Aggregate Consequences of Market Imperfections

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
an der Universität Mannheim

vorgelegt von

Thomas Gall

Mannheim 2005

Referent: Prof. Dr. Hans Peter Grüner
Koreferent: Prof. Felix Kübler, Ph.D.
Dekan: Prof. Dr. Ulrich Falk

Tag der mündlichen Prüfung: 28.7.2005

Meinen Eltern

Acknowledgements

First of all, I wish to thank my supervisor, Hans Peter Grüner, for both his ongoing support and very valuable discussion. Moreover, this work benefitted greatly from the comments of Andrew Newman. In particular, I am grateful for his and Patrick Legros' collaboration on the fourth chapter. Many thanks go also to Felix Kübler for his valuable support.

In addition, I would like to thank my colleagues at the Chair for Economic Policy and at the CDSEM at the University of Mannheim, and at the University College London, for providing such stimulating research environments. In particular, I am indebted to Elisabeth Schulte for her astonishing disposition to discuss my ideas.

Finally, financial assistance is gratefully acknowledged from DFG, the European Commission, and the Swiss National Bank.

Abstract

This dissertation aims to analyze aggregate consequences of market imperfections. In particular, it is interested in the effects of capital market imperfections and contracting constraints on the efficiency of the allocation and, subsequently, the prospects for growth.

In the first chapter, lotteries are introduced into the framework of Galor and Zeira (1993). This is a growth model with imperfect capital markets and indivisible investment permitting an individual poverty trap. Allowing for lotteries leads to a quite remarkable overturn of the original result: more severe capital market imperfections may increase aggregate consumption in finite time. Intuitively, whenever lotteries dominate an imperfect capital market as a means of capital allocation, increasing the relative attractiveness of lotteries also increases allocative efficiency.

The second chapter of this dissertation analyzes the effects of intra-firm bargaining on the formation of firms under imperfect capital markets and contracting constraints. In equilibrium, wealth inequality induces a heterogenous distribution of firm sizes allowing for firms both too small and too large in terms of technical efficiency. The findings connect well to empirical facts as the missing middle of size distributions in less developed countries. The model is able to encompass a non-monotonic relationship between inequality and aggregate wealth, providing a theoretical framework for policy analysis of foreign aid and investment.

The last chapter finds that non-transferabilities of utility in markets, caused by e.g. contracting constraints, may induce inefficient incentives for non-contractible investments taking place prior to markets. The timing of markets with respect to the investment affects their efficiency. Applying this results to a model of education choice and labor markets we find that markets opening only after agents have invested may exhibit inefficient over-education whereas markets that open only prior to investment do not. Hence, enforcing ex post markets by e.g. prolonging compulsory schooling, may lead to over-education and inefficient labor market allocations.

Contents

Chapter 1

Introduction

Still the invisible hand by which markets smoothly and efficiently coordinate productive activity in economies looms large over economic theory. As this classical view holds, competition in markets provide economic agents with the correct opportunity cost as to induce socially efficient decisions by rational individual decision makers. As a corollary then markets solve the problem of allocating any given good as long as there exists a price for it.

However, in practice markets are prone to frictions. Numerous imperfections impede the flawless coordination of economic activity by markets and distort market prices. To name but a few examples, asymmetric information may prevent the formation of a market price, thus leading to a market breakdown (Akerlof, 1970). A lack of well-defined property rights may cause the invisible hand to stagger as externalities invalidate the first welfare theorem (Coase, 1960). Access to markets may depend on agents' characteristics, thus neglecting some resources in the coordination process, the classical task of markets. To what extend is then a theoretical approximation of the world that relies on perfect markets meaningful?

The relevance of any theory hinges, of course, mainly on whether or not it is capable of providing adequate explanations for empirical facts. A brief glance at the empirical literature on the role of market imperfections might prove illuminating at this stage. Less developed countries are of primary concern as these economies are widely held to be affected particularly severely by market frictions that violate the assumptions of classical theory. Hence, they should be expected to be the first place to encounter large scale deviations from the predictions of classical theory.

One of the most thoroughly explored fields in this respect is the market for loans. Banerjee (2002) collects some stylized facts from the empirical literature on this topic. Typically, the borrowing interest rate is significantly higher than the lending interest rate – a finding that is by no means limited to less developed countries – although default rates are generally low. This is in striking contrast to the assumption of perfect markets where the law of one price should hold. Indeed, empirical findings go further than this in that considerable heterogeneity in firms' cost of capital is observed even when the projects being financed do not appear to vary. For instance, it is frequently reported that richer people borrow at lower rates of interest. Interestingly, another common finding is that loans are generally taken in order to finance productive activity, so that a wedge in the cost of capital between agents indeed has the potential to significantly distort investment decisions in an economy.

Banerjee and Munshi (2004) present an empirical example of how the aggregate allocation may be affected by market imperfections. They analyze the knitted garment industry in Tirupur, India. Using data from a natural experiment they find evidence of significant mis-allocation of capital in this particular industry. It is possible to stratify their sample of firms into two separate social communities. Empirical results point to the fact that capital investment is determined significantly by a firm's social community. Firms belonging to one social community have higher investment rates *and* lower productivity compared to the other social community. This distortion may be attributable to different levels of access to suppliers, customers, or credit markets, the last explanation being favored by the authors. Given that at the time of observation Tirupur produced about 70% of India's knitted garment exports and the firm population was split roughly equally into both communities, the mis-allocation of capital found in the study had indeed a sizable effect in aggregate terms.

Field (2003) analyzes the impact of property rights on labor markets. She uses data from Peru benefitting from a natural experiment when the government granted property titles to a vast number of urban squatters. The study finds that in the course of the governmental program total labor supply from the treated households increased by approximately one household member while the probability of working at home decreased by a substantial amount. A tentative explanation for these results is that some household members had to physically guard their home before the titling but were free to seek work outside their home, thereby profiting from the legal security of their premises. Likewise, Besley (1995) provides some evidence that the quality of property rights on land positively affects investment in capital in Ghana.

These examples construct a powerful argument not only that markets are indeed imperfect but also that these imperfections are relevant for the aggregate allocation. This view is shared by Banerjee and Duflo (2004) who demonstrate the shortcomings of neglecting market imperfections in growth models. They highlight the inability of neoclassical growth theory to explain differences in factor productivity both within and across economies by means of a comparison of the US and India.

Consequently, it is of considerable interest to provide insights into the functioning of economies when markets are subject to frictions and imperfections. This dissertation aims to provide a contribution to this analysis focussing on three particular settings where market frictions are relevant in the sense that they determine aggregate allocation. In one setup credit market imperfections combined with indivisible investments generate increasing returns to wealth. Yet imperfect markets may be crowded out by other imperfect means of capital allocation, in particular by lotteries, for instance in the form of property crimes. In the second framework, contracting constraints additionally distort investment incentives, thus leading to an inefficient firm size distribution in the economy. Finally, a reduced form approach on market imperfections is taken and the optimal timing of markets depending on the degree of market imperfection emerges to be of crucial importance for efficiency. It will become clear in the course of this analysis that in all of these settings the nature of market imperfections crucially affects the aggregate outcome and therefore also policy implications.

Galor and Zeira Go Gambling

In the first chapter, "Galor and Zeira Go Gambling", lotteries are introduced into the framework of Galor and Zeira (1993), a growth model with imperfect capital markets and indivisible investment permitting an individual poverty trap. Although agents are averse to risk, admitting for lotteries leads to a quite remarkable overturn of the result in the original work. Not only does the poverty trap vanish for a broad class of parameters, but also more severe capital market imperfections may increase aggregate consumption in finite time. Moreover, the poverty trap may be eliminated by increasing market frictions. The intuition is that whenever lotteries dominate an imperfect capital market as a means of allocation of capital, increasing the relative attractiveness of lotteries increases allocative efficiency.

A natural interpretation for the opportunity to gamble is, of course, conflict. The outcome of conflict is not deterministic and the higher the relative strength of a party, the more likely is success. Eventually, the winner is able to take over the loser's resources. The model then predicts social segregation of the economy into poor, crime-infested communities and rich, relatively secure communities which appears to describe the empirical evidence quite concisely.

Inequality and Firm Size Distribution

The second chapter of this dissertation, "Inequality, Incomplete Contracts, and the Size Distribution of Business Firms", analyzes the formation of firms and the resulting firm size distributions in economies with imperfect capital markets and contracting constraints. Both types of frictions are likely to prevail in less developed countries. When labor contracts are incomplete in the sense that they can be renegotiated at any time, intra-firm bargaining over the joint profit induces firm owners to expand their firm beyond a well-defined technically efficient size (Stole and Zwiebel, 1996). Capital market imperfections drive a wedge between the capital cost of rich and poor agents thus inducing relatively poor firm owners to choose too small firm sizes. Hence, in equilibrium wealth inequality induces a heterogenous distribution of firm sizes allowing simultaneously for firms both too small *and* too large.

The results of the model connect well to empirical facts as the missing middle of firm size distributions in less developed countries and evidence of high volatility of productivity among firms within the same industry in developing economies. Moreover, the model is able to encompass a non-monotonic relationship between inequality and aggregate wealth. While in capital rich economies income inequality decreases as output increases, for capital scarce economies higher output may be associated with higher inequality. Thus, a valid theoretical framework is provided for policy analysis of foreign aid and investment. Indeed, foreign direct investment may hurt an economy as sufficiently wealthy foreign investors need not be affected by the capital market imperfection and then extend firms beyond the efficient size. General equilibrium effects on wages then may reduce both efficiency of the firm size distribution and aggregate output.

Markets and the Hold-up Problem

The last chapter, "Too much Competition: On the Role of Markets in the Hold-up Problem“, takes a more abstract approach to address the question of whether market imperfections are relevant for aggregate outcomes. The classical hold-up problem (see Klein et al., 1978, Williamson, 1975) is revisited where ex-post bargaining over profits from ex-ante non-contractible investments distorts incentives to invest. Recent inquiries in this topic have shown that perfect markets may solve this problem since the market price then fully reflects the social returns to the investment (Cole et al., 2001, Felli and Roberts, 2002). Given the load of the evidence, it is a natural question whether this theoretical finding still holds true once non-transferabilities of utility are taken into consideration. These might stem for instance from contracting constraints, imperfect capital markets or asymmetric information.

It emerges that under non-transferabilities of utility markets can no longer be expected to induce efficient incentives for non-contractible investments taking place prior to markets. The timing of imperfect markets with respect to the investment appears to be of crucial importance for efficiency of the allocation. For instance, whenever investments are strategic substitutes and non-transferabilities are sufficiently severe, markets that open prior to the investment dominate ex post markets in terms of efficiency. Moreover, the efficient timing of markets may reverse as the market frictions decrease thus emphasizing that policy implications depend on the institutional shape of markets. Applying this result to a model of education choice and labor markets with cost-sharing contracts, it turns out that markets opening only after agents have invested may exhibit over-education whereas markets that open only prior to investment provide a more efficient allocation. Hence, prolonging compulsory schooling, and thus effectively opening an ex post market, may lead to over-education and inefficient labor market allocations.

Chapter 2

Lotteries, Inequality, and Market Imperfections: Galor and Zeira Go Gambling

2.1 Introduction

Rigidities in credit markets are widely held to be obstacles for sustainable growth. Thus, policy advice commonly focuses on reducing existing market imperfections in order to achieve long run prosperity. This is also a feature common to most of the recent literature inquiring into cause and effect of wealth and income inequality (this is the case in Aghion and Bolton (1997), Banerjee and Newman (1993), Galor and Zeira (1993), Piketty (1997), and Matsuyama (2000) among others). Typically in this literature the beneficial effects of the reduction of market imperfections are taken for granted. The models build on two ingredients to analyze the effects of income distributions: imperfect capital markets and indivisible investments with locally increasing returns to scale. The first ingredient has the side effect that reducing market imperfections improves efficiency. However, in this chapter this need not be the case and such policy may put in danger long term growth prospects.

Concerning the second ingredient, it is well-known that investment indivisibilities may induce even risk averse agents to demand fair lotteries, an idea that was put forward as early as in Friedman (1953). The indivisibility gives rise to a jump in lifetime income with respect to initial wealth which in turn leads to a non-concavity in lifetime utility. However, it is possible to render the lifetime utility function concave by introducing a suitable lottery, as was pointed out by Ng (1965) in the context of indivisible consumption choice. As a consequence, agents in the neighborhood of the non-concavity, although being averse to risk, prefer participating in the lottery to their certain income.

Given the unbroken demand for lotteries in the world's economies it is hard to argue they are not involved in economic decision making. In a recent study (Garrett, 2001) it is found that substantial amounts of income are invested in gambling activities in countries all over the world. Moreover, it appears that lottery consumption is prevalent especially among the poor (see for instance the studies on state lotteries by Worthington, 2001, Scott and Garen, 1993). It is noteworthy that this does not include the more subtle shifting of activities towards more risky ones on the individual level. Moreover, lotteries are evidently involved in the capital accumulation of credit constrained individuals. Besley et al. (1993) find ample evidence for so called random rotating saving and credit associations. Thus it appears quite justified to consider the possibility of gambling in economic modeling.

In this chapter we look at the model of Galor and Zeira (1993) and analyze the effects of the admission of endowment lotteries. In their setting this may lead to a different dynamic behavior of the model economy. In particular, we are concerned with whether both the two stable steady states found by them still will prevail. We are able to show that the introduction of lotteries always leads to an instantaneous Pareto improvement and we provide a condition to ensure that this is sustainable. Then we analyze the steady state behavior and show under which circumstances the poverty trap emerging in the original work vanishes. It turns out that for a broad parameter range this indeed is the case. In some cases this depends on the scope for moral hazard in the economy being sufficiently great. In contrast to the original results, we come to the conclusion that in these cases amplifying market rigidities may improve welfare and eliminate the poverty trap.

In fact, if the scope for moral hazard on the borrowers' side rises, so does the spread between the lending and borrowing interest rate, thus rationing credits more strictly. Therefore more people prefer gambling to borrowing in spite of the uncertainty. Thus, as market conditions deteriorate, lotteries crowd out credits and it becomes more likely that the redistribution and accumulation of initial endowments by means of lotteries puts winners on the growth path towards the high income steady state. Furthermore, lotteries prove to be rather efficient from a long run point of view. This result indicates that reducing existing capital market frictions always runs the risk of creating or cementing a poverty trap in the presence of lotteries.

In an extension we note that a natural way to encounter opportunities for randomization over different income levels is given by predatory or criminal activities. Suppose agents are able to attempt to expropriate other agents, and the outcome of such a conflict has a random component, such that the probability of success depends positively on an agent's endowments devoted to conflict. Then the opportunity to engage in expropriation is effectively interpretable as a lottery. However, conflict in the sense of fighting typically wastes resources. Extending the model to costly lotteries yields similar conclusions provided lotteries are not too wasteful.

This chapter is related to the work of Garratt and Marshall (1994) who examine a static setting and find that the optimal social contract to provide education has the form of a lottery, similar to the one emerging in this chapter. However, they assume the absence of a market for credits to finance human capital acquisition. Our results show that in the Galor and Zeira (1993) framework there exist cases where allowing people to gamble leads to a complete breakdown of the credit market and that this may actually be desirable. To some extent this might provide a motivation for the assumption and for the results of Garratt and Marshall.

The development economics literature provides a similar finding in the contribution of Ghatak et al. (2001). However, their result is driven by the theorem of the second best whereas the results in this chapter do not depend on the presence of multiple imperfect markets. Ghatak et al. (2001) look at a Banerjee and Newman (1993) style setting where young agents may exert effort to overcome borrowing constraints in order to earn entrepreneurial rents when old. Both the labor market determining effort choice and the capital market are imperfect in the sense that there is moral hazard and transaction cost. Their paper finds that increasing transaction cost may lead to higher aggregate income in the long run. The intuition is that greater income inequality due to more severe market imperfections increases incentives for young agents to choose high effort. As there is under-investment in effort due to the labor market imperfections, greater scope for moral hazard in the credit market may partially offset the under-investment problem and improve welfare.

This chapter proceeds by presenting the model framework and some preliminaries of our analysis in sections 2 and 3. Section 4 analyzes the dynamic implications of lotteries and section 5 conducts a welfare analysis. An application of the model is considered in section 6, while section 7 concludes.

2.2 The Model

We begin by briefly restating the underlying model of Galor and Zeira (1993). In the model economy agents live for two periods in overlapping generations. When young, an individual may acquire human capital at a cost h or work unskilled for a low wage w_n. When old, an agent either works skilled earning the high wage w_s if he invested in human capital, or he works unskilled. Agents are assumed to have 'warm-glow' preferences that give rise to the utility function

$$u = \alpha \ln c + (1 - \alpha) \ln b,$$

where c denotes second period consumption and b the amount bequeathed to their offspring. Let x denote the initial endowment of an agent at the beginning of his life. In the beginning of the economy these endowments are assumed to be continuously distributed on $[0, L]$ with $L > h$.

There exists a credit market where agents can deposit and borrow in order to finance human capital acquisition. It is assumed that the deposit world market interest rate $1 + r$ is determined exogenously because agents live in a small open economy.[1] However, there exists moral hazard on the side of the borrowers, namely the possibility to default on purpose. Lenders may monitor borrowers at cost z, but the latter may choose to default deliberately even if monitored - although facing a cost βz, $\beta > 1$. One way to interpret this is that there is a lack of enforceability of contracts due to the institutional, in particular legal, framework. In the model there exists an incentive compatible debt contract making borrowers indifferent between defaulting and paying back the debt. It requires the borrowing interest rate $1 + i$ to be strictly greater than $1 + r$. In the capital market equilibrium $(1 + i) = (1 + r)\frac{\beta}{\beta-1}$. This creates the rigidity in the credit market. We define the interest rate spread as

$$\gamma \equiv \frac{\beta}{\beta - 1}.$$

It is quite natural to interpret γ as the scope for moral hazard in the economy. As β approaches 1 and the cost of purposeful defaulting decreases, γ increases and as γ approaches 1 the moral hazard problem disappears.

Utility maximization of agents shows that individuals separate into three categories. The first are those agents who prefer to work unskilled both periods thereby earning $(1 + r)x + (2 + r)w_n$. Agents are born with x and earn w_n in the first period of their life and lend this in the capital market. In the second period they earn w_n again, consume and bequeath. But there exists an endowment point

[1]This exogeneity is not crucial to our findings. If decreasing demand for loans induces a decline in the lending interest rate and there exists a a costless storage technology, this would render the results of this model even more pronounced.

such that lifetime income from unskilled working coincides with the lifetime income from borrowing in the first period and working skilled in the second. Equating lifetime incomes yields this point denoted by f:

$$f = \frac{1}{i-r}\left[(2+r)w_n + (1+i)h - w_s\right]$$

So, all individuals with $x < f$ prefer unskilled working. The second category of individuals consists of all agents with $f \leq x < h$ who borrow and work skilled earning $(1+i)(x-h)+w_s$. Finally, all agents with endowments $x \geq h$ need not borrow to acquire human capital, work skilled, and have lifetime income $(1+r)(x-h)+w_s$. This characterizes fully the agents' equilibrium behavior: in any given period t every old individual bequeathes a fixed share $(1-\alpha)$ of his lifetime income to his offspring because of 'warm-glow' preferences. These bequests become the next generation's endowments x_{t+1}. Then there may arise an interesting case permitting figure 2.1.

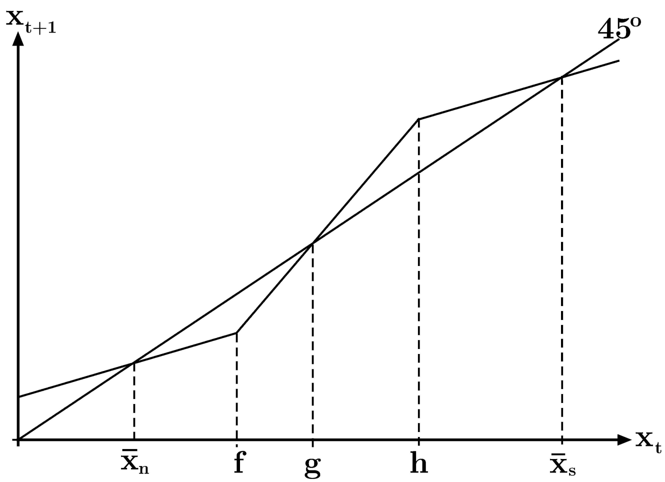

Figure 2.1: The bequest function without lotteries

The individuals' bequest function in period t depending on this period's endowments x_t has two stable steady states. One at $\overline{x}_n$ and another one at $\overline{x}_s$.[2] In the former, people work unskilled and do not wish to undergo education because of the credit market rigidity. In the latter agents work skilled, as they receive sufficient bequests from their parents and do not need to borrow. A third, unstable, steady state separates the agents who will end up rich from those who go into poverty. It is denoted by g and given by the following expression.

$$g = \frac{(1-\alpha)[(1+i)h - w_s]}{(1+i)(1-\alpha) - 1}. \tag{2.1}$$

It now appears appropriate to state the assumptions that are needed in the original work to achieve this dynamic behavior of the economy:

Assumption A 2.1 *In the original work it is assumed that the following holds:*

(i) $(2+r)w_n + (1+r)h \leq w_s < (1+r)\gamma h$,

(ii) $(1-\alpha)(1+r) < 1 < (1-\alpha)(1+i)$, *and*

(iii) $\overline{x}_s > g > \overline{x}_n > 0$.

