline{a}_p = \frac{nw - \theta_p(\underline{a}_p)R}{\phi(1-\tau)}. \tag{14}$$

Substituting this, and $\gamma(\underline{a}_p) = 0$ into the occupational choice condition (13) of the marginal agent who is indifferent, we find at this wealth level, all agents choose entrepreneurship. $\hat{\theta}(\underline{a}_p) = 0$ and

$$\theta_p(\underline{a}_p) = \int_0^1 \theta f(\theta) d\theta. \tag{15}$$

This implies that at the lowest level of wealth that is consistent with the pooling contract, all agents prefer to become entrepreneurs.

Lemma 2. *The lower bound of talent in a pool at any given wealth class is weakly increasing in wealth.*

Proof. In the appendix. □

In words, starting from $\underline{a}_p$, an agent of a higher wealth class receives a lower interest rate but has a greater loss in case of failure, and this second effect always dominates for an agent at the bottom of the talent distribution. Hence entrepreneurship is more attractive to less talented agents when they have less wealth, since they have less to lose in case of default. Since these agents prefer working for a wage at high levels of wealth, the quality of the pool of borrowers is weakly increasing in wealth. The maximum wealth level for which a pooling contract can be acceptable is given by

$$\overline{a} = \frac{nw}{(1-\tau)\phi} \tag{16}$$

such that the pooling interest rate drops to 1. This is the level of collateral that will be charged in a pooling contract when the interest rate equals one.

4 Equilibrium

In the previous section we have discussed the types of credit contracts that can exist in the economy. We are now ready to characterise the equilibrium.

4.1 Equilibrium in the Credit Market

We have shown that both pooling and separating contracts are viable. Given that banks can introduce any contract $(r(a), a)$ we will now characterise the equilibrium in the model. We will use the Rothschild Stiglitz equilibrium concept where an equilibrium is characterised by the conditions: *i)* all the contracts in the equilibrium set make non negative profits and *ii)* non existence of a contract that can be introduced that will make a strictly positive profit. We will assume that $\underline{a}_p > 0$. It is easy to check that $\underline{a}_p < \underline{a}_s < \overline{a}$. Hence there is no contract that can be offered to (and accepted by) an agent with wealth $a < \underline{a}_p$ that will make non negative profits.

Lemma 3. *There exists a level of wealth $\hat{a}_p$ defined by $\underline{\theta}_s = \hat{\theta}(\hat{a}_p)$ where $\overline{a} > \hat{a}_p > \underline{a}_s$ such that the only contract in the equilibrium set for $a < \hat{a}_p$ can be a pooling contract.*

Proof. Recall that $\hat{\theta}(a)$ is the level of talent such that an agent with this talent is indifferent between becoming an entrepreneur with the pooling contract $(r_p(a), a)$ and working for a wage. Since the distribution of wealth is continuous, there exists a level of wealth $\hat{a}_p$ such that an agent with talent $\underline{\theta}_s = \hat{\theta}(\hat{a}_p)$ is indifferent between both these alternatives and the separating contract $(r(a(\underline{\theta}_s)), a(\underline{\theta}_s))$. At $\underline{a}_s$ the agent with type $\underline{\theta}_s$ prefers the pooling to the separating contract since she receives a cross subsidy. At $\hat{a}_p$ the attractiveness of the cross subsidy disappears since the collateral requirement becomes too high. Hence even though a separating contract is feasible at $\underline{a}_s$ it is not incentive compatible for an agent with type $\underline{\theta}_s$ to accept it. It becomes incentive compatible only when the agent has wealth $a \geq \hat{a}_p$ at which point he prefers $(r_s(a(\underline{\theta}_s)), a(\underline{\theta}_s)$ to $(r_p(\hat{a}_p), \hat{a}_p))$ □

Lemma 4. *In the region of wealth $a \in (\hat{a}_p, \overline{a})$ there exists a level of talent $\hat{\theta}_s(a)$ such that agents with talent $\theta > \hat{\theta}_s(a)$ prefer the pooling contract and agents with talent $\theta \leq \hat{\theta}_s(a)$ prefer the separating contract.*

Proof. Note that for $a \in (\hat{a}_p, \overline{a})$ a fully separating contract schedule is not available since the collateral required for full separation of types is $\overline{a}$. $\frac{\partial \hat{\theta}(a)}{\partial a}$ implies that the attractiveness of the pooling contract is increasing in type. This is obvious since it simply captures the fact that more wealth is better for screening than less. This implies the existence of a cutoff talent $\hat{\theta}_s(a)$ for level of wealth $a \geq \hat{a}_p$ such that it becomes possible to offer agents with talent $\theta \leq \hat{\theta}_s(a)$ a separating contract that they prefer to the pooling contract. Note that $\hat{\theta}_s(\hat{a}_p) = \hat{\theta}(\hat{a}_p) = \underline{\theta}_s$ and $\hat{\theta}_s(\overline{a}) = 1$ □

Proposition 1 (Existence and Uniqueness)**.** *A unique credit market equilibrium exists such that agents with wealth a:*

- *$a \geq \overline{a}$: are offered separating contracts*
- *$\overline{a} > a > \hat{a}_p$: are offered both pooling and separating contract*
- *$\hat{a}_p > a > \underline{a}_p$: are offered pooling contracts*
- *$\underline{a}_p > a$: are credit constrained*

Proof. $a < \underline{a}_p$ are credit constrained since no contract that makes non negative profits can be offered to these agents. Lemma 3 shows that only a pooling contract can exist in the region of wealth $a < \hat{a}_p$. Lemma 4 shows that in the region of wealth $\overline{a} > a > \hat{a}_p$ agents with talent $\theta \leq \hat{\theta}_s(a)$ a separating contract and $\theta > \hat{\theta}_s(a)$ accept a pooling contract. For the region of wealth $a \geq \overline{a}$ a fully separating schedule of contract exists that is offered and accepted by agents. This is a unique equilibrium since the zero profit pooling and separating contract schedules are unique. □

Proposition 2 (Occupational Choice)**.** *Agents with wealth:*

- *$\underline{a}_p > a$ become workers*

- *$\hat{a}_p > a \geq \underline{a}_p$ and talent $\theta \geq \hat{\theta}(a)$ accept the pooling contract and become entrepreneurs and the rest become workers*

- *$\overline{a} > a > \hat{a}_p$ and talent $1 \geq \theta > \hat{\theta}_s(a)$ accept the pooling contract and become entrepreneurs; and talent $\hat{\theta}_s(a)) \geq \theta \geq \underline{\theta}_s$ accept the separating contract and become entrepreneurs, and the rest become workers*

