Università degli Studi di Catania

Università degli Studi di Messina

Incentive mechanisms in strategic decision making: Applications in the credit and labor market

Doctoral Thesis

by

Rosaria Distefano

Ph.D. in Economics, Management and Statistics

XXXIV *cycle*

Department of Economics

SSD SECS-P/01

Under the supervision of

Prof. Francesco Reito
University of Catania
Coordinator
Prof. Edoardo Otranto
University of Messina

University of Messina

2020/2021

Acknowledgments

I express my deepest gratitude towards my Supervisor **Prof. Francesco Reito** for having guided and encouraged me since the beginning of my academic path.
Thank you for having faith in me.

Rosaria Distefano

Contents

Introduction

In incentive theory, the main focus is on how a principal or principals can motivate self-interested agents in order to maximize their expected payoffs.

The principal-agent paradigm applies to a waste range of circumstances. In the credit market, for example, lenders/banks with considerable market power try to extract the highest rent possible from the projects of borrowers. In turn, the latter can implement perverse strategies which, in some circumstances, can even reduce the profitability of their investments. This may have a deleterious impact on social welfare. Similarly, in the labor market, private information on skills or abilities of workers, or the unobservability of their effort levels, can influence the structure of optimal contracts and have negative consequences on the performance of organizations. In these settings, contracts entail a principal-agent problem, and information asymmetries can create frictions and necessitate a system of incentives to be addressed.

This thesis analyzes incentive mechanisms and optimal responses in agency relationships in different asymmetric-information settings.

The thesis begins with a model of the credit market, with a single lender and many potential entrepreneurs, who need external funds to carry out their projects. Some of the borrowers may have the propensity to use the loan received for other illegal or non-contractible purposes, diverting funds from the project financed. Then, an extension of a standard business lending model to the consumer credit market is presented. Agents (households) need loans to bring forward the purchase of a durable good and differ in the likelihood of paying back their debt. The third chapter applies the banking model to the occupational choice of an individual between entering the labor market as a wage worker or as a self-managed entrepreneur. The final chapter presents a labor-market model and analyzes the incentive effects when workers are heterogeneous in their skill levels, unobservable to the employer. In this model, some workers have other-regarding preferences and, in particular, are envious of their more-talented colleagues. Envy can lead to a utility loss for the less-talented and affect the system of incentive contracts.

In the thesis, the design of incentive-compatible contracts is analyzed, and policy implications are drawn. In the first model, the tendency of borrowers to divert funds is increasing in the interest rate that the lender charges. The optimal contract would require borrowers to allocate a certain amount of resources to their projects. In the consumer-credit setup, adverse selection can result in a credit-rationed and simultaneously a socially

inefficient equilibrium. In the third chapter, the occupational alternatives available for the individual may depend on both the level of reservation utility and the loan profitability of the bank. In the labor market setting, both the size of the ability gap among workers and the magnitude of the envy cost of the less-skilled can have ambiguous effects on the payoffs of workers.

As for efficiency, the structure of equilibria often does not lead to the socially optimal outcome. In the business lending model, the bank may find it profitable to let borrowers divert funds towards less-productive purposes. In the consumer credit model, it is not just the volume of lending that is inefficient but also the quality of borrowers that demand loans. In the occupational-choice model, the lack of initial wealth may prevent projects with a positive net present value from being undertaken. In the labor-market model, wage disparities among workers with slightly different abilities can yield a larger utility loss.

For each model proposed, the need and the desirability of regulatory interventions are analyzed. As it will be shown, policy can often increase social efficiency but not always achieve the first best.

Chapter 1

The social cost of playing by the rules in the credit market

We present a model of the credit market under imperfect information, with a lender and many would-be entrepreneurs who need external funding for their projects. Some borrowers may have the incentive to divert part of the loan received to other, illegal or non-contractible, uses. We first show that the equilibrium is more likely to be efficient when there is a high proportion of potential diverters. Another result is that, if diversion output is included in the social well-being function, equilibrium welfare can be higher than under symmetric information. When there is inefficiency, a regulatory intervention can be welfare improving but, the cost and desirability of the policy depend on whether the proceeds from diversion are classified as a contribution to social welfare or not.

Introduction

In the credit market, borrowers are expected to adopt all kinds of strategies to outwit creditors and maximize the value or utility of the loan money received. One case is when borrowers have access to business or personal activities other than the projects financed, and creditors have considerable market power that reduces the profitability of formal investment. This is the scenario we explore in this paper. We propose a simple banking model under imperfect information with a single lender and many would-be entrepreneurs who need outside finance to undertake their investment projects. Some borrowers may have the incentive to divert part of the loan received to other private and less-productive uses. We will refer to this misallocation of resources as loan diversion.

Loan diversion is a typical moral-hazard problem, which is generally more relevant for small and medium-sized enterprises because of their informational opacity, lack of initial capital, and risky nature (Berger and Udell, 2002). Here, we model the relationship between cashless entrepreneurs and a profit-maximizing lender. Based on the empiri-

cal observations of Petersen and Rajan (1995), Petersen and Rajan (1995), Beck et al. (2004), Petersen and Rajan (1995) Beck et al. (2004), and Howorth and Moro (2012), the rent-extraction ability may result from specific institutional and industry conditions or alternative lending arrangements, such as relationship or shadow banking (usually provided by non-standard financial institutions).

It is useful to illustrate the logic of the model by means of a simple example. Consider many farmers in a rural context. Each is endowed with two different and independent projects, one formal and one informal. Farmers have no liquid wealth, but can borrow from a local lender or microfinance bank. The formal project involves the cultivation of a piece of land for the production of a cereal, and requires a specific quantity of pesticides, chemicals and protective devices to prevent infestations or other crop damages. A lower application increases the possibility of insect or animal attacks and reduces the probability of a good harvest. The informal project consists of the cultivation of an illicit drug crop, and thus is impossible to formalize in financial contracts. This production requires a variable investment and yields the farmer some private and non-transferable benefits. Farmers are distinguished in compliant and (potentially) defiant: compliant farmers carry out only the formal project, while defiant farmers may be tempted to use part of the loan for the informal investment. The lender cannot observe either the farmer type or the formal planting investment, and so whether the crop has been contaminated and destroyed by pests or wild animals because the appropriate combination of pesticides and protection was missing.

In the paper, we first show that, contrary to expectations, the higher (lower) the proportion of defiant types, the higher (lower) the payoff obtained by both types of borrowers. The reason is that the fraction of loan diverted is increasing in the interest rate. So, when there is a high proportion of potential diverters, the lender is forced to hedge against the risk of massive default by charging a low interest rate, at the cost of giving up more rent to borrowers. The novel result is that, thanks to the propensity of defiant types, even compliant borrowers are able to keep some of their project's rent, despite the monopolistic bank. In our paper, it is "good" types who may end up benefiting from the wrong behavior of "bad" types and from their numerosity. This result is in contrast to most of the literature on credit market imperfections. For instance, in Akerlof (1978), an equilibrium in which only lemons are sold is more likely to occur when their proportion is relatively high. In Stiglitz and Weiss (1981) and de Meza and Webb (1987), a higher fraction of high-risk borrowers leads to an increase in the pooling interest rate, which is also charged to low-risk borrowers. The same conclusion is reached in Minelli and Modica (2009), where the credit sector is in monopoly.

For the same reasons described above, in terms of social welfare, the equilibrium is more likely to be efficient when there is a large number of defiant entrepreneurs. Specifically, if their proportion is higher than a certain threshold, the interest rate is such that no fraction of the loan will be diverted, and the equilibrium is efficient. Conversely,

the bank will tend to charge increasingly higher interest rates for lower proportions of defiant types. The latter will respond by shifting more and more resources in informal investments so that loan diversion will result in socially inefficient use of resources.

The improper use of financial resources is one of the main causes of loan default. The data in Table 1 appears to support our result that interest rates are positively correlated with the proportion of funds diverted and, thus, to the share of non-performing loans.[1]

Table 1. Lending interest rate and non-performing loans

Country	Lending interest rate (%)	Non-performing loans (%)
Congo	26.0	23.1
Tajikistan	23.5	20.4
Iraq	12.4	16.2
Kenya	12.0	14.1
Armenia	11.6	6.6
Lebanon	10.5	15.2
Zambia	9.5	11.6
India	9.2	7.9
Bangladesh	8.3	7.7
China	4.3	1.8
United States	3.5	1.1
Switzerland	2.6	0.7
Italy	2.3	4.4
Netherlands	1.5	1.9
United Kingdom	0.5	1.2

Source: Bankscope
Countries ranked by lending interest rate and non-performing loans. In 2020, the average of bad loans based on 102 countries was 5.86%.

Another result of the model is that, when diversion utility is factored in the social well-being function, equilibrium welfare can be higher than under symmetric information. This result is obtained when diversion is particularly "attractive", or when the productivity of formal projects is relatively small. The logic is very simple. If information were symmetric, the profit-maximizing lender would not allow any diversion activity, regardless of its value for borrowers. This implies that, if diversion is not observable and the pro-

[1] In a report of the European Central Bank, in 2020, non-performing loans of the European banks worth over 550 billion euro, nearly 3% of their total loan investment.

ductivity of formal projects is relatively less valuable than informal projects, equilibrium welfare can be higher than under perfect information.

To our knowledge, the only theoretical works in which equilibrium welfare can be larger than under full information are Shneyerov and Wong (2011) and Lauermann (2012). In an analysis of bilateral trade, Shneyerov and Wong (2011) show that, when sellers and buyers have private valuations, social welfare can be larger than under full information. The reason is that asymmetric information can deter entry in the market and, if there are matching frictions, such an effect can be welfare enhancing. Similarly, and independently, Lauermann (2012) comes to the same conclusion, but extends the analysis to the effect of different levels of frictions and concludes that whether asymmetric information is welfare superior or not depends on the exit rate of sellers, that is on the individual specific discount rate. As in their models, we consider a "seller" making take-it-or-leave-it offers with full bargaining power, but the source of trade inefficiency is different: in our case, it simply depends on the profit-maximizing objective of the lender, which would prefer to perfectly observe the behavior of borrowers, neglecting the benefits they would derive from diversion.[2]

When the equilibrium is inefficient, a regulatory intervention, for instance an interest rate cap, can be welfare improving. However, the cost and desirability of the policy depend on whether the proceeds from informal projects are counted as a contribution to the welfare of the economy or not. The question of whether introducing illegal activities in the statistical measures of national accounts has long been debated by academics and policymakers. In 2010, the European System of National and Regional Accounts (ESA 2010) established a new methodology that allows EU countries to include estimates of part of the black economy (prostitution, drug trafficking, counterfeit goods, smuggled cigarettes) in their gross domestic product to reduce the underground or informal sectors. In a report of Eurostat, published in 2018, it is made clear that: "Illegal transactions are treated the same way as legal actions. Illegal transactions are those that are forbidden by law. Illegal economic actions are transactions only when the institutional units involved enter the actions by mutual agreement".[3] For example, according to the OECD, the inclusion of illegal activities in national accounts has produced an increase in the GDP of 1% in Italy, 0.9% in Spain and, 0.6% in the United Kingdom. Among the reasons that push institutions towards this direction, there are the need for a more accurate and reliable measure of the GDP (including all economic activities, whether legal or illegal), comparability issues across countries, and inconsistency in national accounts.

The main prediction of the model is that credit is more likely diverted when there is a high proportion of borrowers who comply with the rules, that is essentially the clause

[2]This conclusions can be related to the strand of literature on the welfare costs of asymmetric information. See the recent contribution by de Meza et al. (2021).

[3]Handbook on the compilation of statistics on illegal economic activities in national accounts and balance of payments - Sections 3.4-3.6.

that loan money cannot be generally used for other purposes than those stated in the financial contract. This result is supported by the data in Table 2. In the presence of high bank concentration, countries with a high crime index report low levels of non-performing loans. For instance, Argentina has a very high crime index (63.8), associated with a share of non-performing loans below the average (3.9%). Whereas, in countries like Cyprus and the United Arab Emirates, where the crime index is low (31.3 and 15.2), the share of bad loans is above the average (15.0% and 8.2%).

Table 2. Crime index, non-performing loans and bank concentration

Country	Crime index	Non-performing loans (%)	Bank concentration
Argentina	63.8	3.9	41.4
Guatemala	58.7	1.8	65.9
Malaysia	57.3	1.6	55.4
Mexico	54.2	2.4	49.4
France	52.0	2.7	57.1
Ethiopia	49.3	3.4	80.8
Sweden	48.0	0.5	91.1
United States	47.8	1.1	34.8
United Kingdom	46.1	1.2	49.1
Italy	44.9	4.4	71.1
Canada	41.9	0.5	60.5

Source: Bankscope, World Population Review.
Countries ranked by crime index. The crime index is expressed per 100,000 people.
Bank concentration is the percent of bank assets held by top three banks.

In terms of our stylized model, loanable funds might also be diverted into non-contractible socially acceptable ends, such as schooling for children or medical expenditures, whose benefits are not directly observable or measurable and thus cannot be specified in financial contracts. When diversion output is excluded from welfare accounting and is relatively small, a policy can achieve a Pareto efficient equilibrium, whereas when it is included in welfare, an improvement is often not possible and, in any case, this would lessen the need for intervention. This result suggests that the decision to factor in non-contractible activities in welfare may serve as a basis for maintaining the status quo and not changing the regulatory framework. Moreover, if the informal project is highly productive, the socially optimal fraction of loan diverted can even be positive. In such cases, a policy is never Pareto improving, despite social optimality would require a positive informal production.

Related literature

The article by Bhat (1971) is one of the first to offer a detailed analysis of loan diversion. He identifies several reasons for borrowers to misuse financial funds, such as inadequate assessment of project risk or weak correlation between investment required and amount lent. Von Pischke and Adams (1980) add that the fungibility of money makes it easier to shift resources from productive to non-productive activities. Our paper is in line with the theoretical literature on the welfare implications of credit market frictions. On this topic, Burkart and Ellingsen (2004) analyze a banking model in which borrowers can divert investment resources to personal uses. They argue that trade credit can be an efficient substitute for financial loans, thanks to input illiquidity and the monitoring advantage of suppliers over banks. Similarly, Repullo and Suarez (2000) consider both market and bank finance, and assume that the latter involves a higher monitoring activity that may reduce the moral-hazard problem and thus loan diversion. They show that the choice between market and bank finance depends on the ratio of internal funds to investment needed. Bougheas (2004) shows that if the bank's monitoring activity is missing, firms may have the incentive to misuse resources intended to finance investments in intangibles, like R&D. Madestam (2014) focuses on the coexistence of formal and informal lenders and shows that the inability of the legal system to enforce contracts may increase loan diversion and even lead to credit rationing. Our results are also close to those of Navajas et al. (2003), who analyze the impact of interest rates on the strategic behavior of borrowers and show that "unfair" rates may increase the incentive to misallocate financial flows.[4]

On the empirical side, measuring loan diversion is not straightforward or even practically feasible. Many papers focus on the legal protection of investors and try to provide indirect proxies and indicators of diversionary tactics. La Porta et al. (2002) find that the lower the legal protection, the lower the valuation of firms on financial markets, and this may be interpreted as a signal of resource misallocation. Durnev and Kim (2005) report that, as protection becomes weaker, managers tend to divert a higher share of returns to their own advantage, at the expense of shareholders. Other papers find evidence that loan diversion is more widespread in less-developed contexts and for small enterprises, which usually tend to have loose accounting standards. Garikipati (2013) and Mungai et al. (2014) show that, although microfinance institutions provide credit and financial support for income-generating activities, acute poverty may force households to use borrowed money for basic family needs, such as consumption goods, health care, children education and repayment of old loans. In addition, poor property rights can induce borrowers to misallocate funds to protect the property (Besley et al., 2012).

Our conclusions are also consistent, empirically, with the findings in Banerjee et al. (2015), which show that loan diversion is significantly correlated with the interest rates charged by local monopolistic moneylenders. On the topic of market structure, part of the

[4]See Coco (2000) for a brief review on diversionary models.

literature reports that the interest rates charged by most microfinance institutions are far higher than normal bank rates (Kar and Swain, 2014). The global average interest rate is around 35-40%, and it is not uncommon to observe rates above 80%. The riskiness of clients and small-sized loans are the most cited reasons for the strong difference between microcredit and standard bank interest rates, though default can usually explain only a small part of observed spreads (Banerjee, 2013). This is the reason why some recent theoretical and empirical works claim that the focus of many microfinance institutions has shifted from social outreach to financial performance, a process known as mission drift (Madajewicz, 2003; Cull et al., 2007). In terms of our paper, this means that microcredit banks may simply be taking advantage of their monopoly position. Other empirical evidence on monopolistic, for-profit microfinance institutions is reported in Armendáriz and Szafarz (2011), and de Quidt et al. (2012). Both papers also present a theoretical model with a single lender (for other theoretical papers with profit-motivated microfinance institutions, see Guha and Chowdhury (2013); Caserta et al. (2018)). Loan diversion is a serious and widespread problem of moral hazard that limits and tightens credit access. In some countries, the government has introduced measures of personal identification through fingerprints of loan applicants to subject future lending conditions on borrowers' credit history, as documented by Giné et al. (2012).

The next section introduces the model. Then, we describe the equilibrium and discuss the welfare and policy implications. Finally, some brief conclusions are drawn.

1.1 The Setup

Consider a one-period, risk-neutral credit market with a large number of cashless would-be entrepreneurs. At the beginning of the period, each entrepreneur is endowed with two different and independent projects, which we will refer to as formal and informal projects. The formal project requires 1 unit of capital to be properly executed, and yields a deterministic output, y. The informal project requires a variable amount of capital and it too yields a deterministic output, $u_D(\delta)$, increasing and concave in the amount of funds invested. Entrepreneurs have no initial wealth and, to undertake their projects, they need external financing, which can be provided by a single lender/bank.[5] The returns of informal projects are either illicit or unobservable, so the bank will only consider loan applications for formal projects. We restrict attention to debt contracts, which specify the loan advanced, equal to the capital cost of 1, and the repayment (principal plus interest), r, made in case of project success. Implicitly, the contract contains the clause that loan money cannot be used for other purposes than the formal investment. To simplify, we consider limited liability on the part of borrowers, and a risk-free rate normalized to 0.

Entrepreneurs are of two types, differentiated by their attitudes towards the organization and implementation of the two projects. Specifically, there are potentially "defiant"

[5] In Remark 2, we will briefly discuss the case of perfectly competitive lenders.

(D) borrowers, who have the ability to scale down the investment size of the formal project, at the cost of higher failure risk. That is, they can invest a fraction, δ, of the loan in the informal project, deriving private utility $u_D(\delta)$, and lowering the probability of success for the formal project down to $1-\delta$. In other words, loan diversion makes the formal project stochastic. The other borrowers are "compliant" (C), in that they do not possess the downsizing ability and do not derive utility from the informal activity, so they are expected to undertake only the formal project.[6] This setup is consistent with the argument of Hart and Moore (1998), and Diamond (2004), in which earnings are not contractible when borrowers can easily divert or hide their cash flow. In our case, loan diversion is more (infinitely) costly for C types and less for D types.

We will use the following form for the diversion utility of D types:

$$u_D(\delta) = \tfrac{\theta}{2}[1-(1-\delta)^2], \quad \text{with } \theta > 0. \tag{1.1}$$

The parameter θ can be interpreted as the utility weight or productivity parameter of the informal project and can reflect several aspects, such as profitability, social acceptance, stigma, security, and legal risks, associated with this activity. The assumption in (1) is made to allow for an intermediate fraction of funds diverted other than the corner solutions. The functional form of the diversion utility is similar to that in the paper by Repullo and Suarez (2000) to make the analysis more tractable. The key difference between our and their setup is that they consider a single type of borrowers, so there is no hidden information.

The model involves both hidden action and hidden information. The bank cannot distinguish C from D types, but knows their proportions, $1-\lambda$ and λ. We do not consider monitoring activity during the investment stage of the formal project, so the quantity of funds diverted by D borrowers is unobservable.[7] But, at the end of the period, the bank can perfectly observe the formal project output. To simplify, we assume that, for institutional constraints, no penalty can be imposed on defaulters, so contracts based on *ex-post* verification would be equivalent to the debt structure introduced above (as long as the bank can pre-commit to an *ex-ante* fixed payment rule).

The timing of the game is: 1) nature determines y, θ and λ; 2) the bank offers the contract, loan-repayment, $(1, r)$; 3) C and D individuals decide whether to accept or not; 4) if D individuals accept, they choose δ and obtain $u_D(\delta)$ from informal projects; 5) the formal project output is realized and state-contingent payments are made.

[6] The combination of fixed loan size and potential diversion of investment funds is used, among others, by Carter (1988), Repullo and Suarez (2000), Wydick (2001), and Diamond (2004).

The justification for our assumption of fixed loan size may be that formal projects must meet some minimum technical requirements to qualify for financing, without compromising the basic standards established for the legal or material operation of the business. So, C types are simply unable or unwilling to adopt those riskier production technologies.

[7] Using data on mills' contracts in the colonial Taiwan, Koo et al. (2012) show that, when monitoring is possible, high interest rates can deter borrowers from improper uses of funds.

Had we symmetric information, the bank would force D borrowers not to divert ($\delta = 0$) and set $r = y$ (a higher repayment would not be accepted). Each borrower would obtain the reservation payoff, $\pi_C(y) = \pi_D(y) = 0$, and the bank the full-information profit, $\pi_B(y) = y - 1$, which we assume is positive (otherwise, the bank would not finance any project). Under perfect information, the bank is able to extract all the rent from every contract.

The result under perfect information may provide a rationale for why D borrowers may want to activate the informal project if information is asymmetric, as shown below.

1.2 Equilibrium

In what follows, we describe the equilibrium properties under the pooling contract, debt repayment, r (in Remark 1 below, we will show that a separation, in which D types are excluded, cannot be achieved).

The expressions for the expected profits by C and D entrepreneurs are:

$$\pi_C(r) = y - r; \tag{1.2}$$

$$\pi_D(r) = u_D(\delta) + (1 - \delta)(y - r). \tag{1.3}$$

The fraction of funds diverted derives from the maximization of (3), yielding

$$\delta = 1 - \tfrac{y-r}{\theta}, \tag{1.4}$$

so, as r increases, D types will choose to divert more resources (the second-order condition is satisfied).

The expression for the lender's expected profit (per borrower) is

$$\pi_B(r) = \lambda(1 - \delta)r + (1 - \lambda)r - 1, \tag{1.5}$$

where, if the borrower is of type C (D), the lender obtains r with probability 1 ($1 - \delta$). The lender maximizes (1.5), taking into account the participation constraints and the reaction function of D types.

Remark 1. With a large enough penalty in case of default, equal or above the diversion productivity parameter θ, the lender would be able to separate the two types of borrowers. Denote the penalty by τ. The first-best profit is obtained when the participation constraints of C and D types are satisfied with equality, that is $\pi_C(r, \tau) = y - r = 0$ and $\pi_D(r, \tau) = u_D(\delta) + (1 - \delta)(y - r) - \delta\tau = 0$. The system is satisfied when $r = y$ and $\tau = \theta/2$. This solution requires that the penalty must be large enough, and this might be difficult to implement for legal or other institutional restrictions.

There are three relevant levels for the loan repayment charged by the moneylender. These will be derived and discussed below.

The first is the repayment such that D borrowers have no incentive to divert any fraction of the loan, that is $\delta = 0$, which gives

$$r = y - \theta \equiv r_{\min}.$$

If $r < r_{\min}$, D borrowers would still choose not to divert, but the expected profit to the bank would be lower. Hence, $r_{\min}$ is the lowest repayment in this model setup. Note that $r_{\min}$ can be positive or negative depending on whether the diversion productivity parameter is lower or higher than the formal project's output (with no diversion on the part of D types). Specifically, in the latter case, $r_{\min} < 0$, and a repayment for which there is no diversion is not possible. The reason is that, when θ is relatively high, diversion is particularly attractive, so it is never profitable for the lender to charge a repayment such that the informal project is not worth undertaking. Below, we will show that for values of θ even lower than y, an equilibrium at $r_{\min}$ may not exist.