The first assumption is needed in order to ensure the separation of agents into the three groups. The other two are crucial for the dynamics and admit figure 2.1. Note that (iii) implies $g > f > \overline{x}_n$ and $(1-\alpha)w_s > h$. They have to hold to ensure that the lifetime income curve in fact intersects the 45° line three times thus permitting the three steady states $\overline{x}_n$, g, and $\overline{x}_s$. These assumptions already imply that the case Galor and Zeira analyze is characterized by a fairly high scope for moral hazard γ and a sufficiently high bequest ratio $(1-\alpha)$.

Below it will be of considerable interest how the individuals' well being is affected by changes in the scope for moral hazard. In order to analyze this, it suffices to examine the points f and g. An individual with endowment f is indifferent between borrowing and investing in human capital and unskilled working. Note that f increases as the

[2] We will refer to the first one using the term unskilled worker equilibrium or low income steady state, and to the second by skilled worker equilibrium or high income steady state.

scope for moral hazard grows, meaning that the measure of people earning an unskilled worker's wage increases as f moves to the right. Now we turn our attention towards the unstable steady state g, which separates the two growth paths found in the original work. Also g is found to be increasing in γ, thus implying that the measure of individuals choosing the growth path towards the unskilled worker equilibrium increases in the scope for moral hazard.[3]

As g and f increase in the scope for moral hazard this means that the fraction of both the people who find it profitable to acquire human capital and work skilled and those who choose the growth path to the high income steady state decreases. The former implies that the aggregate income curve shifts downwards in a period of an increase of the degree of moral hazard. The latter means that aggregate income from steady state endowments decreases as probability mass is shifted to the low income steady state. Thus, without lotteries the presence of moral hazard is unambiguously socially wasteful.

2.3 Preliminary Results

In this section we allow for lotteries in the original model's setting, characterize the optimal lotteries, and discuss some of their properties. To be precise, we are looking for the lottery preferred by all agents in the set of fair lotteries over initial endowment levels. A fair lottery means in this context that a lottery ticket costs its expected value. The first observation is that an adequate lottery, i.e. a convex combination of different utility levels, should render the lifetime utility function concave as shown in Figure 2.2. Graphically, we are interested in the convex hull of the lifetime utility function and particularly in the linear part of the convex hull. This is in fact the optimal actuarially fair lottery for all agents.[4] Note that between the points of tangency, say between endowments x^* and x^*+y^*, the con-

[3] Both statements can be verified by inspection of the first derivatives.

[4] A thorough analysis of the general underlying problem including a result on Pareto optimality of the tangential lottery can be found in Marshall (1984) or in Garratt and Marshall (1994).

vex combination of the respective utility levels is higher than utility from certain income. Hence, individuals with endowments between the points of tangency prefer to have a certain income of x^* and to invest their remaining wealth in the lottery. Individuals outside the non-concavity region do not wish to gamble.

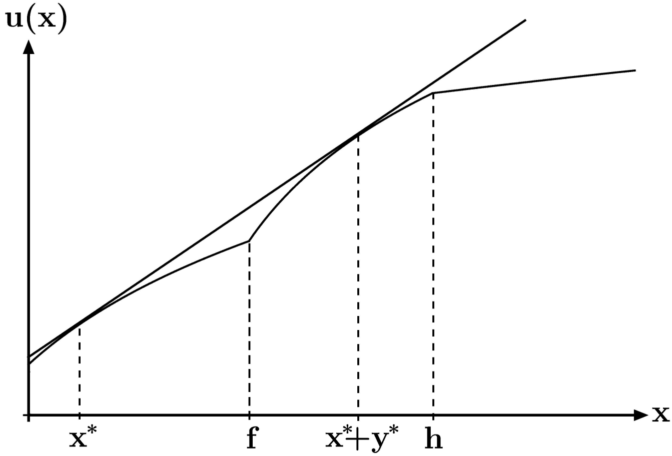

Figure 2.2: An interior solution tangential lottery

To find a Pareto optimal lottery, one has to look for the tangent to both the function $U_n(x) = \ln((2+r)w_n + (1+r)x)$ for $0 < x \leq f$ and the function $U_s(x) = \ln(w_s + (1+i)(x-h))$ for $f < x < h$. From $x = h$ onwards the lifetime utility function changes into $\ln((1+r)(x-h) + w_s)$ which is flatter than U_s and U_n. That means the second point of tangency cannot correspond to endowments greater than h, i.e. $x+y \leq h$. The second property of the tangent is that it connects the points $U_n(x^*)$ and $U_s(x^* + y^*)$. That is, it must hold that:

$$
\begin{aligned}
U_n(x^*) + \frac{\partial U_n(x^*)}{\partial x}\, y^* &= U_s(x^* + y^*) \text{ and} \\
\frac{\partial U_n(x^*)}{\partial x} &= \frac{\partial U_s(x^* + y^*)}{\partial x}
\end{aligned}
\tag{2.2}
$$

These are two equations in two unknowns having a unique solution.

Uniqueness follows from strict concavity of the utility function and existence additionally from inequality of the income slopes $(1+r) < (1+i)$. The first case, which we will call the *interior solution*, is that the unique solution to (2.2), (x^*, y^*), lies in the feasible set defined by:

$$(x^*, y^*) \in \left\{(x, y) : -\frac{2+r}{1+r} w_n < x \leq f;\ 0 \leq x + y \leq h\right\} \tag{2.3}$$

Then solving the system yields the expressions for x^* and y^*:

Lemma 2.1 *The interior solution optimal lottery for $x^* + y^* < h$ is given by:*

$$y^* = h + \frac{2+r}{1+r} w_n - \frac{w_s}{1+i} \quad \textit{and}$$
$$x^* = \frac{(1+i)h - w_s + (1 - \ln(\gamma))\gamma(2+r)w_n}{(1+i)\ln(\gamma)}$$

with the property $x^* + y^* > f$.

This fully characterizes the optimal lottery and the set of agents who want to participate. It is a raffle paying out a fixed prize y^* and participants buy their desired winning probability. Individuals invest an amount x of their initial endowments in lottery tickets and receive in return the probability $\frac{x}{y}$ of winning the prize y.[5] This is preferred by all agents with endowments $x \in (x^*,\ x^* + y^*)$ easily to be seen in Figure 2.2. These individuals wish to invest $x - x^*$ of their endowments in lottery tickets.

As Assumption 1.(i) postulates that $(1+i)h > w_s$, it must be true that $y^* > \frac{2+r}{1+r} w_n > 0$. This implies that there always exists a variety of endowment levels such that agents are better off participating in the lottery, namely all $x \in (\max\{x^*; 0\}; x^* + y^*)$. Note that x^* may be negative implying that all the poor prefer the lottery. Additionally, it holds that $x^* < f$ and $x^* + y^* > f$, so all lottery winners invest

[5]This lottery is analogous to the one described in Friedman (1953), footnote 13. Note that this lottery is equivalent to one paying out x^* to the losers and $x^* + y^*$ to the winners and minimum investment of x^*. Then all individuals with $x \in (x^*,\ y^* + x^*)$ would want to invest all their wealth in the lottery.

in human capital and become skilled workers and all losers become unskilled workers.

If the unique solution of the equation system is not contained in the feasible set (2.3), the point of tangency to $U_s(x)$ is to the right of endowment h. In this case, which we will call the *corner solution*[6], it must hold that:

$$x^* + y^* = h \text{ and}$$
$$\frac{(1+r)y^*}{(1+r)x^* + (2+r)w_n} + \ln((1+r)x^* + (2+r)w_n) = \ln(w_s)$$

which can be rearranged to yield

Lemma 2.2 *The corner solution optimal lottery is given by:*

$$y^* = \ln\left(\frac{w_S}{(1+r)x^* + (2+r)w_n}\right)\frac{(1+r)x^* + (2+r)w_n}{1+r} \quad \textit{with } x^*+y^* = h$$

If there is a corner solution and winners' endowments after the lottery are equal to h this corresponds to a situation where the market for educational loans collapses as all agents in need of capital choose to gamble. Thus lotteries may crowd out loans up to a complete breakdown of the credit market. Clearly, in this case the interval $(x^*, x^* + y^*)$ is non-empty, too. This means that there are always endowment levels for which it is profitable to participate in the lottery. Then some agents are better off gambling while the other agents behave as in the original work, so we can state the following preliminary:

Lemma 2.3 *In the setting of Galor and Zeira there always exists a fair lottery with a prize y^* such that a non-empty set of agents characterized by their initial wealth $x \in (x^*; x^* + y^*)$ participates. This leads to a Pareto improvement from an ex ante point of view.*

Moreover, we are able to state a property of the solution to (2.2) that will be very useful in the further course of the paper. Indeed,

[6] As a matter of fact there is another corner solution at $x^* = 0$. Because this case does not alter our results at all, for notational convenience we will limit our attention without loss of generality to cases where $x^* \geq 0$.

the corner solution is the limit of the interior solution as the degree of moral hazard γ increases.

Lemma 2.4 *An interior solution* (x^*, y^*) *of the system of equations (1) converges to the corner solution as the scope for moral hazard increases. In the case of an interior solution* x^* *and* y^* *increase with moral hazard.*

Proof: In Appendix.

With an optimal lottery there appear to be two ways of human capital acquisition in the model. Firstly by paying h, which applies to the rich individuals. Secondly by buying a lottery ticket which pays partly for the education for the winners. As noted by Garratt and Marshall (1994) this bears certain resemblance to a private and a public schooling sector. Publicly funded schools are financed by parental income or bequest tax giving each contributor a probabilistic right of admission with the success probability increasing in parental income.

2.4 Lotteries and Intergenerational Dynamics

It is a natural question whether the instantaneous Pareto improvement is sustainable. Sustainability will be used in this context in the sense that the lottery winners' increase in utility carries over to the next generation putting the entire dynasty on the growth path to the skilled worker equilibrium with certainty. Winning the lottery at the beginning of the period changes the winners' endowments from $x_t = x$, $x^* < x < x^* + y^*$ to $x_t = x^* + y^*$. In order to put the winners on the growth path to $\overline{x}_s$ the point $x^* + y^*$ must lie to the right of g. Following the argumentation in Galor and Zeira (1993) the agents then bequeath more to their offspring than they had themselves after winning the lottery. This pattern continues for generations and finally bequests converge to $\overline{x}_s$. This argumentation gives rise to the following definition:

Definition 2.1 *Sustainability is said to hold if $x^* + y^* > g$.*

Obviously, this implies also that whenever sustainability holds, the unstable steady state g ceases to exist as people in g prefer playing the lottery. We will be able to provide a condition to ensure that lucky gamblers in fact bequeath enough to their offspring in order to put them on the desirable growth path. Then we look at the robustness of the low income steady state, i.e. the unskilled worker equilibrium, and provide conditions for eliminating this steady state. To ensure that $\overline{x}_n$ is not a steady state, individuals with this endowment must prefer the optimal lottery to certain income from unskilled working. This is the case if and only if $\overline{x}_n > x^*$ which motivates the following definition:

Definition 2.2 *Elimination of the low income steady state is said to hold if $x^* < \overline{x}_n$.*

Note that full benefits of the introduction of lotteries arise only if both sustainability holds and the low income steady state ceases to be one. In this case the introduction of lotteries not only creates a Pareto improvement but also erases the presence of a long run poverty trap at the endowment point $\overline{x}_n$ (see figure 2.1).

Before we begin the analysis of the steady states let us briefly state the bequest function in the presence of lotteries. It depends linearly on lifetime income and fully characterizes the dynamics of the model. As stated in Lemma 2.3, any agent with initial wealth $x \in [x^*; x^* + y^*]$ (weakly) prefers to participate in the lottery. The other agents choose to behave as in the original model. This allows us to write down the individuals' bequest functions as follows:

$$x_{t+1} = (1-\alpha) \begin{cases} ((2+r)w_n + (1+r)x) & \text{for } 0 \leq x < x^* \\ \left\{ \begin{array}{l} ((2+r)w_n + (1+r)x^*) \text{ or} \\ (w_s + (x^*+y^* - h)(1+i)) \end{array} \right. & \text{for } x^* \leq x < x^*+y^* \\ (w_s + (x-h)(1+i)) & \text{for } x^*+y^* \leq x < h \\ (w_s + (x-h)(1+r)) & \text{for } h \leq x. \end{cases}$$

With lotteries available the poorest individuals still tend to prefer the certain income from unskilled working and the rich ones with sufficient endowments to pay the acquisition cost of human capital still choose the skilled worker's income with certainty. And still there may be some who want to borrow to become skilled workers. Yet there appear some individuals preferring to gamble, drawn from the middle income group. Their endowments are almost or just sufficient to find investing in human capital profitable without lotteries but they face a high borrowing interest rate due to the moral hazard. Note that this behavior appears to be consistent with the data collected by Garrett (2001). He finds in a cross section analysis that gambling expenditure as a percentage of GDP is highest for intermediate income countries. Elasticity of demand for gambling is positive and appears to increase for low and intermediate countries whereas it is negative for high income countries.

2.4.1 Sustainability

Analyzing the individuals' bequest behavior in the presence of lotteries in the Galor and Zeira setting we are able to state our first result about the intergenerational dynamics:

Proposition 2.1 *For any set of parameters there exists some $\hat{\gamma}$ such that for all $\gamma \geq \hat{\gamma}$, such that the increase in lottery winners' lifetime income is sustainable. For $\gamma < \hat{\gamma}$, (i) there exists an interval $(h_1, h_2]$ such that for the cost of acquiring human capital $h \in (h_1, h_2]$ sustainability holds, and (ii) there exists an interval $[w_1, w_2)$ such that for skilled workers' wage $w_s \in [w_1, w_2)$ sustainability holds.*

Proof: Assume parameters h, w_s, w_n, r, α and γ such that Assumption A2.1 holds. Using Lemma 2.1 and the definition of the point g from (2.1), the condition $x^* + y^* - g > 0$ can be restated for interior solutions[7] as:

$$\begin{aligned}\left(\frac{\ln\gamma+1}{\ln\gamma}-\frac{(1-\alpha)(1+i)}{(1-\alpha)(1+i)-1}\right)\left[h-\frac{w_s}{1+i}\right]+\frac{1}{\ln\gamma}\frac{2+r}{1+r}w_n &> 0\\ \Leftrightarrow \frac{(1-\alpha)(1+i)-(\ln\gamma+1)}{\ln\gamma[(1-\alpha)(1+i)-1]}\left[h-\frac{w_s}{1+i}\right]+\frac{1}{\ln\gamma}\frac{2+r}{1+r}w_n &> 0\\ \Leftrightarrow \left(1-\frac{\ln\gamma}{(1-\alpha)(1+i)-1}\right)\left[h-\frac{w_s}{1+i}\right]+\frac{2+r}{1+r}w_n &> 0.\end{aligned} \tag{2.4}$$

For $x^* + y^* > g$ to hold it suffices that the first term in the last expression is positive. This is the case if:

$$\frac{\ln\gamma+1}{\gamma} < (1-\alpha)(1+r).$$

This shows clearly, as the LHS decreases with γ, that for any set of parameters there exists always some $\hat{\gamma}$ sufficiently large such that the above condition holds for all $\gamma \geq \hat{\gamma}$.[8] The constraints on γ stated in Assumption 2.1, in particular $\frac{w_s}{(1+r)h} < \gamma < \frac{1}{(1-\alpha)(1+r)}$, are obviously consistent with the sufficient condition and the implication. This in turn implies that if there is sufficient scope for moral hazard, inequality (2.4) holds and lottery winners' endowments $x^* + y^* > g$ after the lottery took place. Thus, winners are set on the growth path to $\overline{x}_s$. As stated in Lemma 2.4, an increase in γ drives the interior solution towards the corner solution where sustainability holds trivially. It is thus always the case that there exists some γ sufficiently great such that lottery winnings are sustainable.

To prove the second part of the proposition we proceed with the analysis of inequality (2.4). Let $\frac{1}{\ln\gamma} < \frac{1}{(1-\alpha)(1+i)-1}$ and the sufficient condition does not hold. Then $x^* + y^* > g$ is equivalent to:

$$(2+r)\gamma w_n > \left(\frac{\ln\gamma}{(1-\alpha)(1+i)-1}-1\right)((1+i)h-w_s) \tag{2.5}$$

[7] Note that this holds by definition if there is a corner solution to the maximization problem, i.e. $x^* + y^* = h$.

[8] Note also that then the LHS of inequality (2.4) increases with γ.

where both sides are positive. It cannot be concluded that this condition generally holds from the Assumption A2.1 alone. However, it can be easily shown by looking at (2.5) that (i) given all other parameters there exists $\hat{h} > \frac{w_s}{1+i}$ such that for $\frac{w_s}{1+i} < h \leq \hat{h}$, consistent with Assumption A2.1, condition (2.5) holds, and that (ii) given all other parameters there exists $0 < \hat{w}_s < (1+i)h$ such that for all $(1+i)h > w_s \geq \hat{w}_s$, consistent with Assumption A2.1, condition (2.5) holds. On the other hand, if γ happens to be very close to its lower bound, then the inequality need not hold.[9] That means there are parameter values such that a lottery winner's income is not sustainable, but there exists always some γ sufficiently high such that inequality (2.4) holds. ■

To sum up, we have shown that in case of an interior solution the long run effect on lottery winners' income paths is ambiguous. However, if the scope for moral hazard captured in the parameter γ is sufficiently great, sustainability can always be achieved. Sustainability trivially holds if there is a corner solution as then lottery winners have endowments equal to the acquisition cost of human capital.

2.4.2 The Low Income Steady State

The next issue is the robustness of the unskilled worker steady state found in the original work. Taking into account that lotteries might change the behavior of the agents situated in that steady state, namely inducing them to gamble for a better life. Formal analysis shows that this indeed may be the case and the low income steady state may disappear:

Proposition 2.2 *Allowing for lotteries in the setting of Galor and Zeira the dynamics of the model has the following properties:*

(i) *In case of an interior solution, $\overline{x}_n$ ceases to be a steady state iff:*

$$\frac{w_s}{(2+r)w_n} > \gamma\left(\frac{(1+r)h}{(2+r)w_n} + 1 - \frac{\ln\gamma}{1-(1-\alpha)(1+r)}\right).$$

[9] An example where γ is consistent with Assumption A2.1 but sufficiently small not to induce sustainability can be found in the Appendix.

This condition tightens in γ for interior solutions.

(ii) *In case of a corner solution, $\overline{x}_n$ ceases to be a steady state iff:*

$$\frac{(1+r)h}{(2+r)w_n} \leq \ln\left(\frac{[1-(1-\alpha)(1+r)]w_s}{(2+r)w_n}\right)\frac{1+(1-\alpha)(1+r)^2}{1-(1-\alpha)(1+r)}.$$

Condition (ii) implies condition (i).