- *$a \geq \overline{a}$ and talent $\theta \geq \underline{\theta}_s$ accept the separating contract and become entrepreneurs and the rest become workers*

Proof. Follows from Lemma 2 to 4 and Proposition 1 □

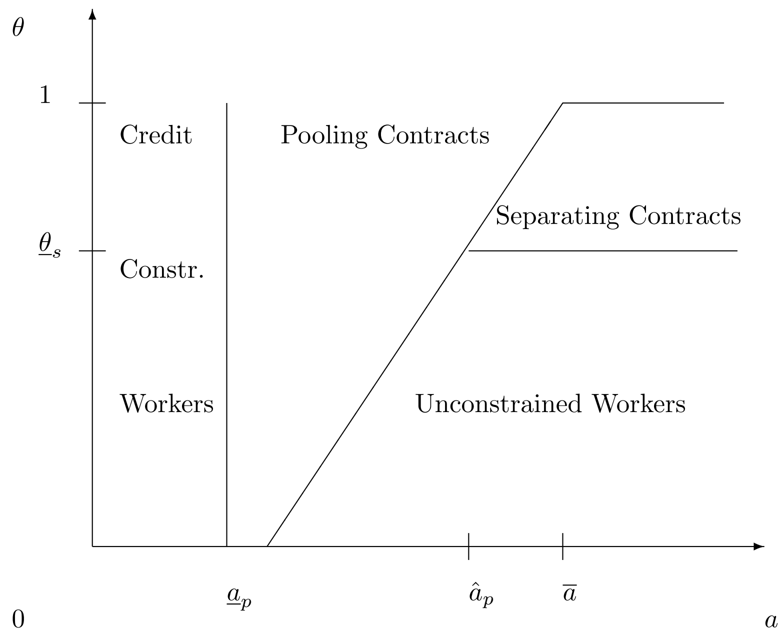
The following figure presents a graphical representation of the equilibrium. As seen in the figure, we can conveniently analyse the equilibrium in terms of four regions of wealth:

- Region 1: $\underline{a}_p > a$ are credit constrained and become workers;

- Region 2: $\hat{a}_p > a \geq \underline{a}_p$ and talent $\theta \geq \hat{\theta}(a)$ accept the pooling contract and become entrepreneurs and the rest become workers;

- Region 3: $\overline{a} > a > \hat{a}_p$ and talent $1 \geq \theta > \hat{\theta}_s(a)$ accept the pooling contract and become entrepreneurs; and talent $\hat{\theta}_s(a)) \geq \theta \geq \underline{\theta}_s$ accept the separating contract and become entrepreneurs, and the rest become workers;

- Region 4: $a \geq \overline{a}$ and talent $\theta \geq \underline{\theta}_s$ accept the separating contract and become entrepreneurs and the rest become workers.



At wealth level $\underline{a}_s$ it is possible to offer a separating contract to the agent with talent $\underline{\theta}_s$. However at this wealth level the agent with talent $\underline{\theta}_s$ will always accept the pooling contract since he pledges the same level as collateral but receives a lower interest rate with the pooling contract because of the cross subsidy. As the wealth of this agent increases, the pooling contract that is offered becomes less attractive since the pooling contract always requires an agent to pledge all his wealth as collateral. At wealth level $\hat{a}_p$ the agent prefers to take the separating contract with collateral $\underline{a}_s$ rather than take the pooling contract with wealth $\hat{a}_p$. Hence though separating contract is feasible from wealth level $\underline{a}_s$, in equilibrium they are only seen from wealth level $\hat{a}_p$. Because of this reason, as the level of wealth rises the talent of the least talented agent who accepts the pooling contract also rises. Similar to region 2 where there is no separating contract, this happens because agents with higher wealth prefer to become workers due to the high collateral requirement for being an entrepreneur. In region 3 however, this happens due to the high collateral requirement of the pooling contract relative to the separating contract.

In region 3 take a specific wealth level a. The agent with talent $\hat{\theta}_s(a)$ is indifferent between the separating contract that is offered to him and the pooling contract and accepts the separating contract. The agents with talent $\hat{\theta}_s(a) > \theta > \underline{\theta}_s$ strictly prefer the respective separating contracts they are offerred. Now consider an agent with talent greater than $\hat{\theta}_s(a)$. I will show that all agents in this group prefer the pooling contract rather than accepting the separating contract offered to the agent with talent $\hat{\theta}_s(a)$. Agent with talent $\hat{\theta}_s(a)$ (lets call this $\hat{\theta}$ to ease notation) who is indifferent between pooling and separating contract implies:

$$v_s(\hat{\theta}) = \hat{\theta}(R - r_s(a_s)nw) + M + \hat{\theta}(1-\tau)a_s + (1-\tau)(a - a_s) \qquad (17)$$

where a_s and $r_s(a_s)$ are the collateral and interest rate for the separating contract defined by (8). The value for this agent from the pooling contract is

$$v_p(\hat{\theta}) = \hat{\theta}(R - r_p(a)nw) + M + \hat{\theta}(1-\tau)a. \qquad (18)$$

Equating $v_s(\hat{\theta}) = v_p(\hat{\theta})$ we get

$$r_p(a)nw = r_s(a)nw - \frac{(1-\tau)(1-\hat{\theta})}{\hat{\theta}}(a - a_s). \qquad (19)$$

Take a $\tilde{\theta} > \hat{\theta}$. I will now show that agent with talent $\tilde{\theta}$ will prefer the pooling contract to the separating contract offered to agents with talent $\hat{\theta}$. Using the expression for the pooling interest rate in equation (19) we have:

$$v_p(\tilde{\theta}) = \tilde{\theta}\left(R - r_s(a_s)nw + \frac{(1-\tau)(1-\hat{\theta})}{\hat{\theta}}(a - a_s)\right) + M + \hat{\theta}(1-\tau)a. \qquad (20)$$

On the other hand the value from mimicking $\hat{\theta}$ and accepting the pooling contract is

$$v_s^{\hat{\theta}}(\tilde{\theta}) = \tilde{\theta}(R - r_s(a_s)nw) + M + \tilde{\theta}(1-\tau)a_s + (1-\tau)(a - a_s). \qquad (21)$$

Equating $v_p(\tilde{\theta})$ and $v_s^{\hat{\theta}}(\tilde{\theta})$ we find that the pooling contract dominates for all $\tilde{\theta}$. Hence types greater than $\hat{\theta}$ prefer to post the higher collateral and get the pooling contract rather than take the separating contract offered to agent $\hat{\theta}$.