The second repayment derives from the maximization of the expected profit by the bank in the range where $\delta \in (0, 1)$. The solution is

$$r = \tfrac{\lambda y + (1-\lambda)\theta}{2\lambda} \equiv \widetilde{r},$$

with $\partial\widetilde{r}/\partial\lambda < 0$, so the higher the proportion of D types, the lower the repayment. Similarly, $\widetilde{r}$ is increasing in θ, as the higher the productivity of informal projects, the higher the fraction diverted and thus the repayment burden on formal loans.

The third repayment is such that $\delta = 1$, which gives

$$r = y \equiv r_{\max}.$$

At this repayment, $\pi_C(r_{\max}) = 0$, so C borrowers obtain an expected profit of 0, and D types take all the loan money and run. This means that $r_{\max}$ is the highest possible repayment level, and implies the following

Lemma 1. *In equilibrium, there is no adverse selection, as the lender will never set a repayment such that C types drop out of the market.*

The equilibrium repayment will depend on the relative proportion of C types. The lender must take into account that, depending on λ, $\widetilde{r}$ may be either lower than $r_{\min}$ or higher than $r_{\max}$. We derive two critical thresholds for λ:

$$\widetilde{r} = r_{\max} \Leftrightarrow \lambda = \tfrac{\theta}{y+\theta} \equiv \underline{\lambda};$$

$$\widetilde{r} = r_{\min} \Leftrightarrow \lambda = \tfrac{\theta}{y-\theta} \equiv \overline{\lambda}.$$

with $\underline{\lambda} < \overline{\lambda}$. Hence, we have the following three possibilities.

a) $\lambda < \underline{\lambda} \implies \widetilde{r} > r_{\max}$. In this case, $\widetilde{r}$ would be above the highest acceptable by C types, so the equilibrium repayment is $r_{\max}$

b) $\underline{\lambda} < \lambda < \overline{\lambda} \implies r_{\min} < \widetilde{r} < r_{\max}$. In this case, $\delta \in (0,1)$ and the maximizing repayment is $\widetilde{r}$.

c) $\lambda > \overline{\lambda} \implies \widetilde{r} < r_{\min}$. For any repayment lower than $r_{\min}$, D types would still choose $\delta = 0$, so it is profit maximizing to set $r_{\min}$.

Using δ in (1.4), we have $\pi_D(r) - \pi_C(r) = (\theta + r - y)^2/2\theta = 0$ when $r = r_{\min}$, and $\pi_D(r) > \pi_C(r)$ when $r > r_{\min}$. Hence, we have the following

Lemma 2. *In equilibrium, if $r = r_{\min}$, all borrowers will earn the same payoff, whereas if $r > r_{\min}$, D types will obtain a payoff higher than C types.*

The threshold $\overline{\lambda}$ is equal to 1 when $y = 2\theta$. Hence, if $y < 2\theta$ the inequality $\widetilde{r} > r_{\min}$ always holds and the bank would never find it profitable to choose $r_{\min}$. If, instead, $y > 2\theta$, an equilibrium with $r = r_{\min}$ is possible. Hence, depending on y, there can then be either two or three relevant intervals of λ characterizing the equilibrium. These possibilities will be discussed in the following three sub-sections.

1.2.1 Equilibrium at $r_{\max}$

When $\lambda \in (0, \underline{\lambda}]$, $\widetilde{r} \geq r_{\max}$ and the repayment charged by the bank is $r = r_{\max}$, such that $\delta = 1$.

In equilibrium, the bank receives the loan repayment from C borrowers (D types divert the entire loan), and its expected profit is

$$\pi_B(r_{\max}) = (1 - \lambda)y - 1, \tag{1.6}$$

linear and decreasing in λ, as in figure 1*a* and 1*b*. The profit does not depend on the diversion utility weight, θ, so we do not have to distinguish between the cases $y > 2\theta$ and $y \leq 2\theta$.

The payoff by C borrowers at $r_{\max}$ is

$$\pi_C(r_{\max}) = 0,$$

whereas D types obtain

$$\pi_D(r_{\max}) = \tfrac{\theta}{2}, \tag{1.7}$$

corresponding to the highest possible output from diversion and no output from the formal project.

The bank extracts all the rent from the formal projects of C types, and no rent from D types, as they do not even activate them. By shifting resources into informal investments, D borrowers are able to earn a positive payoff, although a relatively low λ, that is a high average quality of the pool of entrepreneurs, is the worst possible scenario for them, and especially for C types.

1.2.2 Equilibrium at $\widetilde{r}$

If $y > 2\theta$, there is an interval, $(\underline{\lambda}, \overline{\lambda})$, in which $r_{\min} < \widetilde{r} < r_{\max}$, so it is profit maximizing for the lender to choose $\widetilde{r}$, such that $\delta \in (0,1)$. When $y \leq 2\theta$, the interval in which an equilibrium at $\widetilde{r}$ exists is $(\underline{\lambda}, 1)$, and we will show that this case implies that the equilibrium described in the following subsection (1.2.3) is never possible.

When $r = \widetilde{r}$, in equilibrium, the lender obtains

$$\pi_B(\widetilde{r}) = \tfrac{[\lambda y + (1-\lambda)\theta]^2}{4\lambda\theta} - 1, \tag{1.8}$$

which can be shown is a non-rectangular hyperbola with vertical asymptote at $\lambda = 0$, and with the right branch decreasing in the range $(\underline{\lambda}, \overline{\lambda})$. A numerical example is illustrated in figure 1a. As expected, the higher the proportion of D types, the lower the lender's profit.

To restrict the analysis, we assume that $\theta < y^2/4$, which implies that $\pi_B(\widetilde{r})$ is always positive, as in the example of figure 1a and 1b, and an equilibrium at $\widetilde{r}$ exists, whether in the interval $(\underline{\lambda}, \overline{\lambda})$ or $(\underline{\lambda}, 1)$. This assumption implies that also the equilibrium at $r_{\max}$, analyzed in subsection (1.2.1) exists in the interval $(0, \underline{\lambda}]$.

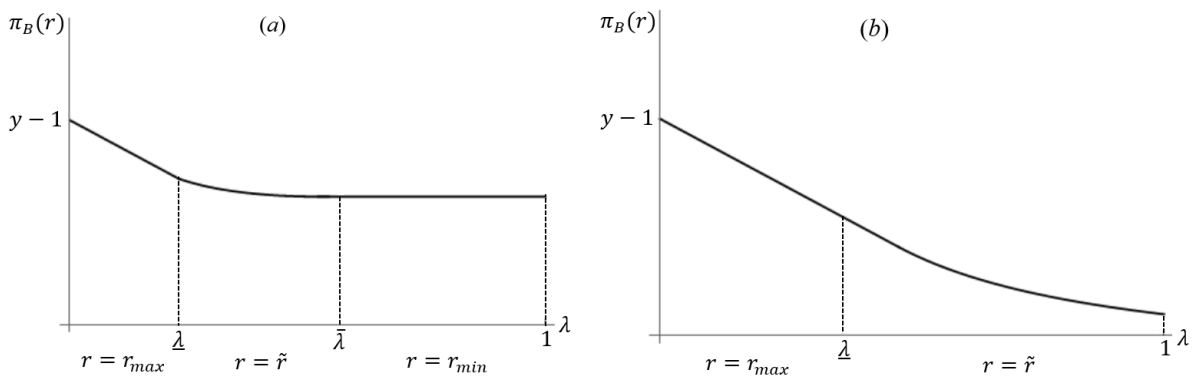

Figure 1. Bank's profit. Numerical example.
a) $y > 2\theta$. Parameters: $y = 6$; $\theta = 2.5$.
b) $y \leq 2\theta$. Parameters: $y = 6$; $\theta = 4.5$.

The equilibrium payoff by C borrowers is

$$\pi_C(\widetilde{r}) = \tfrac{\lambda y - (1-\lambda)\theta}{2\lambda}, \tag{1.9}$$

equal to 0 if $\lambda = \underline{\lambda}$, increasing in λ, and equal to θ if $\lambda = \overline{\lambda}$, as in figure 2$a$. Hence, in this interval C types obtain a positive rent.

At $\widetilde{r}$, the diversion fraction of funds diverted by D types is

$$\delta = \tfrac{1}{2}\left(1 + \tfrac{1}{\lambda} - \tfrac{y}{\theta}\right) \equiv \widetilde{\delta}, \tag{1.10}$$

decreasing in λ, so we obtain the following key result for our analysis.

Lemma 3. *The higher the proportion of D entrepreneurs, the lower the fraction of loan diverted.*

Using $\widetilde{\delta}$, the equilibrium payoff of D borrowers is

$$\pi_D(\widetilde{r}) = \frac{\theta}{2} + \frac{[\lambda y - (1-\lambda)\theta]^2}{8\lambda^2\theta}, \tag{1.11}$$

positive and above $\pi_C(\widetilde{r})$ for Lemma 2. The shape of the function $\pi_D(\widetilde{r})$ is similar to $\pi_C(\widetilde{r})$ in (9), as shown in figure 2*a*.

The utility from diversion,

$$u_D(\tilde{\delta}) = \frac{\theta}{2} - \frac{[\lambda y - (1-\lambda)\theta]^2}{8\lambda^2\theta},$$

reaches a maximum when $\lambda = \underline{\lambda}$, is decreasing in λ, and is equal to 0 when $\lambda = \overline{\lambda}$.

In equilibrium, the profit of the bank is decreasing in λ, and the payoff of both D and C types is increasing, as shown in figures 1 and 2. Compared to the case of subsection (1.2.1), the bank can take advantage of cross-subsidization between the two types of entrepreneurs in a pooling equilibrium. The interest rate decreases in λ and, while D types are able to shift part of their debt liabilities, C types bear most of the repayment burden.

1.2.3 Equilibrium at $r_{\min}$

If $y > 2\theta$ and $\overline{\lambda} < 1$, we can obtain an interval, $[\overline{\lambda}, 1)$, where $\widetilde{r} < r_{\min}$. In this case, the equilibrium repayment is $r_{\min}$, such that $\delta = 0$. When, instead, $y \leq 2\theta$, the equilibrium at $r_{\min}$ does not exist.

The expected profit to the lender is

$$\pi_B(r_{\min}) = y - \theta - 1, \tag{1.12}$$

With $r = r_{\min}$, D borrowers do not divert any fraction of the loan, and this is why the profit in (3.6) does not depend on λ and the probability of project's success is equal to the prior, 1.

The equilibrium payoff of both C and D types is

$$\pi_C(r_{\min}) = \pi_D(r_{\min}) = \theta. \tag{1.13}$$

As this equilibiurm exists when $y > 2\theta$, borrowers obtain less than their project's output. But, due to the relatively large presence of D types in the population, all types of borrowers can earn a positive payoff, more than under perfect information where $\pi_C(y) = \pi_D(y) = 0$.

Therefore, when there is a high number of individuals who do not play by the rules, and thus do not comply with the contract requirement on loan diversion, the lender is

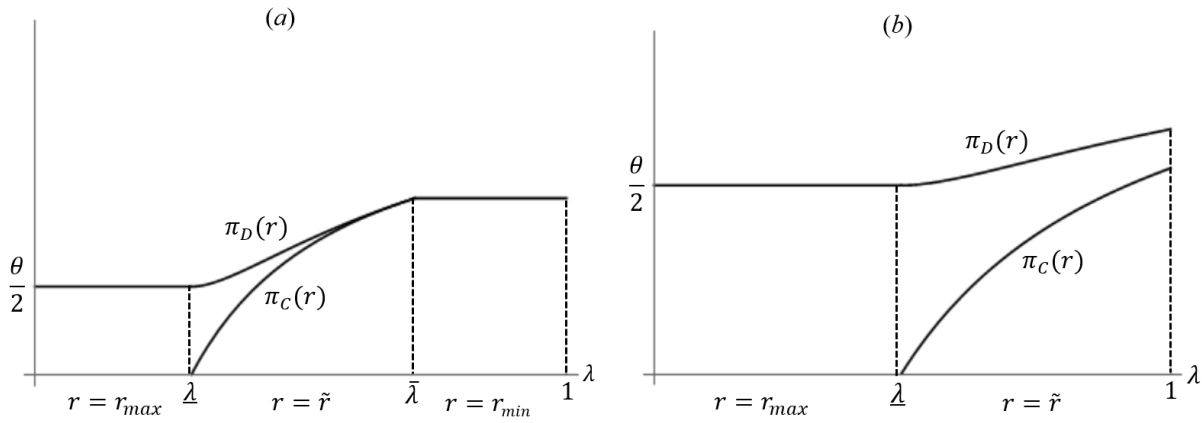

Figure 2. Payoffs of C and D borrowers. Numerical example.
a) $y > 2\theta$. Parameters: $y = 6$; $\theta = 2.5$.
b) $y \leq 2\theta$. Parameters: $y = 6$; $\theta = 4.5$.

forced to charge the lowest interest rate. Setting a repayment higher than $r_{\min}$ is not profitable when most borrowers are prepared to divert their loans and thus can lead to a relevant increase in the default rate. A relatively high θ means that diversion is highly profitable for D types, so the lender will try to compensate for the risk of default by choosing a higher repayment, either $\widetilde{r}$ or $r_{\max}$.

The discussion of this section leads to the following result.

Proposition 1. *The higher the proportion of D entrepreneurs, the higher the surplus that both types of borrowers can obtain.*

As noted in the introduction, this result is in contrast to the main body of literature on credit markets under informational asymmetries. In our case, the propensity of "bad" borrowers to divert their loans forces the bank to charge a low lending rate, and this benefits also "good" borrowers who obtain a payoff larger than under symmetric information. This effect is intensified when bad borrowers are particularly numerous. Except when the number of D types is very low, both types of borrowers earn a positive expected payoff, despite the presence of a monopolistic lender. This would not be possible under perfect information.

Remark 2. Diversion would be possible also with a perfectly competitive credit market. Consider initially that D borrowers choose not to divert their loans. In this case, the equilibrium repayment derives from the zero profit of each lender, $\pi_B(r) = \lambda r + (1-\lambda)r = 1$, yielding $r = 1$. However, if D borrowers can divert funds, and if $r_{\min} < 1$, that is $\theta > y - 1$, the equilibrium repayment derives from $\pi_B(r) = \lambda r + (1-\lambda)(1-\delta)r = 1$. In this case, using the reaction function in (1.4), the equilibiurm repayment[8], r, can be

[8] $r = \frac{(1-\lambda)\theta + \lambda y - \theta\sqrt{\left[1+\left(\frac{y}{\theta}-1\right)\lambda\right]^2 - \frac{4\lambda}{\theta}}}{2\lambda}$

higher than r_{min} if $\theta > y - 1$.[9] Thus, the misallocation of funds can also occur under a competitive equilibrium. Using the setup of this paper, with perfectly competitive banks, the algebraic expressions would become extremely cumbersome, but it can be shown that the conclusions would not differ from the "expected" results: the fraction of loan diverted would be increasing in λ, whereas the payoff of both C and D borrowers decreasing (the complete description is available upon request).

Remark 3. In the paper, we focus on debt contracts and do not consider equity financing. In our case, with asymmetric information and where monitoring is unfeasible, equity contracts between the bank and each entrepreneur would leave the theoretical conclusions unchanged. In the very brief presentation that follows, consider that s is the share of the firm's return that accrues to the bank. We maintain the assumption that, in the partnership, the member "bank" has all bargaining power and that no penalty can be imposed if the output realized with a member who turns out to be a D type is 0. We assume that the bank cannot participate in the diversion activity and, to simplify, we restrict the discussion to the case $\theta < y$.

The contract must satisfy the participation constraints of C and D types,

$$\pi_C(s) = (1-s)y \geq 0;$$
$$\pi_D(s) = u_D(\delta) + (1-s)(1-\delta)y \geq 0.$$

The fraction of funds diverted by D types is $\delta = 1 - (1-s)y/\theta$, increasing in s.

The expression for the bank's expected share-finance profit (on a partnership with an unknown member type) is

$$\pi_B(s) = \lambda s y + (1-\lambda)s(1-\delta)y - 1.$$

From the bank's maximization, we obtain three relevant levels of the participation share, $s_{\min} = (y-\theta)/y$, $\widetilde{s} = [\lambda y + (1-\lambda)\theta]/2\lambda y$ and $s_{\max} = 1$, which have an analogous interpretation to their repayment counterparts. It can be shown that $\pi_B(s_{\min}) = \pi_B(r_{\min})$, $\pi_B(\widetilde{s}) = \pi_B(\widetilde{r})$, and $\pi_B(s_{\max}) = \pi_B(r_{\max})$. Therefore, with equity contracts, the equilibrium configurations of the analysis of this section would remain unchanged.

1.3 Welfare

In this section, we determine the equilibrium and the socially optimal welfare, and analyze the potential divergences between these two measures. From the analysis of Section 3, the optimal welfare depends on whether the utility from diversion, $u_D(\delta)$, is included or not in the accounting system and, in the following, we will discuss how this distinction may lead to different levels of δ maximizing social welfare. We will denote welfare (per borrower),

[9] $r - r_{min} = \frac{(1+\lambda)\theta - \lambda y - \theta\sqrt{\left[1+\left(\frac{y}{\theta}-1\right)\lambda\right]^2 - \frac{4\lambda}{\theta}}}{2\lambda}$.

when diversion is included and excluded, by $\omega_{in}(\delta)$ and $\omega_{ex}(\delta)$. We assume that the institutional decision to factor in informal activities does not affect the diversion utility weight, θ. In addition, we will not model the potential negative externalities associated with the production of informal projects (which is true for cultivating drug crops, but probably not for entrepreneurs investing in small merchandise or health medication). Informal projects and their outputs will still be considered illegal or non-contractible, otherwise they might be embedded into formal financial contracts. This means that C borrowers will continue to comply with all contract requirements and not misallocate their loans. The bank will still be unable to finance informal investments, otherwise it might take advantage, especially if its output is high.

This conceptual ambivalence, illegal but productive in economic terms, may reflect the recent reform ESA 2010 which allows EU countries to record illicit production within the national accounts. This may give a chance for governments to kill two birds with one stone: on the one hand, there can be a significant (and perhaps more accurate) increase in gross domestic product; on the other, it allows to secure political integrity and legitimacy.

If $u_D(\delta)$ is excluded from welfare, the socially efficient diversion fraction derives from

$$\max_{\delta} \omega_{ex}(\delta) = (1 - \lambda\delta)y - 1.$$

Since $\partial\omega_{ex}(\delta)/\partial\delta < 0$, the solution is $\delta = 0$, and optimal welfare is equal to full-information welfare,

$$\omega_{ex}(0) = y - 1 \equiv \omega^{FI}(0). \tag{1.14}$$

If $u_D(\delta)$ is included in welfare, the optimal diversion fraction derives from

$$\max_{\delta} \omega_{in}(\delta) = \lambda u_D(\delta) + (1 - \lambda\delta)y - 1. \tag{1.15}$$

The solution depends on the utility weight θ and, specifically, on whether it is higher or lower than the formal project's output, y (with no diversion on the part of D borrowers). Thus, we need to distinguish between two further cases.

Case $\theta \leq y$. If the informal activity is relatively less productive than the formal project, the welfare-maximizing diversion share is again $\delta = 0$. Efficient welfare is $\omega_{in}(0) = \omega^{FI}(0)$, thus equal to the full-information welfare.

Case $\theta > y$. If diversion is relatively more productive, the solution to (1.15) is

$$\delta = 1 - \tfrac{y}{\theta} \equiv \delta^*, \tag{1.16}$$

with $\delta^* \in (0, 1)$. This means it is socially optimal to let D types divert part of their loans. Namely, social efficiency would require the presence of some entrepreneurs who engage in informal activities, even if illegal (and even if, in theory, they may end up being punished by the law). Welfare is

$$\omega_{in}(\delta^*) = y + \tfrac{\lambda(\theta - y)^2}{2\theta} - 1. \tag{1.17}$$

It is easily verified that $\omega_{in}(\delta^*) > y - 1 = \omega^{FI}(0)$, so we have the following

Proposition 2. *When $\theta > y$ and $u_D(\delta)$ is included, the optimal diversion fraction is $\delta^* > 0$, and the socially efficient welfare is larger than under full information.*

Therefore, optimal welfare is equal to full-information welfare if the proceeds from diversion are excluded or, if included, the productivity of informal projects is relatively low. When the productivity is relatively high, welfare can be larger than under full information. But, as shown below, we will not derive a situation in which the equilibrium welfare under asymmetric information is higher than the socially efficient level.

1.3.1 Equilibrium welfare

Welfare at $r_{\max}$

When the equilibrium is $r_{\max}$, then $\delta(r_{\max}) = 1$. If $u_D(1)$ is excluded, social welfare is

$$\omega_{ex}(1) = (1 - \lambda)y - 1, \tag{1.18}$$

the expected output produced by C borrowers and transferred to the bank, as the loans received by D borrowers are entirely diverted. Since $\omega_{ex}(1) = (1-\lambda)y-1 < y-1 = \omega^{FI}(0)$ for all $\lambda \in (0, \underline{\lambda}]$, welfare is lower than under full information (and thus inefficient).

If the informal output produced by D types is included, welfare is

$$\omega_{in}(1) = \tfrac{\lambda\theta}{2} + (1 - \lambda)y - 1, \tag{1.19}$$

below optimal welfare, $\omega_{in}(\delta^*)$. We can thus conclude that the equilibrium at $r_{\max}$ is socially inefficient.

However, the difference between $\omega_{in}(1)$ and welfare under full information is $\omega_{in}(1) - \omega^{FI}(0) = (\lambda\theta - 2\lambda y)/2$, which is positive if $\theta > 2y$, as depicted in figure 3a. Even if the proportion of funds diverted is above the socially efficient level, welfare including the proceeds of the informal project is larger than that of full information for all $\lambda \in (0, \underline{\lambda}]$.

Welfare at $\widetilde{r}$

If the equilibrium is at $\widetilde{r}$, and $u(\widetilde{\delta})$ is excluded, equilibrium welfare is

$$\omega_{ex}(\widetilde{\delta}) = \tfrac{[\theta+\lambda(y-\theta)]y}{2\theta} - 1, \tag{1.20}$$

which is lower than $\omega^{FI}(0)$, and inefficient, since $\widetilde{\delta} > 0$.

When $u(\widetilde{\delta})$ is included, welfare is

$$\omega_{in}(\widetilde{\delta}) = \tfrac{(2\lambda-1)\theta^2+3\lambda[2y\theta+\lambda(\theta-y)^2]}{8\lambda\theta} - 1, \tag{1.21}$$

again lower than optimal welfare, $\omega_{in}(\delta^*)$. The reason is that $\widetilde{\delta} - \delta^* > 0$, so D borrowers tend to divert too much resources compared to what social efficiency requires. Therefore, the equilibrium at $\widetilde{r}$ is inefficient, whether or not $u_D(\tilde{\delta})$ is counted as social welfare.

It can be shown that the difference between equilibrium welfare in (1.21) and full-information welfare can be positive for all $\lambda > \theta/3(\theta - y) \equiv \widetilde{\lambda}$. The threshold $\widetilde{\lambda}$ is smaller than one if $\theta > 3/2y$, and equal to $\underline{\lambda}$ if $\theta = 2y$. For $\theta \geq 2y$, then $\widetilde{\lambda} < \underline{\lambda}$ and equilibrium welfare is higher than the full information in the interval $[\underline{\lambda}, 1)$. For instance, in the numerical case of figure 3*b*, if $y = 2$ and $\theta = 4$, then $\omega_{in}(\widetilde{\delta}) > \omega^{FI}(0)$ for all $\lambda \in (0.67, 1]$. If $\lambda = 0.8$, then $\omega_{in}(\widetilde{\delta}) - \omega^{FI}(0) = 0.175$.