(iii) *If condition (ii) holds, then there always exists a γ sufficiently high, such that lotteries both achieve sustainability and eliminate the low income steady state.*

Proof: Writing down Definition 2 as $\overline{x}_n - x^* > 0$ leaves us with:

$$\begin{aligned} \frac{(1-\alpha)(2+r)w_n}{1-(1-\alpha)(1+r)} - \frac{1}{\ln\gamma}\left(h - \frac{w_s}{1+i} + (1-\ln\gamma)\frac{2+r}{1+r}w_n\right) &> 0 \\ \Leftrightarrow \frac{w_s}{1+i} - h + (2+r)w_n\left(\frac{(1-\alpha)\ln\gamma}{1-(1-\alpha)(1+r)} - \frac{1-\ln\gamma}{1+r}\right) &> 0 \end{aligned}$$

This can be simplified and rearranged to yield the condition in number (i):

$$\begin{aligned} \frac{w_s}{(1+r)\gamma} - h + \frac{2+r}{1+r}w_n\left(\frac{\ln\gamma - [1-(1-\alpha)(1+r)]}{1-(1-\alpha)(1+r)}\right) &> 0 \\ \Leftrightarrow \qquad \frac{w_s}{(2+r)w_n} > \gamma\left(\frac{(1+r)h}{(2+r)w_n} + 1 - \frac{\ln\gamma}{1-(1-\alpha)(1+r)}\right). \quad (2.6) \end{aligned}$$

By closer inspection we find that the RHS of this inequality is increasing in γ for sufficiently small γ.[10] Note that for sufficiently high γ the corner solution obtains. However, for interior solutions we know from Lemma 2.4 that x^* is strictly increasing in γ. This implies that inequality (2.6) unambiguously tightens as γ increases for x^* interior. For corner solutions x^* remains constant as does $\overline{x}^*$.

Turning to the case of a corner solution note that $x^* \leq \overline{x}_n$ does not automatically hold. It is, however, equivalent to the following

[10] In fact, one can easily show by differentiating the RHS of (2.6) that it has a maximum for 'small' γ which may or may not be smaller than the LHS of (2.6). As γ goes out of bounds the RHS approaches minus infinity.

weak inequality:

$$\begin{aligned}
& \frac{\partial U_n(\overline{x}_n)}{\partial x}(h-\overline{x}_n)+U_n(\overline{x}_n)\leq U_s(h) \\
\Leftrightarrow \quad & \frac{(1+r)(h-\overline{x}_n)}{(1+r)\overline{x}_n+(2+r)w_n}-\ln\left(\frac{w_s}{(1+r)\overline{x}_n+(2+r)w_n}\right)\leq 0
\end{aligned}$$

Note at this point that ln(.) must be strictly positive. Then, applying the definition of $\overline{x}_n$ to the last expression we obtain:

$$\frac{(1+r)\left[h-\frac{(1-\alpha)(1+r)(2+r)w_n}{1-(1-\alpha)(1+r)}\right]}{(1+r)\frac{(1-\alpha)(2+r)w_n}{1-(1-\alpha)(1+r)}}\leq \ln\left(\frac{[1-(1-\alpha)(1+r)]w_s}{(2+r)w_n}\right)$$

Getting rid of the fractions and rearranging yields the following expression:

$$\frac{(1+r)h[1-(1-\alpha)(1+r)]-(1+r)(1-\alpha)(2+r)w_n}{(2+r)w_n}\leq \ln(.)$$

This allows us to write down the condition (ii):

$$\frac{(1+r)h}{(2+r)w_n}\leq \ln\left(\frac{[1-(1-\alpha)(1+r)]w_s}{(2+r)w_n}\right)\frac{1+(1-\alpha)(1+r)^2}{1-(1-\alpha)(1+r)}. \tag{2.7}$$

If this inequality holds then the poverty trap vanishes in case of a corner solution. By Lemma 2.4 we know that x^* strictly increases in γ for interior solutions. This means that x^* is maximal in case of the corner solution. In contrast to this, $\overline{x}_n$ does not depend on γ which allows the conclusion that if $x^*\leq\overline{x}_n$ for the corner solution which is equivalent to the last inequality, this must hold for interior solutions as well. This implies that whenever inequality (2.7) is true, inequality (2.6) must hold as well.

To show the last part of Proposition 2.2, note first that the corner solution condition (ii) implies the interior one (i) and, by Lemma 2.4, the interior solution converges to the corner solution as γ increases. Together this means (ii) implies that $\overline{x}_n$ ceases to be a steady state for all levels of moral hazard. Now it is possible to apply Proposition 1. ■

The essence of Proposition 2.2 is to characterize the class of parameters for which the unskilled worker steady state disappears. Note that condition (i) becomes less likely to hold as γ increases. This is quite disappointing given the results in the last subsection, because there now appears to be a trade-off between sustainability and the elimination of $\overline{x}_n$ as a steady state. Yet the critical condition is number (ii) stating under which circumstances the corner solution lottery induces people with endowment $\overline{x}_n$ to gamble for any scope for moral hazard. In the following corollary we give a sufficient condition for this.

Corollary 2.1 *A sufficient condition for the elimination of the low income steady state is given by:*

$$1 \leq \frac{w_s + (1+r)h}{w_s - (1+r)h} \, \frac{(2+r)w_n}{(1+r)h} \ln \left(\frac{[1-(1-\alpha)(1+r)]w_s}{(2+r)w_n} \right).$$

This is implied by using Assumption A2.1.(iii) on condition (ii) in Proposition 2.2. Note that only wages and education cost enter the sufficient condition. This means the class of parameter sets consistent with (ii) appears to contain all the interesting cases, mainly those where lifetime income from unskilled working is comparable to the cost of human capital acquisition. It seems that mainly polar cases of parameter sets, in the sense that one parameter takes an extreme value compared to the others, violate (ii).

So far, we have shown that there are always agents who demand lotteries. Then we have characterized the optimal lottery. With respect to the dynamics of the model it is of considerable interest whether the optimal lottery leads to a sustainable increase in income and if it erases the low income steady state. We found in Proposition 2.1 that the former can always be achieved provided the scope for moral hazard is great enough. Yet introducing lotteries may or may not achieve the latter, although in the more plausible classes of parameter sets the low income steady state is eliminated. In Proposition 2.2, the critical condition is number (ii). Not only does it imply number (i), it also determines the circumstances under which moral hazard in the economy may increase without causing harm. Lotteries extend their beneficiary effects fully only if both sustainability holds and the low income steady state vanishes. Number (ii) in Proposition 2 is a sufficient condition for both of this and thus for the elimination of the long run poverty trap, given a degree of moral hazard sufficiently high.

2.5 Welfare Analysis

This section aims to provide an analysis of the effects of moral hazard on welfare in the long run. It will focus on those individuals who in fact choose to participate in the lottery. Those who do not gamble behave as in the original work and can be excluded from a welfare analysis. Therefore we limit our attention to the *gamblers*, that is the individuals who prefer to participate in the lottery given some degree of moral hazard in the economy. To make the analysis meaningful, we assume that these individuals have positive measure at the beginning of our analysis. At the beginning of each period the gamblers invest all their wealth surpassing x^* in lottery tickets, so that the individual winning probability is given by $\frac{x-x^*}{y^*}$. Winners receive $x^* + y^*$, whereas the unlucky ones keep x^* before deciding whether to work skilled or unskilled. Losers prefer to work unskilled, whereas we know from Lemma 2.1 that winners invest in human capital and work skilled. This means the expected lifetime income $I^G(x)$ of a gambler with endowment x is:

$$\begin{aligned} E[I^G(x)] &= \frac{x-x^*}{y^*}((1+i)(x^*+y^*-h)+w_s)+ \\ &\quad +(1-\frac{x-x^*}{y^*})((1+r)x^*+(2+r)w_n) \end{aligned}$$

Then it follows immediately that:

Lemma 2.5 *The individual expected income allowing for lotteries is strictly greater than the certain income without lotteries for almost all endowment levels among gamblers.*

The proof is a straightforward application of Jensen's inequality and can be found in the Appendix. Lemma 2.5 holds for any distribution of endowments and thus for any period t. This implies that if lotteries are allowed for in any given period, within that period they move probability mass to the right on the endowment line. If sustainability holds, this carries over to the following periods. As individual expected lifetime income is greater with lotteries, so is aggregate lifetime income because the lotteries are actuarially fair and there is no aggregate uncertainty.

As shown above, it may be the case that allowing for lotteries does not lead to sustainability but this might be achieved, were there more scope for moral hazard in the economy. How would the individuals' well-being be affected then? Note first that all individuals with initial endowments $x \in (x^*, h)$ obviously would face a decrease of utility if γ increased. By the increase, risk averse agents would be made worse off in certain income so that they then would prefer a convex combination of utilities, which they did not before the increase.

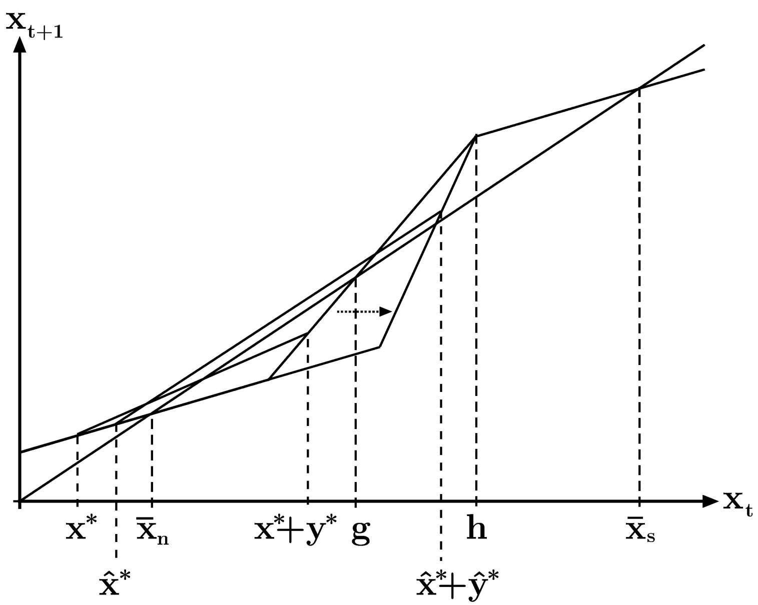

Figure 2.3: Increasing the scope for moral hazard

Assume for now that there is an exogenous deterioration of capital markets in the economy. As can be seen in Figure 2.3, lotteries crowd out borrowing as the scope for moral hazard rises. An increase of γ shifts f to $\hat{f}$, so that the optimal lottery before the increase, given by the flatter dotted line, translates into the steeper dotted line, characterized by $\hat{x}^*$ and $\hat{y}^*$. The interval of endowments such that gambling is preferred extends. Conversely, the interval of endowments such that borrowing is preferred shrinks.

Continuing the welfare analysis, note that because of 'warm glow' - preferences an increase in utility of the following generations does not enter the present generation's utility function. Hence, we have to focus on future period aggregate utility. We begin by assuming that in some period t γ rises to the level $\hat{\gamma}$ where sustainability just holds, i.e. $\hat{x}^* + \hat{y}^* = \hat{g} + \epsilon$, $\epsilon > 0$. Denote the gamblers' initial conditional distribution of initial endowments in period t by[11]

$$F_t(x) := F_t(x_t \,|\, x_t^* \leq x_t \leq \hat{x}_t^* + \hat{y}_t^*)$$

which is assumed to be integrable. Aggregate utility in any period $t + k$ is given by

$$U_{t+k} = \int_{x^*}^{\overline{x}_s} EU(I(x)) dF_{t+k}(x)$$

where $EU(I(x))$ denotes the expected utility of an agent depending on endowments and $F_{t+k}(x)$ the period-$t+k$-distribution of endowments of period-t gamblers. On the one hand, an increase of γ strictly lowers utility in the same period and may lower aggregate income. But in the long run, in particular in the steady states, aggregate utility may be higher if there is more scope of moral hazard. We are able to state the following.

Proposition 2.3 *Assume an interior solution lottery, such that sustainability does not hold, i.e. $x^* + y^* < f$. Then a sufficiently great increase of the scope moral hazard always leads to higher aggregate utility and income in the economy after finitely many periods if*

[11] All endogenous variables will be denoted by hats for the high moral hazard case.

(i) *condition (ii) in Proposition 2.2 holds, or*

(ii) *condition (i) in Proposition 2.2 does not hold, but it holds that*

$$\int_{\hat{x}^*}^{\hat{x}^*+\hat{y}^*} (F_t(g) - F_t(x))dx > 0.$$

Proof: In Appendix.

Note that conditions (i) and (ii) in Proposition 2.2 are sufficient conditions for $x^* < \overline{x}_n$. Intuitively, part (iii) in Proposition 2.2 or part (ii) of Proposition 2.3 ensure that y^* increases sufficiently with γ so that both sustainability and elimination of the poverty trap can be obtained. The main point of Proposition 3 is to provide conditions for sustainability to be more important for long run prospects than the elimination of the low income steady state. As it turns out, this is true whenever Proposition 2.(ii) holds which is depicted in Figure 2.3, independently of the distribution of endowments in the initial period or, if this is not the case, if there is sufficient probability mass to the left of the point g. The latter means that if there are many gamblers with endowments less than g, then sustainability is more important. Loosely speaking, this happens whenever the poor outnumber the middle income class individuals.

Yet sustainability may always be achieved by a sufficiently high scope for moral hazard. The intuition is that borrowing becomes less attractive and more and more agents prefer gambling to borrowing as credit market conditions deteriorate and the interest rate spread rises. Then the prize payed out by the lottery rises and so does lifetime income of the winners. Thus it becomes more likely that winners are able to overcome the barrier that separates rags from riches in the long run. In these cases long term perspectives of the economy depend positively on the scope for moral hazard. Thus reducing market imperfections runs the danger of placing the economy on a growth path towards less aggregate utility, or even of creating the poverty trap found by Galor and Zeira (1993).

Now consider the case of a corner solution, which implies that market imperfections are sufficiently severe to cause a credit market breakdown, being an Pareto improvement as opposed to the case without lotteries. If Proposition 3 holds, then this appears to be at least not harmful if not desirable from a long term point of view. Here appears a nice connection to the work of Garratt and Marshall (1994) who analyze a case where the credit market to finance human capital acquisition is assumed not to exist. In an otherwise very similar setting they find that an optimal social contract involves a fixed prize lottery. This chapter might offer a theoretical motivation for their assumption, explaining why the economy might be better off without a credit market for educational loans once lotteries are taken into account.

2.6 Application

A natural opportunity to gamble that does not rely on third party enforcement is, of course, conflict. Typically the outcome of conflict between two parties does reflect the relative strength of each party, yet the outcome is subject to a random component. The choice of weapons affects success probabilities but victory is far from certain. Incorporating the possibility of costly conflict in the framework, the model predicts a certain segregation within the economy. While at least some of the poorer agents find it desirable to engage in conflict, richer agents do not. Moreover, those agents willing to engage in conflict do not necessarily seek to fight rich agents. This draws a familiar picture of socioeconomic segregation within an economy where poverty and criminal activity are highly correlated. Some evidence for this is provided, for instance, by Cullen and Levitt (1999). They find that although migration into urban areas does not appear to be affected by changes in crime rates, richer people tend to be very responsive to higher crime rates leaving areas where crime rates have increased. One could argue that socioeconomic segregation is largely due to social interaction, e.g. via negative role models, and that this in turn determines criminal activity. However, the study by Ludwig et al. (2001) casts some doubt on this argument. It finds that in an randomized housing experiment property crime rates after relocation are positively affected by being treated.

The main departure of an economy with conflict from our model lies in the waste of resources due to fighting. This section aims to provide an illustration of how conflict may serve as a means to provide agents with their desired opportunities for convexification of their consumption sets. Suppose two agents fight and the winner has the opportunity to expropriate the loser. Suppose further the result is affected by the agents' individual strength represented by the amount of endowments they carry with them. However, the outcome of conflict is not deterministic. Being stronger than one's opponent does not ensure victory, it increases the chances of winning. Therefore we use a stochastic contest technology, the ratio form contest success function (see Hirshleifer, 1989, Skaperdas, 1996, Neary, 1997).[12]

To be precise, let us assume that an agent i employs part of his initial endowment $x_i - x$ for use in the conflict. Let y denote the sum of all endowments used by parties in the conflict. Then agent i's probability of winning the conflict is given by

$$\pi(x, y) = \frac{x_i - x}{y} \tag{2.8}$$

The prize for winning the conflict is determined by the sum of endowments used in the conflict y. However, during the conflict part of the endowments is wasted such that the winner conflict obtains the prize βy where $\beta < 1$ is a cost parameter. This means that for an agent i a conflict is equivalent to a costly ex-ante lottery (x, y) over endowment levels $x + \beta y$ with probability $\pi(x, y)$ and x with probability $1 - \pi(x, y)$.

The following proposition states that given market imperfections are sufficiently severe and the percentage cost of conflict is not too high there are indeed agents preferring to gamble.

[12]Under the ratio form contest success function a party that does not dedicate any endowments to conflict loses with certainty. Hirshleifer (2000) argues that the difference form appears to be more consistent with fighting as a non-resisting losing party need not lose everything due to real-world frictions. In our case we assume that conflicting parties are able to hide away part of their initial endowment they do not want to risk in a fight before conflict thus incorporating the desired property.

Proposition 2.4 *Agents with initial endowments x_i in the neighborhood of f prefer a costly lottery to their certain endowment if (i) the scope for moral hazard is sufficiently high, or if (ii) the cost of lotteries is sufficiently low. Agents with endowments $x_i < x^*$ or $x_i > x^* + y^*$ prefer their certain endowment to any costly lottery. The endowment of a winner of a costly lottery that is the optimal choice for some agent, $x + \beta y$, has the property $f < x + \beta y \leq x^* + y^*$.*

Proof: In Appendix.

Firstly, this proposition reinforces the intuition of the previous sections. There is scope for conflict in the economy if either the capital market is sufficiently imperfect or the cost of fighting sufficiently low. The agents willing to engage in conflict are a subset of the gamblers of the sections above. More interestingly, there is an upper bound on a winner's endowment depending on the cost of conflict. Combined with the fact that winners will become skilled workers, this points to the fact that for sufficiently low cost of conflict the analysis of the previous sections goes through as well.[13] On the other hand, sufficiently rich agents are not threatened by conflict as fighting occurs mainly among the poor. This appears to be consistent with the empirical evidence.

2.7 Conclusion

We find in this chapter that allowing for lotteries in the setting of Galor and Zeira (1993) leads unambiguously to a Pareto improvement. Consistent with existing literature it is found that mainly agents with intermediate incomes want to gamble whereas rich individuals never do. For wide and the more plausible classes of parameters also the dynamics of the model economy changes fundamentally, erasing the possibility of a poverty trap emerging in the original work. Intuitively, agents previously caught in the poverty trap now like to gamble and

[13] Indeed, we conjecture that the equilibrium under actuarially fair lotteries is the limit of an equilibrium with costly lotteries as β approaches 1.

are thus able to escape towards the desirable high income steady state independent of the initial distribution of endowments. Hence, the finding of persistent inequality due to the presence of capital market frictions may be reversed once the possibility of gambling is taken into account.

In the presence of lotteries there emerges another interesting trade off: From a long term perspective it would be desirable both that lottery winnings are sustainable and that agents in the poverty trap want to escape by gambling. The former appears to be achievable as the moral hazard problem becomes more severe, while for the latter the reverse is found to be true. A great scope for moral hazard in the economy may be interpreted as high monitoring costs, due to e.g. a legal system favoring borrowers over lenders, lack of property rights, or difficulties in contract enforcement and the like. As it turns out, the goal of sustainability seems to be more important for a large class of parameters from a long run perspective. Hence, in contrast to the original work, it may in fact be desirable to have more scope for moral hazard thus reducing the demand for inefficient loans. That is, given perfect capital markets cannot be achieved, capital market frictions actually need not be bad news and in this model policies aiming to ameliorate market conditions may put in danger long run prosperity indeed.

The underlying mechanism of the model, namely the crowding out of inefficient means to accumulate sufficient capital, seems to apply to far more general settings of dynamic inefficiency. Introducing a less inefficient accumulation instrument to circumvent barriers to extra rents, similar to this model's human capital acquisition cost, may in general lead to similar findings.

Taking a positive point of view, the far from perfect markets for financing human capital acquisition might state a long run optimum. Then nothing could be gained by adopting political measures aiming at the reduction of rigidities in this market. Quite the contrary, such policy would run the risk of worsening long term perspectives of the economy. Of course, robust empirical evidence is necessary to be able to derive political implications. Unfortunately, existing empirical literature on the effects of lotteries as a means to bypass wealth constraints tends to be scarce. Adopting the interpretation of lotteries as public funding of universities, it should be expected that greater initial endowment inequality goes hand in hand with a less pronounced public higher educational system. This points to an obvious extension of the model delivering testable predictions for empirical research. On the other hand, interpreting property crime as an opportunity to gamble, the analysis predicts that higher population shares of poor agents induce higher crime which is consistent with most empirical observations. However, this in turn affects the institutional shape of markets, namely the security of property rights in an economy. Thus higher crime rates could be interpreted as higher scope for moral hazard. Subsequently, poor agents are attracted more to conflict further crowing out capital market. Growth prospects then depend mainly on the efficiency of the conflict technology.