In the region of wealth $a < \hat{a}_p$ the talent of the least talented agent is $\hat{\theta}(a)$. However when a separating contract becomes feasible the nature of this function that determines the talent of the least talented agent in the pool changes somewhat hence we call it $\hat{\theta}_s(a)$. $\hat{\theta}_s(a)$ is defined by 2 conditions: The first condition determines the feasibility of the separating contract, i.e equation (8). Since equation (8) is monotonically increasing in θ, it is invertible. Expressing equation (8) as a function of theta we have: $\theta_s(a)$. The second condition that determines $\hat{\theta}_s(a)$ is the condition that determines the indifference between the payoffs from the pooling and separating contract for the agent. Let us call this $\hat{\theta}(a)$. Hence in region 3 we have:

$$\hat{\theta}_s(a) = \max\{\hat{\theta}(a), \theta_s(a)\} \tag{22}$$

This is because it is possible for either of the two constraints to be slack in this region. It is possible that agents prefer the separating contract but the separating contract is simply not feasible in which case $\theta_s(a)$ would bind. Alternatively it is possible that the separating contract is feasible but agents at the lower end prefer the pooling contract. In this case $\hat{\theta}(a)$ would bind.

Now consider the threshold at which region 3 begins. If $\theta_s(a)$ binds here then we have $\hat{a}_p = \underline{a}_s$. However this is not possible for the following reason. If $\theta_s(a)$ binds at the threshold this implies that $\hat{\theta}(\underline{a}_s) \leq \theta_s(\underline{a}_s)$. This implies that even though the agent with talent $\underline{\theta}_s$ is made to post the same level of collateral and receives a lower interest rate with the pooling contract, he still prefers the separating contract. This is not possible. Hence at the beginning of region 3 $\hat{\theta}(a)$ binds and we have $\hat{a}_p > \underline{a}_s$. Note that as $\hat{\theta}(a)$ follows continuously from from region 2, the transition from region 2 to 3 from $\hat{\theta}(a)$ to $\hat{\theta}_s(a)$ is continuous. Thereafter $\hat{\theta}_s(a)$ is continuous since both $\hat{\theta}(a)$ and $\theta_s(a)$ are continuous. Note that it is possible that there could be finite points where $\hat{\theta}(a)$ and $\theta_s(a)$ cross each other making $\hat{\theta}_s(a)$ non differentiable. However this does not affect the monotonicity property of $\hat{\theta}_s(a)$ since both $\hat{\theta}(a)$ and $\theta_s(a)$ are monotonically increasing in a.

4.2 Equilibrium in the Labour Market

The labour market is perfectly competitive. An equilibrium is characterised by the demand equalling supply. It is much easier to characterise the equilibrium by thinking of the labour demand of a firm instead of the labour demand by an entrepreneur. A firm demands 1 unit of entrepreneurial and n units of non entrepreneurial labour. Supply is 0 for wage $w < \underline{w}$, and 1 at $w = \underline{w}$. Labour demand is given by:

$$L_d = (n+1)\left(\int_{\underline{a}_p}^{\hat{a}_p} \left(1 - F(\hat{\theta}(a))\right) g(a)da + (1 - F(\underline{\theta}_s))(1 - G(\hat{a}_p))\right) \tag{23}$$

Proposition 3. *The equilibrium wage is $\underline{w}$ when $L_d(\underline{w}) \leq 1$ $w > \underline{w}$ when $L_d(\underline{w}) > 1$*

Proof. Note that Labour demand is monotonically decreasing in the wage:

$$\frac{\partial L_d}{\partial w} = (n+1)\left(-g(\underline{a}_p)\frac{\partial \underline{a}_p}{\partial w} - f(\underline{\theta}_s)\frac{\partial \underline{\theta}_s}{\partial w}(1 - G(\hat{a}_p)) - \int_{\underline{a}_p}^{\hat{a}_p} f(\hat{\theta}(a))\frac{\partial \hat{\theta}(a)}{\partial w}g(a)da\right) < 0 \tag{24}$$

since

$$\frac{\partial \underline{a}_p}{\partial w} > 0 \qquad \frac{\partial \underline{\theta}_s}{\partial w} > 0 \qquad \frac{\partial \hat{\theta}(a)}{\partial w} > 0 \tag{25}$$

If Labour demand is less than 1, there is excess supply of labour in the economy and the wage must equal $\underline{w}$ which is the outside option to working for a wage. If the labour demanded at $w = \underline{w}$ is more that 1, then the economy is tight in the sense that no one is engaged in the subsistence sector, and the wage must increase to equilibriate demand and supply.

□

The proof shows that there are two effects that create a labour demand that is monotonically decreasing in wage.

Firstly as the wage increases, the amount that entrepreneurs need to borrow also increases. This drives up the credit constraint.

Secondly, as wage increases, the agent with the lowest talent that was previously indifferent between entrepreneurship and paid employment now prefers paid employment. This is because the increase in wage tips his occupational choice constraint. Both these effect imply a reduction in labour demand for an increase in wage.

Note that in this economy $M > w$ is necessary and sufficient for there to be a credit constraint. If the equilibrium wage rises above this then the bank's zero profit condition is satisfied even at 0 wealth. We will assume that the equilibrium wage is lower than M since the problem without credit constraint is not interesting to analyse.[8]

[8] It should be noted that in contrast to Ghatak et al. (2007) there are no multiple equilibria since firm level labour demand is constant at n. This implies that in our model the what drives the labour demand is the extensive margin effect.

5 Credit Market Institutions

The argument we make is that when interest groups are created in an imperfect market, then this can lead to an inefficient choice of institutional reform. In the first best world where talent is observable, the best institutions are chosen. As we move away from the first best world, there is not only a market inefficiency created by the unobservability of talent, but also a political inefficiency through the creation of class structure in the electorate that votes in favour of inefficient institutions.

The parameter τ captures the strength of enforcement of property rights. A high τ implies that law enforcement is poor and assets are likely to be stolen by thieves or taken over by the local strongman. Hence a straightforward way to think about τ is how tough government is on property related crime and how well it enforces the claims of someone dispossessed of their property. Alternatively, τ can also be thought of as how well the titling system works. To the extent it is easy to bribe the local bureaucrat to get the name on someone's land title changed, τ would be high and vice versa.