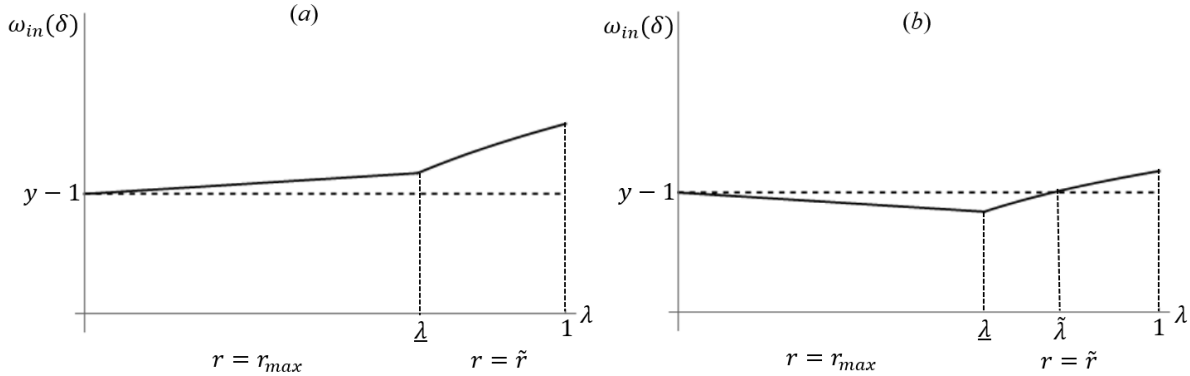

Figure 3. Equilibrium and full-information welfare (diversion output included). Numerical example.
a) Parameters: $y = 2$; $\theta = 4.5$.
b) Parameters: $y = 2$; $\theta = 4$.

Welfare at $r_{\min}$

If $y > 2\theta$, in equilibrium, the repayment is $r_{\min}$, for $\lambda \in [\overline{\lambda}, 1)$. Since $\delta(r_{\min}) = 0$, social welfare does not depend on whether the output from diversion is included or not, so

$$\omega_{ex}(0) = \omega_{in}(0) = y - 1 = \omega^{FI}(0).$$

Thus, if an equilibrium at $r_{\min}$ exists, it is socially efficient, and welfare is equal to the full-information level. In contrast, when $y \leq 2\theta$, both the interval $(\overline{\lambda}, 1]$ and the equilibrium at $r_{\min}$ do not exist.

From the analysis of this section, we can state the following.

Proposition 3. *The credit market equilibrium is socially efficient only when the equilibrium is at $r_{\min}$, when the proportion of D types is relatively high and the utility weight of diversion is small. The equilibria at $\widetilde{r}$ or $r_{\max}$ are socially inefficient.*

From this proposition, we come to the rather counterintuitive conclusion that the equilibrium is more likely to be efficient when there is a high proportion of potential diverters. When both $y > 2\theta$ and $\lambda \in [\overline{\lambda}, 1)$, the equilibrium at $r_{\min}$ exists and this is the only situation in which equilibrium welfare is socially efficient.

Proposition 4. *In the social inefficient equilibria at $\widetilde{r}$ and $r_{\max}$, when diversion output is included and the productivity of informal projects is relatively high, equilibrium welfare can be higher than under full information.*

1.3.2 Policy

When the equilibrium is inefficient, a social planner (government) might want to implement policies to improve the allocation of resources in the credit market. To restrict the discussion below, we will only present what would happen if the planner were to implement a policy when diversion output is particularly valuable, that is $\theta > y$, and is included in the measure of social welfare.[10] In this situation, the socially optimal welfare would be $\omega_{in}(\delta^*)$ in (1.17), which would call for a diversion fraction equal to δ^*, as derived in (1.16).

We follow the Kaldor-Hicks compensation principle, so the policy should be implemented only if borrowers, who will gain from the intervention, can potentially compensate the lender and still be better off. To be more specific, for Pareto efficiency, the lender's post-policy payoff must be equal to pre-policy equilibrium profit, based on the (inefficient) equilibrium repayment. The difference between these two profit levels will correspond to the policy cost. The *gross* expected benefit of the policy is the difference between optimal and equilibrium welfare. Therefore, the *net* expected benefit is the difference between gross benefit and cost.

From the welfare analysis above, including the utility from diversion in welfare means that socially efficient diversion fraction *should* be equal to the positive level, δ^*. And, if we follow this thought-provoking line of reasoning, this welfare maximizing fraction of loan diverted can be obtained by introducing an interest rate cap. In the specific case in which diversion output is included in welfare and $\theta > y$, using the reaction function of D borrowers in (1.4), the cap such that $r = 0$. At this interest rate, the bank makes a loss of $\pi_B(0) = -1$.

In the following, we will derive the net benefit of the policy in the two inefficient equilibria at $r_{\max}$ and $\widetilde{r}$.

Equilibrium at $r_{\max}$

From (1.17) and (1.19), the difference between socially efficient and equilibrium welfare is

$$\omega_{in}(\delta^*) - \omega_{in}(1) = \tfrac{\lambda y^2}{2\theta}.$$

The policy cost is the difference between the pre-policy profit in (4.11) and $\pi_B(0) =$

[10] In the other cases in which $\theta < y$ and diversion output is excluded, there may be a range where a policy can be welfare improving (a complete description of the policy is available upon request).

-1, that is

$$\pi_B(r_{\max}) - \pi_B(0) = (1-\lambda)y.$$

Comparing policy gain and policy cost, we obtain

$$[\omega_{in}(\delta^*) - \omega_{in}(1)] - [\pi_B(r_{\max}) - \pi_B(0)] = \left[\lambda\left(1+\tfrac{y}{2\theta}\right) - 1\right] y, \tag{1.22}$$

increasing in λ and equal to 0 when $\lambda = 2\theta/(y+2\theta)$, which is above the lower threshold $\underline{\lambda}$. So, the net benefit in (1.22) is negative in the interval $(0, \underline{\lambda}]$, that is when the equilibrium at $r_{\max}$ exists.

Equilibrium at $\widetilde{r}$

The gross expected benefit of the policy, using the optimal welfare in (1.17) and the equilibrium welfare in (1.21), is

$$\omega_{in}(\delta^*) - \omega_{in}(\widetilde{\delta}) = \tfrac{[\lambda y + (1-\lambda)\theta]^2}{8\lambda\theta}.$$

In equilibrium, the bank obtains the profit in (1.8), the policy cost is

$$\pi_B(\widetilde{r}) - \pi_B(0) = \tfrac{[\lambda y + (1-\lambda)\theta]^2}{4\lambda\theta}.$$

Therefore, it is immediate that the net benefit of the policy would be

$$[\omega_{in}(\delta^*) - \omega_{in}(\widetilde{\delta})] - [\pi_B(\widetilde{r}) - \pi_B(0)] = -\tfrac{[\lambda y + (1-\lambda)\theta]^2}{8\lambda\theta} < 0. \tag{1.23}$$

Proposition 5. *If diversion output is included in welfare and $\theta > y$, a policy intervention is never Pareto improving.*

This is a key result for our analysis, and will be true also for the equilibrium at $r_{\max}$ below. When the informal activity is extremely productive and included in welfare, it is better for the government to let D borrowers divert their loans and not intervene with a policy. Even though diversion will result in a relatively unproductive investment for formal projects, the government can exploit the welfare gains achievable through the illegal but tolerated informal activities. Of course, this conclusion does not involve any issue of ethics and morality. The reason is simply that achieving the socially efficient diversion production would require a target repayment of zero on formal contracts, and this would be too costly to implement.

This type of policy analysis is in line with the literature on the effects of public interventions in credit markets under imperfect information (see Innes (1991)). In particular, our policy setup is close to the moral-hazard section of Minelli and Modica (2009), where they analyze a series of public interventions in a monopolistic credit market. They discuss the effects of two of the most widely used policy instruments, the interest-rate subsidy

and the investment subsidy, and show that the former is optimal, as it maximizes net welfare benefits. In contrast, we compare the cost with the efficiency gain that the regulator can expect to obtain and, in particular, we focus on the quantity and "quality" of the output produced before the policy intervention.

1.4 Conclusion

We model the strategic interaction between a monopolistic lender and a large number of cashless entrepreneurs, in the presence of asymmetric information. Some of the borrowers may have the incentive to use part of the loan received for other projects, which may be or not more productive than the formal business. There are two main results. The first is that the equilibrium is more likely to be efficient when there is a high proportion of borrowers who can potentially divert funds into informal projects. The reason is that this propensity of defiant individuals forces the bank to reduce the loan repayment, which otherwise would be equal to the entire output produced. The second result is that, when funds are diverted into highly productive projects and the informal output is included in the social well-being function, the equilibrium welfare, although inefficient, can be higher than that of full information.

When the equilibrium is socially inefficient, we analyze the impact and, in particular, the desirability of policy interventions. The key determinant of desirability is the productivity of the alternative projects and, in particular, whether their output is included or not in the social welfare accounting. We show that, when the informal project output is included in welfare, the need for public intervention is weaker, and even absent if the informal activity is highly productive.

Bibliography

Akerlof, G. (1978). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84:235–251.

Armendáriz, B. and Szafarz, A. (2011). On mission drift in microfinance institutions. In *Handbook of Microfinance*, pages 341–366. World Scientific.

Banerjee, A. (2013). Microcredit under the microscope: What have we learned in the past two decades, and what do we need to know? *Annual Review of Economics*, 5:487–519.

Banerjee, T., Roy, M., Raychaudhuri, A., and Ghosh, C. (2015). What drives households to divert loans? A village level study. *Asia-Pacific Social Science Review*, 15:33–55.

Beck, T., Demirgüç-Kunt, A., and Maksimovic, V. (2004). Bank competition and access to finance: International evidence. *Journal of Money, Credit and Banking*, 36:627–648.

Berger, A. and Udell, G. (2002). Small business credit availability and relationship lending: The importance of bank organisational structure. *Economic Journal*, 112:F32–F53.

Besley, T., Burchardi, K., and Ghatak, M. (2012). Incentives and the de soto effect. *Quarterly Journal of Economics*, 127:237–282.

Bhat, M. (1971). Diversion of long-term agricultural finance: A study of past trends and future strategy. *Economic and Political Weekly*, pages 2151–2158.

Bougheas, S. (2004). Internal vs external financing of R&D. *Small Business Economics*, 22:11–17.

Burkart, M. and Ellingsen, T. (2004). In-kind finance: A theory of trade credit. *American Economic Review*, 94:569–590.

Carter, M. (1988). Equilibrium credit rationing of small farm agriculture. *Journal of Development Economics*, 28:83–103.

Caserta, M., Monteleone, S., and Reito, F. (2018). The trade-off between profitability and outreach in microfinance. *Economic Modelling*, 72:31–41.

Coco, G. (2000). On the use of collateral. *Journal of Economic Surveys*, 14:191–214.

Cull, R., Demirgu¨ ç Kunt, A., and Morduch, J. (2007). Financial performance and outreach: A global analysis of leading microbanks. *The Economic Journal*, 117:F107–F133.

de Meza, D., Reito, F., and Reyniers, D. (2021). Too much trade: The hidden problem of adverse selection. *Journal of Public Economics*, 204.

de Meza, D. and Webb, D. (1987). Too much investment: A problem of asymmetric information. *Quarterly Journal of Economics*, 102:281–292.

de Quidt, J., Fetzer, T., and Ghatak, M. (2012). Microfinance, social capital and for-profit lending.

Diamond, D. (2004). Presidential address, committing to commit: Short-term debt when enforcement is costly. *Journal of Finance*, 59:1447–1479.

Durnev, A. and Kim, E. (2005). To steal or not to steal: Firm attributes, legal environment, and valuation. *The Journal of finance*, 60:1461–1493.

Garikipati, S. (2013). Microcredit and women's empowerment: Have we been looking at the wrong indicators? *Oxford Development Studies*, 41:S53–S75.

Giné, X., Goldberg, J., and Yang, D. (2012). Credit market consequences of improved personal identification: Field experimental evidence from Malawi. *American Economic Review*, 102:2923–2954.

Guha, B. and Chowdhury, P. (2013). Micro-finance competition: Motivated micro-lenders, double-dipping and default. *Journal of Development Economics*, 105:86–102.

Hart, O. and Moore, J. (1998). Default and renegotiation: A dynamic model of debt. *Quarterly Journal of Economics*, 113:1–41.

Howorth, C. and Moro, A. (2012). Trustworthiness and interest rates: An empirical study of Italian SMEs. *Small Business Economics*, 39:161–177.

Innes, R. (1991). Investment and government intervention in credit markets when there is asymmetric information. *Journal of Public Economics*, 46:347–381.

Kar, A. and Swain, R. (2014). Interest rates and financial performance of microfinance institutions: Recent global evidence. *European Journal of Development Research*, 26:87–106.

Koo, H., Huang, C., and Kan, K. (2012). Interlinked contracts: An empirical study. *Economica*, 79:350–377.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. (2002). Investor protection and corporate valuation. *Journal of Finance*, 57:1147–1170.

Lauermann, S. (2012). Asymmetric information in bilateral trade and in markets: An inversion result. *Journal of Economic Theory*, 147:1969–1997.

Madajewicz, M. (2003). Does the credit contract matter? The impact of lending programs on poverty in Bangladesh. *Working Paper*.

Madestam, A. (2014). Informal finance: A theory of moneylenders. *Journal of Development Economics*, 107:157–174.

Minelli, E. and Modica, S. (2009). Credit market failures and policy. *Journal of Public Economic Theory*, 11:363–382.

Mungai, J., Maingi, J., and Sma, M. (2014). Loan repayment and sustainability of government funded micro-credit initiatives in Murang'a county, Kenya. *International Journal of Business and Social Science*, 5.

Navajas, S., Conning, J., and Gonzalez-Vega, C. (2003). Lending technologies, competition and consolidation in the market for microfinance in Bolivia. *Journal of International Development: The Journal of the Development Studies Association*, 15:747–770.

Petersen, M. and Rajan, R. (1995). The effect of credit market competition on lending relationships. *Quarterly Journal of Economics*, 110:407–443.

Repullo, R. and Suarez, J. (2000). Entrepreneurial moral hazard and bank monitoring: A model of the credit channel. *European Economic Review*, 44:1931–1950.

Shneyerov, A. and Wong, A. (2011). The role of private information in dynamic matching and bargaining: Can it be good for efficiency? *Economics Letters*, 112:128–131.

Stiglitz, J. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Economic Review*, 71:393–410.

Von Pischke, J. and Adams, D. (1980). Fungibility and the design and evaluation of agricultural credit projects. *American Journal of Agricultural Economics*, 62:719–726.

Wydick, B. (2001). Group lending under dynamic incentives as a borrower discipline device. *Review of Development Economics*, 5:406–420.

Chapter 2

Consumer credit under asymmetric information: The wrong types apply

Published in Economics Letters, 2020, 197.

We present a simple asymmetric information model of the consumer credit with perfectly competitive banks and many borrowers who wish to advance the purchase of an indivisible good. We show a credit rationing equilibrium is possible. Unlike business lending models, it is not just the volume of lending that is wrong but also who gets loans. A policy that targets the interest rate can increase social efficiency, but not achieve the first best.

Introduction

This paper extends the models of business lending under asymmetric information, notably Stiglitz and Weiss (1981) and de Meza and Webb (1987), to the consumer credit market. In the business lending context, the volume of lending is inefficient, either too high or too low. With consumption loans there is an added problem. It is the wrong borrowers that are served. This applies whether or not there is rationing.

In Stiglitz and Weiss (1981), borrowers can be ranked by second-order stochastic dominance so are equally valuable from a social perspective. When the interest rate rises, it is the low-risk types that drop out and these are the least profitable. As a result, too few loans are made whether or not credit rationing applies. In de Meza and Webb (1987), projects' returns are ranked by first-order stochastic dominance, so high-risk borrowers are more likely to drop out as the interest rate rises. This may lead to investment in excess of social efficiency since some projects with negative net present value are funded.

Low-default types are the ones that should get loans, but these are the ones that do get loans. So, in either formulation, the inefficiency is in the volume of lending.

Jaffee and Russell (1976) seem to be the first to analyze the effect of asymmetric information on consumer credit. They show how loan size can be used by lenders as a screening device. Borrowers obtain smaller loans than they would like at the ruling interest rate and are not excluded from the market, but the authors do not examine market efficiency. Here, consumers wish to purchase an indivisible durable good and have no initial wealth to use as collateral or down-payment. Varying loan size is not an option. The theoretical model is similar to Inderst (2008), although the purpose of his paper is to explain how a monopolistic lender with an informational advantage over borrowers may engage in predatory lending. In our competitive model, consumers can use a loan to bring forward the purchase of a durable good before they receive income. Their income flow is risky and, under a low realization, the loan is not repaid and the good foreclosed. Lenders only know the distribution of types in the population, but do not observe the riskiness of each borrower. An equilibrium is possible in which some borrowers are randomly rationed and others receive a loan. High repayment types do not apply for loans though they would contribute most to social efficiency[1]. Even if there is no credit rationing, it remains the case that, from the perspective of social efficiency, the wrong types are the borrowers.

2.1 The model

Consider a simple two-period consumer credit market with a continuum of individuals/households and a large number of lenders. There are three time points, $t = 0, 1, 2$. Individuals have no endowment in $t = 0$. In $t = 1$, they receive a binary[2] stochastic income, which can be equal to W or 0. Individuals are indexed by their probability, $p_i \in [0, 1]$, of obtaining positive income, so they can be ordered in the sense of first-order stochastic dominance. Households and lenders are risk neutral[3]. Every individual wants to buy an indivisible good, whose price is $L < W$. The good can be purchased in $t = 0$ by means of a financial loan, or in $t = 1$ without intermediation if income is positive. The utility derived from the good in $t = 0$ is θu, with $\theta \in [0, 1)$, while the utility in $t = 1$ is u. As in Inderst (2008), θ can be considered small for the purchase of durable goods (which would imply that the second period can be considered as the sum of several periods, or simply longer than the first). The good does not produce utility after $t = 2$. Lenders

[1]de Meza and Webb (2000) show in a model with both hidden types and hidden action there may be excessive lending despite credit rationing.

[2]The theoretical results would also hold under the more general assumption of continuous income distribution (results available upon request).

[3]The assumption of risk neutrality is quite standard when screening instruments are not available, and the market equilibrium is characterized by a pooling of borrowers (as in Stiglitz and Weiss, 1981, de Meza and Webb, 1987, and Inderst, 2008).

are perfectly competitive maximizers. While households know their own probability of receiving the income, lenders only know the distribution of individuals in the population, $F(p_i)$. The risk-free rate is zero, and lenders compete by setting the (pooling) interest rate, r, on the loan contract (L, r). In case of default, loans are foreclosed, and the resale value for each seized good is $\beta < L$ (we assume β so low that the resale value is lower than the loan repayment). We consider two types of default penalties. To prevent strategic default, we assume that borrowers incur a non-pecuniary cost (for example, social stigma and psychological stress) if, in $t = 1$, they have positive income but choose to default and repurchase the good. Deliberate default does not occur if the penalty is equal or larger than rL, which we assume throughout the paper. To simplify the exposition, we do not consider penalty costs associated with unintentional default.

2.2 Equilibrium

The expected payoff of household i, if the good is purchased through a loan in $t = 0$ is

$$E[u_i]_0 = p_i[W + \theta u + u - (1 + r)L] + (1 - p_i)\theta u, \tag{2.1}$$

where: with probability p_i, the borrower obtains a positive income, repays the loan and receives the first- and second-period utilities; with probability $1-p_i$, the borrower defaults and pays the penalty, but receives the first-period utility even if the loan is foreclosed in $t = 1$.

We assume that the price of the good does not change in $t = 1$ so, if household i decides to purchase the good in $t = 1$, the expected payoff is

$$E[u_i]_1 = p_i(W + u - L). \tag{2.2}$$

From (2.1) and (2.2), households will accept the credit contract if $E[u_i]_0 \geq E[u_i]_1$. The expected payoff from borrowing, net of opportunity costs, is

$$E[u_i]_0 - E[u_i]_1 = \theta u - p_i rL$$

which is lower the higher the probability p_i. Thus, individuals with low probabilities of repayment are those who benefit the most from the financial contract. Marginal borrower have a repayment probability denoted by p_M, and are characterized by

$$E[u_M]_0 - E[u_M]_1 = \theta u - p_M rL = 0. \tag{2.3}$$

The expected net social value of the loan granted to household i is

$$E[v_i]_0 = p_i(W + \theta u + u) + (1 - p_i)(\theta u + \beta) - (1 + \rho)L, \tag{2.4}$$

where ρ is the safe rate. The social value of purchasing the good in $t = 1$ is

$$E[v_i]_1 = p_i(W + u - L), \tag{2.5}$$

corresponding to the payoff in (2.2).

The net expected social value of borrowing, that is, the difference between receiving credit in $t = 0$ and buying the good in $t = 1$, is

$$E[v_i]_0 - E[v_i]_1 = \theta u - (1 - p_i)(L - \beta) - \rho L,$$

increasing in p_i. This expression corresponds to the expected payoff of borrower i under symmetric information. We indicate with p_{SE} the borrower such that

$$E[v_{SE}]_0 - E[v_{SE}]_1 = \theta u - (1 - p_{SE})(L - \beta) - \rho L = 0, \tag{2.6}$$

which is the least efficient borrower from a social perspective (from (2.3) and (2.6), $p_{SE} > 0$ is compatible with $p_M > 0$).

If $\theta u \leq rL$, and thus $p_M \leq 1$, it follows that the lender's expected return on each loan is

$$E[\pi] = (1 + r)L \int_0^{p_M} p_i \tilde{f}(p_i) dp_i + \beta \int_0^{p_M} (1 - p_i) \tilde{f}(p_i) dp_i, \tag{2.7}$$

where $\tilde{f}(p_i) = f(p_i)/F(p_M)$.

As in Stiglitz and Weiss (1981), the banks' expected return may not be monotonically increasing in the interest rate. At a higher interest rate, the return from successful individuals increases but the borrowers with the highest probability of repayment drop out. This adverse selection effect drives down returns. There may therefore be a turning point to the return function, say at r^*, yielding expected return $E[\pi^*]$. If the supply of funds when depositors are paid ρ is below the number of borrowers applying for loans, some individuals with $E[u_i]_0 - E[u_i]_1 \geq 0$ are denied credit.

Proposition 1. *If the supply of funds to the banking sector is upward sloping, the market equilibrium may be characterized by credit rationing.*

Proof. Each bank maximizes (4.9). The first-order condition is

$$\frac{dE[\pi]}{dr} = \frac{dp_M}{dr} \frac{f(p_M)}{F(p_M)} (E[\pi_M] - E[\bar{\pi}]) + \frac{L}{F(p_M)} \int_0^{p_M(r)} p_i f(p_i) dp_i = 0, \tag{2.8}$$

where $E[\pi_M] = p_M(1 + r)L + (1 - p_M)\beta$ is the expected return on marginal households, and $E[\bar{\pi}] = \bar{p}(1 + r)L + (1 - \bar{p})\beta$ the return on average households ($\bar{p}$ is the average probability in $[0, p_M]$). From (2.3), $d(E[u_M]_0 - E[u_M]_1)/dr < 0$, so the positive effect of an increase in r, the second term in (4.10), is offset by the negative adverse-selection effect, as $dp_M(r)/dr < 0$. Since $E[\pi_M] > E[\bar{\pi}]$, credit rationing may exist in equilibrium.

High-probability individuals do not apply for a loan because they are discouraged by the high interest rate, and thus prefer to postpone the purchase at $t = 1$.