Chapter 3

Inequality, Incomplete Contracts, and the Size Distribution of Business Firms

3.1 Introduction

The size distribution of business firms has proved to be one of the more enduring objects of economic research as it reflects the organization of production in an economy. Naturally, it is one of the major concerns of economic analysis whether production is organized efficiently, be it within plants, firms, or industries. Within an industry, that is among firms producing the same output good, presumably with access to the same technology, there is little theoretical reason to expect a non-degenerate distribution of sizes in classical competitive equilibrium theory. Empirical evidence, however, tends to confirm heterogeneous size distributions of firms. A possible way of reconciliation is that firms differ with respect to unobservable technological characteristics and firm sizes are chosen optimally in response to those characteristics. Yet judging from observations it is far from obvious that all firms are indeed efficiently sized (see e.g. Banerjee and Duflo, 2004).

Moreover, when efficient organization of production given technology means optimal factor input choice, this primarily delivers a theory of plant size as put forward by Hart (1995). It is not obvious how technical efficiency at plant level might be connected to technical efficiency at the firm level. Integration of technically efficient plants might for instance result in undesirable market concentration. That is, focussing on input-output analysis of production in plants precludes a possibly interesting and relevant aspect of the firm, namely strategic considerations. Firm size may not only affect production but also output markets or the distribution of profits among stakeholders, an issue this chapter will emphasize. Hence, in order to explain size distributions of firms it is desirable to employ a proper notion of the boundaries of a firm. As a consequence, this chapter follows along the lines of property rights theory, explicitly modeling ownership rights on productive assets thereby allowing for a well-defined boundary of the firm.

Departures from a degenerate firm size distribution have also been explained by introducing market imperfections or missing markets. Often this comes in conjunction with assumptions on technology that ensure market imperfections on one market affect another. In this chapter markets are complete and imperfections are assumed explicitly in that contracts within firms are incomplete, thus leading to renegotiations, and there is a spread between lending and borrowing interest rate. There is an immediate application as these assumptions can be expected to hold in developing countries, for instance due to enforcement problems. The production technology is assumed to be deterministic, to have strict complementarity between labor and capital, and a well-defined optimum at a certain firm size such that average output is maximized. This has the virtue of providing a clear-cut efficiency benchmark as opposed to most of the literature where size choice is always constrained efficient given a firm's characteristics, for instance technology or wealth level, and equilibrium prices. Complementarity of inputs to production is needed primarily to ensure that at least one agent is indispensable. However, in this model an agent always has the incentive to choose a technology that makes him indispensable because incompleteness of labor contracts implies renegotiations within firms.

Agents in the model economy meet in a matching market deciding on coalition membership, ownership rights on assets and debt or deposit and are allowed to make instantaneous side payments. Then agents engage in productive activities in their coalitions. Due to the incompleteness of contracts, output shares of agents are determined by renegotiation, consistent with intra-firm bargaining in Stole and Zwiebel (1996). Agents' payoffs then depend positively on their ownership rights on assets.

The equilibrium firm size distribution turns out to be heterogenous and is capable of simultaneously allowing for firms both too small *and* too large compared to the firm size that maximizes average output. For this both intra-firm bargaining and imperfect capital markets are needed. Inequality in initial endowments among agents and capital market imperfections induce a wedge in the cost of capital between agents of different wealth levels. Subsequently, poorer agents choose less investment in assets than rich agents do. The bargaining induces owners of firms to employ more agents than technically efficient as their share of joint profit increases with the number of employees and assets owned. Hence, the desired firm size by a firm owner exceeds the efficient one, but wealth-constrained owners choose smaller sizes. In addition, the model generates a relation between wealth and income inequality reminiscent of the dual economy. Whereas wealth inequality between workers and owners exceeds income inequality, income inequality exceeds wealth inequality between firm members and unmatched agents. As the measure of unmatched agents negatively depends on capital stock the model is able to incorporate a non-monotonic relationship between inequality and growth.

In equilibrium, the firm size distribution tends to be bimodal whenever market imperfections and endowment inequality are sufficiently severe and the endowment distribution has a Paretian tail. This is consistent with empirical findings as bimodal size distributions appear to be typical for developing countries (e. g. the survey by Tybout, 2000), countries generally deemed to be prone to market imperfections and high wealth inequality. In contrast, size distributions of firms in developed countries are commonly found to be unimodal (see Cabral and Mata, 2003). A number of empirical studies (e. g. Little et al., 1988, Biggs et al., 1995, Hallward-Driemeier et al., 2002) provide some evidence of an inverted U-shaped relationship between efficiency and size of firms in terms of output per worker in countries with bimodal firm size distributions, indicating that the presence of both too large and too small firms is indeed common. This suggests that the model is capable of providing a viable framework for policy analysis in developing countries.

The literature on size distributions of firms is abundant, yet tends to focus on input-output analysis. A research agenda can be traced back at least to Simon and Bonini (1958). Lucas (1978) introduces a general equilibrium model with agents heterogenous with respect to productivity. Firms sizes are chosen optimally based on complementarity of agents' productivity type with input factors. In equilibrium, firms always have a technically efficient size and the size distribution depends on the distribution of types. In Kihlstrom and Laffont (1979) agents differ in their aversion to risk. In a general equilibrium model without insurance markets less risk averse agents become entrepreneurs employing more risk averse agents. Labor demand increases and equilibrium firm size decreases in the entrepreneur's risk aversion such that firms tend to be too small in equilibrium compared to the first best. The size distribution of firms then depends on the distribution of risk aversion among agents.

Another strand of the literature concerned with firm size heterogeneity is on entry and exit of firms. Among other notable contributions are the selection models in Jovanovic (1982) and Hopenhayn (1992). They introduce dynamic partial equilibrium models of firms that are subject to productivity shocks and decide on market entry or exit based on the properties of the stochastic process governing productivity shocks. Firm sizes are always efficient given their productivity state and the size distribution depends exclusively on the properties of the productivity shocks.

These contribution share a neo-classical view of the firm as a black box combining input factors to produce output. More recent contributions tend to move away from this standpoint, as for instance, Caballero and Hammour (1996). They assume specificity of factors to production and include bargaining over joint profits in their analysis. This leads to a better bargaining position and a higher profit share of the factor that is less specific in the sense that it may be more easily employed elsewhere. Consequently, they find overemployment in equilibrium when labor is less mobile than capital and thus the equilibrium wage rate is below its first best level. However, given equilibrium prices, firms choose their factor inputs efficiently. Another approach, by Cooley et al. (2004), is to model financial constraints on firms' investments which results in undersized firms compared to the average output maximizing firm size. In a general equilibrium model where contracting constraints prevent efficient lending and firms differ in wealth, poorer firms are undercapitalized. Given a Leontief production technology demand for labor is below its first best level and so are the wages. If there are firms sufficiently rich not to be financially constrained, these choose a size above its first best level. Yet, given the equilibrium wage rate, these firms are efficiently sized. However, with the possible exception of Caballero and Hammour (1996), the mentioned contributions share a potential deficiency in that it is not straightforward whether they are on plant or firm sizes.

Methodologically this chapter is related to Legros and Newman (2004b), who find too large firms in a matching model between technologically complementary firms using an incomplete contracting framework. In their model monetary side payments between coalition members are limited by liquidity constraints, but agents can be compensated by changing the organizational structure. This may lead to integration although this type of ownership in their model is inefficient in a well-defined sense. In a partial setting close to Hart and Moore (1990), Bolton and Whinston (1993) find a similar over-integration result.

This chapter will proceed by introducing the theoretical model in section 2. Section 3 gives an equilibrium existence result, and section 4 identifies some properties of the matching equilibrium with respect to the endowment distribution. In section 5 we identify circumstances leading to a bimodal firm size distribution and discuss applications of the model. Section 6 concludes with a discussion while the more cumbersome proofs are confined to the appendix.

3.2 The Model

3.2.1 Agents

There is a single-good economy populated by a continuum of agents $I \equiv [0, 1]$. The single good can be used both for consumption and investment. Agents are heterogenous in their initial endowments of the good which is given by the one to one mapping $\omega : I \to [\underline{\omega}, \overline{\omega}]$, $\underline{\omega} < \overline{\omega}$, assumed to be continuously differentiable. The function $\omega(i)$ is then the inverse of the atom-less wealth distribution function in the model economy. All agents are endowed with the same amount of human capital h, so $h_i = h$ for all $i \in I$. We assume that an agent's human capital becomes specific to the task that it is employed in.[1] Agent i's utility function $u_i(c_i)$ is linear in consumption of the single good c_i which is given by final payoffs and assumed to be $u_i = c_i$.

3.2.2 Assets

In the economy there exist two types of assets, A^I and A^{II}. These assets are complementary in the sense that combined output exceeds the sum of individual output and one asset without the other is useless. Imagine these assets to be a factory building and a machine, for instance. Thus, under our assumptions a factory cannot produce with-

[1] While this assumption appears to be not entirely innocuous, it has the virtue that outside options in the equilibrium of the renegotiation game do not depend on equilibrium profits in coalitions. An alternative assumption is that the matching process involves frictions, for instance in the form of costs associated with switching coalitions.

out a machine and vice versa. The production technology incorporated in the assets is denoted by $f(A, h_n)$ where the set A denotes the assets used in production and $h_n = (h_0, h_1, ..., h_n)$ a vector of human capital. We follow the convention of writing $A_n = \{A^I, A_1^{II}, ..., A_n^{II}\}$. Assumptions on the technology are

Assumption A 3.1 *(Production technology) Let h_n denote a $(1 \times n{+}1)$ vector of hs.*

(i) $f(\emptyset, h_0) = f(\{A^I\}, h_0) = f(\{A^{II}\}, h_0) = 0.$

(ii) $f(A_n, h_n) - f(A_{n-1}, h_{n-1}) > f(A_{n+1}, h_{n+1}) - f(A_n, h_n) \geq 0 \; \forall n > 1.$

(iii) $\lim_{n\to\infty}(f(A_n, h_n) - f(A_{n-1}, h_{n-1})) = 0.$

(iv) $\frac{1}{n+1} f(A_n, h_n) < \frac{1}{K+1} f(A_K, h_K) \; \forall n \neq K.$

(v) $f(A_n, h_n) = f(A_n, h_{n+1}) = f(A_{n+1}, h_n) \; \forall n \geq 1.$

(vi) *Investment cost of an asset is c regardless of the type.*

As this model is concerned with the output from human capital only we assume for convenience that production cost is zero. Thus we limit our attention to additional output achievable by the strict complementarity of human and physical capital.[2] Assumption (i) gives the output of an agent who is not member of a coalition which is normalized to zero thus giving the agents' outside option. By Assumption (ii) the production technology has diminishing returns to scale and Assumption (iii) will guarantee a finite bound on coalition sizes. Assumption (vi) ensures that there is a firm size K that is efficient in the sense that it maximizes average surplus. Note that it is implied by (ii) and (iii). Assumption (v) states that human capital and assets are strict complements, so that a factory and a machine need an operator each to produce output. By Assumption (vi) setting up an asset involves the investment of c units of the good regardless of the asset type. This assumption is made to avoid that exogenous

[2]Note that this interpretation is also consistent with the costly acquisition of specific human capital provided the acquisition cost is constant across agents.

asset cost asymmetries drive the results of the analysis. Wherever the level of human capital is constant across agent we will drop it as an argument to the production function and write $f(A_n)$ for $f(A_n, h_n)$.

3.2.3 Coalitions and Ownership Structures

All agents $i \in I$ may form finite coalitions to jointly produce output. Coalitions or teams of agents are assumed to be of finite size to capture the notion of small units in a large economy. Within a team of agents each team member matters, but the impact of a team on the whole economy is negligible. Let $\mathcal{F}(I)$ denote the set of finite subsets of I. A coalition will be denoted by $N \in \mathcal{F}(I)$ consisting of $n + 1 = |N|$ members. M will denote the member of coalition N owning A^I. We will call M the owner and the remaining agents the workers, although this proves to be appropriate only for coalitions with $n > 1$ workers.[3] The distribution of ownership rights on assets in a set of agents $L \in \mathcal{F}(I)$ with $L \subseteq N_L$ for some coalition N_L, is captured by $\theta(L)$ giving the assets of one or more agents M:

Definition 3.1 *Let $\theta(L)$ be the mapping $\theta : \mathcal{F}(I) \to \{A^I; A^{II}; \emptyset\}^{\mathbb{R}}$, assigning a set of agents to their assets.*

To familiarize the reader with the concept, note that e.g. $\{A^I\} \subseteq \theta(M)$ by convention. We follow the convention of referring to a coalition that is a singleton by simply writing the identity of the agent. The combination of a coalition and the allocation of ownership rights within that coalition (N, θ) is called a *firm*. At this point we limit our analysis for simplicity to a limited set of admissible ownership structures:

Assumption A 3.2 *For all coalitions N, $|\theta(i)| = 0 \; \forall \; i \neq M \in N$, i.e. no worker holds any assets and $\theta(i) \neq \theta(j)$ for $|\theta(i)|, |\theta(j)| > 0$ and $i, j \in N$.*

[3] In the course of this chapter we will refer to a coalition N by the number of its workers, n, and use the term n firm as a synonym.

This assumption excludes ownership structures that assign joint ownership of assets and those that allow workers to hold assets.[4] A discussion on this assumption is postponed to section 6 where we also present an example where Assumption 3.2 does not hold. However, it is best motivated by putting more structure on the technology incorporated in the assets. Suppose for instance that asset type A^I represents a non-physical asset as for instance trust in the firm embodied in its physical presence, while asset type A^{II} represents the size of the productive asset. This way, the holder of asset A^{II} must at the same time hold A^I. This could describe the setup of a shop where the store building itself represents the firm as perceived by customers and suppliers.

Indeed, Chowdhury (1994) provides a case study of small grocers in Dhaka, Bangladesh, that appears to be consistent with our assumptions. A grocery store typically consists of an owner-manager assisted by employees where the owner-manager owns both the building and the stock. The physical size of the shop determines the number of assistants. Interviewed entrepreneurs emphasize the need for visibility in order to obtain trade credit. Almost all case studies in the same survey find also that owners place huge importance on personal trust between their firm and both suppliers and customers.

Another case study in the same survey, on women entrepreneurs in the textile industry (Zohir, 1994), finds a similar pattern. The typical firm consists of one cutting master who provides design and supervision and several operators of sewing machines. Interestingly, assets, i.e. the sewing machines, are usually owned by the cutting master. Asset A^I would represent the human capital for designing clothes and asset A^{II} a sewing machine. Strikingly, a common feature among small firms in different industries is that initial capital investment is to a large extent financed by equity with typical debt-equity ratios of 0.25. Only after setting up firms begin to regularly use trade credit for financing working capital.

[4] It is possible to relax the assumption to allow workers to own one asset A^{II} without changing the analysis.

3.2.4 Sequence of Events

The timing of events in the model economy is given as follows.

1) Matching market takes place, agents simultaneously decide on coalitions, investment plans, and their capital market position.

2) Production takes place and firm members may renegotiate the distribution of profits at any time.

3) Payoffs take place.

3.2.5 Contractual Environment and Renegotiations

The exposition proceeds backwards in time starting with intrafirm renegotiations. We assume that labor contracts can be renegotiated at any time. One possible foundation for this assumption lies in the non-contractability of claims to future output whereas contracts that determine ownership rights on assets and instantaneous side payments in the course of coalition formation are feasible in this economy.

A reason for the lack of contracts conditioning on output may reside in the legal institutions of an economy and the nature of output. It seems reasonable that at one time or another in the course of production any team member has the opportunity to hide units of output or profits due to its liquid nature.[5] In contrast to this, assets are viewed as machinery or buildings and are somewhat more protected from adverse behavior. Thus output may not be verifiable especially in small and medium businesses lacking a sophisticated accounting scheme. Moreover, even if output is verifiable, in order to enforce the contract a court will be needed. It will then depend on the ability of this court to reveal the true circumstances within a given team whether a certain sharing rule is contractible or not.

[5] This points to another motivation for our assumptions. Suppose for instance that team members' actions are not observable and the amount of output units that can be diverted depends on the individual's access to output. Access in turn is likely to depend on individual ownership rights. A Nash bargaining solution model where threat points are determined by the amount of output agents can secure themselves in case renegotiations break down yields similar conclusions.

Non-binding labor contracts can be caused by weak legal institutions as well.

An alternative set of assumptions also giving rise to renegotiations is that labor contracts are non-binding and contracts on future output only possible within coalitions. This means that contracts on coalitions cannot be written in the matching market. However, agents may decide in the matching market to become specific to each other. This describes an economy where written labor contracts and worker protection laws are not frequent. This appears to hold for less developed countries, and even in some developed countries their effectiveness is not beyond doubt. Outsiders' inability to write contracts on future output may be explained by an asymmetric information setting where firm members' actions are observable only within firms.

Whenever contracts are incomplete in one of these senses renegotiations will determine individual shares of joint profit. Fortunately, the literature on renegotiations with property rights is abundant and provides some more robust results. Let $\pi_i(n, \theta)$ denote the payoff of agent i in team N of size $n+1$ given an allocation of ownership rights θ. The arguments capture joint output $\Pi(n) = f(A_n)$ and agent i's relative bargaining power based on the importance of his assets to the team. Formally, we assume

Assumption A 3.3 *Individual payoffs from renegotiations within the team have the following properties provided individual outside options are worse for all team members.*[6]

(i) *Asset ownership increases the individual profit share:*
$\pi_i(n, \theta) \geq \pi_i(n, \theta')$ *for* $|\theta(i)| > |\theta'(i)|$ *with strict inequality for* $|\theta(i)| > 1$.

(ii) *Profits are split equally in symmetric teams:* $\pi_i(1, \theta) = \pi_j(1, \theta) \forall\, \theta$.

(iii) *The indispensable agent's payoff share increases in firm size:*
$\frac{\pi_i(n,\theta)}{\Pi(n)} > \frac{\pi_i(n-1,\theta)}{\Pi(n-1)}\ \forall\, n$ *if* $\theta(i) \supset A^I$.

[6]It is straightforward that in a matching equilibrium this must hold by stability.

Assumption A 3.3.(ii) is straightforward as within symmetric firms both agents are indispensable for production. Parts (i)[7] and (iii) are central results from most of the literature on bargaining under the possibility of exclusion and compatible with production that has diminishing returns to team size.

Notably our set of assumptions is consistent with a number of extensive-form bargaining games, in particular with those yielding the Shapley value as an outcome. For instance, Hart and Moore (1990) use this concept for modeling renegotiations with the result that holding property rights increases an agent's share of joint profit.[8] Intuitively, this is due to possibility to use the right to withdraw assets from production as a threat. Obviously, if one agent's profit shares increases in the number of assets then the share of some other agents must decrease. The fact that the Shapley value of an indispensable agent increases in the number of substitutable co-workers can be checked by calculations as in Rajan and Zingales (1998).

Another prominent example is the non-cooperative bargaining game proposed by Stole and Zwiebel (1996). They analyze renegotiations within a firm such that an indispensable agent bargains bilaterally with the remainder of the team. The bargaining game used is that of Binmore et al. (1986). In a generalized version they find that individual profits are given by the Shapley value and not less than the individual outside option, that is an agent's maximum profit obtained outside the team. That intra-firm bargaining allowing for property rights on assets is indeed consistent with the assumptions can be concluded from Propositions 1-3 in Gall (2004). Even when

[7]Part (i) is superfluous under centralized ownership as assumed in Assumption A3.2. However, at a later stage of the paper, Assumption A3.2 will be relaxed as suggested in footnote 4. Note that then strict equality in (i) for $|\theta(i)| = 1$ implies that workers holding one A^{II} have the same renegotiation payoff as workers holding no assets. This in turn implies that both roles are equivalent in this model up to an interpretation of the side payments, i.e. whether they include investment cost for asset-less workers or not for asset-owning workers.

[8]Interestingly, de Meza and Lockwood (1998) find that the indication for the property rights allocation is not robust to minor alterations in the extensive form of the underlying bargaining game. The finding that agents' profit shares are non-decreasing in asset ownership carries over nevertheless.

allowing for minor alterations in the extensive form, as in Westermark (2003), Assumption A 3.3 continues to be consistent with this bargaining game.