The parameter ϕ measures the efficiency of contractual institutions. The treatment of ϕ is somewhat different since it is the proportion of collateralized wealth that can be liquidated. If an agent pledges wealth a as collateral to become an entrepreneur, and his project fails, the bank only recovers ϕa. Hence $(1-\phi)a$ is pure inefficiency and consequently there is a strong case for thinking that $\phi = 1$ will be the surplus maximising policy. However under certain conditions, this effect may be dominated through the inefficiencies caused in the occupational choices since a high ϕ can end up making entrepreneurship attractive to agents who should optimally become workers.

ϕ and τ are parameters that capture institutional frictions that reduce the efficiency of market transactions involving wealth.[9] This can be illustrated with the following example. To fix ideas let us think of wealth as land. Consider a scenario where there's an agent who wishes to rent out his land. This landlord would consider two things when entering into a rental contract with a potential tenant. Firstly he would consider how secure his property rights are. When τ is high, the landlord realises that his property rights over the land he is renting out are not very secure. This dampens the incentives for renting the land since

[9] We have focused only on institutional frictions involving wealth because wealth is the instrument that banks can use to mitigate the inefficiencies due to the unobservability of talent, and we want to show that the political process can fail to choose the right reforms even when there is no redistributive objective.

the landlord worries about a potential capture by the tenant. Independently, a low ϕ implies that enforcement of contracts is costly. The landlord anticipates that in the event a tenant refuses to vacate the land as per the terms of the rental contract, the landlord would need to approach the courts for enforcement of his contractual rights. Even if property rights are fully secure, if ϕ is low, the court costs would be substantial. Therefore a low ϕ would also dampen the incentives to put land to its productive use.

The distinction between the two institutions is heuristic.[10] In most applications one can think of, ϕ and τ would interact together creating aggregate transaction costs that would dampen the incentives for market transactions involving wealth. For example in the model presented here, both enter multiplicatively when agents post their wealth as collateral to become entrepreneurs. The credit market takes into account both the insecurity of the property right over the collateral and the costs of enforcing the credit contract in case of default.

5.1 Institutions in the First Best World

We now show that in the first best world the surplus maximising institutions are chosen.

Proposition 4. *When talent is observable, voters unanimously choose surplus maximising institutions.*

Proof. Total surplus in the economy is maximized when the most talented agents become entrepreneurs regardless of their wealth. This is equivalent to the quality of the pool of entrepreneurs being maximised. Under the first best the total surplus in the economy is:

$$W_{fb} = R\int_{\theta^*}^{1} \theta f(\theta)d\theta + \frac{M}{n+1} + \int_{0}^{\infty}(1-\tau)ag(a)d(a) \tag{26}$$

By inspecting this expression it is clear that the total surplus is decreasing in τ. Hence $\tau = 0$ is the surplus maximising. Since all agents lose a part of their wealth as τ increases, agents unanimously vote for τ equal to zero. Since

[10] In Besley (1995) three channels through which property rights affects investment incentives are laid out. These are the security of tenure, the use of property as collateral, and the benefits of gains from trade. Of these we feel that the first and the third are channels through which τ would affect investment incentives whereas the second channel relating to the use of land as collateral is affected by an interaction of τ and ϕ as is the case in the model. Of course wealth in our model is exogenous and therefore the issue of investment incentives does not arise.

ϕ does not appear in (26), all values of ϕ are surplus maximising, and hence the proposition is trivially true for ϕ. □

When talent is observable, the preferences of the electorate are unanimously aligned with surplus maximisation. Hence a $\tau = 0$ is chosen because better property rights increase the expected payoff of all agents. Similarly the optimal ϕ would be chosen to the extent there are any contractual transactions involving wealth. Note that in the first best in our model there are no contractual transactions involving wealth since talent is observable and wealth has no use as a screen. Hence all values of ϕ are optimal in the first best world.

5.2 Institutions in the Second Best World

In the last subsection we showed that in the first best world the preferences of the electorate are unanimously aligned with surplus maximisation. We will show that as soon as there's a departure from the first best, the inefficiency of the market gets further amplified by the choices of the electorate that is created in the inefficient market. In the second best world with unobservable talent, the total surplus is:

$$W_{sb} = R\left(\int_{\underline{a}_p}^{\hat{a}_p}\int_{\hat{\theta}(a)}^{1} \theta f(\theta)g(a)d\theta da + \int_{\hat{a}_p}^{\infty}\int_{\underline{\theta}_s}^{1} \theta f(\theta)g(a)d\theta da\right) \tag{27}$$

$$+M\left(\int_{\underline{a}_p}^{\hat{a}_p}(1-F(\hat{\theta}(a)))g(a)da + (1-G(\hat{a}_p))(1-F(\underline{\theta}_s))\right)$$

$$+\underline{w}\left(1-(n+1)\int_{\underline{a}_p}^{\hat{a}_p}(1-F(\hat{\theta}(a)))g(a)da + (1-F(\underline{\theta}_s))(1-G(\hat{a}_p))\right)+\int_0^{\infty}(1-\tau)ag(a)da$$

$$-(1-\tau)(1-\phi)\left(\int_{\underline{a}_p}^{\hat{a}_p}\int_{\hat{\theta}(a)}^{1}(\theta(1-\gamma(a))+(1-\theta))\,af(\theta)g(a)d\theta da + \int_{\hat{a}_p}^{\infty}\int_{\underline{\theta}_s}^{1} a(1-\theta)f(\theta)g(a)d\theta da\right)$$

In this economy there are two productive activities, the subsistence sector where a worker produces $\underline{w}$, and the hi tech sector where n workers and 1 entrepreneur of ability θ produce R with probability θ and 0 with probability $(1-\theta)$. The project also yields a non expropriable return M to the entrepreneur.

The wage paid to the worker in the hi tech sector is simply a transfer from the entrepreneur to the worker which doesn't enter the total surplus. In the world with full information, the first best is guaranteed, where all agents are engaged in the hi tech sector either as a worker or entrepreneurs. This is what equation (26) captures.

In the second best world it is possible that there are agents that work in the subsistence sector. The mass of agents engaged in the hi tech sector is $n+1$ times the mass of entrepreneurs. The rest of the agents work in the subsistence sector where they produce $\underline{w}$. This is captured in the third part of equation (27) which takes a positive value when $w = \underline{w}$ and 0 otherwise.

The fourth part of the expression captures the loss of wealth when τ is greater than 0. Similarly when ϕ is less than one there is some loss of collateral in case of default. The first best could be achieved if $\underline{a}_p = 0$ and $\hat{\theta}(a) = \underline{\theta}_s = \theta^*$. In such a case none of the agents in the economy are engaged in the subsistence sector and hence the second term in the expression drops out.