We now show that a subset of borrowers in $[0, p_M)$ are characterized by negative net social value. Namely, from a social viewpoint, they should postpone the purchase of the good in $t = 1$, and not ask for a loan in $t = 0$.

Proposition 2. *The market equilibrium involves (socially) inefficient borrowers for which* $E[v_i]_0 - E[v_i]_1 < 0$.

Proof.[4] Denote by $\bar{p}$ the average probability of borrowers in $[0, p_M)$. The bank's expected return on each household with $p_i = \bar{p}$ is

$$E[\bar{\pi}] = \bar{p}(1 + r)L + (1 - \bar{p})\beta = (1 + \rho)L. \tag{2.9}$$

Since $\bar{p} < p_M$, individuals with $p_i = \bar{p}$ strictly prefer to borrow rather than waiting until $t = 1$ to purchase the good, i.e.

$$E[\bar{u}]_0 - E[\bar{u}]_1 = \theta u - \bar{p}rL > 0. \tag{2.10}$$

Substituting $r = [\rho L + (1 - \bar{p})(L - \beta)]/\bar{p}L$ from (2.9) into (2.10), we obtain

$$E[\bar{u}]_0 - E[\bar{u}]_1 = \theta u - (1 - \bar{p})(L - \beta) - \rho L = E[\bar{v}]_0 - E[\bar{v}]_1 > 0,$$

so, the expected net social value of average borrowers is positive. It follows that $E[\bar{v}]_0 - E[\bar{v}]_1 > E[v_{SE}]_0 - E[v_{SE}]_1$, which implies $p_{SE} < \bar{p}$ and $p_{SE} < p_M$. Thus, all loans granted to borrowers with $p_i \in [0, p_{SE}]$ are such that $E[v_i]_0 - E[v_i]_1 < 0$.

As in de Meza and Webb (1987), the contracts offered to a fraction of borrowers with high-default risk have a negative net social value and, thus, result in inefficient lending. Figure 1 plots the net private and social value against p_i. The private net benefit is decreasing in p_i, whereas the social value is increasing.

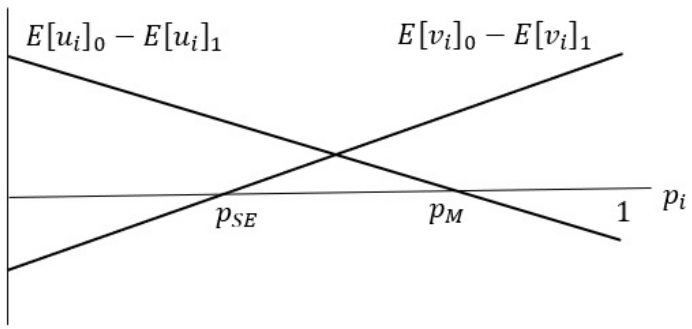

Figure 1: Equilibrium configuration.
Parameters: $L = 1$, $r = 0.6$, $\theta u = 0.5$, $\rho = 0.1$, and $\beta = 0.3$.
$p_{SE} = 0.4$ and $p_M = 0.8$.

From the above two propositions, we obtain the following

Proposition 3. *It is the highest default probability types that apply for loans although it is the lowest default types that generate the most social surplus.*

[4] We can obtain the same result using the proof by contradiction of de Meza and Webb (1987).

2.3 Policy implications and concluding remarks

In Stiglitz and Weiss (1981), an interest-rate subsidy increases the supply of funds, and can achieve a first-best outcome. To eliminate overlending in de Meza and Webb (1987), taxing interest income is sufficient to restore efficiency.

In this paper, an interest rate subsidy can enhance efficiency, but does not provide the incentive for borrowers with negative net social value to delay their purchases.

Proposition 4. *Whether or not there is credit rationing, there is a lending subsidy that enhances efficiency although it cannot achieve a first-best outcome.*

Within the rationing regime, a subsidy (or a tax) does not change the interest rate, so there is no effect on loan applications. As it is random who gets a loan, there is also no effect on the quality of loans. As the average borrower gets a surplus, a small subsidy must raise welfare[5]. In the absence of rationing the marginal borrower is profitable. Inducing another borrower is therefore beneficial especially as are of even higher quality.

On the empirical side, it is generally accepted that informational asymmetries may have adverse consequences in the market for consumer loans. However, it is difficult to find clear evidence in the literature of credit rationing and especially of socially inefficient lending in this particular market. Zinman (2014) reviews the empirical literature and argues that previous studies do not provide very convincing answers to whether the consumer loan market over- or under-supplies credit. The analysis of this paper shows this is not the only inefficiency contributed by hidden types.

Bibliography

de Meza, D. and Webb, D. (1987). Too much investment: A problem of asymmetric information. *Quarterly Journal of Economics*, 102:281–292.

de Meza, D. and Webb, D. (2000). Does credit rationing imply insufficient lending? *Journal of Public Economics*, 78:215–234.

Inderst, R. (2008). "Irresponsible lending" with a better informed lender. *Economic Journal*, 118:1499–1519.

Jaffee, D. and Russell, T. (1976). Imperfect information, uncertainty, and credit rationing. *Quarterly Journal of Economics*, 90:651–666.

Stiglitz, J. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Economic Review*, 71:393–410.

[5] If the supply of funds is upward sloping, a large subsidy may drive up the opportunity cost of loans so high that the net benefit is zero.

Zinman, J. (2014). Consumer credit: Too much or too little (or just right)? *Journal of Legal Studies*, 43:S209–S237.

Chapter 3

Occupational choice and entrepreneurship: From necessity to opportunity

We present a simple agency model with an individual who faces an occupational choice between starting a self-managed business or entering the labor market as a wage worker. To set up the entrepreneurial activity, the individual needs external funding. We show that, when the reservation utility is low, the lender may not find it profitable to offer any contract and the individual is constrained to become a wage worker out of necessity. Whereas, when the reservation utility is high, the bank may be forced to give up part of the contract rent to induce the individual to become an entrepreneur out of opportunity.

Introduction

Entrepreneurship is the main driving force of economic growth in both developed and less developed countries (Valliere and Peterson, 2009). Understanding the motives behind the occupational choice of becoming an entrepreneur helps evaluate the contribution of entrepreneurship to economic growth, distribution of wealth, and income inequality, within and across countries (Naudé, 2010). But despite its relevance, there is still limited comprehension of which factors determine the birth of new businesses, especially for small enterprises.

Becoming an entrepreneur is an occupational choice made as an alternative to wage employment. In the Flash Eurobarometer survey conducted in 2012, over 42,000 respondents from European and non-European countries, 37 percent of the European respondents prefer to be self-employed. Among the non-European countries, Turkey and Brazil report the highest level of respondents who prefer self-employment to wage employment

(respectively, 82% and 63%). However, starting a business from scratch is not always a viable choice. The majority of the European respondents perceive self-employed, in terms of entrepreneurship, as an unfeasible alternative (67%), with 21 percent stating that the lack of enough financial resources represents the main constraint.

Based on the motivations surrounding the start-up decision, the Global Entrepreneurship Monitoring (GEM) program distinguishes between necessity and opportunity entrepreneurs (Reynolds et al., 2005). Necessity entrepreneurs are individuals with a low outside option, pushed to entrepreneurship to start an income-generating activity when other employment alternatives are missing. In contrast, opportunity entrepreneurs are individuals with a relatively high reservation utility, pulled to entrepreneurship to pursue advantageous business ventures. A low outside option can be interpreted as an indicator of limited job alternatives due to economic downturns or, more generally, to inequality of opportunities, such as access to education, related to family background, race, and poverty (Roemer, 1998). But it can also be a reflection of glass ceilings in developed countries, that is, invisible barriers that deter female employment or empowerment. In such circumstances, setting up a business may also be associated with non-monetary rewards, such as personal fulfillment and autonomy.

According to the Flash Eurobarometer survey, necessity entrepreneurs were mainly unemployed people with a low education level (they drop out of school at less than age 15). In contrast, before starting the activity, opportunity entrepreneurs were employed individuals with a high education level (they finished school at more than age 20). The majority of the European respondents to the survey claimed that their enterprises were pulled by opportunity (49%), with the highest share in Denmark, the Netherlands, Finland, and Luxembourg (above 60%); whereas, a lower share claimed that their enterprises were driven by necessity (29%), with the highest share in Romania, Estonia, and Greece (above 40%). Among the extra-European countries, the highest percentage of opportunity entrepreneurs are reported in China and Israel (above 50%) whereas, India and South Korea recorded the highest share of necessity entrepreneurs (above 60%). However, the distinction between opportunity and necessity entrepreneurs is blurry. Williams (2007) and Williams and Williams (2014) assert that the opportunity versus necessity dichotomy is a simplistic way to categorize entrepreneurship, not only because motives change over time but also because the two conditions may coexist. So, entrepreneurship requires a more detailed and realistic description that considers both the motives and the objective conditions behind new businesses creation.

This paper aims to provide a different interpretation of this necessity and opportunity dichotomy. We show that depending on both the reservation utility and the loan profitability of the bank, individuals may or may not have the possibility to choose the occupation that makes them better off. We present a simple model in which an individual faces an occupational choice between starting an owner-managed business and entering the labor market as a wage worker. The individual is endowed with a project requiring

a fixed investment and some level of effort, and has some initial wealth, which cannot be invested into the project but can be used as collateral. Credit is provided by a single lender/bank, which aims to extract the highest possible surplus from the entrepreneurial project. The credit contract entails a principal-agent problem, as the probability of project success depends on the effort level, unobservable to the bank.

The reservation utility consists of the sum of the initial wealth and the opportunity cost of becoming a wage worker. Our two main results are: *i*) if the reservation utility is low, the individual earns a higher profit from the entrepreneurial project so that he would be worse off in the labor market as a worker. But, if the expected output from the project and the initial wealth fall short the costs of lending, the bank will be unwilling to lend, and the individual forced to become a worker; *ii*) if the reservation utility is high, the individual is indifferent between the two employment alternatives, but the bank may not find it profitable to offer a credit contract. If the opportunity cost is higher than a given threshold, the individual will strictly prefer to be a wage worker. But, if the project's expected output is sufficiently high, the bank will be willing to reduce the repayment and, thus, give up part of its monopolistic rent to encourage the entrepreneurial activity.

Therefore, when the prospects in the labor market are poor, the individual would prefer to set up a business out of necessity but, the presence of credit constraints may preclude the optimal occupational choice. We interpret the concept of necessity as a circumstance with limited employment alternatives, where the individual is made worse off because the chance to choose the best occupation is missing. According to our results, in necessity conditions, the individual would be better off by running a business but, unless the initial wealth is sufficiently high, is credit constrained and forced to work for a wage. In contrast, in opportunity conditions, the individual can choose indifferently, regardless of the level of the initial endowment. And, even in the case in which the bank may deny credit, the individual is equally better off by working for a wage. In opportunity conditions, borrowing constraints may be binding because the expected return by the bank is decreasing in the individual's reservation payoff. Indeed, if the reservation payoff is high, the individual has more options in the labor market, and it would be too costly for the bank to provide the loan. If the reservation payoff is higher than a certain threshold, the individual may prefer to be a subordinate worker even when offered advantageous financial terms in the credit market. This result is in line with Iyigun and Owen (1998), who observe that, during economic expansions, individuals may prefer to work for a safe wage rather than setting up a risky entrepreneurial activity.[1] Our results are also in line with **?**, who, in the attempt to estimate the magnitude of latent entrepreneurship across countries, find evidence that liquidity constraints are the main obstacle that hinders potential entrepreneurs from running their businesses.

In the second part of the paper, we briefly analyze some policy interventions to ad-

[1] Ahunov and Yusupov (2017) show that risk-tolerant individuals are more likely to become self-employed than workers.

vocate entrepreneurship in necessity conditions. However, the desirability of a policy depends on the social efficiency of the entrepreneurial project. When the project's expected output is higher than the resources employed, the equilibrium in which borrowing constraints prevent the individual from making the optimal occupational choice is inefficient. We will show that a policy intervention aimed to either reduce the lending costs or endow the individual with the wealth needed to borrow can restore efficiency. Whereas, when the business project is socially inefficient, a policy would never encourage the entrepreneurial activity, and thus a no-intervention would be preferable.

Related Literature

The necessity and opportunity dichotomy is at the basis of the motives that push and pull individuals towards business creation. According to the entrepreneurship literature, pull factors include the desire for independence, success, and the expectation of high financial returns (Carter et al., 2003; Clark and Drinkwater, 2000). As far as push factors are concerned, the most important are generally family commitment, risk of unemployment, and dissatisfaction with the current standard of living (Rotefoss and Kolvereid, 2005).

Our paper is related to the literature that investigates the relationship between economic development and entrepreneurship as an occupational choice.[2] In developing countries, Yamada (1996) finds evidence of countercyclical behavior in the rate of new owner-managed businesses. Carree et al. (2002) provide a cross-sectional analysis by using data from 23 OECD countries from 1976 to 1996. They show that the number of business activities declines as the economy grows, but it rises again in highly developed countries. Similarly, Wennekers et al. (2005), using data from the Global Entrepreneurship Monitor (GEM), confirm such a U-shaped relationship between the birth rate of new firms and the level of the per capita income. Economies of scale and better employment opportunities can explain the preference of being a wage-earner during expansions. By contrast, high levels of economic development can create an increasing demand for variety, which prompts the birth of new market opportunities (Jackson, 1984; Baker et al., 2005). Table 1 shows the relationship among per-capita GDP, the number of employees, and new owner-managed firms in Italy. The case of Italy is appealing because it consists of two radically different sub-economies: a developed north and a less-developed south. The data show the countercyclical trend of new small business activities. Specifically, high levels of GDP in the north are associated with a higher number of employees. Whereas, in the south, low values of GDP are related to higher levels of owner-managed firms.

Throughout the model, we assume the presence of a single monopolistic lender. This assumption is made to represent rural credit markets where small entrepreneurs rely on informal lenders, such as money lenders or microfinance institutions, with considerable market power. Especially in developing countries, poor individuals have little access to

[2]For the counter-cyclicality of cooperative firms see Monteleone and Reito (2018).

Table 1. GDP, employees and new individual firms in Italy.

	GDP per capita (EUR)	Employees	Employees to total population (%)	Owner-managed firms	Owner-managed firms to total population (%)
North	34,487.9	7,167,884	25.8	1,378,903	5.0
Center	30,473.8	2,633,976	21.8	631,973	5.2
South & Islands	18,159.2	2,285,713	11.5	1,199,640	6.0
Italy	83,120.9	12,087,573	59.0	3,210,516	16.0

Source: Eurostat, Istat, Unioncamere.
Employees refers to the number of workers aged between 15-64 years old in active enterprises. Owner-managed firms refer to individual enterprise where the owner is natural person.

the formal banking sector and often turn to informal lenders, which, thanks to their ability to obtain superior information on local borrowers' reliability, charge individuals with very high or prohibitive interest rates (Hossein, 2013; Kar and Swain, 2014; Mookherjee and Motta, 2016). The assumption of monopolistic lender is in line with the argument of Besley (1994) and supported empirically by the analysis of Beck et al. (2004) and Delis et al. (2017). By using a cross-country analysis, Beck et al. (2004) show that high bank concentrations obstacle access to credit to small firms.[3] Delis et al. (2017) analyze US data and show that firms with poor investment returns are more likely to get credit from monopolistic banks. Other theoretical and empirical works claim that microfinance institutions are shifting from non-profit to for-profit status, a process known as mission drift (Cull et al., 2007; Mersland and Strøm, 2010; Serrano-Cinca and Gutiérrez-Nieto, 2014). Roberts (2013) reports empirical evidence that the for-profit status has led to an increase of about 4 percent of effective interest rates charged by microfinance lenders. The global average interest rate is about 35 percent, but it is possible to observe microcredit interest rates even above 80 percent. Table 2 reports some examples. Argentina has one of the highest interest rate (67%), although the bank concentration is relatively low (41%). Whereas, other countries, like Gambia or Malawi, have lower interest rates (28% and 25%), although an extremely high bank concentration.[4]

Among the reasons that explain the spread between microcredit and standard bank interest rates, Banerjee (2013) reports the riskiness of clients and small-sized loans. In terms of our paper, this means that money lenders and microcredit banks may take advantage of their monopoly position and deny credit to low-wealth-low-return individuals.[5] Even in developed countries, it is not uncommon to observe very high interest rates. For instance, in Table 3, we report the spread between borrowing and lending interest rates in Italy and show that, in the less-developed regions of the south, it is about 20% higher

[3]The authors also posit that the presence of government bank ownership tightens borrowing restrictions.

[4]Data also show that it is possible to observe countries with high bank concentration (Switzerland 71.52%, Hungary 58.89%) associated with low interest rates (Switzerland 2.63%, Hungary 1.79%).

[5]For other theoretical papers with profit-motivated microfinance institutions, see Guha and Chowdhury (2013), and Caserta et al. (2018).

Table 2. Lending interest rate and banking system concentration

Country	Lending interest rate (%)	Bank concentration (%)
Argentina	67.3	41.4
Madagascar	49.0	74.8
Brazil	46.9	56.6
Gambia	28.0	100.0
Malawi	25.7	86.8
Uzbekistan	23.6	59.5
Tajikistan	23.6	95.8
Ukraine	19.8	39.2
Angola	19.3	58.3
Peru	16.8	72.7

Source: Bankscope, World Bank
The lending interest rates refer to the interest rates on bank credit to the private sector. The bank concentration refers to the percent of bank assets held by top three banks.

than in the north. However, even if less developed regions strongly rely on debt finance, this does not necessarily imply that they are poor in terms of wealth. Table 3 also reports the territorial disparities in terms of loan-deposit ratio in Italy. In the south, the low value of ratio indicates that not all deposits translate into local investments. In the north, local investment exceeds the amount of local liquid assets, and this may mean that these regions borrow from the south or abroad.

Table 3. Loan-deposit ratio and interest rate spread in Italy

	Loans	Deposit	Loan/Deposit Ratio	Interest Rate Spread
North	542,975	650,186	1.2	2.2
Center	215,448	286,338	1.3	2.0
South & Island	255,039	197,971	0.8	2.7
Italy	1,134,494	1,013,463	1.1	2.2

Source: Bank of Italy.

Our paper also contributes to the literature that analyzes the role of liquidity constraints on new businesses creation. Myers (1977), Stiglitz and Weiss (1981) and Beck and Demirgüç-Kunt (2006) identify the difficulty to raise external capital as the main obstacle to the birth of new businesses, especially for poor individuals (Ghatak and Jiang, 2002). The inability to borrow may stem from market imperfections, such as informational asymmetries, which restrict credit for potential entrepreneurs, with the consequence that, in equilibrium, the volume of lending is inefficient (Stiglitz and Weiss, 1981; Minelli and Modica, 2009). Our results are close to those in Banerjee and Newman (1993), who ar-

gue that, during economic downturns, capital market frictions hinder access to credit for poor people, who thus are forced to become workers or remain unemployed. So, wage contracts may act as substitutes for entrepreneurial activities when individuals cannot borrow.[6]

Credit constraints may be softened in the presence of some initial wealth to be invested in the activity. Many works analyze the relationship between an individual's initial wealth and business entry. Evans and Jovanovic (1989) show that, when credit constraints bind, wealthier people are more likely to become entrepreneurs than less wealthy individuals, who often set up small business activities with suboptimal amounts of capital.[7] Similarly, Holtz-Eakin et al. (1994) posit that entrepreneurs who have personal resources, such as an inheritance, are more likely to survive in the market and earn higher returns (Evans and Leighton, 1989). Thus, credit constraints are critical to the creation and survival of new businesses. Our result is close to that in Ghatak and Jiang (2002), in which the individual's occupation depends on the endowment level. Unlike his work, in our model, the individual's wealth cannot be invested in the entrepreneurial project but can be used as collateral in the financial contract. If the initial wealth is high enough, the bank will find it profitable to offer the credit contract. Otherwise, the individual will have no choice but to work for a wage.

As for the policy implications, we follow the literature on the effects of public interventions in credit markets under imperfect information (Mankiw, 1986; Innes, 1991). Our policy setup is, in part, close to the moral-hazard section of Minelli and Modica (2009), in which they analyze a series of public interventions in a monopolistic credit market. They discuss the effects of two of the most widely used policy instruments, the interest-rate subsidy and the investment subsidy, and show that the former is optimal, as it maximizes net benefits for the government.[8]

The rest of the paper is as follows. Section 1 introduces the model. Section 2 characterizes the equilibrium. Section 3 draws policy implications. Section 4 presents a brief discussion. Section 5 concludes.

[6]Limited financial resources also explain why necessity entrepreneurs are more likely to compete by pursuing a cost leadership strategy (Dencker et al., 2009).

[7]The author also claims that small businesses grow faster than larger firms because of the tendency of small entrepreneurs to reinvest the returns on capital into their activity.

[8]They also argue that providing collateral directly to borrowers, before the contract is signed, is sub-optimal. In addition, they propose an innovative policy, "money in a savings account", in which the government directly provides the necessary collateral either to firms or banks depending on whether projects succeed or fail. They show that this policy has the same expected cost of the interest-rate subsidy.

3.1 The Setup

Consider a one-period, risk-neutral economy where an individual faces an occupational choice between paid employment and entrepreneurship. The individual has an initial (illiquid) wealth, W, and an exogenous reservation wage, ω, which can be interpreted as the employment options available in the labor market. Hence, as a wage worker, the individual has an outside option yielding utility $W+\omega \equiv u_R$. As a would-be entrepreneur, the individual is endowed with a project that requires a fixed investment, I, and yields a stochastic output, Y in case of success, and 0 in case of failure, with probabilities specified below. The initial wealth W cannot be invested in the project, but can be used as collateral in a financial contract. Thus, to undertake the entrepreneurial activity, the individual needs external credit, which can be provided by a single monopolistic bank. We restrict attention to a debt contract[9], which specifies the loan advanced, I, and the couple (R, S), where R is the amount the entrepreneur has to pay back in case of success, and S is the security (collateral) transferred to the bank in case of failure. The probability of project success depends on the effort, high or low, that the entrepreneur chooses to exert. With high effort, the project succeeds with probability p_H, whereas with low effort, with probability p_L, with $p_H - p_L = \Delta p > 0$. Low effort entails no cost to be implemented, whereas high effort requires a cost of e.

We assume that

$$Y > W + \tfrac{e}{\Delta p}. \tag{3.1}$$

This assumption implies that the maximum profit the bank can extract from the project, when high effort is exerted, is higher than that with low effort. However, the bank cannot observe the effort provided by the agent, and this implies a hidden action problem.[10] From Equation (4.1), it follows that, if a financial contract is offered, it will always promote the high-effort strategy.

Throughout the paper, we also assume limited liability and a risk-free rate normalized to 0.

3.2 Equilibrium

If a debt contract is signed, the individual's expected utility is

$$u(R, S) = W + p_H(Y - R) - (1 - p_H)S - e. \tag{3.2}$$

The agent accepts the contract if the project's expected payoff satisfies the participation constraint, that is

$$u(R, S) \geq u_R. \tag{PC}$$

[9] In the Remark below, we show that the theoretical conclusions would not change with equity financing.

[10] We exclude monitoring activity by the bank.

If we had full information, the effort would be observable by the bank, and the individual would choose the high-effort strategy. The full-information contract is any linear combination of R and S that satisfies (PC). For instance, the pair (R^{FI}, S^{FI}), with $R^{FI} = Y + W - (e + u_R)/p_H$ and $S^{FI} = W$. Under full information, the lender would extract all the rent from the entrepreneurial project, and the individual would obtain just the outside payoff, u_R. This means that the individual would be equally better off in the two employment alternatives, and the bank's expected profit would be

$$\pi(R^{FI}, S^{FI}) = p_H Y - e - \omega - I, \tag{3.3}$$

which, as we will see in subsection 3.2, is positive for $\omega < p_H Y - e - I$.