3.2.6 The Matching Stage

In the matching stage a matching market takes place. This means that individuals decide simultaneously on their investment in assets, on the coalition they will take part in, and their activity on the capital market. When choosing a course of action, agents treat the interest rate as exogenous. In the economy the interest rate is determined endogenously by credit market clearing asset investment plans. At the matching stage agents know their continuation payoffs from the subsequent stage of being member in some coalition N with an allocation of property rights θ and holding ownership rights on assets $\theta(i)$. This continuation profit is given by the individual payoffs from renegotiations, $\pi_i(n, \theta)$.

Let the individual valuation function $v_i(n, \theta, \omega(i), r)$ denote this continuation payoff from being in coalition N net of capital investment costs and interest rate payments depending on the number of workers in a coalition n, asset ownership θ, individual endowments $\omega(i)$ and interest rate r:

$$\begin{aligned} v_i(n, \theta, \omega(i), r) &= \pi_i(n, \theta) + (1+r)(\omega(i) - |\theta(i)|c) \text{ if } \omega(i) \geq |\theta(i)|c \\ v_i(n, \theta, \omega(i), r) &= \pi_i(n, \theta) + (1+i)(\omega(i) - |\theta(i)|c) \text{ if } \omega(i) < |\theta(i)|c. \end{aligned}$$

Feasible allocations of ownership rights θ associated to any coalition N can be restricted to require agents to be able to repay their debts and technical efficiency. The set of feasible allocations of ownership rights on assets in a coalition N, $\Theta(N)$, is then given by

$$\begin{aligned} \Theta(N) &\equiv \{(\theta(i))_{i \in N} : \theta(i) \in \{\emptyset; A_n\} \text{ and} \\ &\quad \sum_{i \in N} |\theta(i)| = |N|; \pi_i(n, \theta) + (1+i)(\omega(i) - |\theta(i)|c) \geq 0\}, \\ \Theta(\{i\}) &\equiv \emptyset. \end{aligned}$$

That means that only allocations that are consistent with Assumption A3.2, that employ the proper number of assets, and whose members are able to pay back their debt are considered feasible except for singletons where only the empty set, i.e. owning no asset and receiving the outside option, is deemed feasible.

3.2.7 Preferences over Ownership and Firm Size

We first determine some properties of the optimal individual asset ownership $\theta(i)$ depending on the interest rate r and the coalition N individual i is member of. Let $\pi_1 \equiv \frac{f(A_1,h_1)}{2}$ abbreviate average profit in a symmetric $n = 1$ firm which is equal to the renegotiation game payoff for a $n = 1$ firm member given Assumption A 3.3. The following lemma states some properties of agents' preference orderings over ownership rights.

Lemma 3.1 *Agents' preferences over individual ownership rights are well-defined and monotone in the interest rate and wealth:*

(i) For any two pairs of ownership rights and coalition size there is at most one interest rate r_0, such that an agent with $\omega(i)$ is indifferent at r_0, prefers the pair involving more asset ownership at $r < r_0$, and prefers the other one at $r > r_0$.

(ii) For any two pairs of ownership rights and coalition size there is at most one endowment level ω_0, such that an agent endowed with ω_0 is indifferent, an agent with $\omega(i) > \omega_0$ prefers the pair involving more asset ownership, and an agent with $\omega(i) < \omega_0$ prefers the other one. Moreover, ω_0 is non-decreasing in the interest rate r.

Proof: In Appendix.

This means that agents' bilateral preference orderings on asset ownership may switch exactly once both in the interest rate and in initial endowments. Moreover, a monotonicity result holds: more assets are preferred as the individual cost for capital decreases, namely whenever the interest rate decreases or wealth increases. It is possible to show that indeed for all agents $\theta(i) = \emptyset$ is strictly preferred to $\theta'(i) = \{A^{II}\}$ in coalitions with $|N| > 1$ and side payments from workers to owners $|t(i) - t'(i)| < c$.

In order to determine an upper bound on firm size we look at asymmetric firms with $n > 1$ where members obtain payoffs from either owning nothing, that is $\theta(i) = \emptyset$ which is given by

$$v_i(n, \theta, \omega(i), r) = \pi_i(n, \theta) + (1 + r)\omega(i),$$

or from owning everything, with $\theta(i) = A_n$, which amounts to

$$v_i(n, \theta, \omega(i), r) \quad = \quad \pi_i(n, \theta) + (1 + r)(\omega(i) - (n + 1)c).$$

It is not necessary to include side payments in these equations as we are interested in the optimal firm size an owner with unlimited resources would choose and side payments must be less than c. Clearly, the workers' payoffs are decreasing in n, net of side payments, whereas a sufficiently rich owner's payoff increases in n as long as

$$\begin{aligned} & v_i(n{+}1, \theta_{n+1}, \omega(i), r) - v_i(n, \theta_n, \omega(i), r) > 0 \\ \Leftrightarrow\ & \pi_i(n + 1, \theta_{n+1}) - \pi_i(n, \theta_n) > (1{+}r)c, \end{aligned} \tag{3.1}$$

where $\theta_n(i) = A_n$. This inequality implies there exists a cutoff firm size $\overline{n}$ so that further increasing firm size leads to a decrease in the owner's payoff even for an owner who has sufficient endowments not to need to borrow. Independently of the owner's asset A^{II} holdings, marginal profit from increasing firm size by one must outweigh the cost for an additional asset. Furthermore, this cutoff firm size must be finite because of the assumptions on the production function. Therefore $\bar{n}$ provides a general finite upper bound on firm sizes. To make things interesting we limit our analysis to the case where investment costs are sufficiently small so that at least for an agent who is not credit constrained it pays to choose a firm size $n > K$ at $r = 0$, that is

Assumption A 3.4 *(sufficiently small investment cost)*

$$\exists n > K : \pi_i(n, \theta_n) > \pi_i(n{-}1, \theta_{n-1}) + c > \pi_1 + nc.$$

where $\theta_n(i) = A_n$. Moreover, $\pi_i(n{+}1, \theta') < \pi_i(n, \theta')$ for all $n \geq 1$ and $\theta(i) = \emptyset$.

This assumption holds whenever profits from renegotiation are increasing sufficiently in firm size despite diminishing returns to scale compared to asset investment cost. As M's renegotiation payoff increases in firm size the maximal firm size to expect is determined by (3.1) and by equilibrium side payments. Let us collect the results on preferred firm sizes in the following lemma.

Lemma 3.2 *Preferences over firm size have the following properties:*

(i) *For $n > 1$, $v_i(n, \theta(i) = \emptyset, \omega(i), r)$ is strictly decreasing in n and*

(ii) *any firm size $n > \overline{n} \in \mathbb{N}$ cannot result from profit-maximizing behavior.*

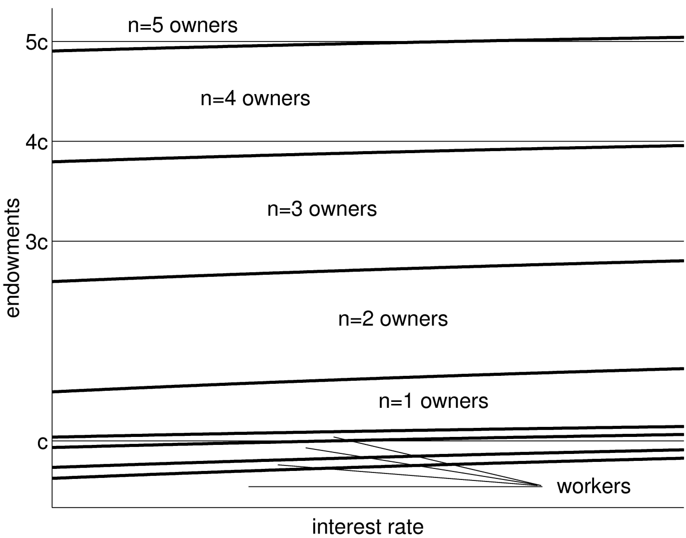

Figure 3.1: Ownership preferences depending on endowments and interest rate

Let us represent Lemmata 3.1 and 3.2 by Figure 3.1. It is possible to partition the endowment space into a finite number of subsets so that all endowment levels in each subset induce the same preference ordering. The support of the endowment distribution is on the vertical axis and the interest rate on the horizontal axis with the thin lines representing multiples of investment cost c. The bold lines correspond to cutoff values for endowments given r, so that the labeled areas denote the intervals of endowments at a given interest rate for which the labeled action is optimal. For workers the bold lines represent indifference between owning one asset in a symmetric $n = 1$ firm and working in some $n > 1$ firm. The support of the endowment distribution is on the vertical axis and the interest rate on the horizontal axis with the thin lines representing multiples of investment cost c. This describes a class society of agents as found in the work by Evans and Jovanovic (1989) and Banerjee and Newman (1993).

3.2.8 Side Payments

Let $t(i, N)$ denote individual i's side payment in coalition N. We allow only for balanced side payments within a coalition, that is $\sum_{i \in N} t(i, N) = 0$ for all coalitions N.[9] Side payments are subject to the capital market imperfection as well. This introduces a crucial non-transferability of utility into the matching model that will prevent the allocation from generally maximizing aggregate utility. Then the individual valuation function $v_i(.)$ from being member of firm (N, θ) incorporating individual side payments $t(i, N)$ can be written as

$$v_i(n, \theta, \omega(i), r, t(i, N)) = \begin{cases} \pi_i(n, \theta) + (1+r)(\omega(i) + t(i, N) - |\theta(i)|c) \\ \quad \text{if } \omega(i) \geq |\theta(i)|c + t(i, N) \\ \pi_i(n, \theta) + (1+i)(\omega(i) + t(i, N) - |\theta(i)|c) \\ \quad \text{if } \omega(i) < |\theta(i)|c + t(i, N). \end{cases} \tag{3.2}$$

To put this into context with figure 3.1, note that side payments may shift upwards or downwards the curves indicating indifference between different investment plans.

[9] We follow the convention of interpreting $t(.)$ as the payments from workers to owners which may be negative, of course.

3.2.9 Capital Market

Investment in assets may be financed by individual endowments or by loans taken on the capital market. Endowments not used for assets may be lent on the capital market. The capital market is assumed to be imperfect such that there is an interest rate spread between lending interest rate $1+r$ and borrowing interest rate $1+i \equiv (1+r)\gamma$, $\gamma > 1$.[10] The lending interest rate is endogenously defined by market clearing. We assume there exists a costless storage technology providing a rate of return of 0 which will provide a lower bound for the equilibrium interest rate.

The excess demand for loans is given by:

$$L^D(\theta, r) = \int_I [|\theta(i)|c - \omega(i)]di.$$

Agent i's preferred ownership choice $\theta(i)$ depends on r and thus excess demand for loans does so, too. However, preferred ownership choices also depend on the side payments. Side payments are balanced within coalitions, so they do not enter excess demand. Due to feasibility of ownership right allocations for all $n > 0$ coalitions investment plans involve exactly $n + 1$ assets. Whenever aggregate endowments are scarce the equilibrium interest rate is determined by inducing a sufficiently large measure of agents to weakly prefer their outside option to any role in any equilibrium firm. In addition, these agents have to prefer their outside option to anything they could possibly get in any non-equilibrium firm at side payments that induce any agent to own this particular firm.

To illustrate this let us look more closely at a symmetric firm and suppose side payments are 0 for the moment. For an agent with endowments $\omega(i) < c$ not to prefer to join a symmetric firm it must hold that

$$(1 + r^*)\omega(i) \geq \pi_2 + (1 + r^*)\gamma(\omega(i) - c).$$

[10]This is a simple and straightforward way to model imperfect credit markets that has frequently been used in the literature. It may be generated by the presence of moral hazard on the borrower's side, see for instance Galor and Zeira (1993) or chapter 2 of this dissertation for a similar setup.

Define the endowment level such that an individual is indifferent between the outside option and being member of a symmetric firm, $\hat{\omega}$, accordingly by

$$\hat{\omega} = \frac{1}{\gamma - 1}(\gamma c - \frac{\pi_1}{1+r}). \tag{3.3}$$

Using Lemma 3.1 and continuity of $\omega(i)$ the measure of agents preferring their outside option to the symmetric firm, μ_U, is given by

$$\hat{\mu} = \mu(i \in I : \omega(i) \leq \hat{\omega}),$$

which is continuous and strictly increasing in r on $(\underline{\omega}, c)$. Returning to the general case let $w = \int_I \omega(i)di$ denote aggregate endowments. Define the cutoff endowment level *omega_U* such that the measure of agents poorer than ω_U equals the difference between the capital stock necessary to provide every agent with an asset and aggregate endowments implicitly by

$$\mu(\omega(i) \leq \omega_U) = \max\{c - w; 0\}.$$

Finally, let $\hat{t}(n)$ denote the side payment that makes the richest unmatched agent with wealth ω_U exactly indifferent between the outside option and working in an n firm. Then the interest rate is determined by the following set of weak inequalities:

$$\hat{r} : \begin{cases} \hat{\mu}(r) \geq c - w, \text{ and} \\ (1+r)\omega(i) \geq v_i(n, \emptyset, \omega(i), r, t^*) \,\forall\, n : \\ \qquad \mu_n > 0, \omega(i) \leq \omega_U, \text{ and} \\ \forall\, n' \leq \overline{n} \text{ s.t. } \mu_{n'} = 0 \;\exists\, n \text{ s.t. } \mu_n > 0 : \\ \qquad v_i(n, A_n, \overline{\omega}, r, t^*) \geq v_i(n', A_{n'}, \overline{\omega}, r, \hat{t}(n')) \end{cases} \tag{3.4}$$

with at least one expression holding with equality. That is, agents with endowments less than ω_U must prefer their outside option to being member of a symmetric firm, to working in an equilibrium firm at equilibrium side payments, and to bribe the richest agent in the economy into a non-equilibrium firm. Otherwise the equilibrium interest rate is zero because of the costless storage technology. Formally, the equilibrium interest rate is given by

$$r^* = \begin{cases} \hat{r} \text{ s.t. } (3.4) & \text{if } w < c \\ 0 & \text{if } w \geq c. \end{cases} \tag{3.5}$$

3.2.10 Equilibrium Concept

Formally, a non-transferable utility matching equilibrium with limited side payments similar to Legros and Newman (2004a) and Legros and Newman (1996) is defined as follows

Definition 3.2 *A matching equilibrium with limited side payments denoted by $(P^*, \theta^*, r^*, t^*)$ is a measure consistent partition P^* of the agent space, a mapping θ^* of individuals to their assets, an interest rate r^* and transfers t^* such that*

- *condition (3.5) holds (credit market clearing),*
- *for all $P_i \in P^*$ it holds that $\sum_{j \in P_i} t_j^* = 0$ for all $i \in I$ (within coalition transfers),*
- *$\nexists P_i' \in \mathcal{F}(I)$ such that $v_i(|P_i'|, \theta', \omega(i), r^*, t_i') > v_i(|P_i^*|, \theta^*, \omega(i), r^*, t_i^*) \; \forall \, i \in P_i', \, i \in P_i^*$ and feasible $\theta' \in \Theta(P_i')$ such that $\sum_{j \in P_i'} t_j' = 0$ (stability).*

Measure consistency intuitively requires that the measure of first members of any coalition must equate the measure of the second members which must equate the measure of third members etc.[11] This equilibrium concept postulates that there exists no blocking coalition which is feasible with respect to the distribution of ownership rights and aggregate endowments and makes every member strictly better off using side payments. It coincides with the f-core with limited side payments (see Kaneko and Wooders, 1996).

[11] The reader is referred to Kaneko and Wooders (1986) for the formal definition and extensive discussion.

3.2.11 Feasibility and Labor Demand

Stability and measure consistency of a matching equilibrium translate into a condition equating supply and demand for labor for each firm size. That is, given equilibrium payoffs for each firm size n the measure of agents weakly preferring to be owners of such firms must equate n times the measure of agents weakly preferring to be workers in such firms. Let us first define the measures μ_{n_M} by

$$\begin{aligned}\mu_{n_M} &= \mu(\{i : v_i(n, \theta, \omega(i), r, t(i)) \geq v_i(n', \theta', \omega(i), r, t'(i)) \,, \\ &\quad \forall\, (n', \theta') \neq (n, \theta)\}), \end{aligned} \tag{3.6}$$

and μ_{n_W} by

$$\begin{aligned}\mu_{n_W} &= \mu(\{j : \omega(i) \geq \omega_U, v_j(n, \theta, \omega(j), r, t(j)) \geq \\ &\quad v_j(n', \theta', \omega(j), r, t'(j)) \,, \forall\, (n', \theta') \neq (n, \theta)\}). \end{aligned} \tag{3.7}$$

Monetary side payments $t(.)$, $t'(.)$ are those associated to the respective coalition and ownership rights combination with positive measure in equilibrium.[12] These expressions give the measures μ_{n_M} and μ_{n_W} of all agents weakly preferring to be owners of n firms to all other ownership rights and other firm sizes, and, respectively, to be workers in an n firm.

Let μ_n denote the measure of n firms in equilibrium. Feasibility requires aggregate investment demand to equate aggregate endowments whenever the latter are not abundant

$$\sum_{n=1}^{n(r^*)} (n+1)\mu_n c \leq w. \tag{3.8}$$

[12]Strictly speaking, side payments in firm types of measure zero in equilibrium need not be defined. However, due to monotonicity found in Lemma 3.1 and the fact that firms consist only of workers and owners, if there exist side payments $t(n)$ for a firm size n such that $\mu_{n_M} = \mu_{n_W} = 0$, then this firm size must have measure zero, ie. there exist no side payments making both workers and owners in n firms better off than some other firm size with positive measure given equilibrium side payments. Therefore we can assign these side payments $t(n)$ to firm types of measure zero as shadow prices which will be quite useful when simulating.

In equilibrium for all firm sizes $n > 0$ it must hold that

$$\mu_n \leq \mu_{n_M} \text{ and } n\mu_n \leq \mu_{n_W}. \tag{3.9}$$

This means that in this model side payments are used to equate labor supply and demand for any firm size thus functioning similar to market prices in general equilibrium models. It is noteworthy though, that in equilibrium a positive measure of agents typically will be indifferent between activity in several labor markets.

3.2.12 First Best Benchmark

Concluding this section, we will provide an efficiency benchmark in order to enable comparison with later results. In a first best world where everything is verifiable coalition size should be expected to maximize the marginal increase in net output of an additional worker subject to efficient human capital investments. This is – given the assumption on the production function – equivalent to maximizing average output per coalition member. Hence, in a first best world coalition size would be equal to K throughout the economy provided sufficient aggregate endowments.

3.3 Existence

As a first step let us sum up some important preliminaries that will hold for any equilibrium.

Lemma 3.3 *Monetary side payments workers pay to owners within firms depend only on the size of the firm and decrease in firm size for all firm sizes that have positive measure in equilibrium. If aggregate endowments are scarce, side payments are positive for all firm sizes.*

Proof: In Appendix.

The finding that side payments depend only on firm sizes means that side payments function in fact like market wages for labor in a firm of certain size and do not depend on individual wealth. That bigger firms pay higher wages seems to be consistent with much of the empirical findings, yet some caution is appropriate. Side payments only constitute part of a worker's payoff, in addition they receive the payoff from intra-firm renegotiations of the joint profit.

Moreover, Lemma 3.3 implies that in equilibrium at least for scarce aggregate endowments the poorest agents match into the largest firms requiring the smallest investment in form of side payments. On the other hand, richer agents match into firms that require more investment, that is into firms with more assets, as can be verified by comparing investment costs for an n and an $n+1$ firm and using Lemma 3.3 and the fact that differences in workers' side payments can never be higher than the cost of an asset. Hence, matching is negative assortative.

Now let us verify that the matching equilibrium indeed exists and induces a unique distribution of firm sizes. The full proof is somewhat involved and therefore refined to the appendix but a sketch of it can be given here. First we show that there exists an r^* not dependent on side payments that induces credit market clearing. Given r^*, existence of a matching equilibrium follows using a well known result by Kaneko and Wooders (1996). Then we establish uniqueness of equilibrium payoffs which follows almost by definition of the equilibrium. Finally, it is shown by contradiction that uniqueness of payoffs pins down coalition size and allocation of ownership rights in almost all coalitions.

Proposition 3.1 *A matching equilibrium with limited side payments exists and induces a unique distribution of firm sizes.*

Proof: In Appendix.