It is easy to see why the first best is never possible when talent is unobservable. Even when there is no credit constraint, at low enough levels of wealth, separation is not possible. At the bottom of the wealth distribution where $a = 0$, the credit market can only offer a pooling contract. With a pooling contract at $a = 0$, the talent of the least talented agent that chooses entrepreneurship is always lower than θ^* since θ^* is the talent of the least talented agent that accepts her actuarially fair contract in the full information case. Since the least talented agent receives a cross subsidy with the pooling contract but not a separating contract, the talent of the marginal agent with 0 wealth is lower when talent is unobservable. But since the mass of entrepreneurs is bounded at $\frac{1}{n+1}$, and at the lower end of the wealth distribution agents with talent less than θ^* are entrepreneurs, then at wealth $a \geq \hat{a}_p$, $\underline{\theta}_s$ must be greater than θ^*. That is, agents that would become entrepreneurs in the first best world, choose to work for a wage. This drives the inefficiency in the model. If credit constraint exists then there is the added inefficiency of agents with high talent but low wealth that are excluded from entrepreneurship.

The first best can only be replicated in the world with incomplete information if all agents have sufficient wealth and can be offered a separating contract. Therefore if the average wealth in this economy is greater than the threshold level of wealth required for separation, a policy of redistribution can restore full efficiency in this economy. If the total level of wealth is insufficient or if the instruments for conducting such a redistribution are unavailable then there will always be some inefficiency since there would at the same time be agents with talent less than θ^* who choose entrepreneurship and talent greater than θ^* that choose working for a wage.

Observation 1. *A non-zero level of credit constraint may be optimal in this economy.*

Given this discussion, it is possible to envisage distributions of wealth and talent such that there exists a non zero "natural level of credit constraint". That is, the total surplus may not always be maximised when the credit constraint is pushed down. Though reducing the credit constraint allows agents with low wealth to become entrepreneurs, this has an effect through the labour market of increasing the wage. Increasing the wage may in turn reduce the number of high type entrepreneurs with high wealth.

To discuss whether endogenous institutions can bring the economy in the direction of higher welfare or not, suppose that all agents can vote in a binary election between a status quo institution (status quo ϕ or τ) and an alternative. When faced with a binary choice, each agent votes sincerely.

One obvious remark we will make, without making distributional assumptions, is that an alternative policy that is aimed at maximising total surplus may not win when put to majority vote. This result in itself is not particularly surprising. Since redistributive instruments are lacking it is to be expected that agents inefficiently use institutions to redistribute rather than to maximise surplus. Indeed such a choice of institutions is not inefficient in the paretian sense. What is interesting here however is that the alignment of interest groups is itself created by the existence of market failure and this alignment takes the economy away even from the second best world with market failures. In other words, the inefficiency of market failure is further amplified by the political alignments it creates.

The cornerstone to understanding why agents choose non surplus maximising institutions is the following: in this economy there are always at least $\frac{n}{n+1}$ workers. Since $n \geq 1$, a policy that increases wage has support of at least half the population. However policies that increase the wage may not increase the quality of the pool of entrepreneurs. This is the insight that we will use to generate the results in the rest of this section. Thus efficient institutions are those that increase the quality of the pool of entrepreneurs whereas institutions that increase wage are politically feasible.

5.2.1 Support for improvement in judicial enforcement

The parameter ϕ in the model denotes the amount of collateral that banks can liquidate in case of default and is the parameter that denotes the quality of the judiciary. Instead of a cost that is proportional to the collateral in dispute, the quality of the judiciary could be modelled as a fixed cost that need to be paid for approaching the judiciary. In such a model ϕ would be a fixed cost and interest rate would instead be determined by the following zero profit condition:

$$r(a)nw\theta + ((1-\tau)a - \phi)(1-\theta) = nw \tag{28}$$

The idea we wish to capture with ϕ is the efficiency of the judiciary in expropriating assets of a defaultor and handing them over to the creditor at the least possible cost. This idea is captured in both these formulations. Given the discussion on efficiency and political feasibility, we have:

Proposition 5. *A policy aimed at increasing ϕ is guaranteed majority support but may not always be surplus maximising.*

Proof. There are two parts to this proposition. The first part is that a policy of increasing ϕ is guaranteed majority support. This is proven in the appendix. The second part is that such a policy is not guaranteed to be surplus maximising. This is proven by construction of an example in the final extension where increasing ϕ reduces total surplus. □

The intuition for the result is the following. It is easy to show that the equilibrium wage is non decreasing in ϕ, and hence the proposal for increasing ϕ is supported by the majority. However, total surplus may not be increasing in ϕ since the effect of an increase in ϕ on the quality of the pool of entrepreneurs is ambiguous.

This result is quite striking when contrasted against the standard intuition about contracting institutions. Here improving the quality of contracting institutions (increasing ϕ) is not always good since that makes entrepreneurship more attractive and this induces low types to become entrepreneurs. This result arises because there are inherent externalities when agents borrow money: the low type entrepreneurs by their very existence impose an externality on the high types. Our result can be easily understood when seen in the light of the theory of second best.

5.2.2 Support for Improvement in Property Rights

Imperfect protection of property rights reduces the value of wealth. This in turn makes entrepreneurship more attractive since agents do not place as much weight on default and consequent loss of collateral.

The political support for a change in τ is ambiguous because the effect on the wage is ambiguous. We can see this from the following:

$$\frac{\partial L_d}{\partial \tau} : (n{+}1)\left(-g(\underline{a}_p)\frac{\partial \underline{a}_p}{\partial \tau}(1-F(\underline{\theta}_p)) - f(\underline{\theta}_s)\frac{\partial \underline{\theta}_s}{\partial \tau}(1-G(\hat{a}_p)) - \int_{\underline{a}_p}^{\hat{a}_p} f(\hat{\theta}(a))\frac{\partial \hat{\theta}(a)}{\partial \tau}g(a)da\right) \tag{29}$$

The sign of this expression is ambiguous. This is because:

$$\frac{\partial \underline{a}_p}{\partial \tau} > 0 \qquad \frac{\partial \underline{\theta}_s}{\partial \tau} = 0 \qquad \frac{\partial \hat{\theta}(a)}{\partial \tau} < 0 \tag{30}$$

$$\frac{\partial \underline{a}_p}{\partial \tau} = \frac{nw - \theta_p(\underline{a}_p)R}{\phi(1-\tau)^2} - \frac{\partial \theta_p(\underline{a}_p)}{\partial \tau}\left(\frac{R}{\phi(1-\tau)}\right) > 0 \tag{31}$$

since

$$\frac{\partial \underline{\theta}_p}{\partial \tau}|_{\underline{a}_p} = 0 \tag{32}$$

The credit constraint is increasing in τ. When τ increases, the effective wealth of an agent decreases, and the interest rate at all levels of wealth increases. This is intuitive since an increase in τ decreases the value of wealth as a screen. Since agents are likely to have their wealth expropriated anyway, posting a high collateral is less effective in revealing an agent's type. Take the limiting case where τ goes close to 1, in this case, the credit market correctly anticipates that all agents are equally eager to post any collateral since they know that their wealth will be expropriated and hence don't attach any value on recovery of collateral in the event of success and consequent repayment of the loan.