If information is asymmetric, the lender designs the financial contract to maximize the expected profit. From (4.1), if low effort is exerted, the bank would obtain a profit lower than with high effort. So, the bank will set R such that it is in the interest of the borrower to pursue the high-effort strategy, and maximize

$$\pi(R, S) = p_H R + (1 - p_H)S - I. \tag{3.4}$$

The agent chooses high effort if the incentive compatibility constraint,

$$p_H(Y - R) - (1 - p_H)S - e \geq p_L(Y - R) - (1 - p_L)S, \tag{IC}$$

is satisfied, that is

$$R \leq Y + S - \tfrac{e}{\Delta p} \equiv R^*,$$

where $e/\Delta p$ is the information rent the bank must give up to induce the "right" level of effort.

Since the bank's profit in (4.4) is increasing in both R and S, the bank will set the repayment and the collateral at the highest possible level, satisfying both (PC) and (IC). To promote the high-effort strategy, the lender will set $R = R^*$, which implies that the incentive constraint is binding. As for the collateral, it will be set by taking into account the repayment R^*, and the constraint $S \leq W$. The equilibrium collateral will be $S = \min\{W, S_{PC}\}$, where S_{PC} is obtained from the agent's participation constraint evaluated at R^*, that is

$$S = \bar{u} - \omega \equiv S_{PC},$$

with $\bar{u} = p_L e/\Delta p$. As we will show below, the value $\bar{u}$ is the project's expected payoff when the equilibrium collateral is W.

Under the contract $(R^*, \min\{W, S_{PC}\})$, the individual's expected payoff is

$$u(R^*, S) = W + \bar{u} - \min\{W, S_{PC}\}, \tag{3.5}$$

and, the bank's expected profit,

$$\pi(R^*, S) = p_H Y + \min\{W, S_{PC}\} - I - \tfrac{p_H e}{\Delta p}, \tag{3.6}$$

where I and $p_H e/\Delta p$ are, respectively, the loan cost and the expected information rent. For the sake of the exposition, we define the total lending costs as

$$I + \frac{p_H e}{\Delta p} \equiv C.$$

It is interesting to note that, (4.5) and (3.6) imply that the amount specified in the $\min\{W, S_{PC}\}$ is transferred to the lender regardless of whether the project succeeds or fails.[11]

The equilibrium collateral will depend on the values of u_R, that is

$$u_R = \bar{u} \implies W = \bar{u} - \omega \implies W = S_{PC}. \tag{3.7}$$

To derive the equilibrium contract, if any, we will distinguish two subcases. The first subcase refers to a condition of *necessity*, in which $u_R < \bar{u}$, and thus $W < S_{PC}$. The second subcase refers to a condition of *opportunity*, in which $u_R \geq \bar{u}$, and thus $W \geq S_{PC}$.

Note that, since u_R and $\bar{u}$ both depend on exogenous variables, the two subcases are not endogenously determined. As we will show, the possibility to choose the optimal employment alternative will depend on the loan profitability of the bank. Specifically, in necessity conditions, the willingness of the bank to offer a credit contract will depend on the value of the individual's initial wealth, W, whereas, in opportunity conditions, it will depend on the value of the reservation wage, ω.

3.2.1 Choice by necessity

If $u_R < \bar{u}$, then $W < S_{PC}$, and the equilibrium contract is (R^*, W). From (4.5), the agent's expected utility is

$$u(R^*, W) = \bar{u}. \tag{3.8}$$

Since $u_R < \bar{u}$, it follows that the entrepreneurial project yields a payoff higher than the reservation utility. So, when the job alternatives in the labor market are poor, the individual will find it more profitable to become an entrepreneur out of necessity. Note that, in this case, the individual would obtain a payoff higher than under the full-information contract.

As for the bank's profitability, under the contract (R^*, W), the bank's expected profit in (3.6) can be rewritten as

$$\pi(R^*, W) = p_H Y + W - C, \tag{3.9}$$

which is positive if

$$p_H Y + W \geq C. \tag{3.10}$$

[11] This result is similar to Minelli and Modica (2009). In their model, the individual has a reservation utility equal to zero, so that the bank must give up a lower fraction of its profit to induce the agent to participate.

The inequality in (3.10) implies that the bank will be willing to lend if $W \geq C - p_H Y$. The individual's initial wealth must be at least sufficient to compensate the bank when the expected project's return falls short of the lending costs. This represents the minimum amount of wealth needed to borrow. The assumption in (4.1) does not exclude that the expected output from the project can be lower than the sum of the expected information rent and the loan cost. If $p_H Y < C$, then the credit contract is offered, provided the condition in (3.10) holds. But, if $W < C - p_H Y$, then $\pi(R^*, W) < 0$, so the bank will be unwilling to lend, and the agent forced to become a wage worker due to the lack of other viable alternatives. In this case, credit constraints make the individual worse off as they result in a sub-optimal occupational choice. If $C - p_H Y \leq W < S_{PC}$, then the bank's profit in (3.9) is positive, and the individual will have the opportunity to start an owner-managed firm. Hence, in conditions of necessity, when the alternatives on the labor market are limited, and the entrepreneurial project is poor, an initial wealth relatively high gives the individual the chance to choose the optimal occupation. Whereas, if $p_H Y \geq C$, that is, if the project's expected output is higher than the costs of lending, then the bank will find it profitable to offer the contract, regardless of the individual's endowment.

Lemma 1. *In equilibrium, if $p_H Y \geq C$ a credit contract will be offered, whereas, if $p_H Y < C$, the bank may prefer to deny credit and offer no contract.*

From the results of this subsection, we can state the following.

Proposition 1. *If $u_R < \bar{u}$, the equilibrium contract is (R^*, W). The individual prefers setting up an owner-managed firm rather than entering the labor market as a worker. But, if $p_H Y + W < C$, the bank will not find it profitable to finance the entrepreneurial project, and the agent is forced to become a worker out of necessity.*

As mentioned earlier, a low outside option represents high unemployment, low education levels, and poverty, mainly present in rural areas. But it can also be the mirror of employment discrimination across gender and race, even in more developed regions. In such contexts, entrepreneurship can represent a way to escape poverty and gain personal fulfillment and autonomy. However, liquidity constraints may hinder the viability of the entrepreneurial project and force individuals to be wage workers out of necessity. Biding financial constraints are the byproduct of wealth inequality that only wealthier people can overcome. Indeed, the presence of a sufficient initial endowment that can be pledged against default can relax credit restrictions, but this may require secure property rights not always feasible in developing areas (Besley and Ghatak, 2008). As we will show, whether entrepreneurship should be encouraged by policy interventions will depend on the efficiency of the project.

Figure 1*a* shows the equilibrium payoffs when $p_H Y < C$. For $W < C - p_H Y$, the bank's profit is negative, and the agent obtains the reservation payoff, u_R, albeit it is lower

than the payoff from the entrepreneurial project (dashed line), which is not attainable. When $C - p_H Y \leq W < S_{PC}$, the bank's profit is positive and increasing in W, and the agent chooses to set up the business, obtaining the higher utility level, $\bar{u}$. Figure 1b shows the equilibrium payoffs when $p_H Y \geq C$. Credit constraints are not binding, and the individual chooses to become an entrepreneur. In the next subsection, the case in which $W \geq S_{PC}$ will be analyzed.

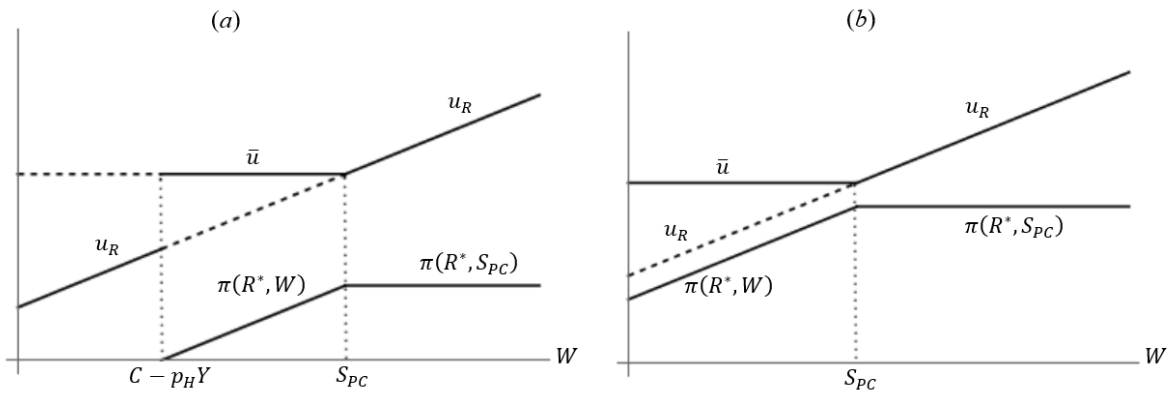

Figure 1: Equilibrium payoffs:
a) $p_H Y < C$. Parameters: $Y = 2$, $p_L = 0.4$, $p_H = 0.7$, $e = 0.7$, $I = 0.5$, $\omega = 0.2$;
b) $p_H Y \geq C$. Parameters: $Y = 4$, $p_L = 0.4$, $p_H = 0.7$, $e = 0.7$, $I = 0.5$, $\omega = 0.2$.

3.2.2 Choice by opportunity

If $u_R \geq \bar{u}$, then $W \geq S_{PC}$. The optimal contract is (R^*, S_{PC}) and the individual's expected utility from the project is

$$u(R^*, S_{PC}) = u_R, \tag{3.11}$$

which implies that the entrepreneurial activity yields a payoff equal to the reservation utility. Thus, when the initial endowment is relatively high, the agent will find it equally profitable to become an entrepreneur or a worker. In this case, the equilibrium collateral enables the lender to extract all the rent from the entrepreneurial project.

Under the equilibrium contract (R^*, S_{PC}), the bank's expected profit is

$$\pi(R^*, S_{PC}) = p_H Y - e - \omega - I, \tag{3.12}$$

equal to the full-information profit in (3.3).

As shown in Figure 1a and 1b, when $W \geq S_{PC}$, the individual obtains the reservation payoff, and the bank obtains a higher profit, which does not depend on the amount of the individual's endowment. However, the function in (4.10) is decreasing in the reservation wage, ω. The higher the employment opportunities on the labor market, the higher the fraction of the rent that the bank must lose to induce the agent to accept the contract. A high reservation wage weakens the monopoly power of the lender and strengthens the

bargaining power of the potential entrepreneur. However, as mentioned before, here we focus on small entrepreneurs, who have limited access to competitive banking sectors, and rely on informal lenders with considerable market power. The bank's expected profit in (4.10) is positive if

$$\omega < p_H Y - e - I \equiv \tilde{\omega}.$$

If $\omega > \tilde{\omega}$, then $\pi(R^*, S_{PC}) < 0$ and the bank will not offer any loan contract. In this case, the bank should provide a level of utility to the individual at least equal to ω. If the reservation wage is relatively high, offering the loan contract would be too costly for the bank, which thus prefers to deny credit. In the condition of opportunity, borrowing constraints do not make the individual worse off. Indeed, becoming an entrepreneur as well as working for a wage makes the agent equally better off.

From the results of this subsection, we can state the following.

Proposition 2. *If $u_R \geq \bar{u}$, the equilibrium contract is(R^*, S_{PC}). The individual is indifferent between setting up a firm and entering the labor market as a worker. But, if $\omega > \tilde{\omega}$, the bank will not find it profitable to finance the entrepreneurial project, and the agent will become a worker.*

It is interesting to note that if $\omega \geq \bar{u}$, then S_{PC} is equal or lower than zero. Since the security cannot be negative, the bank will set $S = 0$. In this case, under the contract $(R^*, 0)$, the individual's payoff from the entrepreneurial project is

$$u(R^*, 0) = W + \bar{u} < u_R. \tag{3.13}$$

The availability of many outside employment opportunities makes the project less attractive. In this case, the individual prefers to be a wage worker and, for the contract $(R^*, 0)$ there will not be a loan demand. If $p_H Y > C$, by Lemma 1, the lender would make a positive expected profit, $\pi(R^*, 0) > 0$. Thus, to attract the potential entrepreneur, the bank will be willing to lower the equilibrium repayment and set it such that the agent is equally better off between starting a firm and working for a wage, that is

$$R(\omega) = Y - \tfrac{e+\omega}{p_H} \equiv R_{PC}.$$

At this new equilibrium repayment, the individual's participation constraint is binding, and the incentive constraint is slack. Under the contract $(R_{PC}, 0)$, the bank will make the profit in (4.10), and again, the agent will be indifferent between the two occupational options. Therefore, if $\omega \geq \bar{u}$, although the equilibrium payoffs do not change, the contracts through which those payoffs are achieved are different.

Proposition 3. *If $p_H Y < C$, then $\tilde{\omega} < \bar{u}$ and the equilibrium contract is (R^*, S_{PC}). If $p_H Y \geq C$, then $\tilde{\omega} \geq \bar{u}$ and the equilibrium contract is $(R_{PC}, 0)$.*

This result implies that in conditions of opportunity, a project with a high expected return enables the potential entrepreneur to get the bank to decrease the equilibrium repayment, in spite of its monopolistic power.

Figure 2 below shows the plot of the bank's profit to ω. The cases where $\omega < \bar{u} - W$ refer to the condition of necessity examined in subsection 3.1, where the equilibrium contract is (R^*, W) and the profit of the lender does not depend on ω. When $\omega \geq \bar{u} - W$, the bank's profit declines as ω increases. From Proposition 2, if $\omega \geq \tilde{\omega}$, the bank is unwilling to lend, and no equilibrium exists. So, to determine whether an equilibrium under the contract $(R_{PC}, 0)$ is feasible, we need to check if $\tilde{\omega} \geq \bar{u}$. It is possible to note that, depending on the value of $p_H Y$, the threshold $\tilde{\omega}$ can be greater or lower than $\bar{u}$. Indeed, if the project expected output is lower than the costs of lending, that is $p_H Y < C$, then $\tilde{\omega} < \bar{u}$, so that the binding threshold will be $\tilde{\omega}$ and the equilibrium contract will be (R^*, S_{PC}), as depicted in Figure 2a. Whereas, if $p_H Y \geq C$, then $\tilde{\omega} \geq \bar{u}$, and the equilibrium with $(R_{PC}, 0)$ will be feasible, as in Figure 2b.

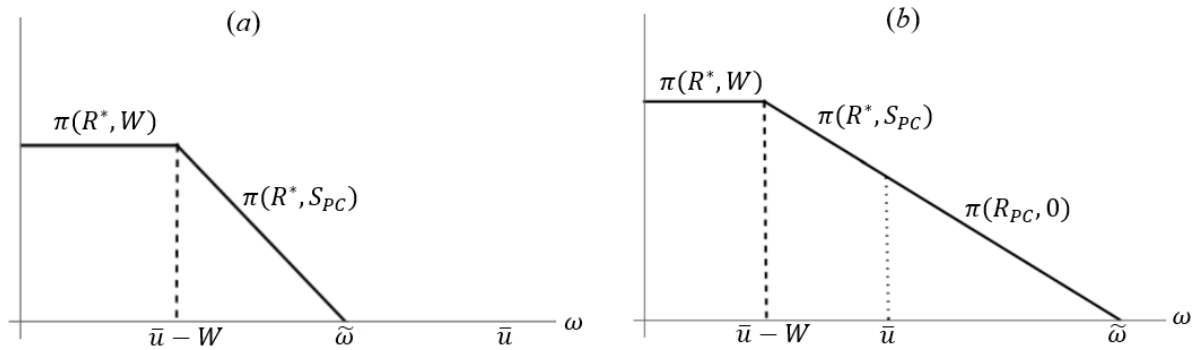

Figure 2: Bank's profit:
a) $p_H Y < C$. Parameters: $Y = 2$, $p_L = 0.4$, $p_H = 0.7$, $e = 0.7$, $I = 0.5$, $W = 0.8$;
b) $p_H Y \geq C$. Parameters: $Y = 4$, $p_L = 0.4$, $p_H = 0.7$, $e = 0.7$, $I = 0.5$, $W = 0.8$.

Remark In the paper we restrict attention to the case in which external funds are raised through a debt contract. This is quite so for small business, whose expected returns are generally low. If the return of the entrepreneurial project is observable, the bank can participate to a fraction of it by financing the indivdual through a combination of debt and equity finance. We would obtain the same result of the Section 3, even in the absence of collateral. Let $W = 0$ and (α, R) be the debt/equity contract, where α is the share of the investment I financed with debt, whereas $(1 - \alpha)$ is the share of I financed with equity. The payoff would be $u(\alpha, R) = \alpha p_H (Y - R) - e$ for the individual, and $\pi(\alpha, R) = p_H[(1 - \alpha)(Y - R) + R] - I$ for the bank. The individual's incentive constraint, $\alpha p_H (Y - R) - e \geq \alpha p_L (Y - R)$, would be satisfied for $R = Y - e/\alpha \Delta p$, which is equal to R^* for $\alpha = 1$ and $S = 0$. For this equilibrium repayment, the individual obtains $u(\alpha, R) = \bar{u}$, which is equal to the payoff derived in the subsection 3.1. Then, Proposition 1 also applies to the case with an equity/debt contract.

3.3 Policy

In the choice by necessity, borrowing constraints prevent the individual from choosing the profitable occupation. In this section, we focus on the case in which $u_R < \bar{u}$ and discuss some policy implications. As mention earlier, if enough initial wealth is available, the project is financed, and the individual is better off by choosing to be an entrepreneur. However, efficiency requires that the expected output from the investment must be at least equal to the resources used for the investment.

If the condition in (3.10) is satisfied, that is $p_H Y + W \geq C$, the lender is willing to fund the project. This inequality implies that the expected output from the project can be either higher or lower than the resources used, that is $I+e$. Depending on the value of $p_H Y$, we can distinguish two cases: $p_H Y < I+e$; $p_H Y > I+e$. In the first, the project is socially inefficient but, from the result in subsection (3.2.1), if the individual has enough endowment, the bank will find it profitable to finance the project, and the equilibrium will be inefficient. From a social perspective, it would be preferable to discourage investment with an interest or income tax, or a subsidy to the inactivity, as in de Meza and Webb (1987). Whereas, if the initial wealth is not sufficiently high, the bank is unwilling to lend, and the individual forced to become a wage worker out of necessity. In contrast to other models of business lending, in this case it would be socially optimal to prevent the individual from implementing the project, as this would result in a waste of resources. Therefore, the desirable scenario is a no-intervention policy.

When $p_H Y > I + e$ but $W < C - p_H Y$, the project is socially efficient, but credit constraints preclude the individual from setting up the business. Policy interventions aimed at encouraging the investment would be beneficial. We analyze below the effects of introducing either an interest-rate or investment subsidy, or a transfer to the entrepreneur that can be used as collateral in the credit contract. We then show that the policy cost does not depend on the instrument chosen to implement the intervention.

With the interest-rate subsidy, the incentive-compatible repayment, R^*, increases up to the level such that the bank makes the same profit as in the inefficient equilibrium, where the project is not financed, that is

$$\pi(R^*(1+\phi), W) = p_H R^*(1+\phi) + (1-p_H)W - I = 0,$$

with

$$\phi = \frac{p_H c - \Delta p(p_H Y + W - I)}{p_H[\Delta p(Y+W) - c]}. \tag{3.14}$$

Since the interest-rate subsidy is only provided if the project succeeds and is proportional to the repayment R^*, the expected cost of this policy is $p_H \phi R^* = C - p_H Y - W$.

With an investment subsidy, the loan investment provided by the bank is reduced by $\phi \in (0,1)$. The loan cost for the bank is thus $(1-\phi)I$. The subsidy is such that

$$\pi(R^*, W) = p_H R^* + (1-p_H)W - (1-\phi)I = 0,$$

that is

$$\phi = \frac{C - p_H Y - W}{I}, \tag{3.15}$$

and the policy cost is $\phi L = C - p_H Y - W$, as in the interest-rate subsidy.

As for the transfer to the entrepreneur, in this case the policy aims to provide the potential entrepreneur with the wealth needed to get the loan. Thus, the optimal size is

$$\hat{S} = C - p_H Y - W. \tag{3.16}$$

In equilibrium, the contract $(R^* + \hat{S}, W + \hat{S})$ is such that the bank will make non-negative profits, and the individual start the business. The cost of the policy is $\hat{S} = C - p_H Y - W$, as in the interest-rate and investment subsidies.

After the policy intervention, the individual can choose to become an entrepreneur and get the project payoff, $p_L e/\Delta p$. By comparing the policy gain with the cost, the net benefit from the intervention is $p_H Y + W - I - e$, which is positive since social efficiency requires that $p_H Y > I + e$.

Proposition 4. *In the choice by necessity, if $p_H Y > I + e$, the project is socially efficient, but if $W < C - p_H Y$, the bank denies credit and the equilibrium is inefficient. A policy intervention on the costs of lending or on the individual's wealth would restore efficiency. If $p_H Y < I + e$, the project is socially inefficient, and it would be optimal to not finance the project.*

But for the cases in which the project is inefficient, a policy intervention that reduces the costs of lending to the bank or provides the wealth needed to collateralize the loan can make the individual better off, becoming a necessity entrepreneur.

3.4 Discussion

In this section, we relax some assumptions made in the model and discuss the implications. We analyze the effects on the occupational choice of wealth investment, competitive banking sector, and individual risk aversion.

5.1 Wealth investment

In many pieces of research, would-be entrepreneurs can invest their initial wealth into the business activity, thereby lowering the loan cost. Here, we demonstrate how wealth is irrelevant to the equilibrium occupational choice.

We adopt the same set-up as before, but now we assume that the individual's initial wealth is liquid and can be invested in the entrepreneurial activity. However, the endowment is lower than I, and the individual needs outside financing to start the business. Funds are provided by the monopolistic lender, which now offers a contract specifying only the repayment to pay back in case of success and the loan advanced, $I - W$. Since W is invested in the business activity, the bank cannot use it as collateral to pledge

against default. Thus, the bank's expected profit is $\pi(R) = p_H R + W - I$, whereas the individual's participation and incentive constraints are $p_H(Y - R) - e \geq u_R$ and $p_H(Y - R) - e \geq p_L(Y - R)$.

If $u_R < \bar{u}$, then the incentive constraint is binding, and the participation constraint is slack. The incentive-compatible repayment thus requires $R \leq Y - \frac{e}{\Delta p} \equiv R^W_{IC}$, which is equal to R^* when $S = 0$. Since the bank's profit is increasing in R, the bank will set the repayment at the highest possible level, that is $R = R^W_{IC}$. The individual's expected utility from the project is $u(R^W_{IC}) = \bar{u}$. Since $\bar{u} > u_R$, the individual prefers starting a business to working for a wage. The bank's expected profit, $\pi(R^W)$, is equal to that in (3.9), which is positive if $p_H Y \geq C - W$. The initial wealth should be high enough to equalize (at least) the difference between the lending costs and the project expected return. Otherwise, the bank would not lend, and the individual forced to be a wage worker. Thus, in choice by necessity, whether the initial wealth is used as collateral or invested in the business activity, as claimed in Proposition 1, the individual can be made worse off by borrowing constraints.

If $u_R \geq \bar{u}$, then the participation constraint is binding, and the incentive constraint is slack. The bank thus sets $R = Y - [(W + e + \omega)/p_H] \equiv R^W_{PC}$. In this case, the individual's expected utility is $u(R^W_{PC}) = u_R$. The entrepreneurial occupation gives the individual the same payoff as working for a wage and the individual will be indifferent between the two employment alternatives. The bank's expected profit is $\pi(R^W_{PC}) = p_H Y - e - \omega - I$, as in (4.10), which is negative if $\omega < \tilde{\omega}$. Thus, as stated in Proposition 2, when the reservation wage is relatively high, the bank can deny credit and constrain the individual, who is never worse off in choice by opportunity.

5.2 Competitive banking sector

If the banking sector were competitive, the individual would always have the opportunity to choose the best occupational alternative.