To illustrate the matching equilibrium a brief informal discussion of two polar examples follows. Consider first an example with perfect equality in the economy. Assume there is an almost degenerate distribution of endowments so that $\omega(i) = c$ and that $K = 1$.[13] This

[13]In this case side payments are zero. However, for $K > 1$ the efficiency result carries over only if K firms are sufficiently efficient compared to other firm sizes.

means any agent has sufficient endowments to prefer owning assets to not owning any and being a worker in a large firm. It follows immediately that the model economy will be populated entirely by small, efficient firms: there will emerge only $n = 1$ coalitions with symmetric asset ownership. This can be interpreted as an economy consisting exclusively of self-employed individuals collaborating via markets.

The second example focuses on an extreme case of endowment inequality and severe credit market imperfections. Let there be a large measure of endowment-less individuals, and a comparatively small measure of rich individuals owning all the endowments in the economy. This is not covered by the assumption of continuous differentiability of $\omega(i)$, however, an equilibrium may exist nevertheless. To be more concrete, assume that the measure of poor individuals exceeds the one of the super-rich more than twofold. Then there will emerge only large firms in the economy and all assets are owned by rich agents, and all poor individuals are workers.

3.4 Firm Size and Endowment Distribution

Having established that a matching equilibrium indeed exists in our model and induces a unique distribution of firm sizes, we are interested in its properties. The examples at the end of the last section only describe polar cases of wealth distributions. In this section we will therefore examine the properties of the matching equilibrium more generally.

Key to the matching equilibrium analysis is that all agents can secure themselves at least the payoff from being member of a symmetric firm, that is $v_i(1, \theta, \omega(i), r)$ where $|\theta(i)| = 1$, or the outside option 0, whichever is higher. This is because all agents have the opportunity to choose symmetric $n = 1$ coalitions and do not need agents from other endowment classes to form their desired coalition. This means there are only agents willing to be workers if the poorest agent has sufficiently bad outside options, that is the cost of capital for the poorest agent is sufficiently high. Equally, there will only be owners if the richest agent has sufficiently bad outside options, that is the cost of capital for the richest agent is sufficiently low. Moreover, poor agents having high cost of capital choose low investment occupations whereas rich agents facing low cost of capital choose high investment occupations. This means large firms will only emerge in equilibrium if there is sufficient heterogeneity in the agents' cost of capital, i. e. there is sufficient endowment inequality.

3.4.1 Properties of the Matching Equilibrium

This subsection is concerned with some preliminary properties of the matching equilibrium, especially with equilibrium side payments. These largely determine the size distribution of firms and thus whether a variety of large firms emerges in equilibrium, and if so under what circumstances. As the endowment distribution becomes more equal and agents' bargaining positions become more equal the firm size distribution may converge to a point measure at the efficient size K. However, this may happen only if the capital market imperfections are not too severe. If this is not the case the firm size distribution converges to a point measure at the minimal size $n = 1$. For instance, with perfect credit markets transferability of utility between coalition members is perfect. Then, of course, the f-core converges to coalitions that maximize average output, namely of size K. Intuitively, as the endowment distribution becomes more equal, coalition members not owning assets are richer and therefore less dependent on the inefficient capital market when transferring utility. The following proposition subsumes some interesting properties of the equilibrium allocation.

Proposition 3.2 *Properties of the matching equilibrium allocation*

(i) When capital is scarce, a necessary condition for only efficient firms to emerge in equilibrium whenever $K > 2$ is that both capital market imperfections are not too severe and average pro-

ductivity in the efficient firm is sufficiently high compared to all other firm sizes.

(ii) *When capital is scarce and the spread between upper and lower bound of the endowment distribution sufficiently large, both* $1 < n' < K$ *and* $n > K$ *firms have positive measure in equilibrium if, in line with Assumption A3.1, average output in* n *firms is sufficiently high compared to* n' *firms and average output in* n *and* n' *firms is sufficiently high compared to all other firm sizes, and capital market imperfections are sufficiently severe.*

Proof: In Appendix.

Part (i) of the proposition gives conditions for the firm size distribution to be degenerate at the efficient firm size. Not surprisingly, the set of necessary condition relaxes if average productivity in K firms is high compared to all other firm sizes. However, it is possible for any technology to find capital market imperfections sufficiently severe such that an equilibrium with only efficiently sized firms cannot occur for any distribution of endowments. Part (ii) states then the desired result of the emergence of both too small and too large firms in equilibrium which demands sufficiently imperfect capital and a relatively high average output in these firms compared to other firm sizes, as long as covered by Assumption A3.1.

3.4.2 Large Firms and Mean Preserving Spreads

The next step is to analyze whether a redistribution of initial endowments can affect technical efficiency in the economy. To this end we determine the effects of a mean preserving spread in the endowment distribution on the size distribution of firms. As probability mass moves from the center of the endowment distribution to its tails, both the measures of prospective owners of and of workers in large firms increase. This in turn leads to an increase of the measures of large firms at the cost of small firms and moves the equilibrium firm size distribution to the right. The following proposition captures this insight but first a technical definition is required.

Definition 3.3 *Let F, G be endowment distribution functions with mean ω_M. G is said to a redistribution from owners to workers from F satisfying the single crossing property if for some $\hat{\omega}$ such that $\hat{n} \equiv |\theta(i)| \neq 0$ for $\omega_i = \hat{\omega}$ in equilibrium under F*

$$F(\omega) \geq G(\omega) \forall \omega < \hat{\omega} \text{ and } F(\omega) \leq G(\omega) \forall \omega > \hat{\omega}$$

holds.

A transition from F to G describes a redistribution of endowments from owners to workers such that owners of $n > \hat{n}$ firms transfers endowments to agents that hold less ownership than $\hat{n}$.[14]

Proposition 3.3 *Let the endowment distribution G be a redistribution from owners to workers from the the endowment distribution F as in Definition 3.3. Let G reduce the measure of agents willing to work in $\hat{n}$ firms under endowment distribution F. In an economy with endowment distribution G the measure of $n > \hat{n}$ firms cannot be greater than in an economy with endowment distribution F. In particular, the the largest firm size $n_{max} > \hat{n}$ in the economy with F cannot be smaller than the largest firm size in the economy with G.*

Proof: In Appendix.

Proposition 3.3 states that suitable redistributions of initial endowments from owners to workers in fact prevents the formation of inefficiently large firms. A reshuffle of measure from the tails of the endowment distribution to the center simply leaves less agents willing to participate in large firms either as workers or as owners. More generally, it follows that the measure of large firms and thus the skewness of the firm size distribution depend on the kurtosis of the wealth distribution. The higher the kurtosis of the endowment distribution the more large firms emerge and the more is the firm size distribution skewed to the right.

On the other hand, Proposition 3.3 tells an important story about the dynamics of inequality in the model economy. Whenever endowments are scarce, a decrease in endowment inequality can lead to

[14]Note that F must be a mean preserving spread of G.

a decrease in the measure of large firms suggesting a more equal distribution of income. Therefore the model is consistent with a non-monotonic evolution of inequality as long as the economy has not grown enough for endowments to be abundant. To characterize the effect of changes in inequality on the size distribution of firms is the aim of the next proposition. A change inducing scarcity of workers (owners) for all firm sizes is typically associated with less (more) inequality as relatively less (more) agents are poor after the distributional change. A special case of such a change in the endowment distribution is a *uniform scarcity of workers (owners)*, that is labor demand exceeds (falls short of) labor supply by the same measure for all firms.

Proposition 3.4 *Assume a unimodal endowment distribution. Let a redistribution from owners (workers) to workers (owners) of the endowment distribution induce a uniform scarcity of workers (owners) given old side payments and the new interest rate. Then all side payments decrease (increase), however side payments in smaller firms increase (decrease) relative to side payments in larger firms.*

Proof: In Appendix.

This proposition tells primarily a story about the evolution of side payments as inequality in endowments changes in a certain way. But it contains a more general insight, in that side payments in small firms are affected relatively less by certain changes in the endowment distribution than those in larger firms. A particular intriguing corollary of Proposition 3.4 is that as inequality decreases sufficiently inducing a uniform scarcity of workers, the firm size distribution converges to a point measure at $n = K$ provided the capital imperfection is not overly severe, see Proposition 3.2, and otherwise to a point measure at $n = 1$.

It has to be emphasized that more equality in the limit tends to be associated with higher aggregate output because output efficiency is not related directly to credit market imperfections. One could argue, for instance, that a move towards more equally distributed initial endowments increases efficiency because an imperfect credit market that wastes resources is needed less for the proper allocation of endowments. Although this line of reasoning applies to *payoffs*, which increase as credit market activity declines, aggregate *output* is only influenced by the friction in the capital market through changes in the equilibrium allocation. On the other hand, the effect on aggregate output of a move towards more equality in the distribution of endowments is ambiguous. Depending on parameters, the resulting shift in the firm size distribution towards larger firms can rise the measure of efficient K firms sufficiently to offset the distorting effect on firms of other sizes.

3.4.3 Heterogeneity of Large Firms

As we have just shown, the measure of large firms that emerge in equilibrium depends on the range and the kurtosis of the endowment distribution. Now we are interested in the properties of the distribution of $n > 1$ firms and which circumstances favor heterogeneity in those large firms, that is the variance of the firm size distribution. Equilibrium firm sizes are determined by the conflicting interests of poor agents and rich agents. Poor agents wish to work in small firms whereas rich agents are likely to prefer greater firm sizes. More exactly, an agent's preferred firm size increases in wealth. This means large firms will be more heterogenous with respect to size the scarcer potential large firm owners are and the more heterogenous these are, that is the more right-skewed the endowment distribution is.

Proposition 3.5 *A necessary condition for a firm of size $n > 1$ to emerge in equilibrium is that (i) the spread between $\underline{\omega}$ and $\overline{\omega}$ is sufficiently large and endowments are abundant, or (ii) endowments are scarce and $\overline{\omega}$ is sufficiently large.*

Proof: In Appendix.

This proposition states effectively that economies having wealth distributions with larger support, that is a larger spread between lower and upper bound of the endowment distribution, $\underline{\omega}$ and $\overline{\omega}$, tend to have more inefficient and therefore larger firms in equilibrium. Whenever endowments are scarce, firms tend to be larger for unequal endowment distributions, that is for those with high skewness and a Paretian right tail. Then the interest rate is high, but endowments of the wealthiest unmatched agent, ω_U, are relatively small due to the large measure of agents in the left tail.

This indicates that the endowment level of the richest individual affects the size distribution of firms beyond the result on mean preserving spreads in the last subsection. Connecting this with Proposition 3.4 implies that the size of the largest firms tends to increase with a change in the endowment distribution such that the gap between rich and poor widens and owners become scarce. The latter is attributable to an increase in skewness of the endowment distribution. Thus skewness appears to favor a heterogenous firm landscape in the sense that there are large firms of different sizes.

3.4.4 The Income Distribution

The income distribution in equilibrium is key to the dynamics of the firm size distribution, as it determines next period's endowment distribution. Consequently, it is of particular interest whether the distribution of income is more or less unequal than the endowment distribution in terms of Propositions 3.3 to 3.5. Note that an agent's income is given by $v_i(n^*, \theta^*, \omega(i), r^*)$. Suppose for the moment that $\underline{\omega} > 0$ in order to properly define the endowment gap for agents i and j, $\omega(i) > \omega(j)$, as $\frac{\omega(i)}{\omega(j)}$. Define the income gap between the same agents i and j likewise as $\frac{v_i(.)}{v_j(.)}$. The following proposition characterizes the link from endowment to income distribution.

Proposition 3.6 *An agent's income is weakly increasing in endowments. Moreover, the income gap between the richest and the poorest agent in the economy exceeds the endowment gap if the poorest agent remains unmatched in equilibrium. However, the endowment gap between the richest and the poorest agent matched into some firm of size n exceeds the income gap.*

Proof: In Appendix.

Proposition 3.6 provides a valuable insight. For capital scarce economies income inequality exceeds endowment inequality between unmatched agents and agents matched into firms but *not* between workers and owners. This finding suggests that a dynamic version of the model will be compatible with a non-monotonic relationship of inequality and growth. On the one hand the income distribution of those agents matched is less unequal than the endowment distribution of those agents incorporating a trickle-down effect of growth. On the other hand the trickle-down effect does not extend to those agents remaining unmatched creating greater income inequality than endowment inequality between agents employed or owning and those agents remaining in autarky. As the capital stock grows less agents remain in autarky and the trickle-down effect dominates. Moreover, there is a dual economy flavor to the dynamics of this model. Whether the economy is able to transform into an economy of firms from an economy consisting of agents living in economic autarky then depends largely on the economies' ability to generate capital stock growth and thus on agents' optimal saving policies.

3.5 Application

3.5.1 Bimodal Size Distributions

In this section we consider the extent to which the model is able to capture empirical findings on firm size distributions in different countries. Firm size distributions in developing countries are typically characterized by the missing middle (see e.g. Tybout, 2000). That is, the share of the work force employed in intermediately sized firms is significantly less than both the share of those employed in small or large firms. A sufficient condition for this is that the size distribution of firms is bimodal with one peak to left and the other one to the right of the mean. In contrast, developed economies typically have single peaked distributions of workforce per firm size with the modus to the left of the mean. Suppose that the endowment density is single peaked and strictly increasing to the left of its peak while strictly decreasing to the right of it.

Proposition 3.4 already suggests how a bimodal size distribution of firms may emerge. Starting from a relatively equal distribution, in the sense that it is almost degenerate at its mean, and reshuffling measure to induce mean preserving spreads, that yield a uniform scarcity of owners, makes side payments in smaller firms decrease relative to those in large firms. For a sufficiently large reshuffle of measure and sufficiently severe capital market imperfections this will induce the measure of some n firm to exceed the measure of some n' firm with $n > n' > 1$ as larger firms become more attractive to potential owners in equilibrium. Given that there still remains sufficient probability mass around $\omega(i) = c$ the measure of symmetric firms exceeds that of smaller $n > 1$ firms and the size distribution of firms is bimodal. On the other hand, necessary conditions for a bimodal distribution are given in the following (heuristic) proposition.

Proposition 3.7 *Necessary conditions for a bimodal firm size distribution in equilibrium are sufficient skewness of the endowment distribution and (i) for the left peak sufficient average output of some small firm compared to larger firms or sufficient mass of the endowment distribution around the mean, or (ii) for the right peak a Paretian right tail or sufficient average output of some large firm compared to smaller firms.*

Proof: In Appendix.

This means that for a bimodal size distribution to prevail either some $n > 2$ firm must be very efficient at least locally, that means compared to the next smaller sized firm, or side payments in that n firm must be particularly high which can only be the case if owners for this firm are scarce and thus the endowment distribution sufficiently skewed. This reflects the simple fact that there must be both enough agents weakly preferring to work in such a firm and enough agents willing to own them. Depending on which market side is relatively scarce, side payments in this firm type must be extreme - either high or low. For the left peak to emerge sufficient endowment measure around c and a sufficiently high degree of capital market imperfections is needed as well. The intuition for this is that the bargaining inefficiency increases with firm size as long as the capital market is not perfect.

In case of particularly high side payments for some $n > 2$ firm, we know from Proposition 3.4 that side payments increase faster for larger firms as the measure of agents willing to work exceeds the measure of owners proportionately for any large firm. This means that the general level of side payments has to be sufficiently high in the economy which is only possible if the endowment distribution is very skewed and there are many poor and few rich agents and capital market imperfections are sufficiently severe. An illustration of this case follows in Figures 3.2 and 3.3.

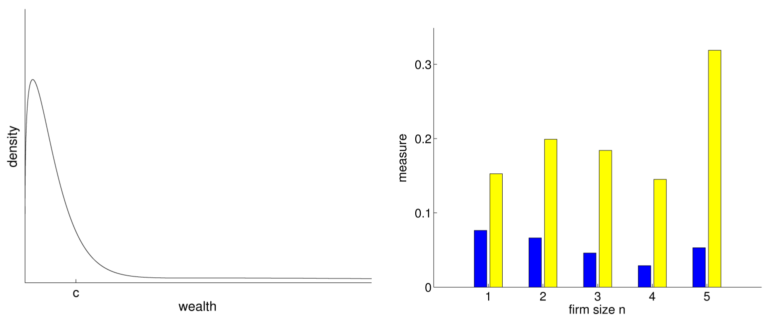

Figure 3.2: Wealth distribution leading to an inverse firm size distribution

Figure 3.2 depicts an endowment distribution that results in a missing middle of firm sizes. The mode is situated far to the left, thereafter the density is a decreasing convex function becoming relatively flat near the mean in order to put sufficient probability mass on the endowment levels where agents prefer to own assets. The latter suggests necessity of a Paretian tail. The firm size distribution represented by the darker bars has a left peak at symmetric $n = 1$ firms and a right peak at $n = 5$ firms. The lighter bars represent the overall workforce allocated to each size category. It is notable

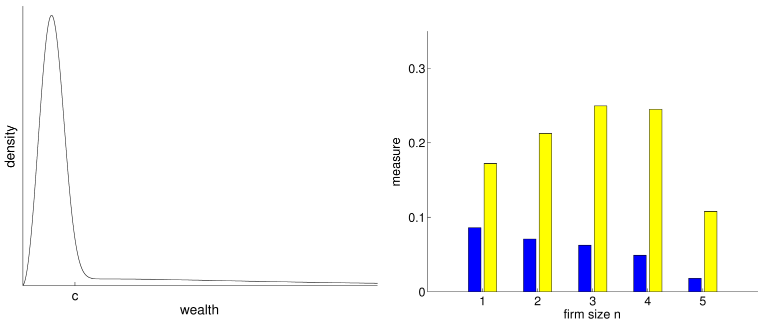

Figure 3.3: Wealth distribution leading to a normal firm size distribution

that the efficient firm size in this case is $K = 3$ and the renegotiation payoffs are determined by intra-firm bargaining (Stole and Zwiebel, 1996).

Now suppose the mode shifts to the right while preserving the mean as depicted in Figure 3.3. As a consequence, probability mass is shifted away from the tails which may result in a loss of the right tail's Paretian properties. Then less agents want to own or work in large firms and the size density moves slightly to the left. A move from a very unequal endowment distribution to a more equal one transforms a bimodal firm size distribution into a single-peaked one.

3.5.2 Inequality and Efficiency

Given an economy has a bimodal firm size distribution in equilibrium what are the implications for the efficiency of the allocation? Unfortunately, they appear to be ambiguous at best as a bimodal firm size distribution in this model may reflect either a very inefficient allocation with the intermediate firm sizes being efficient, as in the example in figures 3.2 and 3.3, or a more efficient allocation with one of the peaks on the efficient size. At the heart of this lies the fact that intra-firm bargaining leads to an unequal distribution of profits within firms which can only be compensated ex ante by side payments using the imperfect capital market. Hence, it depends on the capital market imperfection which firm size turns out to be the most attractive one given that side payments must compensate workers for intra-firm bargaining.

As Proposition 3.2 suggests, a high degree of capital market imperfections precludes the formation of only efficient $K > 1$ firms even when the endowment distribution is degenerate. In this case inequality in endowments can serve to bypass an improperly functioning capital market and therefore more inequality in the sense of skewness and kurtosis may increase aggregate income (see e.g. Grüner and Schils (2002) for a similar finding). Whenever capital market imperfections are not too severe, an unequal distribution of firm profits ex post can successfully be compensated by side payments. In this case a move towards more equality in the endowment distribution drives down the level of side payments required to equate labor supply and demand, so that in the limit only efficient firms emerge.

3.5.3 Increasing Aggregate Endowments

Efficiency and the shape of the firm size distribution also depend on the aggregate wealth of the model economy. This insight lies at the heart of the discussion on desirability of foreign aid and foreign direct investments. This subsection provides examples of the effects of both in our framework. If endowments are scarce, Proposition 3.2 states that side payments are positive for all firm sizes. This implies immediately that for capital market clearing to hold the poorest agents remain unmatched as they need to borrow to pay the side payments. Scarcity of endowments may also be associated with a scarcity of owners. In this case the firm size distribution is largely determined by the firm size choices of rich agents. Adding wealth into this economy does not necessarily lead to a more efficient allocation or an increase in aggregate income. If despite the increase in aggregate endowments workers are still sufficiently abundant, the efficiency of the allocation depends mainly on the distribution of the additional wealth.