There are two opposing effects on wage of a decrease in τ. Firstly decreasing τ reduces the level of credit constraint. This increases the number of entrepreneurs. Decreasing τ also decreases the attractiveness of entrepreneurship for marginal agents ($\hat{\theta}(a)$), who were previously accepting the pooling contract to become entrepreneurs due to the cross subsidy from higher types within their wealth level. Since there are two opposite effects on wage, the precise effect on total surplus of a change in τ would depend on the assumptions on the distribution of wealth and talent. However in case these two effects exactly cancel each other out, it is possible then to characterise the effect on total surplus.

Proposition 6. *If the wage remains unchanged as a result of a change in τ, then decreasing (increasing) τ increases (decreases) total surplus*

Proof. If wage remains unchanged as a result of a decrease in τ then the new equilibrium pareto dominates the previous equilibrium. All agents who remain workers are unaffected, all entrepreneurs are made better off due to a reduction in the interest rate. Additionally there are agents who were previously credit constrained who can now become entrepreneurs for whom the policy is a strict improvement over status quo. Since it is a pareto improvement, it must also increase total surplus. Similarly if an increase in τ keeps the wage unchanged, it must reduce the total surplus since workers are unaffected, entrepreneurs are made worse off due to the increase in the interest rate, and there are at least some agents who are denied credit as a result of the increase in the credit constraint who are made strictly worse off. □

By continuity we can extend this proposition to mean that if the change in wage as a result of an improvement in property rights is small enough, then total surplus must have increased. It is possible to push this result further.

Proposition 7. *If the change in wage as a result of improvement (deterioration) in property right is negative (positive) then total surplus must increase (decrease).*

Proof. Note first that the average quality of the pool of entrepreneurs is a sufficient statistic for gauging changes in total surplus. If the wage decreases as a result of an decrease in τ, it must be the case that the effect on labour demand through $\hat{\theta}(a)$ dominates the reduction in the credit constraint. Now note that the average talent at the lowest level of wealth where a pooling contract is offered is lower than the average talent of the pool. This is true because the distribution of wealth and talent are independent and the talent of the least talented agent within a wealth level is increasing in wealth.

Now note that is always possible to construct a distribution of wealth such that the pre reform average talent is the same but post reform the credit constraint is relaxed more to the extent that the two opposing effects on wage cancel each other out and wage remains unchanged. In this case, the average talent post reform would be lower than the case where the wage went down. However, given the previous result, the total surplus would still increase. Since the initial average quality of the pool of entrepreneurs is the same by construction, this implies that the ex post level of talent must have increased in the case where the wage decreases. □

This result brings into sharp relief the trade-off between political feasibility and efficiency of institutional reform. Only reforms that increase wages are politically feasible but these may not correspond to reforms that are surplus maximising. In case of property rights institutions, when worsening them (increasing τ) is politically feasible, they have an unambiguously negative effect on the total surplus. The political feasibility of τ depends on the distribution of wealth. if the median voter is a worker with very low wealth she would care more about the effect on wage and would therefore vote in favour of worsening property rights.

6 Conclusion

To summarise our result on institutional efficiency and feasibility, we find that improving contractual institutions is always feasible but may not always be efficient since improving contracting induces too many low type agents to choose entrepreneurship. On the other hand we find that if worsening property rights institutions is politically feasible then it unambiguously reduces total surplus. Similarly if improving property rights is politically infeasible then it unambiguously increases total surplus. These results bring into sharp focus the tension between political feasibility and surplus maximisation.

When there's a market failure, the competitive equilibrium is no longer guaranteed to be on the Pareto frontier. Our model makes the point that in the event of a market failure, competitive markets can passively play a political role of creating constituencies. These constituencies can have a preference for inefficient policies. This leads to the inefficiencies of market failure being further amplified by the policy choices that constituencies created in a flawed market make. In this sense our paper provides an additional reason to worry about market failure; market failure may lead to a political failure even in a fully representative democracy.

Appendix

Proof for Lemma 1

Proof. If $R < r_s nw$ then there are insufficient appropriable returns to cover the interest payments from the loan. Hence the only way the banks can break even is if entrepreneurs pledge a portion of their collateral even in the state where the project is successful. Let us call the proportion of collateral that banks seize in the good state $(1-\gamma(a))$. The new zero profit condition for banks when they lend to an agent whose declared type is $\tilde{\theta}$ is

$$\tilde{\theta}(R + (1-\gamma(a))(1-\tau)\phi a) + (1-\tilde{\theta})(1-\tau)\phi a = nw$$

Rearranging this, we get

$$\gamma(a) = \frac{\tilde{\theta}R - nw + (1-\tau)\phi a}{\tilde{\theta}(1-\tau)\phi a}. \tag{33}$$

The left hand side of the occupational choice constraint for an agent with talent θ who accepts a separating contract (designed for an agent with talent $\tilde{\theta}$) is

$$M - (1 - \theta\gamma(a))(1-\tau)a$$

Substituting the value for $\gamma(a)$ from equation (33) we get:

$$v_\theta(\tilde{\theta}) = M + \theta\left(\frac{\tilde{\theta}R - nw + (1-\tau)\phi a}{\tilde{\theta}\phi}\right) - (1-\tau)a.$$

We can now differentiate this equation with respect to the declaration $\tilde{\theta}$ to see whether the agent has an incentive to declare his type truthfully. It is easy to check that in the relevant range, the payoff of the agent $v_\theta(\tilde{\theta})$ is increasing in his declaration $\tilde{\theta}$. Hence agents will always overstate their type and a separating contract cannot exist. □

Derivation of $a(\tilde{\theta})$ from section 3.2.1.