The equilibrium contract would be any linear combination of R and S such that the bank obtains zero profit and (IC) is satisfied. From the zero-profit condition and the binding (IC), the equilibrium repayment is $R_0 = [I - (1 - p_H)S]/p_H$, and the equilibrium security is $S_0 = C - p_H Y$, which is positive if $p_H Y < C$. If $p_H Y \geq C$, then S_0 is negative and the bank sets $S_0 = 0$. At this equilibrium collateral the bank still obtains zero profit and the incentive constraint holds. The individual's expected payoff from the project is $u(R_0, S_0) = W + \tilde{\omega}$. If $\omega < \tilde{\omega}$, then $u(R_0, S_0) > u_R$ and the individual will prefer to start a business. If $\omega > \tilde{\omega}$, then $u(R_0, S_0) < u_R$ and the individual will prefer to become a wage worker to the entrepreneurial occupation. In this case, the bank cannot reduce the repayment further which otherwise would result in a negative profit.

5.3 Risk aversion

Throughout the analysis, we assume universal risk neutrality. Now, we consider a risk-averse individual and a risk-neutral lender and show that the qualitative results remain unchanged. Since a low outside payoff may represent high uncertainty, we restrict atten-

tion to the choice by necessity.

Let the individual's utility function be qausi-linear in the wealth and the effort cost, i.e. $W+u(\cdot)-e$ with $u'(\cdot)>0$, $u''(\cdot)<0$ and $u(0)=0$. Let x be the net return in case of project success, then the participation constraint would be $p_H u(x)-(1-p_H)u(S)-e \geq u(\omega)$ and the incentive constraint $p_H u(x)-(1-p_H)u(S)-e \geq p_L u(x)-(1-p_L)u(S)$. From the binding incentive constraint, the equilibrium repayment is R^*. If $W+u(\omega)<p_L(x)$, which corresponds to $W+u(\omega)<p_L e/\Delta p$, the individual initial wealth is lower than S_{PC}, then $u(x)=c/\Delta p$ and $u(S)=W$. Let $u(x)=u$, it follows that $x=v(u)$, where $v(u)$ is a convex function. The individual's expected utility from the entrepreneurial project is $u(R^*,W)=W+p_H u(x)-(1-p_H)u(S)-e=p_L e/\Delta p>W+u(\omega)$. Thus, under risk aversion, if the outside job opportunities are low, the individual still prefers to become an entrepreneur out of necessity. The bank's expected profit is $\pi(R^*,W)=p_H Y+W-I-pHv(u)=p_H Y+W-I-p_H v(c/\Delta p)$. The loan contract will be profitable for the bank if $p_H Y+W \geq I+p_H v(c/\Delta p)$. Thus, the result in Proposition 1 holds, provided the individual is not too risk averse.

3.5 Conclusions

This paper provides a new interpretation of necessity and opportunity entrepreneurs and focuses on the behavior of a monopolistic lender in the presence of market frictions, such as asymmetric information and liquidity constraints. We analyze the occupational choice of an individual between setting up a firm or supplying the labor force for a wage. We show that borrowing constraints may force the agent to the suboptimal occupation out of necessity. Whereas, in the case of opportunity, the agent can choose indifferently between the two employment alternatives.

Understanding the factors behind the occupational choice is helpful to design adequate policy measures that encourage entrepreneurial activity so as to promote the economic growth. As argued by Blanchflower (2000), a country with high economic growth indicates high job opportunities, both as a self-employed entrepreneur as well as a wage-paid worker. However, because of financial restrictions, entrepreneurship may not be a viable option for wealth-constrained agents, who do not have enough resources to secure against default. This is particularly true for younger people, who have less time to build up the capital needed to borrow and start a business. In addition to liquidity constraints, socio-demographic, cultural, technological, and institutional factors can be the main determinants of the entrepreneurial growth rate across countries (Noorderhaven et al., 2004; Armington and Acs, 2002; Estrin et al., 2013). Many OECD countries have adopted different programs to advocate entrepreneurship among unemployed, such as job traninng, start-up subsidies for the unemployed assistance in job searching (Caliendo and Künn, 2011; Battisti et al., 2019).

Bibliography

Ahunov, M. and Yusupov, N. (2017). Risk attitudes and entrepreneurial motivations: Evidence from transition economies. *Economics Letters*, 160:7–11.

Armington, C. and Acs, Z. (2002). The determinants of regional variation in new firm formation. *Regional Studies*, 36:33–45.

Baker, T., Gedajlovic, E., and Lubatkin, M. (2005). A framework for comparing entrepreneurship processes across nations. *Journal of International Business Studies*, 36:492–504.

Banerjee, A. (2013). Microcredit under the microscope: What have we learned in the past two decades, and what do we need to know? *Annual Review of Economics*, 5:487–519.

Banerjee, A. and Newman, A. (1993). Occupational choice and the process of development. *Journal of Political Economy*, 101:274–298.

Battisti, M., Giesing, Y., and Laurentsyeva, N. (2019). Can job search assistance improve the labour market integration of refugees? Evidence from a field experiment. *Labour Economics*, 61:101745.

Beck, T. and Demirgüç-Kunt, A. (2006). Small and medium-size enterprises: Access to finance as a growth constraint. *Journal of Banking & Finance*, 30:2931–2943.

Beck, T., Demirgüç-Kunt, A., and Maksimovic, V. (2004). Bank competition and access to finance: International evidence. *Journal of Money, Credit and Banking*, 36:627–648.

Besley, T. (1994). How do market failures justify interventions in rural credit markets? *The World Bank Research Observer*, 9:27–47.

Besley, T. and Ghatak, M. (2008). Creating collateral: The de soto effect and the political economy of legal reform. Technical report, Working Paper. London: London School of Economics.

Blanchflower, D. (2000). Self-employment in OECD countries. *Labour Economics*, 7:471–505.

Caliendo, M. and Künn, S. (2011). Start-up subsidies for the unemployed: Long-term evidence and effect heterogeneity. *Journal of Public Economics*, 95:311–331.

Carree, M., Van Stel, A., Thurik, R., and Wennekers, S. (2002). Economic development and business ownership: An analysis using data of 23 OECD countries in the period 1976–1996. *Small Business Economics*, 19:271–290.

Carter, N., Gartner, W., Shaver, K., and Gatewood, E. (2003). The career reasons of nascent entrepreneurs. *Journal of Business Venturing*, 18:13–39.

Caserta, M., Monteleone, S., and Reito, F. (2018). The trade-off between profitability and outreach in microfinance. *Economic Modelling*, 72:31–41.

Clark, K. and Drinkwater, S. (2000). Pushed out or pulled in? Self-employment among ethnic minorities in england and wales. *Labour Economics*, 7:603–628.

Cull, R., Demirgu¨ ç Kunt, A., and Morduch, J. (2007). Financial performance and outreach: A global analysis of leading microbanks. *The Economic Journal*, 117:F107–F133.

de Meza, D. and Webb, D. (1987). Too much investment: A problem of asymmetric information. *Quarterly Journal of Economics*, 102:281–292.

Delis, M., Kokas, S., and Ongena, S. (2017). Bank market power and firm performance. *Review of Finance*, 21:299–326.

Dencker, J., Gruber, M., and Shah, S. (2009). Pre-entry knowledge, learning, and the survival of new firms. *Organization Science*, 20:516–537.

Estrin, S., Korosteleva, J., and Mickiewicz, T. (2013). Which institutions encourage entrepreneurial growth aspirations? *Journal of Business Venturing*, 28:564–580.

Evans, D. and Jovanovic, B. (1989). An estimated model of entrepreneurial choice under liquidity constraints. *Journal of Political Economy*, 97:808–827.

Evans, D. and Leighton, L. (1989). The determinants of changes in US self-employment, 1968–1987. *Small Business Economics*, 1:111–119.

Ghatak, M. and Jiang, N. (2002). A simple model of inequality, occupational choice, and development. *Journal of Development Economics*, 69:205–226.

Guha, B. and Chowdhury, P. (2013). Micro-finance competition: Motivated micro-lenders, double-dipping and default. *Journal of Development Economics*, 105:86–102.

Holtz-Eakin, D., Joulfian, D., and Rosen, H. (1994). Sticking it out: Entrepreneurial survival and liquidity constraints. *Journal of Political Economy*, 102.

Hossein, C. (2013). The black social economy: Perseverance of banker ladies in the slums. *Annals of Public and Cooperative Economics*, 84:423–442.

Innes, R. (1991). Investment and government intervention in credit markets when there is asymmetric information. *Journal of Public Economics*, 46:347–381.

Iyigun, M. and Owen, A. (1998). Risk, entrepreneurship, and human-capital accumulation. *American Economic Review*, 88:454–457.

Jackson, R. (1984). An evaluation of alternative measures of regional industrial diversification. *Regional Studies*, 18:103–112.

Kar, A. and Swain, R. (2014). Interest rates and financial performance of microfinance institutions: Recent global evidence. *European Journal of Development Research*, 26:87–106.

Mankiw, N. (1986). The allocation of credit and financial collapse. *Quarterly Journal of Economics*, 101:455–470.

Mersland, R. and Strøm, R. (2010). Microfinance mission drift? *World Development*, 38:28–36.

Minelli, E. and Modica, S. (2009). Credit market failures and policy. *Journal of Public Economic Theory*, 11:363–382.

Monteleone, S. and Reito, F. (2018). Cooperative firms in hard times. *Small Business Economics*, 51:171–179.

Mookherjee, D. and Motta, A. (2016). A theory of interactions between MFIs and informal lenders. *Journal of Development Economics*, 121:191–200.

Myers, S. (1977). Determinants of corporate borrowing. *Journal of Financial Economics*, 5:147–175.

Naudé, W. (2010). Entrepreneurship, developing countries, and development economics: New approaches and insights. *Small Business Economics*, 34:1–12.

Noorderhaven, N., Thurik, R., Wennekers, S., and Van Stel, A. (2004). The role of dissatisfaction and per capita income in explaining self–employment across 15 European countries. *Entrepreneurship Theory and Practice*, 28:447–466.

Reynolds, P., Bosma, N., Autio, E., Hunt, S., De Bono, N., Servais, I., Lopez-Garcia, P., and Chin, N. (2005). Global entrepreneurship monitor: Data collection design and implementation 1998–2003. *Small Business Economics*, 24:205–231.

Roberts, P. (2013). The profit orientation of microfinance institutions and effective interest rates. *World Development*, 41:120–131.

Roemer, J. (1998). *Equality of opportunity*. Harvard University Press.

Rotefoss, B. and Kolvereid, L. (2005). Aspiring, nascent and fledgling entrepreneurs: An investigation of the business start-up process. *Entrepreneurship & Regional Development*, 17:109–127.

Serrano-Cinca, C. and Gutiérrez-Nieto, B. (2014). Microfinance, the long tail and mission drift. *International Business Review*, 23:181–194.

Stiglitz, J. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Economic Review*, 71:393–410.

Valliere, D. and Peterson, R. (2009). Entrepreneurship and economic growth: Evidence from emerging and developed countries. *Entrepreneurship & Regional Development*, 21:459–480.

Wennekers, S., Van Wennekers, A., Thurik, R., and Reynolds, P. (2005). Nascent entrepreneurship and the level of economic development. *Small Business Economics*, 24:293–309.

Williams, C. (2007). Entrepreneurs operating in the informal economy: Necessity or opportunity driven? *Journal of Small Business & Entrepreneurship*, 20:309–319.

Williams, N. and Williams, C. (2014). Beyond necessity versus opportunity entrepreneurship: Some lessons from english deprived urban neighbourhoods. *International Entrepreneurship and Management Journal*, 10:23–40.

Yamada, G. (1996). Urban informal employment and self-employment in developing countries: Theory and evidence. *Economic Development and Cultural Change*, 44:289–314.

Chapter 4

Better to be in the same boat: Positional envy in the workplace

In a simple agency model of the labor market, we examine how fairness concerns affect the structure of optimal contracts. In the framework, we consider two types of workers, high and low ability, with the latter being envious and incurring a utility loss whenever the more talented earn a surplus from their contracts. We focus on the equilibrium payoff of the envied and show that, when the ability gap is relatively small, it is first increasing and then decreasing in the level of envy cost borne by the envious. In contrast, when the gap is large, the payoff is always decreasing. We also find that the utility loss of the envious is higher the lower the skill heterogeneity between types.

"We envy those who are near us in time, place, age, or reputation"
Aristotle.

Introduction

Human beings are social by nature, and thus they weave a dense network of socioeconomic relationships in which they interact and compare themselves with each other. Many studies suggest that agents derive utility or disutility from peer comparison and care about their social rank (Postlewaite, 1998; Ball et al., 2001). In a wide range of social contexts, competition exacerbates the process of interpersonal comparison and group identification. This is particularly true in workplace settings, where relative-pay inequalities and other-regarding preferences can have substantial incentive effects, positive or negative, especially when workers are heterogeneous in their skill levels (Bolton and Ockenfels, 2000; Fehr et al., 2009; Breza et al., 2018).

This paper examines how fairness concerns can affect the structure of optimal labor contracts when workers have different inherent skills or abilities, and so different productivities. We present a simple agency model with an employer (principal) and many

workers (agents), distinguished into two types, low productive and high productive. Low-productive workers incur a utility loss whenever the more talented earn a surplus from their labor contracts. Namely, the more talented can impose (in our case, inadvertently) a sense of "inferiority" upon the less talented and trigger a feeling of resentment or envy. In turn, the principal has to compensate the less talented for their envy cost but, at the same time, needs to pay the same reward to the more talented to avoid mimicking behavior. This will eventually increase the pay disparity among workers and then the loss from envy. Unlike previous research (Desiraju and Sappington, 2007; Bartling and Von Siemens, 2010; Manna, 2016), we show that both the size of the ability gap between the two types of workers and the magnitude of the disutility cost play a key role in determining the effect of envy on the payoffs of players. Specifically, our model predicts that, when the ability gap is relatively small, the surplus of the envied is first increasing and then decreasing in the disutility cost of the envious. In contrast, when the ability gap exceeds a certain threshold, the payoff of the envied is always decreasing in the envy cost. Since envy translates into a monetary reward for both types of agents, the surplus that more talented obtain depends not only on the information rent but also on the compensation needed to satisfy their incentive compatibility constraint. This conclusion holds even in the most simplifying case where high-ability workers do not directly derive any utility from being envied.

In the model, we also consider the effect of task complexity or difficulty (technology intensity) on the payoff of high-talented workers. The task complexity on the job may intensify or mitigate the effect of social comparison, and thus can be viewed as an alternative measure of ability heterogeneity. We interpret the slope of the effort cost function of workers as task complexity, and show that, if the job complexity is relatively small, the payoff of high-skilled workers will be first increasing and then decreasing in the envy cost, as in the case of low ability gap.[1] If the job complexity is higher, the effect of other-regarding preferences becomes less prominent, and the payoff of high-ability workers will be always decreasing.

In our stylized workplace setting, envy stems from a feeling of unfairness, because high-ability agents receive a wage higher than their marginal productivity and so their surplus is perceived as undeserved. The envy utility loss increases as the productivity gap between worker types shrinks. The reason is that, if abilities are similar, separating the two types of workers is more difficult, and the size of information rent that the principal must give up to the more talented is relatively high. When, instead, workers are considerably different in skills, separation becomes easier, and this lowers the information rent and the disutility of the envious. Therefore, a workplace with small skill heterogeneity does not mitigate the cost resulting from envy but, rather, it will raise the "pain" of enviers. In

[1]This result is in line with Nickerson and Zenger (2008), who posit that some managers have the tendency to increase the physical and relational distance among employees to reduce social comparison costs, even at the expense of production efficiency.

contrast, a higher skill gap leads to a decrease in both the utility loss of the envious and in the payoff of the envied.

Our main conclusion is that envied workers may end up benefiting from the presence of envious colleagues when they are in a superior position, but not too high. This result implies that, in our model, high-productive agents would have no incentive to sort themselves into homogenous groups, as, for example, in the marriage market of Becker (1973) or in the credit market of Ghatak (1999). Individuals with high abilities will prefer a work environment where they can stand out in a team, but not so markedly as one would expect. Being the most productive in a group of low performing and envious workers is likely to be the most undesirable situation in terms of monetary payoff (frog pond effect). Similarly, for the principal, we will show that workers with slightly different abilities, even though not completely uniform, generate higher expected profit.[2] These results can have interesting empirical implications on the effects of job-skill heterogeneity and peer-group composition though, from an observational point of view, it is extremely difficult to conduct an investigation of envy feelings and dynamics in the workplace. Empirically, Mas (2006), Card et al. (2012), Cohn et al. (2014), Song et al. (2019) find evidence of relative-pay disparities and fairness concerns in the labor market. Mas (2006) tests the sensitivity of employee productivity to relative pay variations. By using market data from final offer arbitration for police unions in New Jersey, he shows that the performance of officers declines as wage expectations are disappointed. Card et al. (2012) conduct a randomized experiment at the University of California on peer wages revelation. They find evidence that disadvantageous pay inequality reduces job satisfaction and pushes workers to look for another job. Cohn et al. (2014) report results from a field experiment of cutting wages in Germany and report that an individual rather than a general wage reduction has a strong and negative effect on worker performance and conduct. In addition, using a longitudinal data set of the U.S. labor market, Song et al. (2019) find that equity aversion may have sorting effects on the composition of the labor force within firms. Our result is consistent with that of Breza et al. (2018), who, by using data on seasonal manufacturing jobs in India, document that the ability of workers may provide a rationale for pay disparities. Indeed, they show that the higher wage of more productive workers (when the ability is observable) is perceived as fair, and thus, it does not affect production and group cohesion.

Related Literature

The effect of inequality and positional concerns has been analyzed in the literature from both sociological and economic perspectives. According to the social comparison theory, first introduced by Festinger (1954), individuals have a hardwired tendency to compare

[2]This result is supported by the empirical analysis of Song et al. (2019), in which they identify an increasing trend toward a homogeneous composition of skills in workplace settings.

themselves with others, with the ultimate objective of improving the quality of their standard of living (Ben-Ze'ev, 1992; van de Ven et al., 2009). Fiske (2010) posits that the social comparison is intrinsically inevitable across all relational domains, and involves friends, relatives and colleagues. We confirm here that this tendency may decline as the closeness and similarities with the "target" are less prominent.

From an economic perspective, the relevance of relative wage differentials in the labor market was already implicitly pointed out in the *General Theory*. Explicitly, Duesenberry (1952) was the first to suggest that individuals have interdependent preferences and their utility is affected by the disparity or gap between their own consumption (broadly defined) and that of a reference group. Depending on whether the target is worse or better off, the social comparison may be downward or upward. While individuals engage in downward comparison to improve their self-esteem or gain superiority over their rivals, the goal of upward comparison involves self-evaluation and self-enhancement. The potential disadvantageous position resulting from the comparison can generate feelings of inferiority resentment, which may turn into jealousy or envy towards others (Smith and Kim, 2007). Envy is an emotional state that occurs when a person lacks another's superior quality or possession. This feeling is usually arises when the other's success threatens the self-esteem and in general the well-being of the envious (Rustemeyer and Wilbert, 2001; Grund and Sliwka, 2005). Although envy may sometimes encourage friendly competition among employees, it is often associated with undesirable consequences. The sociological literature indentifies two main behavioral patterns, malicious and benign envy (Crusius and Lange, 2014). In some circumstances, envy can stimulate motivation and work engagement (Grant and Mayer, 2009; Tai et al., 2012), but in other (perhaps most) situations envy can lead to a series of negative consequences, such as personal aggression, sabotage or hostility towards colleagues and the organization as a whole (Silver and Sabini, 1978; Ostell, 1996; Vecchio, 2000; Cohen-Charash and Mueller, 2007; Khan et al., 2014).

In Aristotle's thought, envy is interpreted as the pain caused by "those who have what we ought to have". This implies that envy can be positional, in the sense that it emerges when there is a certain level of proximity with the envied, and when their success is potentially achievable (Solnick and Hemenway, 2005). The relationship between envy and "closeness" of the reference group can also be inferred from the tenth commandment: "You shall not covet [...] anything that belongs to your neighbor". Ben-Ze'ev (1992) argues that the displeasure caused by envy strictly depends on the dichotomy between competition and comparison, and more specifically, on whether the success of our comparative stand is attainable or not. Envy may derive from a feeling of "injustice" because those similar to us, with whom we compete, have obtained something that was not so out of reach. In other words, as claimed by van de Ven and Zeelenberg (2015), it is the thought "It could have been me" that makes us want what others have. Individuals close to us but in a slightly superior position indirectly emphasize our inability more than those distant

from us. Hesiod argues that "potter is furious with potter and craftsman with craftsman, and beggar is envious of beggar and singer of singer". Descartes stresses the undeserved attribute that goes along with the sentiment of envy, "We judge the others unworthy of their good."

In his seminal work, Frank (1984) studies the implications of relative preferences and fairness concerns on the labor market. He suggests that, if workers care about their relative payoffs, wages may not reflect their marginal productivity. He shows that the more talented receive less than their marginal contribution, as they *directly* enjoy being the higher-net earners, whereas the less talented receive more because they need to be compensated for their positional disutility. Fehr and Schmidt (1999) develop an inequity aversion model and show that in competitive environments, such as the labor market, fairness concerns are more likely to lead to equitable outcomes when workers threaten to reduce their effort and thus the principal's material payoff. Loewenstein et al. (1989) estimate how relative payoffs affect individual utility functions and provide evidence that people dislike disadvantageous as well as advantageous inequality, though the latter effect is weaker than the former. In contrast, we will show that, thanks to the structure of the incentive scheme chosen by the principal, the more talented can benefit from their favorable position, even in the case where they do not derive any direct utility or disutility from being envied. Itoh (2004) embeds other-regarding preferences in a standard principal-agent model and investigates the design of optimal incentive contracts. He argues that, in the presence of workers with interdependent preferences, the principal may find it profitable to offer team contracts to reduce the negative impact of envy. Similarly, Desiraju and Sappington (2007) show that, when workers are *ex-ante* different in abilities, equity concerns induce the principal to give up some rent to avoid *ex-post* inequality. Bartling and Von Siemens (2010) introduces a moral-hazard model with unlimited liability and risk-averse agents, envious whenever others receive a higher wage. They show that, since envy increases the cost of the incentive scheme, the principal will elicit cost-minimizing efforts from their agents through pooling contracts. Unlike their papers, in our setup, the principal never has the incentive to offer flat wages. Neilson and Stowe (2010) introduce other-regarding preferences in a principal-agent model with *ex-ante* identical workers, and conclude that piece wages push inequity-averse agents to exert a higher effort than inequity-neutral agents. In our model, workers are *ex-ante* different, and disadvantageous inequality drags down the effort of other-regarding workers.

In other models, such as Dur and Glazer (2008) and Manna (2016), workers can be envious both of their colleagues and their boss. In Dur and Glazer (2008), the solution proposed is a sharing-profit contract, which may reduce the utility loss generated by different employment roles or status. In the principal-agent model of Manna (2016), she shows that the complementary between the two types of envy can mitigate the distortion on the effort exerted by low-productive workers. In our paper, we focus on envy among peers in order to address the issue of closeness or distance, in terms of labor skills, with

targets in comparable job positions (though, in the second part of the model, we briefly analyze the effect on the profit obtained by the principal of pursuing envy-reduction strategies).

The paper proceeds as follows. The next section introduces the setup. Then, we describe the equilibrium. Finally, some brief conclusions are drawn.