To illustrate this idea let us conduct a thought experiment by dropping additional capital on the economy. Formally, we shift the whole endowment density to the right and increase skewness. Thus aggregate endowments increase (by approximately three percent in this case) while holding the cost of an asset constant at c. The efficient firm size is again $K = 3$ in this example. Details of the parametrization can be found in the appendix. Endowment densities are depicted to the left in Figure 3.4 and the corresponding workforce distributions across firms to the right. Note that the bars represent measures of agents being member of a firm of the respective size. The dashed line represents the initial endowment distribution and the endowment distribution after the increase in wealth is depicted by the solid line. Workforce distributions are drawn to the right in Figure 3.4 with the darker bars corresponding to the equilibrium before the increase in wealth.

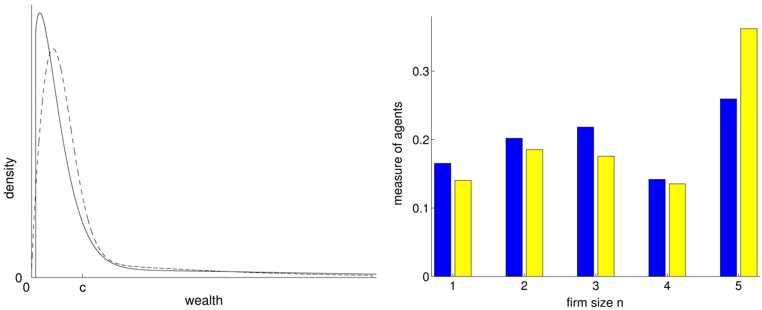

Figure 3.4: An increase in wealth leading to a decrease in income

Given the initial endowment distribution, workers are relatively abundant and owners relatively scarce. As aggregate wealth increases sufficiently, this might reverse. However, if aggregate wealth increases by a moderate amount, workers remain abundant whereas the measure of owners increases and thus the measure of efficient K firms may decrease. In the case of Figure 3.4 more agents become members of inefficiently large $n = 5$ firms. This is favored by the rightward shuffle of a substantial amount of probability mass among the poorly endowed and by the fact that the interest rate has to decrease.

Of course, the effect in the example is of transitory nature only.

If endowments of the poorest agent increase sufficiently so do equilibrium outside options, notably the payoff from being member of a symmetric firm. This drives down side payments and ultimately moves the firm size distribution towards a single peaked one. Whether the modus sits at the efficient firm size or at symmetric firms depends on the severity of the capital market imperfection and on relative productivity as in Proposition 3.2.

To achieve an increase in aggregate income or even productivity in this particular case, aggregate endowments must increase sufficiently or the additional endowments must be distributed appropriately. This means here that the capital injection should induce receivers of capital to own K firms. For economies with scarce endowments and a large efficient firm size, a uniform increase of endowments tends primarily to increase the measure of small and large firms. Only when poor agents become sufficiently scarce may the general equilibrium effect on side payments facilitate the formation of large efficient firms.

Now suppose that we add a small measure of very rich agents to an economy where workers are abundant and side payments are high, as in the example. Given that the additional endowments are sufficiently few then workers may still be abundant for all firm sizes, in the sense of Proposition 3.4. In this case very rich agents face high side payments and find it optimal to set up large firms. Then, as in Figure 3.4, the right tail of the size distribution tends to gain measure. If the efficient firm size is small but greater than 1 this decreases productivity and may decrease even aggregate output. This can be interpreted as an influx of foreign direct investment adding a measure of agents facing foreign capital cost and not being subject to the local credit market imperfections. Therefore economies with high wealth inequality and scarce endowments need not benefit from foreign direct investments unless the amount of investment is sufficiently large.

3.6 Discussion and Conclusion

Allowing for partial and joint ownership, i.e. dropping Assumption 3.2, results primarily in adding another dimension of complexity to notation and analysis of equilibrium allocation. Joint ownership of the factory, that is asset A^I, could be accommodated relatively easily, changing the size of the symmetric firm size to the efficient one and thus replacing symmetric $n = 1$ firms by symmetric $n = K$ firms. A natural application of this is the emergence of cooperatives, for instance. In this case the firm size distribution converges always to the efficient one when capital market imperfections or endowment inequality vanish and effects on efficiency in the previous section become less ambiguous. Otherwise the results carry over.

By introducing partial ownership another set of cutoff endowment levels emerges that allows a finer coarsening of the type space. Therefore the main effect of an extension of the model would be to create a new investment opportunity with cost situated between those of $n = 1$ and $n = 2$ firms. Asset investments that induce this cost to agents involve agents owning one or two assets thus qualifying for $n = 1$ or $n = 2$ firms. While there may emerge some shifting of equilibrium measure between these firm types the general property of the model that poorer agents choose less and richer agents more ownership of assets does not change. As it is exactly this feature that drives the missing middle result in section 6, we conjecture that distributional results will not be affected significantly by admitting partial ownership. As owners of only type II assets too have an incentive of increasing firm size beyond the efficient one (see Gall, 2004) the scope for changing the central finding that both too large and too small firms emerge in equilibrium appears to be quite limited.

We are able to present an example where partial ownership can be fully admitted by capping firm size at $n = 3$ assets A^{II} in order to keep the model numerically solvable. The results of a numerical simulation of this model are shown in figure 3.5.[15] The lighter bars show the distribution of the workforce and the darker bars the distri-

[15] The details of the simulation are provided in the appendix.

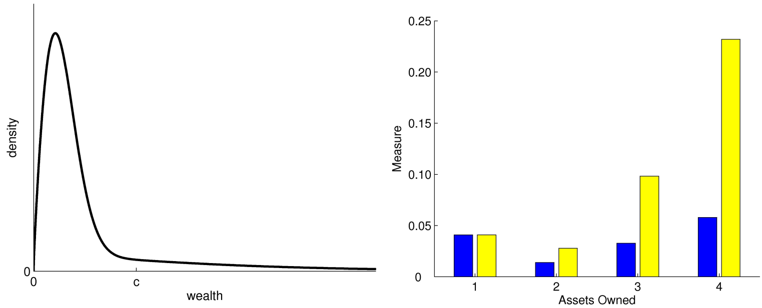

Figure 3.5: An example with partial ownership

bution of firms according to their size. Parameters are chosen such that the efficient coalition size is $K = 2$, that is a coalition with two assets. Indeed, both too small and too large productive coalitions emerge in equilibrium and there appears a missing middle of the size distribution of business firms.

The view that A^I and A^{II} in this model resemble productive non-human assets as a factory building and machines is by no means exclusive. It is consistent with the model to let these assets stand for e.g. plants thus giving an indication inefficient vertical integration and empire building. This points to an interesting reinterpretation of the model: for instance asset A^I might be a brand and asset A^{II} a plant. Then endowment inequality determines the variety of goods produced in an economy. In the examples in section 5 more unequal endowment distributions translate into less variety of goods consistent with empirical facts.

An extension of this work could analyze the effects of the development of contractual and labor institutions, for instance by introducing profit sharing contracts. Here arises a neat connection to the work of Williamson (2000) proposing a theory of developing institutions where secure property rights emerge before contract enforcement. In this sense this work aims to contribute to an institutional explanation of economic development.

How do our findings fit into the big picture of development economics? We already mentioned that within developing economies the distribution of labor force among firm sizes is frequently found to be bimodal, and in that way to lack labor in intermediately sized firms. This feature is not known for industrialized economies. Our model is able to explain this empirical fact as bimodal firm size distributions emerge for very skewed endowment distributions with a Paretian right tail. As endowment inequality decreases and the modus shifts to the right the firm size distribution becomes uni-modal. Moreover, under intra-firm bargaining an increase in endowments does not necessarily lead to an increase of production, especially if endowments are scarce, the poor abundant and the additional endowments are not distributed exactly as to induce additional demand for ownership of efficient firms. Moreover, the model generates a wedge in income of unmatched agents and members of firms thus creating a potential for class societies reminiscent of the dual economy. This suggests that the model provides an adequate instrument for policy analysis of the industrial sector in developing countries.

Chapter 4

On the Role of Markets in the Hold-up Problem

4.1 Introduction

It is a common perception that economic agents face the correct incentives for their choice of actions if only there exist competitive markets and corresponding prices for the consequences of their actions. In particular, it has been argued (Cole et al., 2001, Felli and Roberts, 2002) that the well-known hold-up problem (see e.g. Klein et al., 1978, Williamson, 1975) can be ameliorated, if not solved, by the presence of markets. In the canonical example agents undertake non-contractible complementary investments prior to an opportunity to jointly produce output. At the production stage investments are sunk and agents may have the incentive to renege on their contractual obligation. Renegotiation-proof contracts may then not fully internalize the social benefits of their investments. However, competition on markets ex post may lead to prices capturing the social return of effort, thus inducing first best incentives at the investment stage.

However, in practice only rarely can utility be expected to be transferred perfectly between agents on markets. The presence of e.g. imperfect credit markets as in chapter 3 of this dissertation, contracting constraints, or social norms limits the ability of agents to transfer utility to their co-workers or partners ex ante, when contracting in the matching market. For instance, a spread between borrowing and lending interest induces a wedge in the opportunity cost of capital between rich and poor agents thus precluding a one-to-one transfer of utility. If there is limited liability in an economy, making side payments in the matching market face tightens the incentive constraints of agents subject to moral hazard. Social norms postulating the exchange of gifts rather than payments may drive a wedge between agents' opportunity costs, thus preventing perfect transfer of utility.

In such a framework the intuition that competition sets correct incentives is incomplete. Whenever markets are characterized by less than perfectly transferable utility, market prices no longer have to coincide with the full social benefit of agents' actions. For instance, as Becker (1973) pointed out, non-transferabilities may change the matching pattern compared to the efficient allocation. More recently, results characterizing conditions on the circumstances that lead to inefficient patterns are provided in the work of Legros and Newman (2002, 2004b). Then the desirability of markets as a solution to the hold-up problem may weaken considerably. To test this conjecture we analyze two settings. On the one hand, we let markets open before an investment choice affecting agents' characteristics is made – corresponding to the classical hold-up problem. And on the other hand, we let markets open after agents have made the investment choice. As it turns out, the presence of markets ex post does not generally increase output in an economy, but may be harmful. This result continues to hold if we allow markets to open both ex ante and ex post.

Intuitively, market imperfections that induce non-transferabilities of utility have a twofold effect on the market allocation. On the one hand, they may render the matching pattern inefficient, and on the other hand, they typically induce inefficient incentives for the investment ex ante. Indeed the market allocation may be distorted sufficiently, so that an ex post matching market is no longer desirable both from an individual and a social point of view. It has to be emphasized that our results are caused solely by deterministic effects of non-transferabilities of utility and do not hinge on multiplicity of equilibria and coordination failure as in Mailath et al. (2004).

To provide an example with an immediate application, consider teamwork production where prior to production agents have an investment opportunity to acquire education, thereby augmenting their human capital. Suppose further that human capital positively affects an agent's productivity and that individual inputs to production are strategic substitutes. Efficiency then requires negative assortative matching to maximize aggregate output. Permitting non-transferabilities in form of cost sharing contracts as the only means of utility transfer between agents within teams changes the picture. If markets open after the investment opportunity, matching is positive assortative. The intuition is that in order to be an attractive match, agents need to be able to compensate their matches. In this example, compensation is limited to cost sharing and an agent's human capital. In order to match with a highly productive type, agents have to be sufficiently attractive, meaning that they have to be sufficiently productive themselves. This sets education investment incentives too high, resulting in over-investment in human capital within coalitions. In contrast, if agents match prior to investments, investment incentives are determined by a coordination game within the firm. Then investment incentives are given by the game's payoffs – which may depend on the matching equilibrium payoffs. Despite of usually too low investment incentives compared to the first best investment level, an allocation under ex ante matching may generate higher output and even Pareto dominate an allocation under ex post matching.

The example appears to be consistent with the well-documented observation of over-education within firms (see e.g. the survey by Hartog, 2000), that is the presence of workers having acquired more education than needed for the task they are employed at. A more recent study (Cockx and Dejemeppe, 2002) focuses on the competition for jobs among the unemployed. They find clear evidence of a crowding out of the less educated by the more educated for jobs that are reported to require comparatively low education. This clearly lends some support to the relevance of positive assortative matching in an ex-post labor market. Our result of distorted investment incentives ex ante provides a caveat for prolonging the duration of compulsory schooling. Indeed, empirical studies on the effect of education spells on growth are as plenty as ambiguous (see e.g. the survey by Krueger and Lindahl, 2001). There are some studies finding a negative relationship between duration of education and growth (e.g. Barro and Lee, 1993, Benhabib and Spiegel, 1994), but do not seem overly robust. However, Krueger and Lindahl (2001) find some evidence of an inverted U-shaped relationship, such that education is a significantly positive factor for growth only in the countries with the lowest level of education. This appears to be consistent with our model.

Taking one step further, the optimality of market timing may reverse as market imperfections vanish, thus providing a motivation for the development of market institutions. Moving both markets closer to perfectly transferable utility, the ex post market equilibrium may switch to negative assortative matching, thus erasing the matching inefficiency yet retaining higher investment incentives than ex ante markets.

To put this into context with literature, in particular the one on property rights (Grossman and Hart, 1986, Hart and Moore, 1990), let us emphasize that by no means investments have to be relationship specific as in the property rights literature, while, of course, specificity of investments can be easily accommodated within our framework. Quite the reverse is true, exclusive business relations may arise as a result of inefficient markets. It is therefore possible to interpret our results within the theory of the firm, where the firm emerges to shelter its members' investments from markets.

In the literature on the hold-up problem two recent contributions generate first best investment incentives by assuming markets with perfectly transferable utility. Cole et al. (2001) examine matching equilibria under transferable utility with continua of agents and two-sided investment in complementary attributes. They find inefficiencies due to the possibility of multiple matching equilibria which arise if the optimal investment choice function has discontinuities. However, an agent's equilibrium investments are constrained efficient given the coalition this agent is matched into in equilibrium. The optimal allocation is found to be always in the equilibrium set, and for sufficiently large attribute spaces it is indeed the only equilibrium. Felli and Roberts (2002) analyze the hold-up problem with a finite number of agents. They get efficiency for one-sided investments in attributes. Here, too, investments are constrained efficient with respect to the respective coalition. The intuition is that firms are auctioned off to workers who receive the full marginal benefit while firms receive marginal benefits equivalent to full compensation for the next smaller attribute firm. However, uniqueness of the efficient equilibrium depends crucially on the fact that attributes of agents from the other market side are sufficiently close to each other. Thus, when choosing their attribute, agents have an incentive to choose the optimal investment within the equilibrium match. Peters and Siow (2002) consider a two sided matching model with two-sided investments in a single attribute where utility is strictly non-transferable. In contrast to the contributions cited above which assume supermodularity of the joint production function, agents' attributes are perfect substitutes. This and symmetry in the type distributions lead to efficiency for large economies. However, for finite economies the under-investment problem is ameliorated but efficiency need not be achieved.

Mailath and Postlewaite (2004) present an application of non-transferable utility matching and the theorem of the second best and consider different degrees of non-transferability of utility. An economy may move from a regime with strictly non-transferable utility to one where limited side payments are possible. After the regime change, the matching exhibits no longer positive assortative matching (see the results by Legros and Newman, 2002, 2004b, for a more general analysis). This is constrained efficient given the joint production technology. Due to the absence of insurance and income uncertainty, production efficiency of the allocation may be higher under the regime where some limited means of utility transfer within coalitions is possible. However, their paper does not consider ex ante investments in attributes.

The chapter will proceed by introducing the framework and some technical preliminaries in Section 2. Section 3 provides a simple model of human capital acquisition and labor-sharing contracts. In section 4 we present a more general differentiable model of non-transferable utility matching allowing us to state an impossibility result in Section 5. Section 6 concludes.

4.2 The Framework

4.2.1 Agents

Throughout the chapter we will consider a world populated by a continuum of agents living on A and B, both sets endowed with equal Lebesgue measure. That is, the analysis in this chapter focuses on two-sided matching, although most of our results should carry over to one-sided matching models as well. Agents are characterized by their *type* θ_i, describing the characteristics they are born with. An agent's type may determine for instance cost of productive activity, but also captures an agent's ability to transfer utility to another agent. In addition to an agent's type, an agent also has an *attribute* x_i. In contrast to the type, the attribute may be changed by an agent at some stage, for instance by investing in human capital and thus increasing productivity at a cost determined by the type. An agent's investment in his attribute is non-contractible. Agents derive utility from consumption which depends only on output and possible side payments and is assumed to be additively separable in consumption and side payments.

4.2.2 Coalitions

We focus exclusively on coalitions of pairs of agents $C = (i \in A, j \in B)$ consisting of one agent from each market side. Within a coalition there exists an opportunity of joint production. However, due to the non-transferability, output may not be transferred at a rate of 1 within a coalition. The *utility possibility frontier* of a coalition of agents a and b denotes the Pareto frontier of all attainable combinations of both agents' utilities within a coalition. In particular, the utility possibility frontier of a coalition of agents a and b will be represented by a non-negative real-valued function $\phi(a, b, v)$ denoting a's maximum payoff when b obtains payoff $v \in \mathbb{R}^0_+$.[1]

This means $\phi(.)$ represents the Pareto frontier of feasible distributions of the coalition's payoff between coalition members. In general, joint payoffs can be redistributed between members only by side payments at the matching stage or by appropriately choosing actions in the subgame within a coalition in case the extensive form allows for this. It follows that under perfectly transferable utility the utility possibility frontier always has the slope of -1 whereas this holds no longer for the case of non-transferable utility. An example for $\phi(a, b, v)$ under non-transferabilities is depicted in figure 4.1.

In the graph agents' payoffs without side payments from being member of a coalition (a, b) given their attributes are denoted by U and V. Agent a has inferior means of transferring utility compared to agent b as the utility a has to spend to give b one unit is greater than the utility b has to give up in order to increase a's utility by one unit. Two properties of ϕ are of special interest for our analysis and are given in attribute notation as follows.

Definition 4.1 *$\phi(.)$ is said to have generalized increasing differences (GID) if for all $a, b \in A$ and $c, d \in B$ with $x_a > x_b$ and $x_c > x_d$*

$$\phi(c, a, \phi(a, d, v)) \geq \phi(c, b, \phi(b, d, v))$$

for all $v \in [0, \bar{v}]$ with $\bar{v}$ defined implicitly by $\phi(b, d, \bar{v}) = 0$.
$\phi(.)$ is said to have generalized decreasing differences (GDD) if for all $a, b \in A$ and $c, d \in B$

$$\phi(c, a, \phi(a, d, v)) \leq \phi(c, b, \phi(b, d, v))$$

for all $v \in [0, \bar{v}]$.

[1] Here we slightly abuse notation. $\phi(a, b, v)$ will typically depend either on both agents' types θ_a and θ_b in the case of matching prior to attribute choice, or on both agents' attributes x_a and x_b. However, we use an agent's identity as an abbreviation for the relevant variable.

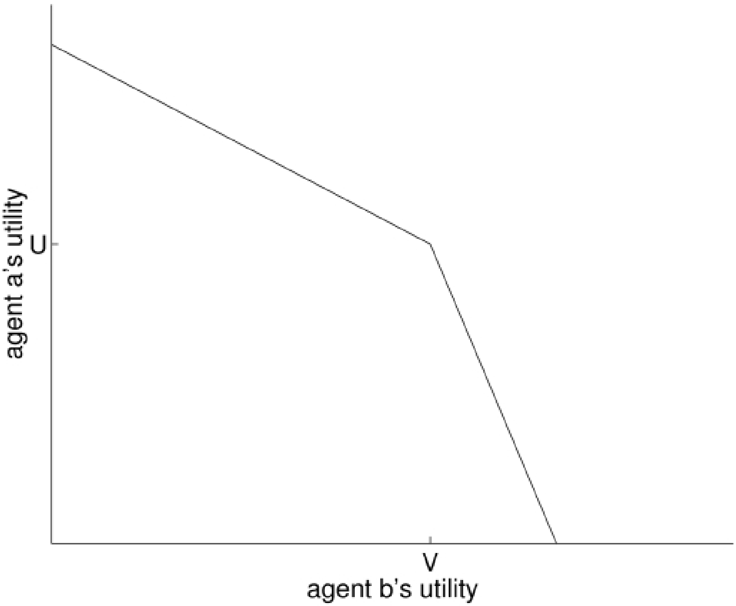

Figure 4.1: Utility possibility frontier at the matching stage

As shown in Legros and Newman (2004a), these conditions determine the matching pattern in equilibrium.

4.2.3 Matching Equilibrium

Coalition formation occurs by matching, and the solution concept is a matching equilibrium. We admit side payments between coalition members albeit of a limited form governed by the agents' transferability type.