An agent of type θ maximises his payoff from entrepreneurship by choosing the declaration that maximises $v_\theta(\tilde{\theta})$.

$$\underset{\tilde{\theta}}{\operatorname{argmax}}\, v_\theta(\tilde{\theta}) = \theta. \tag{34}$$

The first order condition for this problem evaluated at $\tilde{\theta} = \theta$ is:

$$\frac{nw}{\tilde{\theta}(1-\tilde{\theta})(1-\tau)(1-\phi)} - \frac{a(\tilde{\theta})\phi}{\tilde{\theta}(1-\tilde{\theta})(1-\phi)} - a'(\tilde{\theta}) = 0. \quad (35)$$

Let $\frac{nw}{\tilde{\theta}(1-\tilde{\theta})(1-\tau)(1-\phi)}$ be $Q(\tilde{\theta})$ and $\frac{\phi}{\tilde{\theta}(1-\tilde{\theta})(1-\phi)}$ be $P(\tilde{\theta})$. This is a differential equation of the following form:

$$a'(\tilde{\theta}) + P(\tilde{\theta})a(\tilde{\theta}) = Q(\tilde{\theta})$$

which is characterised by the solution

$$e^{\int P(\tilde{\theta})d\tilde{\theta}}a(\tilde{\theta}) = \int e^{\int P(\tilde{\theta})d\tilde{\theta}}Q(\tilde{\theta})d\tilde{\theta} + C$$

Solving this for $a(\tilde{\theta})$ we find that:

$$a(\tilde{\theta}) = \frac{nw}{\phi(1-\tau)}\left(1 + C\left(\frac{1-\tilde{\theta}}{\tilde{\theta}}\right)^{\frac{\phi}{1-\phi}}\right),$$

where C is the constant of integration. Since lower bound values of $\tilde{\theta} = \underline{\theta}_s$ and $a(\tilde{\theta}) = \underline{a}_s$ we can solve for the particular solution. This is

$$a(\tilde{\theta}) = \frac{nw}{\phi(1-\tau)}\left(1 - \frac{(R-nw)}{nw}\left(\frac{\underline{\theta}_s}{1-\underline{\theta}_s}\right)^{\frac{1}{1-\phi}}\left(\frac{1-\tilde{\theta}}{\tilde{\theta}}\right)^{\frac{\phi}{1-\phi}}\right).$$

It is possible to check that the second order condition is satisfied. The second order condition for this problem is:

$$-\frac{1}{\theta^2}\left(nw - \phi(1-\tau)a(\theta)\right) - \frac{2\phi(1-\tau)}{\theta}a'(\theta) - a''(\theta)(1-\tau)(1-\theta)(1-\phi) < 0$$

By inspection it is possible to check that this equation always holds, and hence the function is globally concave.

Proof of Lemma 2

Proof. Let us first consider the region where $R \geq r_p(a)nw$ is satisfied. In this region $\hat{\theta}(a)$. Totally differentiating this equation, and rearranging we find

$$\frac{d\hat{\theta}(a)}{da}\left(R - r(a)nw + (1-\tau)a - \hat{\theta}(a)\frac{dr(a)}{d\hat{\theta}}\right) = (1-\tau)\left(1 - \frac{\hat{\theta}(a)}{\theta_p(a)}(\theta_p(a) + (1-\theta_p(a))\phi)\right). \tag{36}$$

Note that

$$R - r(a)nw \geq 0, \qquad \frac{dr(a)}{d\hat{\theta}} < 0, \qquad \hat{\theta}(a) < \theta_p(a). \tag{37}$$

Hence

$$\frac{d\hat{\theta}(a)}{da} > 0. \tag{38}$$

□

Proof of Proposition 5

Proof. For proving political support it is sufficient to show that wage is non decreasing in ϕ.

$$\frac{\partial L_d}{\partial \phi} : (n+1)\left(-g(\underline{a}_p)\frac{\partial \underline{a}_p}{\partial \phi}(1 - F(\underline{\theta}_p)) - f(\underline{\theta}_s)\frac{\partial \underline{\theta}_s}{\partial \phi}(1 - G(\hat{a}_p)) - \int_{\underline{a}_p}^{\hat{a}_p} f(\hat{\theta}(a))\frac{\partial \hat{\theta}(a)}{\partial \phi}g(a)da\right) \geq 0 \tag{39}$$

This can be demonstrated by showing that

$$\frac{\partial \underline{a}_p}{\partial \phi} < 0, \qquad \frac{\partial \underline{\theta}_s}{\partial \phi} < 0, \qquad \text{and} \qquad \frac{\partial \hat{\theta}(a)}{\partial \phi} < 0. \tag{40}$$

$$\frac{\partial \underline{a}_p}{\partial \phi} = -\frac{nw - \theta_p(\underline{a}_p)R}{\phi^2(1-\tau)} - \frac{\partial \theta_p(\underline{a}_p)}{\partial \phi}\frac{R}{(1-\tau)\phi} < 0 \tag{41}$$

since

$$\frac{\partial \theta_p(\underline{a}_p)}{\partial \phi} = 0. \tag{42}$$

Also

$$\frac{\partial \underline{\theta}_s}{\partial \phi} = -\frac{M - w}{R} < 0. \tag{43}$$

And lastly,

$$\frac{\partial \hat{\theta}(a)}{\partial \phi} < 0 \tag{44}$$

by inspection.

This implies that the equilibrium wage is non decreasing in ϕ. This ensures that the policy enjoys majority support. The inequalities are strict when $w > \underline{w}$. If there is a subsistence sector, then there is no effect of ϕ on the labour demand. The effect of increasing ϕ on total surplus is ambiguous. □

We now construct an example where the credit constraint effect dominates, that is, the surplus maximising ϕ in the economy is less than one. Assume that the distribution of wealth is discrete. There are three classes in the population: the rich, the middle, and the poor of size p_r, p_m, p_p with wealth a_r, a_m, a_p respectively. Assume:

$$q(p_r + p_m) < \frac{1}{n+1}$$

It turns out that if there is a subsistence sector in the economy then it is always surplus enhancing to locally increase ϕ. Hence to make the problem interesting assume that there is no subsistence sector in the economy. The two feasible values for ϕ will be $\{\underline{\phi}, 1\}$. In this economy the credit constraint will be higher with $\phi = 1$.

The change in total surplus as a result of increasing $\underline{\phi}$ to 1 is:

$$\Delta TS = TS(\underline{\phi}) - TS(1) = qp_m(1-\theta)R - (1-\theta)(1-q)(1-\underline{\phi})\big(p_m a_m \lambda_m(\underline{\phi}) + p_r a_r \lambda_r(\underline{\phi})\big) \tag{45}$$

The first term in the expression represents the increase in the total surplus due to replacement of some low type entrepreneurs by high types as a result of access to credit due to reduction in the credit constraint. The second term represents the reduction in the surplus due to destruction of a proportion of assets in case of default due to imperfect judiciary. In the second term $\lambda_(\underline{\phi})$ is the proportion of low type entrepreneurs with wealth i that choose entrepreneurship in equilibrium.