4.1 Setup

Consider a single-period, risk-neutral, labor-market model with a principal and many agents, distinguished in two types, high ability (H) and low ability (L). The principal offers labor contracts that specify the wage, ω_i, with $i \in \{H, L\}$, and the effort exerted, e_i. As in Manna (2016), the effort is assumed to be observable and contractible (for example, the number of hours worked in a day or week). The effort costs depend on the workers' ability, θ_i, and are equal to $\theta_i e_i^{\alpha}/\alpha$, with $\alpha > 1$. The parameter α can be interpreted as the level of task complexity or difficulty of the job position. We will show that the equilibrium effort levels are such that, for both types of workers, $e_i \in (0, 1]$, so the higher α, the lower the curvature of the cost function and thus the complexity of the task. For H workers, $\theta_H = 1$, whereas for L workers, $\theta_L = \theta > 1$. The ability gap between types is thus $\Delta\theta = \theta - 1$. Workers have a reservation wage normalized to 0.

While workers know theirs and each other's productivity, the principal only knows the proportions, μ and $1 - \mu$, of H and L types in the population. So, there is a problem of asymmetric information. We assume that low-ability workers are envious of their high-ability peers whenever the latter are *expected* to receive a positive surplus from their contracts. Envy entails a disutility loss for L types, which is proportional to the parameter $c \in [0, 1]$, and to the expected surplus obtained by H workers. The parameter c measures the intensity of the envy cost and it is assumed to be common knowledge.[3] Throughout the paper, it is also assumed that H types do not derive any *direct* benefit from being envied (there is no envy-enjoyment).[4]

The timing of the game is as follows. 1) Nature determines α, μ and c. 2) The principal offers labor contracts. 3) Workers choose whether to accept or not. 4) If workers accept, production takes place and wages are paid.

[3]This assumption is standard in this type of literature (see, for example, Caserta et al. 2021, Manna, 2016, and Dur and Glazer, 2008). Although envy cannot be measured directly, the empirical and experimental studies use some proxies to estimate the magnitude of its effects (see, for example, Smith et al., 1999).

[4]This assumption is made to simplify the analysis. A direct, positive effect of envy on the well-being of H types would strengthen the qualitative results.

Benchmark: symmetric information

Had we perfect information on each worker's ability, the principal would offer two contracts such that the participation constraint of each type of worker is satisfied with equality. The equilibrium contract would be such that the payoff of low-ability and high-ability agents, under the contracts (ω_L, e_L) and (ω_H, e_H), are

$$u_L(\omega_L, e_L) = \omega_L - \tfrac{\theta}{\alpha} e_L^{\alpha} = 0, \quad \text{and} \tag{4.1}$$

$$u_H(\omega_H, e_H) = \omega_H - \tfrac{1}{\alpha} e_H^{\alpha} = 0, \tag{4.2}$$

yielding $\omega_H = e_H^{\alpha}/\alpha$ and $\omega_L = \theta e_L^{\alpha}/\alpha$.

The principal's expected profit (on the average worker) is

$$\pi = \mu(e_H - \omega_H) + (1-\mu)(e_L - \omega_L), \tag{4.3}$$

where we assume that the return is equal to the effort exerted by workers.

By replacing ω_H and ω_L in (4.3), and taking the first-order conditions, the full-information effort levels are

$$e_L^{FB} = \left(\tfrac{1}{\theta}\right)^{\sigma} \quad \text{and} \quad e_H^{FB} = 1,$$

with $\sigma = 1/(\alpha - 1)$ and with $e_L^{FB} < e_H^{FB}$. The first-best wages are

$$\omega_L^{FB} = \tfrac{1}{\alpha\theta^{\sigma}}, \quad \text{and} \quad \omega_H^{FB} = \tfrac{1}{\alpha},$$

with $\omega_L^{FB} < \omega_H^{FB}$.

The expected profit of the principal is

$$\pi = \tfrac{(\alpha-1)[\theta^{-\sigma}(1-\mu)+\mu]}{\alpha} \equiv \pi^{FB}.$$

Welfare is the sum of the utility of both types of workers, weighted for their fraction in the population, and the principal's profit. If the workers' ability is observable, agents receive a wage equal to their marginal productivity, so that the only surplus produced is the principal's profit. So, the full-information welfare W^{FB} is equal to π^{FB}.

When workers' ability is *ex-ante* observable, full-information contracts do not yield any contract rent for H workers, so that no envy feeling arises for L types.[5] The higher wage of high-ability workers is simply the "fair" reward for the higher effort they exert.

4.2 Equilibrium

With asymmetric information, the principal can offer one of three contract types: incentive-compatible separating contracts; pooling (flat-wage) contracts; screening contracts to H

[5] In Manna (2016), even under full information, workers may feel envy towards the principal.

workers (keeping L types out). In what follows, we analyze the first type of offer and then we show that the other two are less profitable for the principal.

If abilities are unobservable, the effect of envy is such that the participation constraints of L and H agents, under the separating contracts (ω_L, e_L) and (ω_H, e_H), can be written as:

$$u_L(\omega_L, e_L) = \omega_L - \tfrac{\theta}{\alpha} e_L^{\alpha} - c \cdot \max\{0, u_H(\omega_H, e_H)\} \geq 0; \qquad (PC_L)$$

$$u_H(\omega_H, e_H) = \omega_H - \tfrac{1}{\alpha} e_H^{\alpha} \geq 0. \qquad (PC_H)$$

In (PC_L), envy entails a utility cost, which is proportional to the parameter c and to the surplus H types receive from their contract.

The incentive compatibility constraints are:

$$u_L(\omega_L, e_L) = \omega_L - \tfrac{\theta}{\alpha} e_L^{\alpha} \geq \omega_H - \tfrac{\theta}{\alpha} e_H^{\alpha} = u_L(\omega_H, e_H); \qquad (IC_L)$$

$$u_H(\omega_H, e_H) = \omega_H - \tfrac{1}{\alpha} e_H^{\alpha} \geq \omega_L - \tfrac{1}{\alpha} e_L^{\alpha} = u_H(\omega_L, e_L). \qquad (IC_H)$$

As said, being envious is an intrinsic characteristic of L workers, whereas H types are not envious, nor they derive utility from being envied. Hence, if H workers obtain a surplus, this gives rise to an envy loss for L workers even in the case the latter were to accept the contract (ω_H, e_H). So, in both right- and left-hand side of (IC_L), the envy-cost terms simplify. In (IC_H), H workers they do not incur any utility loss (even if they were to choose the contract designed for L types). From the participation constraint in (PC_L), it is possible to note that c cannot be larger than 1, otherwise L agents would receive a surplus higher than H types and the latter would prefer the contract (ω_L, e_L).

In this class of incentive problems, the binding constraints are the participation constraint of the "low" type and the incentive constraint of "high" type (in the Appendix we show that the other constraints are satisfied in equilibrium). Hence, from the binding (PC_L) and (IC_H), and from the first-order conditions of the principal's profit function in (4.3), the equilibrium effort levels are

$$e_L = \left[\tfrac{(1-\mu)(1-c)}{\theta - c(1-\mu) - \mu}\right]^{\sigma} \equiv \widehat{e}_L \quad \text{and} \quad e_H = 1 = e_H^{FB} \equiv \widehat{e}_H,$$

where $\widehat{e}_L > 0$, as $\theta > 1$, and $\widehat{e}_L < e_L^{FB}$.

If L workers had no fairness concerns, that is if $c = 0$, the effort of L types would be $\widehat{e}_L = (1-\mu)/(\theta - \mu)$, lower than e_L^{FB}. If the envy cost is positive, there is a further distortion in the effort of L types. In addition, the more envious L workers are, the lower their equilibrium effort, as $d\hat{e}_L/dc < 0$. At the extreme, $\widehat{e}_L = 0$ when $c = 1$. In contrast, the effort of H types does not depend on c and is equal to the full-information level.

The equilibrium wages are:

$$\omega_L = \tfrac{\theta - c}{\alpha(1-c)} \cdot \widehat{e}_L^{\alpha} \equiv \widehat{\omega}_L; \qquad (4.4)$$

$$\omega_H = \tfrac{1}{\alpha} + \tfrac{\Delta\theta}{\alpha(1-c)} \cdot \widehat{e}_L^{\alpha} \equiv \widehat{\omega}_H. \qquad (4.5)$$

In the wage in (4.4), L workers are compensated for their utility loss, and the compensation is increasing in c. The higher the envy cost, the higher the material rent the principal has to give up. In (4.5), H workers obtain an information rent, as the principal must reward them in order to prevent mimicking. This rent is increasing in θ, so the higher the ability gap between types, the higher the surplus of H types, as is standard in asymmetric information problems.

The equilibrium payoff of L and H workers are:

$$\widehat{u}_L(\omega_L, e_L) = 0; \tag{4.6}$$

$$\widehat{u}_H(\omega_H, e_H) = \tfrac{\Delta\theta}{\alpha(1-c)} \cdot \widehat{e}_L^{\alpha} \equiv \widehat{u}_H. \tag{4.7}$$

The payoff of H is positive for each $c \in (0,1)$, with a maximum for

$$c = 1 - \tfrac{\Delta\theta}{(1-\mu)(\alpha-1)} \equiv \widehat{c}.$$

It can be shown that, if $\theta < \alpha$, the critical value $\widehat{c}$ is above zero, and below one when $\mu < (\alpha - \theta)/(\alpha - 1)$, where the right-hand side of this inequality is positive. The second-order condition, evaluated at $\widehat{c}$, is negative. The function $\widehat{u}_H$ is thus increasing for $c < \widehat{c}$ and decreasing for $c > \widehat{c}$. As $d\hat{c}/d\theta = -1/(1-\mu)(\alpha-1) < 0$, this means that, when skill heterogeneity increases, the critical value for which the function $\widehat{u}_H$ reaches a maximum decreases. If $\theta > \alpha$, $\widehat{c}$ is negative and $\widehat{u}_H$ is decreasing for all $c \in [0,1]$.

We derive the following result.

Proposition 1. *In equilibrium, the payoff of H workers is:*
1) *first increasing and then decreasing in c if $\theta < \alpha$;*
2) *always decreasing in c if $\theta > \alpha$.*

Note that, since $\theta_H = 1$, on the basis of Proposition 1, the ratio between workers' abilities must be lower than α for $\widehat{u}_H$ to reach a maximum in $c \in (0,1)$. That is, when $\theta < \alpha$, that is when H types are not so much more productive than L types, the payoff in (4.7) is increasing in the interval $c \in [0, \widehat{c}]$ and decreasing in $c \in [\widehat{c}, 1]$, as depicted in figure 1a. When $\theta > \alpha$, that is, when H types are relatively much more productive than L types, the critical value $\widehat{c}$ is negative, and $\widehat{u}_H$ is decreasing for all $c \in [0,1]$, as in figure 1b.

This result suggests that, when skill heterogeneity is small, high-ability workers benefit from being envied. This may explain why, as empirically reported by Duffy et al. (2012), high-performing workers would prefer a workplace where the ability gap among colleagues is relatively low . Whereas, when skill heterogeneity is large, envy has always a negative effect on the payoff of H types.

To understand the shape of the payoff of H workers, note that in the function $\widehat{u}_H$ in (4.7), an increase in c, as well as in θ, pushes upward the first term, $\Delta\theta/\alpha(1-c)$. This represents the incentive effect, consisting in the ability reward. The second term, $\widehat{e}_L^{\alpha}$, is

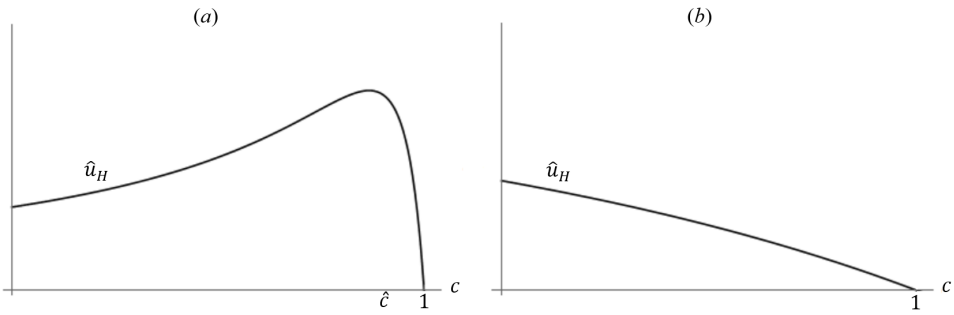

Figure 4.1. Payoff of high-ability workers.
a) $\theta < \alpha$. Parameters: $\mu = 0.4$, $\alpha = 2$, $\theta = 1.05$.
b) $\theta > \alpha$. Parameters: $\mu = 0.4$, $\alpha = 2$, $\theta = 5.00$.

the equilibrium effort of L types. Their effort pushes the payoff of H types downward, as $\widehat{e}_L$ is decreasing in c and in θ. When $\theta < \alpha$, and $c < \widehat{c}$, the positive effect of the incentive reward prevails, and $d\hat{u}_H/dc$ is positive. When, instead, $\theta > \alpha$, the negative effect of $\widehat{e}_L$ dominates, and $d\hat{u}_H/dc$ is negative for all $c \in (0, 1)$. In the numerical example of figure 1a, it is shown that, when $c > 0$, envy allows H types to obtain a payoff higher than without envy ($c = 0$). In figure 1b, H workers receive a payoff lower than in the case $c = 0$. From equation (4.7), it follows that the information rent the principal pays to H types is increasing in $\widehat{e}_L$, and this explains why, in equilibrium, the principal will want to drag down the effort level of L types. Figure 2 illustrates how the payoff of H types varies for small changes in the parameter θ.

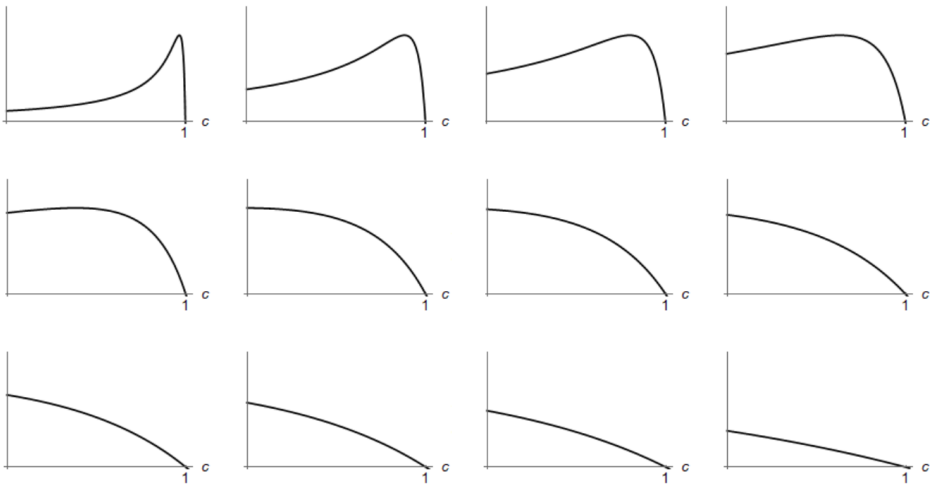

Figure 2. Payoff of the high-skilled workers for variations in θ. Ranging from 1.02 to 5.

The critical value $\widehat{c}$ depends also on the parameter α, the task complexity of the job. Our conjecture is that, when performing "easy" tasks, workers are more likely to engage in social comparison and thus envy gives rise to a larger utility loss for L workers, whereas H workers can benefit from the presence of envious peers. Previous works, like Manna (2016), Barigozzi and Manna (2020) and Caserta et al. (2021), assume a quadratic form

of the effort cost (in our setup, if $\alpha = 2$, the effect of envy on the payoff of H types would only depend on the ability gap between workers, as depicted in figure 1). In our equilibrium, $\widehat{e}_L < \widehat{e}_H = 1$, so the two effort levels are both contained in the interval $(0, 1]$, so the higher α the lower the task difficulty (the effort function becomes flatter). Since $d\hat{c}/d\alpha = \Delta\theta/(1-\mu)(\alpha-1)^2 > 0$, the less difficult the task, the higher the value for which the function $\widehat{u}_H$ reaches the maximum. Besides, the sign of $d\hat{u}_H/dc$ depends on α. An increase in α produces two opposite effects on the function (4.7). The first is a reduction in the incentive reward, and the second is an increase in $\widehat{e}_L^{\alpha}$, as $d\hat{e}_L^{\alpha}/d\alpha > 0$. As in Proposition 1, if $\alpha < \theta$, then the first effect dominates, and $d\hat{u}_H/dc$ is negative for any c, as in figure 3a. If $\alpha > \theta$, then the positive effect on the effort of L types prevails, and $d\hat{u}_H/dc$ is positive, provided $c < \widehat{c}$, as in figure 3b. This result suggests that task complexity is equivalent, in its effects, to skill heterogeneity. The intuition is that easy tasks can emphasize the productivity discrepancies between worker types, whereas difficult tasks may make skill heterogeneity less relevant. Figure 4 shows the payoff of high skilled for small variations in the parameter α.

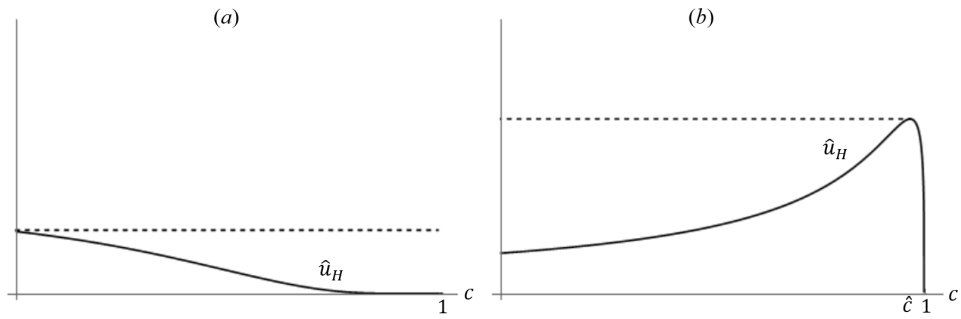

Figure 3. Payoff of high-ability workers.
a) $\theta < \alpha$. Parameters: $\mu = 0.4$, $\theta = 1.05$, $\alpha = 1.05$.
b) $\theta > \alpha$. Parameters: $\mu = 0.4$, $\theta = 1.05$, $\alpha = 5.00$.

The envy cost incurred by L is $c \cdot \widehat{u}_H$, which is positive for all $c \in [0, 1]$, and reaches a critical value for $c = 1 - \Delta\theta/[\alpha\theta - (\alpha-1)\mu - 1] \equiv \bar{c}$ (the second-order condition, evaluated at $\bar{c}$ is negative). In addition, $\bar{c}$ is below 1 and above 0 when $\mu > (\alpha\theta - 1)/(\alpha - 1)$, where this threshold is positive if $\theta > 1/\alpha$, which is always satisfied as $\alpha > 1$ and $\theta > 1$ by assumption. It can also be shown that $\bar{c}$ is decreasing in θ, and $d\bar{c} \cdot \widehat{u}_H(\bar{c})/d\theta < 0$, so the envy disutility of L workers, evaluated at $\bar{c}$, is decreasing in θ.

Proposition 2. *In equilibrium, the utility loss of L workers decreases as the skill heterogeneity increases.*

We can interpret this result as follows. When agents have similar abilities, the social comparison becomes more salient. Namely, L workers will perceive the positive payoff received by their slightly superior H colleagues as "unfair", and suffer a substantially higher disutility, as shown in the numerical example of figure 5a. In turn, the principal

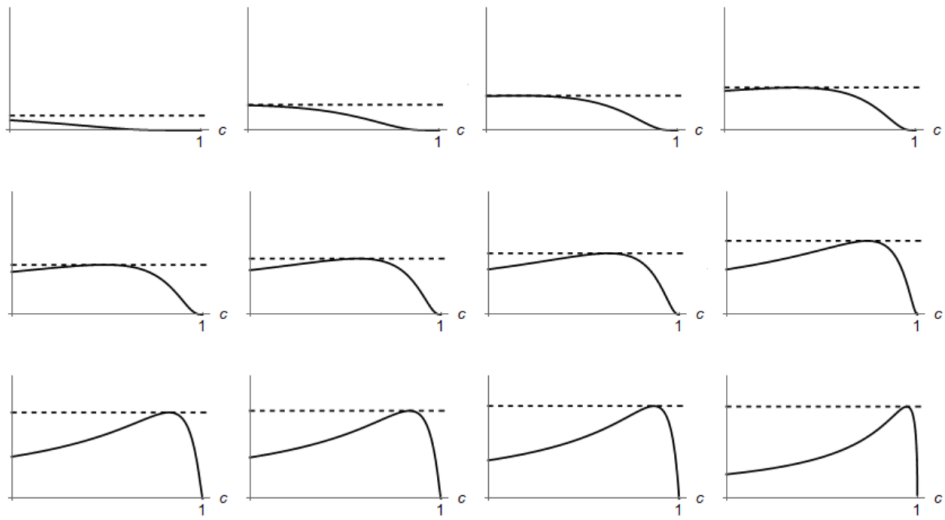

Figure 4. Payoff of the high-skilled workers for variations in α. Ranging from 1.05 to 5.

has to pay a higher envy compensation to both types of agents to satisfy (PC_L) and (IC_H). This explains why, when the ability gap is low, the positive effect of the incentive compensation prevails and $d\hat{u}_H/dc > 0$ for all $c < \widehat{c}$. As θ increases, and the ability gap is larger, the impact of social comparison is weaker. That is, this time higher surplus of H workers is perceived as a "fair" reward, and the utility loss of L workers is lower, as in figure 5b.

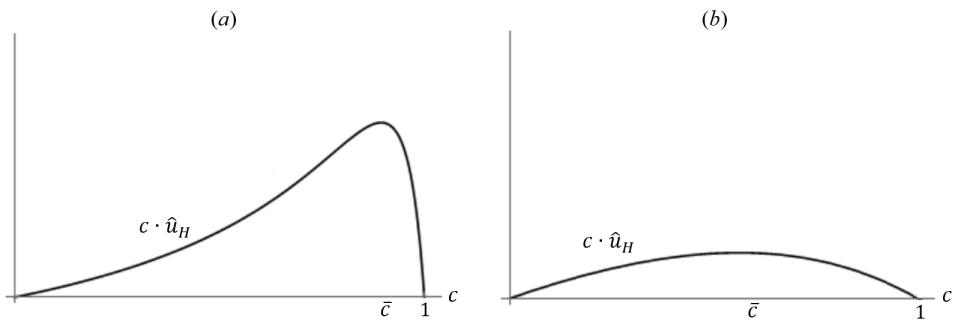

Figure 5. Disutility from envy.
a) $\theta < \alpha$. Parameters: $\mu = 0.4$, $\alpha = 2$, $\theta = 1.05$,.
b) $\theta > \alpha$. Parameters: $\mu = 0.4$, $\alpha = 2$, $\theta = 3.05$.

In words, figures 5a shows the utility loss deriving from the feeling associated with the van de Ven and Zeelenberg's "It could have been me", that is the resentment due to the success that other people have, although their ability is somewhat comparable. In contrast, figure 5b shows the loss from what it can be synthesized as the individual thought "I wish it could be me", that is the desire to have what people with very distant abilities or skills possess.

The equilibrium profit of the principal is

$$\pi = \tfrac{\alpha-1}{\alpha}[\mu + (1-\mu)\widehat{e}_L] \equiv \widehat{\pi}, \tag{4.8}$$

which is positive for all $c \in [0,1]$ and decreasing in c. An increase in envy produces two effects, both negative, on the principal's profit. It reduces the effort of the envious and increases the compensation needed not only to satisfy their participation constraint but also the incentive constraint of the envied. As $d^2\widehat{\pi}/dc^2 < 0$, envy has an increasingly negative effect on the expected profit. As for the ability gap between worker types, the principal's profit is decreasing in θ, as $d\widehat{\pi}/d\theta < 0$. This because, for any level of the envy cost, c, the higher the skill heterogeneity, the higher the information rent and the envy compensation paid to workers. This implies that the principal would prefer to hire workers with the lowest degree of heterogeneity.[6]

As expected, envy leads to a social welfare loss, increasing in c. For a given envy cost, welfare (per average worker) is the sum of the payoff of H types and the principal's profit,

$$W(c) = \mu \cdot \widehat{u}_H + \widehat{\pi} = \frac{\mu\Delta\theta(\widehat{e_L})^{\alpha}}{\alpha(1-c)} + \frac{\mu+(\alpha-1)(1-\mu)\widehat{e}_L}{\alpha}.$$

Welfare under asymmetric information but without envy ($c = 0$) is

$$W(0) = \frac{\alpha-1}{\alpha}\left[\mu - (1-\mu)\left(\frac{1-\mu}{\theta-\mu}\right)^{\sigma}\right] + \frac{\mu\Delta\theta}{\alpha}\left(\frac{1-\mu}{\theta-\mu}\right)^{\alpha\sigma}.$$

It can be shown that the difference $W(0) - W(c)$ is positive and increasing in c. Compared to the full-information case discussed above, we have that $W^{FI} > W(0) > W(c)$ thus, envy causes a further loss of efficiency to welfare already twisted by information asymmetries.

In our model, relative-pay inequalities reduce both the productivity of other-regarding workers and the profit of the principal, especially when agents have slightly different abilities. This provides a rationale for the aim of many employers at creating a friendly work environment. As we will discuss in the following subsection, many studies suggest that interpersonal relationships and reciprocal support among workers can mitigate the detrimental effects of envy. In a different direction, other studies, as for example Nickerson and Zenger (2008) and Obloj and Zenger (2017), show that increasing the structural distance among differently rewarded employees can reduce fairness concerns and increase productivity. This is consistent with the idea that closeness and similarity are crucial moderators of the comparative process.

Remark 1: Shut down of L workers.
The principal can in theory find it profitable to modify the contract terms so that (PC_H) is binding and only H workers apply for the job. In this case, $\omega_H = e_H^{\alpha}/\alpha$, so H types do not derive any surplus, and the principal's profit is $\pi = \mu(e_H - \omega_H) = \mu(e_H - e_H^{\alpha}/\alpha)$. From the first-order condition, the equilibrium effort level is $e_H = 1 = e_H^{FB}$, and the wage $\omega_H = 1/\alpha = \omega_H^{FB}$. Under the contract $(\omega_H^{FB}, e_H^{FB})$, L workers would not accept the contract, as $u_L(\omega_H^{FB}, e_H^{FB}) = -(\theta-1)/\alpha < 0$. The equilibrium profit of the principal is

[6] $\frac{d\widehat{\pi}}{d\theta} = -\frac{1-\mu}{\alpha(\theta-\mu)}\left[\frac{1-\mu}{\alpha(\theta-\mu)}\right]^{\sigma} < 0.$

$\pi = \mu(e_H^{FB} - \omega_H^{FB}) = \mu(\alpha - 1)/\alpha$. As mentioned before, in order to IC_H be satisfied, c is assumed to be below 1. Thus, it is possible to prove that the difference between the separating profit in (4.8) and the expected profit with only H workers is

$$\widehat{\pi} - \tfrac{\mu(\alpha-1)}{\alpha} = \tfrac{(\alpha-1)(1-\mu)}{\alpha} \cdot \widehat{e}_L > 0,$$

so for the principal it is never profitable to screen out L workers.[7]

Remark 2: Pooling equilibrium.
If the principal offers flat wages under the pooling contract (ω, e), from the binding (PC_L), the wage is $\omega = (\theta - c)\, e^{\alpha}/\alpha(1-c)$. The equilibrium effort is $e = [(1-c)/\,(\theta - c)]^{\sigma}$, and the principal's profit is

$$\pi = \tfrac{\alpha-1}{\alpha}\left(\tfrac{1-c}{\theta-c}\right)^{\sigma} \equiv \pi^P.$$

It can be shown that the difference between the profit under separating contracts in (4.8) and π^P is

$$\widehat{\pi} - \pi^P = \tfrac{\alpha-1}{\alpha}\left[\mu + (1-\mu)\widehat{e_L} - \left(\tfrac{1-c}{\theta-c}\right)^{\sigma}\right],$$

which is positive for all $c \in [0,\, 1]$. So, the principal will never offer flat wages.

Remark 3: Envy and net wage comparison.
In our setup, we assume that the envy loss of L workers depends on the surplus obtained by H workers from their contracts. But, as shown by Manna (2016) and Barigozzi and Manna (2020), envy can also be triggered by the comparison of the payoffs that both types receive in equilibrium. Under this alternative specification, the envy cost would be $c \cdot max\{0, \widetilde{\omega}_H - \widetilde{\omega}_L\}$, where $\widetilde{\omega}_H = \omega_H - 1/\alpha e_H^{\alpha}$ and $\widetilde{\omega}_L = \omega_L - \theta/\alpha e_L^{\alpha}$. The equilibrium payoff of H types would be

$$u_H = \tfrac{(1+c)\Delta\theta}{\alpha}\left(\tfrac{1-\mu}{\Delta\theta+\theta-\mu}\right)^{\alpha\sigma} \equiv \widetilde{u}_H,$$

positive for each $c \in (0, 1)$, with a maximum at $\widetilde{c} = [\alpha(1-\mu) + \mu - \theta]/\Delta\theta > \widehat{c}$ (the second-order condition, evaluated at $\widetilde{c}$, is negative). Similarly to the case analyzed in the main model, the sign of $\widetilde{c}$ depends on the ability gap between employees. If $\theta < \alpha$, the critical value $\widetilde{c}$ is above zero, and below one when $\mu < (\alpha - \theta)/(\alpha - 1)$, where the right-hand side of this inequality is positive. The function $\widetilde{u}_H$ is thus increasing for $c < \widetilde{c}$ and decreasing for $c > \widetilde{c}$. Whereas, if $\theta > \alpha$, then $\widetilde{c}$ is negative and the payoff of H types is decreasing for all $c \in [0, 1]$. Therefore, the results of Proposition 1 above would hold and our qualitative results would remain unchanged.

[7] Of course, an equilibrium in which a contract is offered just to L types is not possible in this framework.

4.2.1 Envy-reduction strategy

In this section, we assume that the principal can introduce envy-reduction activities to mitigate the envy cost of low-productive workers.[8] A large literature suggests that employers can achieve significant advantages by creating a "pleasant" work environment for their employees (Langlieb and Kahn, 2005). Many empirical studies report that interpersonal relationships among workers increase motivations and job satisfaction, as a friendly workplace can elicit cohesion, support, and information sharing among employees (Riordan and Griffeth, 1995; Krueger and Schkade, 2008; Chiaburu and Harrison, 2008; Mas and Moretti, 2009). For instance, Berman et al. (2002) survey US managers and find evidence that they actively promote social activities among workers and encourage both horizontal and vertical communications. The authors also show that these strategies have beneficial effects on employee performance. Bandiera et al. (2010) add that these effects depend on workers' ability. Specifically, while having high-ability colleagues can foster job productivity, having low-ability colleagues can reduce it, but the first effect dominates the second. Chen et al. (2016) show that firms with a high-quality workplace better resist periods of economic distress. This is because friendship among workers increases risk tolerance, which prompts innovation investments. Another possible strategy for the principal may be trying to make the attributes and rewards of high-skilled workers less visible or even secret, as argued in the paper by Bebchuk and Fried (2003).[9]

In terms of our model, if the principal follows an envy-reduction strategy, the aim is reducing the envy cost, c, incurred by L workers. We assume that the envy cost can be reduced by the factor $\delta \in (0,1)$, at the cost of $\beta\delta^2/2$, with $\beta > 0$, and that the activity of envy "manipulation" is observable by all workers.

The principal's expected profit under the envy-reduction strategy can be rewritten as

$$\pi = \mu(e_H - \omega_H) + (1-\mu)(e_L - \omega_L) - \tfrac{\beta\delta^2}{2}. \tag{4.9}$$

The incentive constraints of H types is the same as in (IC_H), whereas the participation constraint of L workers is

$$u_L(\omega_L, e_L) = \omega_L - \tfrac{\theta}{\alpha} e_L^{\alpha} - c(1-\delta)\max\{0, u_H(\omega_H, e_H)\} \geq 0. \tag{PC_L^{ER}}$$

From the binding (PC_L^{ER}) and (IC_H), and from the first-order conditions of (4.9), the equilibrium effort levels are

$$e_L = [1 - \tfrac{\Delta\theta}{\theta - c(1-\delta)(1-\mu) - \mu}]^{\sigma} \equiv e_L^{ER} \quad \text{and} \quad e_H = 1 = e_H^{FB}.$$

The effort of H workers is not distorted compared to the first best, whereas the effort of L workers is lower and decreasing in the manipulation activity, as $de_L^{ER}/d\delta < 0$. However, the strategy of envy-reduction increases the effort of L workers, as $e_L^{ER} > \widehat{e}_L$.

[8] See Caserta et al. (2021) for envy manipulation by high-skilled workers.

[9] However, it is important to stress that this type of strategies may easily prove to be ineffective or even backfire, as it may lead workers to form wrong beliefs and estimates about the abilities and compensation of high-skilled colleagues. The same may be true of other types of envy-reduction activities.

The equilibrium wages are:

$$\omega_L = \frac{\theta - c(1-\delta)}{\alpha[1-c(1-\delta)]} \cdot \left(e_L^{ER}\right)^{\alpha} \equiv \omega_L^{ER}; \tag{4.10}$$

$$\omega_H = \frac{1}{\alpha} + \frac{\Delta\theta}{\alpha[1-c(1-\delta)]} \cdot \left(e_L^{ER}\right)^{\alpha} \equiv \omega_H^{ER}. \tag{4.11}$$

The equilibrium payoffs of L and H workers are:

$$u_L(w_L^{ER}, e_L^{ER}) = 0;$$
$$u_H(w_H^{ER}, e_H^{FB}) = \frac{\Delta\theta}{\alpha[1-c(1-\delta)]} \cdot \left(e_L^{ER}\right)^{\alpha} \equiv u_H^{ER}.$$

The payoff of H types is positive for each $c \in (0,1)$ and $\alpha > 1$, with maximum at

$$c = \frac{\alpha(1-\mu)+\mu-\theta}{(\alpha-1)(1-\mu)(1-\delta)} \equiv c^{ER}.$$

The results of Proposition 1 applies to this extension. If $\theta < \alpha$, c^{ER} is positive, and the payoff of H workers is first increasing and then decreasing in c. Besides, $dc^{ER}/d\delta > 0$, so the value for which the function u_H^{ER} reaches a maximum is increasing in the envy-reduction activity. If $\theta > \alpha$, then c^{ER} is negative and the utility of H types is always decreasing in c.

It can be shown that the equilibrium principal's profit with envy manipulation is always positive[10], and is larger than the separating profit when the envy cost, c, is relatively high. Figure 6 shows a numerical example. The parameters are $\mu = 0.4$, $\alpha = 2$, $\delta = 0.4$, $\theta = 1.2$, $\beta = 0.2$, and $\pi^{ER} > \widehat{\pi}$ when $c > 0.40$. Note that H workers can also benefit from the activities of envy manipulation carried out by the principal (in the numerical, this happens when $c > 0.80$).

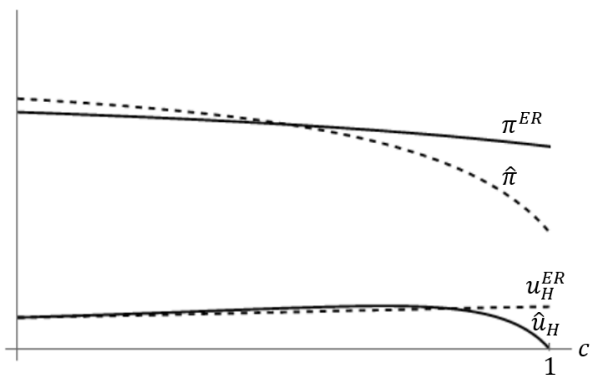

Figure 6. Equilibrium with envy-reduction interventions.
Parameters: $\mu = 0.4$, $\alpha = 2$, $\delta = 0.4$, $\theta = 1.2$, $\beta = 0.2$

This conclusion is in line with Dur and Sol (2010), in which managers may prefer to mix workers and use team incentives to promote altruism and cohesion. And, with Larkin et al. (2012), where compensation schemes based on worker seniority or group performance can mitigate the psychological costs caused by wage comparison.

[10] $\pi^{ER} = \frac{[1-c(1-\delta)][\mu(\alpha-1)+\alpha(1-\mu)e_L^{ER}]-[\theta-c(1-\delta)(1-\mu)-\mu](e_L^{ER})^{\alpha}}{\alpha[1-c(1-\delta)]} > 0.$

4.3 Conclusions

When agents have other-regarding preferences, wage differences may not reflect the true disparity in the productivity among workers. In this paper, we contribute to the literature on inequity aversion by investigating the effect of pay-rent structures on contract design. We examine the upward social comparison in a principal-agent model with asymmetric information on worker abilities, and our results suggest that skill heterogeneity is pivotal to define the effect of envy on the structure of optimal incentive schemes. Specifically, we show that the utility loss caused by envy is higher (lower) when skill heterogeneity is low (high). The intuition behind this result is that skill similarity may emphasize the "inferiority" of low-performing workers, as they can perceive as undeserved the surplus that their slightly more talented colleagues earn. By contrast, when heterogeneity is high, low-skilled workers are less likely to feel inferior and thus may judge the surplus as a fair reward. Since the loss turns into compensation for both types of agents, this result explains why the payoff of the more talented can be first increasing and then decreasing in the envy cost of the less talented. This conclusion holds when the ability gap among worker types is lower than a certain threshold. Whereas, when the gap is large, envy always translates into a reduction of well-being for the more talented.

The literature on managerial and organizational strategies suggests that wage compression, pay secrecy, and office relocation policies can mitigate the costs deriving from interpersonal conflicts. On this topic, Cohen-Charash and Mueller (2007) posit that secrecy about labor contract terms actually reduces envy and harmful behavior. Nickerson and Zenger (2008) argue that social comparison and envy among employees may lead managers to make inefficient productive decisions. Ockenfels et al. (2015) find evidence that, in countries and sectors where legal rules impose the disclosure of employment contracts, wage compression is a widespread practice used to minimize the cost of interpersonal comparisons. Finally, other studies show that fairness concerns can arise also in non-market activities, as in Barigozzi and Manna (2020) who show that, in mission-oriented organizations, envy in the workplace depresses labor donations and volunteering from employees.

Appendix

If PC_L is binding, then

$$\omega_L = \tfrac{\theta}{\alpha} e_L^{\alpha} + c \cdot \max\{0, u_H(\omega_H, e_H)\}.$$

Replacing ω_L in IC_H, it follows that

$$\omega_H - \tfrac{1}{\alpha} e_H^{\alpha} \geq \tfrac{\theta}{\alpha} e_L^{\alpha} + c \cdot \max\{0, u_H(\omega_H, e_H)\} - \tfrac{1}{\alpha} e_L^{\alpha}.$$

From the binding IC_H, then

$$\omega_H = \tfrac{1}{\alpha} e_H^{\alpha} + \tfrac{\Delta\theta}{\alpha} e_L^{\alpha} + c \cdot \max\{0, u_H(\omega_H, e_H)\}.$$

Thus, H workers obtain the marginal cost of their productivity, the information rent, and the envy compensation of L types to avoid mimicking behavior. So, $\omega_H > 1/\alpha e_H^\alpha$, which means that PC_H is slack. As for the incentive constraint of L workers, keeping PC_L and IC_H binding, then IC_L becomes

$$c \cdot \max\{0, u_H(\omega_H, e_H)\} > \tfrac{1}{\alpha} e_H^\alpha + \tfrac{\Delta\theta}{\alpha} e_L^\alpha + c \cdot \max\{0, u_H(\omega_H, e_H)\} - \tfrac{\theta}{\alpha} e_H^\alpha.$$

After simplifying and rearranging, we obtain

$$\tfrac{\Delta\theta}{\alpha} e_H^\alpha > \tfrac{\Delta\theta}{\alpha} e_L^\alpha,$$

which is always true in equilibrium, since $\widehat{e}_H^\alpha > \widehat{e}_L^\alpha$. So, if PC_L and IC_H are binding, the other constraints, PC_H and IC_L, are satisfied as well.

Bibliography

Ball, S., Eckel, C., Grossman, P., and Zame, W. (2001). Status in markets. *Quarterly Journal of Economics*, 116:161–188.

Bandiera, O., Barankay, I., and Rasul, I. (2010). Social incentives in the workplace. *Review of Economic Studies*, 77:417–458.

Barigozzi, F. and Manna, E. (2020). Envy in mission-oriented organisations. *Journal of Economic Behavior & Organization*, 179:395–424.

Bartling, B. and Von Siemens, F. (2010). The intensity of incentives in firms and markets: Moral hazard with envious agents. *Labour Economics*, 17:598–607.

Bebchuk, L. and Fried, J. (2003). Executive compensation as an agency problem. *Journal of Economic Perspectives*, 17:71–92.

Becker, G. (1973). A theory of marriage: Part I. *Journal of Political Economy*, 81:813–846.

Ben-Ze'ev, A. (1992). Envy and inequality. *Journal of Philosophy*, 89:551–581.

Berman, E., West, J., and Richter, M. (2002). Workplace relations: Friendship patterns and consequences (according to managers). *Public Administration Review*, 62:217–230.

Bolton, G. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90:166–193.

Breza, E., Kaur, S., and Shamdasani, Y. (2018). The morale effects of pay inequality. *Quarterly Journal of Economics*, 133:611–663.

Card, D., Mas, A., Moretti, E., and Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, 102:2981–3003.

Caserta, M., Ferrante, L., and Reito, F. (2021). Envy manipulation at work. *The B.E. Journal of Theoretical Economics*, 1:287–314.

Chen, J., Leung, W., and Evans, K. (2016). Are employee-friendly workplaces conducive to innovation? *Journal of Corporate Finance*, 40:61–79.

Chiaburu, D. and Harrison, D. (2008). Do peers make the place? Conceptual synthesis and meta-analysis of coworker effects on perceptions, attitudes, OCBs, and performance. *Journal of Applied Psychology*, 93:1082–1103.

Cohen-Charash, Y. and Mueller, J. (2007). Does perceived unfairness exacerbate or mitigate interpersonal counterproductive work behaviors related to envy? *Journal of Applied Psychology*, 92:666–680.

Cohn, A., Fehr, E., Herrmann, B., and Schneider, F. (2014). Social comparison and effort provision: Evidence from a field experiment. *Journal of the European Economic Association*, 12:877–898.

Crusius, J. and Lange, J. (2014). What catches the envious eye? Attentional biases within malicious and benign envy. *Journal of Experimental Social Psychology*, 55:1–11.

Desiraju, R. and Sappington, D. (2007). Equity and adverse selection. *Journal of Economics & Management Strategy*, 16:285–318.

Duesenberry, J. (1952). *Income, saving, and the theory of consumer behavior*. Cambridge MA: Harvard University Press.

Duffy, M., Scott, K., Shaw, J., Tepper, B., and Aquino, K. (2012). A social context model of envy and social undermining. *Academy of Management Journal*, 55:643–666.

Dur, R. and Glazer, A. (2008). Optimal contracts when a worker envies his boss. *Journal of Law, Economics, & Organization*, 24:120–137.

Dur, R. and Sol, J. (2010). Social interaction, co-worker altruism, and incentives. *Games and Economic Behavior*, 69:293–301.

Fehr, E., Goette, L., and Zehnder, C. (2009). A behavioral account of the labor market: The role of fairness concerns. *Annual Review of Economics*, 1:355–384.

Fehr, E. and Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114:817–868.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7:117–140.

Fiske, S. (2010). Envy up, scorn down: How comparison divides us. *American Psychologist*, 65:698–706.

Frank, R. (1984). Interdependent preferences and the competitive wage structure. *RAND Journal of Economics*, 65:510–520.

Ghatak, M. (1999). Group lending, local information and peer selection. *Journal of Development Economics*, 60:27–50.

Grant, A. and Mayer, D. (2009). Good soldiers and good actors: Prosocial and impression management motives as interactive predictors of affiliative citizenship behaviors. *Journal of Applied Psychology*, 94:900–912.

Grund, C. and Sliwka, D. (2005). Envy and compassion in tournaments. *Journal of Economics & Management Strategy*, 14:187–207.

Itoh, H. (2004). Moral hazard and other-regarding preferences. *Japanese Economic Review*, 55:18–45.

Khan, A., Quratulain, S., and Bell, C. (2014). Episodic envy and counterproductive work behaviors: Is more justice always good? *Journal of Organizational Behavior*, 35:128–144.

Krueger, A. and Schkade, D. (2008). Sorting in the labor market do gregarious workers flock to interactive jobs? *Journal of Human Resources*, 43:859–883.

Langlieb, A. and Kahn, J. (2005). How much does quality mental health care profit employers? *Journal of Occupational and Environmental Medicine*, 47:1099–1109.

Larkin, I., Pierce, L., and Gino, F. (2012). The psychological costs of pay-for-performance: Implications for the strategic compensation of employees. *Strategic Management Journal*, 33:1194–1214.

Loewenstein, G., Thompson, L., and Bazerman, M. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology*, 57:426–441.

Manna, E. (2016). Envy in the workplace. *Economics Letters*, 142:18–21.

Mas, A. (2006). Pay, reference points, and police performance. *Quarterly Journal of Economics*, 121:783–821.

Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, 99:112–45.

Neilson, W. and Stowe, J. (2010). Piece-rate contracts for other-regarding workers. *Economic Inquiry*, 48:575–586.

Nickerson, J. and Zenger, T. (2008). Envy, comparison costs, and the economic theory of the firm. *Strategic Management Journal*, 29:1429–1449.

Obloj, T. and Zenger, T. (2017). Organization design, proximity, and productivity responses to upward social comparison. *Organization Science*, 28:1–18.

Ockenfels, A., Sliwka, D., and Werner, P. (2015). Bonus payments and reference point violations. *Management Science*, 61:1496–1513.

Ostell, A. (1996). Managing dysfunctional emotions in organizations. *Journal of Management Studies*, 33:525–557.

Postlewaite, A. (1998). The social basis of interdependent preferences. *European Economic Review*, 42:779–800.

Riordan, C. and Griffeth, R. (1995). The opportunity for friendship in the workplace: An underexplored construct. *Journal of Business and Psychology*, 10:141–154.

Rustemeyer, R. and Wilbert, C. (2001). Jealousy within the perspective of a self-evaluation maintenance theory. *Psychological Reports*, 88:799–804.

Silver, M. and Sabini, J. (1978). The perception of envy. *Social Psychology*, 41:105–117.

Smith, R. and Kim, S. (2007). Comprehending envy. *Psychological Bulletin*, 133:46–64.

Smith, R., Parrott, W., Diener, E., Hoyle, R., and Kim, S. (1999). Dispositional envy. *Personality and Social Psychology Bulletin*, 25:1007–1020.

Solnick, S. and Hemenway, D. (2005). Are positional concerns stronger in some domains than in others? *American Economic Review*, 95:147–151.

Song, J., Price, D., Guvenen, F., Bloom, N., and Von Wachter, T. (2019). Firming up inequality. *Quarterly Journal of Economics*, 134:1–50.

Tai, K., Narayanan, J., and McAllister, D. (2012). Envy as pain: Rethinking the nature of envy and its implications for employees and organizations. *Academy of Management Review*, 37:107–129.

van de Ven, N. and Zeelenberg, M. (2015). On the counterfactual nature of envy: "It could have been me". *Cognition and Emotion*, 29:954–971.

van de Ven, N., Zeelenberg, M., and Pieters, R. (2009). Leveling up and down: The experiences of benign and malicious envy. *Emotion*, 9:419–429.

Vecchio, R. (2000). Negative emotion in the workplace: Employee jealousy and envy. *International Journal of Stress Management*, 7:161–179.