Lemma 5. *If credit constraint worsens as a result of an increase in ϕ from $\underline{\phi}$ to 1, then $\lambda_m(\underline{\phi}) = 1$*

Proof. Assume this is not true. Then there are two possibilities: either $\lambda_m(\underline{\phi}) = 0$ or $0 < \lambda_m(\underline{\phi}) < 1$. Consider $\lambda_m(\underline{\phi}) = 0$. This implies that $\lambda_r(\underline{\phi}) = 0$ since $\frac{\partial \lambda}{\partial a} < 0$. This implies that all entrepreneurs are high types. This contradicts Assumption 1. If $0 < \lambda_m(\underline{\phi}) < 1$ then the interest rate for agents with wealth a_m is:

$$r_s(a_m,(\underline{\phi}) = \frac{\theta R + M - w(\underline{\phi}) - (1-\theta)a_m}{nw(\underline{\phi})\theta}$$

Substituting this into the equation that determines the credit constraint:

$$R - r_s(a_m, \underline{\phi})nw(\underline{\phi}) \geq 0$$

it is easy to check that the credit constrain is decreasing in equilibrium wage. Since the equilibrium wage is monotonically increasing in ϕ and hence the credit constraint with $\phi = 1$ must be lower than the credit constraint with $\underline{\phi}$ but this is a contradiction. □

Hence the change in total surplus simplifies to:

$$\Delta TS = qp_m(1-\theta)R - (1-\theta)(1-q)(1-\underline{\phi})\big(p_m a_m + p_r a_r \lambda_r(\underline{\phi})\big) \tag{46}$$

Now we can back out $\lambda_r(\underline{\phi})$ since we know that the proportion of entrepreneurs in the economy is $\frac{1}{n+1}$.

$$\lambda_r(\underline{\phi}) = \left(\frac{1}{(n+1)p_r} - \frac{p_m}{p_r} - q\right)\frac{1}{1-q}$$

Substituting this into the expression for the change in total surplus, we find that $\Delta TS > 0$ if:

$$R > \frac{1-\underline{\phi}}{q}\left((1-q)a_m + \left(\frac{1}{(n+1)p_m} - 1 - q\frac{p_r}{p_m}\right)a_r\right) \tag{47}$$

This equation ensures that if the credit constraint worsens as a result of an increase in ϕ, the loss of efficiency through reduction in the quality of the pool of entrepreneurs dominates the loss of collateral during recovery with a lower ϕ. The credit constraint worsens if:

$$R - nw(\underline{\phi})r_p(a_m, \underline{\phi}) > 0 > R - nw(1)r_p(a_m, 1)$$

Solving the model to derive the equilibrium wage rate, and interest rate at wealth level a_m for both values of ϕ we find:

$$n\frac{\theta(\theta R + M) + (n+1)p_r q(1-\theta)(\theta R + M - a_r)}{(n+1)(\theta + (1-\theta)p_r q)} - (1-\theta)(1-q)a_m >$$

$$(q+(1-q)\theta)R >$$

$$n\frac{\theta\big(\theta R + M - (1-\underline{\phi})a_r\big)(1-p_m(n+1)) + p_r q(n+1)\big((1-\theta)(\theta R+M) - (1-\theta(1-\underline{\phi}))a_r\big)}{(n+1)\big(\theta(1-p_m(n+1)) + (1-\theta)qp_r\big)}$$

$$-(1-\theta)(1-q)\underline{\phi}a_m \tag{48}$$

Proposition 8. *For any constellation of parameter values for which* (48) *and* (47) *are satisfied,* $\phi = 1$ *is suboptimal.*

Proof. (48) implies that the credit constraint worsens as a result of an increase in ϕ from $\underline{\phi}$ to 1. This implies that there are fewer high type entrepreneurs with $\phi = 1$. (47) ensures that assuming the credit constraint worsens, the change in total surplus is negative for an increase in ϕ from $\underline{\phi}$ to 1. Taken together they imply that the credit constraint worsens, and enough high type entrepreneurs are credit constrained such that total surplus is diminished. □

References

Acemoglu, Daron, "The Form of Property Rights: Oligarchic vs. Democratic Societies," *MIT Department of Economics Working Paper*, September 2003, *3* (34).

Banerjee, Abhijit, "Contracting Constraints, Credit Markets and Economic Development," *MIT Department of Economics Working Paper*, September 2001, (02-17).

_ **and Andrew Newman**, "Occupational Choice and the Process of Development," *Journal of Political Economy*, 1993, *101.*

Bardhan, Pranab, *The Economic Theory of Agrarian Insititutions*, Oxford University Press, 1989.

Besley, Tim, *Principled Agents? The Political Economy of Good Government*, Oxford University Press, 2006.

Besley, Timothy, "Property Rights and Investment Incentives: Theory and Evidence from Ghana," *Journal of Political Economy*, 1995, *103* (5).

Biais, Bruno and Thomas Mariotti, "Credit, Wages and Bankruptcy Laws," *C.E.P.R. Discussion Papers*, 2003, *3996.*

Boyer, Marcel and Jean-Jacques Laffont, "Toward a Political Theory of the Emergence of Environmental Incentive Regulation," *Rand Journal of Economics*, 1999, *30* (1).

Caselli, Francesco and Nicola Gennaioli, "Economics and Politics of Alternative Institutional Reforms," *Quarterly Journal of Economics*, 2008, *123* (3).

Djankov, Simeon, Rafael La Porta, Florencio Lopez-De-Silanes, and Andrei Shleifer, "The Regulation of Entry," *Quarterly Journal of Economics*, February 2002, *117* (1).

Galor, Oded and Joseph Zeira, "Income Distribution and Macroeconomics," *Review of Economic Studies*, 1993, *60.*

Ghatak, Maitreesh and Timothy Guinnane, "Economics of Lending with Joint Liability: Theory and Practice," *Journal of Development Economics*, 1999, *60* (1).

_ , Massimo Morelli, and Tomas Sjostrom, "Entrepreneurial Talent and Occupational Choice," *Journal of Economic Theory*, 2007.

Perotti, Enrico C. and Paolo F. Volpin, "Lobbying on Entry," *CEPR Discussion Paper*, August 2004, *4519.*

Rajan, Raghuram and Luigi Zingales, "The Persistence of Underdevelopment: Institutions, Human Capital, or Constituencies?," *NBER Working Paper*, March 2006, *12093.*