Definition 4.2 *A matching equilibrium is a one-to-one mapping $\mathcal{M}$: $A \rightarrow B$ that assigns any $a \in A$ to some $b \in B$ and payoffs $\{u_i\}_{i \in A,B}$ such that*

- *$\mathcal{M}$ is measure-consistent,*

- $u_a = \phi(a, b, u_b)$ *for all* $a \in A$ *and* $b \in B$,
- *there does not exist* $a \in A$ *and* $b \neq \mathcal{M}(a) \in B$ *with* $u_a < \phi(a, b, v)$ *and* $u_b < v$ *for all* $v \in [0, \bar{v}]$ *with* $\bar{v}$ *defined implicitly by* $\phi(b, d, \bar{v}) = 0$.

Note that by definition the matching equilibrium coincides with the f-core which has the virtue of existence under relatively mild conditions (see Kaneko and Wooders, 1986, 1996). It is notable at this point that an agent's attribute and the degree of utility transferability act as substitutes with respect to an agent's attractiveness towards other agents in an ex post market. For instance, an agent from market side A may be indifferent between matching with an agent b, who has high attribute and low transferability of utility, and an agent b' with low attribute but excellent means of transferring utility. This is the only source of distortions in this matching model.

Of special interest are monotone matching patterns such as *positive assortative matching* (PAM) which is said to hold if the equilibrium matching function $\mu(.)$ has the following property:

$$x_a \geq x_b \Leftrightarrow x_{\mu(a)} \geq x_{\mu(b)} \; \forall \; a, b \in A,$$

for matching taking place after attribute choice. Likewise, *negative assortative matching* (NAM) holds for $\mu(.)$ such that

$$x_a \leq x_b \Leftrightarrow x_{\mu(a)} \geq x_{\mu(b)} \; \forall \; a, b \in A.$$

Legros and Newman (2004a) show that generalized increasing differences of the utility possibility frontier implies positive assortative matching independently of the distribution of types. Correspondingly, generalized decreasing differences induce negative assortative matching in equilibrium.

4.2.4 The Timing of Markets

A central issue of this chapter will be to compare efficiency of two different regimes of market activity. Under one regime, markets will be allowed to open before agents invest in their attributes, a case we will refer to as *ex ante markets*. Under another regime, markets will be allowed to open after investments in attributes have been chosen by agents, a case that will be referred to as *ex post markets*.[2] That means the sequence of events is given by

(i) Nature chooses agents' types

(ii) Matching markets may or may not open ex ante

(iii) Agents decide on their attribute investment

(iv) Matching markets may or may not open ex post

4.2.5 A Numerical Example

Let us look now at a simple numerical example in the spirit of the one in Becker (1973) where constraints on contracting render matching markets outcomes inefficient. In particular, we assume complete non-transferability of utility, so that markets both ex ante and ex post exhibit positive assortative matching despite the substitutability of investments in the production function. However, ex post markets in this example generate over-investment whereas ex ante markets do not. Agents are born without physical endowment and there is no capital market. Although agents are free to contract on future output, when deviating from splitting the surplus this incurs an efficiency loss which we impose on the model. In line with incomplete contracts theory this efficiency loss is assumed to be sufficiently severe as to generate completely non-transferable utility.

The economy is populated by two market sides represented by two continua A and B of measure 1. Agents' types consist of the investment cost to increase their attribute, θ_i, and of their lack of ability to transfer utility. Assume there are only two cost types, $1 < \underline{\theta} < \frac{3}{2}$ and $\overline{\theta} > 3$, and the measure of agents on market side A with $\underline{\theta}$, α, is equal to the one on market side B. Let $\alpha \in (0, 1)$. Agents have

[2] This is equivalent to a regime where markets open both before and after attribute investments, but commitment of agents to remain in the coalition chosen ex ante is excluded.

the possibility to form partnerships consisting of two members, one from each market side. Depending on whether agents have invested in their attribute (I) or not (NI) the payoffs in a coalition (a, b) are given by

	I	NI
I	$4, 4$	$3.5, 3.5$
NI	$3.5, 3.5$	$2, 2$

That is, payoffs within coalition are split equally and agents have no means to deviate from that distribution nor to make side payments when forming a coalition in the market. Moreover, payoffs have decreasing returns to the human capital investment. This implies immediately that efficient allocations are characterized by negative assortative matching.

Ex post market

We are analyzing two different regimes. First assume there is a matching market after agents invest in their attributes, but not before they invested. We already know that, due to the structure of payoffs, for sufficiently severe non-transferable utility positive assortative matching will obtain. A low cost agent a who invests, either matches with an agent b who has invested leading to a payoff of $4 - \underline{\theta}$, or matches with an agent b who has not invested, giving agent a a payoff of $3.5 - \underline{\theta}$. Hence, low cost agents a always invest whereas high cost agents never do.[3] The same reasoning applies to agents b, so that under our assumption of completely non-transferable utility all low cost agents invest before the matching market, whereas high cost agents never do. This means under the second regime measure α of coalitions consist of two low cost agents who both invest and measure $1 - \alpha$ of coalitions consist of two high cost agents who do not invest.

[3]However, there is the possibility of asymmetric equilibria where low cost agents on one market side expect low cost agents on the other market side to invest. These asymmetric equilibria are, however, an artefact of the two-sided matching and disappear when matching is one-sided.

Ex ante market

Let us now look at the ex ante matching market when the ex post market does not open. Then within a coalition agents split the surplus. In this case an agent a invests if $4 - \theta_a > 3.5$ given his match b invests and if $3.5 - \theta_a > 2$ given his match does not invest. This means, within coalitions a low cost agent a invests if agent b does not and vice versa. High cost agents never invest. Assume for the sake of simplicity that agents have access to a correlated randomization device when deciding on investments in attributes. Then expected payoffs of two low cost agents in a coalition are $3.5 - \frac{1}{2}\underline{\theta}, 3.5 - \frac{1}{2}\underline{\theta}$. Given the payoffs from the investment game within the coalition the payoffs in the ex ante matching market are given by the following matrix.

	$\underline{\theta}$	$\overline{\theta}$
$\underline{\theta}$	$3.5 - \frac{1}{2}\underline{\theta}, 3.5 - \frac{1}{2}\underline{\theta}$	$3.5 - \underline{\theta}, 3.5$
$\overline{\theta}$	$3.5, 3.5 - \underline{\theta}$	$2, 2$

Given that the means of transferring utility is sufficiently inefficient the matching will exhibit positive assortative matching in equilibrium. This pins down an allocation where measure α of equilibrium coalitions consists of low cost agents with exactly one agent investing and measure $(1-\alpha)$ of coalitions consists of high cost agents that do not invest.

Discussion

Calculating the difference in aggregate output between ex ante and ex post markets yields $\alpha(\underline{\theta}-1)$ which is positive under our assumptions. This means that aggregate output is greater when markets open prior to the investment only. Moreover, ex ante markets Pareto dominate ex post markets from an ex ante perspective in this example as low cost agents' expected payoffs are strictly higher in ex ante markets whereas high cost agents' payoffs are exactly the same under both regimes.

Hence, in this example competition ex post reduces efficiency. The intuition is that in order to become a more attractive match in the market, an agent has an incentive to invest in the attribute whenever this is cheaper than transferring utility. This holds quite generally for matching markets with non-transferable utility and the efficiency result can be expected to carry over to other cases where efficient matching is not positive assortative.

Introducing transferability in this example leads to interesting dynamics (see Gall et al. (2005) for a more thorough analysis of this example). Assume that agents have access to a perfect means of utility transfer, called liquidity. However, the maximum amount of utility they are able to transfer is bounded above by $L > 0$. Then, for $L < \frac{1}{2}$ the analysis remains unchanged for sufficiently high $\overline{\theta}$. For $\frac{1}{2} < L < \frac{\underline{\theta}}{2}$ agents of type $\overline{\theta}$ in the ex post market are able to compensate agents of type $\underline{\theta}$ and negative assortative matching emerges, whereas the results for the ex ante market continue to hold for sufficiently high $\overline{\theta}$. In this case, aggregate output with ex post markets exceeds aggregate output with ex ante markets. Thus the constrained efficient timing of markets is reversed. For $L > \frac{\underline{\theta}}{2}$ both markets yield negative assortative matching and efficient investments.[4]

The following illustration provides some reason to believe in the relevance of this last finding. Suppose that the non-transferabilities of utility are caused by capital market imperfections. Agents may have to finance side payments in the matching market by loans which are subject to moral hazard. Absent non-pecuniary punishments this drives a wedge between the borrowing and the deposit rate (see e.g. Stiglitz and Weiss, 1981). The development of credit market institutions may then be represented by the introduction of liquidity in the market. As a consequence, the optimal timing of markets switches from ex ante to ex post. An ex ante market in our terminology is equivalent to a commitment ex ante not to go on the market ex post, a commitment that can be solved by a firm, for instance. As market imperfections decline the value of this commitment becomes obsolete and productive activity increasingly relies on spot markets. Hence, our example may offer some insight into the raise and the fall of the

[4]For this result we need the correlated equilibrium in the investment game within a coalition of two $\underline{\theta}$ agents with ex ante markets.

corporation.

4.3 Cost-sharing and Educational Attainment

This section examines the decision of economic agents whether to invest in human capital in order to increase productivity. After the investment in human capital, agents pursue productive activities yielding output. Production occurs in teams of two agents and has the property that attribute investments are substitutes. In particular, it will never be efficient that both team members undertake the human capital investment. We analyze two possible regimes: one in which the labor market takes place prior to human capital investment and one in which the labor market opens after the human capital investment. This may be interpreted as different durations of compulsory schooling as by prolonging compulsory schooling agents have an additional opportunity to invest in human capital before they are allowed to enter the labor market. It turns out that prolonged compulsory schooling may indeed deteriorate aggregate output.

Suppose the economy consists of two market sides, each one populated by a continuum of agents of measure 1. The agents differ with respect to their type θ_i which can be high, $\overline{\theta}$, or low, $\underline{\theta} < \overline{\theta}$.[5] Both types of agents have measure $\frac{1}{2}$ each on both market sides. Agents have the opportunity to form teams in order to produce joint output. The production opportunity may be thought of as a local public good giving both agents the same payoff π from its existence. To produce the good a fixed amount of labor K is required, and inputs to production are perfect substitutes.[6]

[5] We follow the convention of denoting a high type by an upper bar which corresponds to low cost.

[6] Empirically, some support is lent to this assumption in a study by van Smoorenburg and van der Velden (2000) who find evidence that overeducated individuals are less likely to receive on-the-job training in the Netherlands, suggesting that there is some substitutability of educational skills acquired before entering a firm and those acquired within firms which supports the specification of our example.

We assume that only the division of labor within a team of agents is contractible. Moreover, agents have the opportunity to invest in their attributes prior to production. An agent who invests has the attribute $\overline{c}$, an agent who does not has $\underline{c} < \overline{c}$. The attributes can be interpreted as productivity, as they affect both unit labor cost and value of output. The investment in attributes comes at a cost, that is either high, $\underline{\theta}^{-1}$, or low, $\overline{\theta}^{-1}$. Finally, the attributes of agents within a team have some influence on the payoff they receive from the good:

$$\pi(\overline{c}, \underline{c}) - \pi(\underline{c}, \underline{c}) > \pi(\overline{c}, \overline{c}) - \pi(\overline{c}, \underline{c}) > \frac{\underline{c}}{\overline{c}}[\pi(\overline{c}, \underline{c}) - \pi(\underline{c}, \underline{c})], \tag{4.1}$$

and $\pi(.)$ is symmetric. That is, we assume decreasing returns to attributes in the payoff. An agent i has the utility function

$$u_i = \pi(c_i, c_{-i}) - \frac{x_i}{c_i} - k(c_i, \theta_i), \tag{4.2}$$

where $k(.) = \frac{1}{\theta_i}$ if $c_i = \overline{c}$ and $k(.) = 0$ otherwise. Given the fixed amount of labor required for the production of the good, $x_i = K - x_{-i}$. Finally, we need to state assumptions on the investment cost to generate the desired results. We assume that $\underline{\theta}$ is prohibitively low, i.e.

$$\frac{1}{\underline{\theta}} > \pi(\overline{c}, \overline{c}) - \pi(\underline{c}, \underline{c}). \tag{4.3}$$

The high type is assumed to have low investment cost, but not too low, that is

$$2\frac{\underline{c}}{\overline{c}}[\pi(\overline{c}, \underline{c}) - \pi(\underline{c}, \underline{c})] > \frac{1}{\overline{\theta}} > \pi(\overline{c}, \overline{c}) - \pi(\overline{c}, \underline{c}) + K\left(\frac{1}{\underline{c}} - \frac{1}{\overline{c}}\right). \tag{4.4}$$

Moreover, we assume that $\overline{c} \geq 2\underline{c}$ to save notation.

4.3.1 Ex post Market

Suppose for now that the matching market takes place after investments in attributes. In this case investment costs are sunk, and from (4.2) we can easily derive a utility possibility frontier for this market. First note that u_a in terms of u_b is given by

$$u_a(c_a, c_b, u_b) = \pi(c_a, c_b) - \frac{K - c_b}{(\pi(c_a, c_b) - u_b)} c_a.$$

A similar expression can be derived for $u_b(c_b, c_a, u_a)$. Both expressions define the utility possibility frontier $\phi(i, j, v) = u_i(c_i, c_j, v)$. Now we are in a position to check whether one of the generalized difference conditions applies to this case. Indeed, generalized increasing differences, that is in this case

$$u_b(\overline{c}, \overline{c}, u_a(\overline{c}, \underline{c}, v)) > u_b(\overline{c}, \underline{c}, u_a(\underline{c}, \underline{c}, v)) \, \forall v \in [0; \phi(\overline{c}, \underline{c}, 0)],$$

holds if

$$\pi(\overline{c}, \overline{c}) - \pi(\overline{c}, \underline{c}) > \frac{\underline{c}}{\overline{c}}[\pi(\overline{c}, \underline{c}) - \pi(\underline{c}, \underline{c})]. \tag{4.5}$$

This is assumed to be the case in (4.1). Therefore, in the ex post market positive assortative matching occurs. Figure 4.2 provides an illustration of the generalized increasing difference condition. It contains the utility possibility frontiers of heterogeneous coalitions (dashed lines) and homogeneous coalitions (solid lines). The arrows represent the generalized increasing difference condition.

Finally, we need to determine ex ante investment behavior of agents. Consider agent i's decision ex ante: i invests if

$$\pi(\overline{c}, c_{-i}) - \pi(\underline{c}, c'_{-i}) - \left(\frac{x}{\overline{c}} - \frac{x'}{\underline{c}}\right) > \frac{1}{\theta_i}.$$

c_{-i} and x, as well as their starred counterparts, denote expected attribute and labor share in a match conditional on the own attribute. Then in the matching equilibrium all $\overline{c}$ agents match with one another and all $\underline{c}$ do likewise. Hence, the attribute investment condition reduces to

$$\pi(\overline{c}, \overline{c}) - \pi(\underline{c}, \underline{c}) - \frac{K}{2}\left(\frac{1}{\overline{c}} - \frac{1}{\underline{c}}\right) > \frac{1}{\theta_i}.$$

This holds for $\theta_i = \overline{\theta}$, but not for $\theta_i = \underline{\theta}$ under assumptions (4.1) and (4.4), and (4.3). To invest is indeed a dominant strategy for the high type for all equilibrium labor shares as can be seen by assumption (4.4). Hence, with an ex post matching market, high type agents invest, whereas low type agents do not, and the matching is positive assortative.

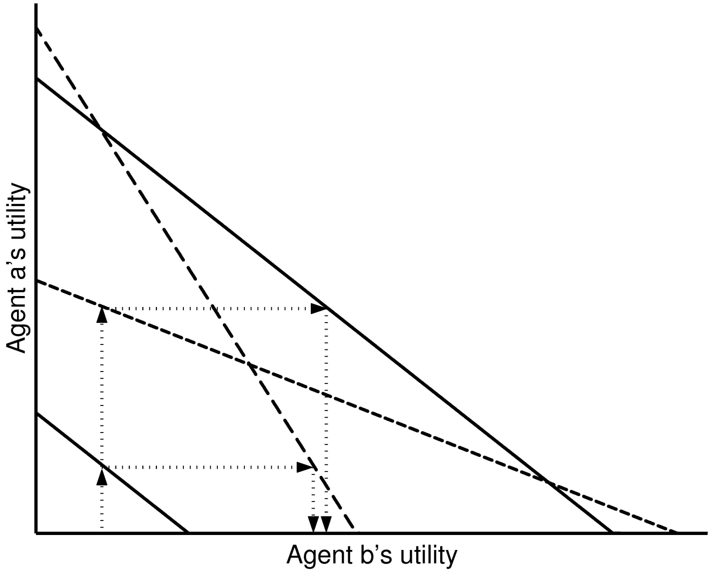

Figure 4.2: Generalized increasing differences in the ex post market

4.3.2 Ex ante Market

For the ex ante market case agents match before investing in attributes while being able to contract on the labor share in production. Within the team agents play a simultaneous coordination game by deciding on their attribute investment. Low type agents have sufficiently high cost so that not investing is a dominant strategy for them. High type agents' investment decisions depending on their match's type are determined by the following inequalities:

$$\begin{aligned} \pi(\bar{c}, \underline{c}) - \pi(\underline{c}, \underline{c}) - x_i \left(\frac{1}{\bar{c}} - \frac{1}{\underline{c}} \right) &> \frac{1}{\bar{\bar{\theta}}} \text{ and} \\ \pi(\bar{c}, \bar{c}) - \pi(\underline{c}, \bar{c}) - x_i \left(\frac{1}{\bar{c}} - \frac{1}{\underline{c}} \right) &< \frac{1}{\bar{\bar{\theta}}}. \end{aligned}$$

x_i is the labor share agent of i in the matching equilibrium. First suppose agent i matches with another high type agent. Then $x = \frac{K}{2}$ because of symmetric distributions. This means for

$$\pi(\overline{c},\underline{c}) - \pi(\underline{c},\underline{c}) - \frac{K}{2}\left(\frac{1}{\overline{c}} - \frac{1}{\underline{c}}\right) > \frac{1}{\overline{\overline{\theta}}} > \pi(\overline{c},\overline{c}) - \pi(\underline{c},\overline{c}) - \frac{K}{2}\left(\frac{1}{\overline{c}} - \frac{1}{\underline{c}}\right), \tag{4.6}$$

in the investment game among high type agents there exist three equilibria. To make this matching as attractive as possible, we assume that agents have access to a correlated randomization device so that the expected payoff from a high type match is

$$Eu_i(\overline{\theta},\overline{\theta}) = \pi(\overline{c},\underline{c}) - \frac{1}{2}\left(\frac{1}{\overline{\overline{\theta}}} + x_i\left(\frac{1}{\overline{c}} + \frac{1}{\underline{c}}\right)\right).$$

In the investment game in a heterogeneous team there is a unique equilibrium where the high type invests and the low type does not for all x_i if

$$\pi(\overline{c},\underline{c}) - \pi(\underline{c},\underline{c}) > \frac{1}{\overline{\overline{\theta}}} > \pi(\overline{c},\overline{c}) - \pi(\overline{c},\underline{c}) - K\left(\frac{1}{\overline{c}} - \frac{1}{\underline{c}}\right). \tag{4.7}$$

This complies with our assumptions determining outcome and payoffs in a heterogeneous match. For a coalition of two low type agents, the unique equilibrium is that neither invests. Using these results we can determine the utility possibility frontier for the ex ante matching market for each possible combination of types (θ_a, θ_b):

$$\begin{aligned}
Eu_a(\overline{\theta},\overline{\theta},u_b) &= 2\pi(\overline{c},\underline{c}) - \frac{1}{\overline{\overline{\theta}}} - \frac{K}{2}\left(\frac{1}{\overline{c}} + \frac{1}{\underline{c}}\right) - u_b, \\
u_a(\overline{\theta},\underline{\theta},u_b) &= \pi(\overline{c},\underline{c}) - \frac{1}{\overline{\overline{\theta}}} - \frac{1}{\overline{c}}(K - \underline{c}(\pi(\overline{c},\underline{c}) - u_b)), \\
u_a(\underline{\theta},\underline{\theta},u_b) &= 2\pi(\underline{c},\underline{c}) - \frac{K}{\underline{c}} - u_b.
\end{aligned} \tag{4.8}$$

Checking whether the ex ante matching has some generalized differences property, we find that indeed generalized decreasing differences, that is

$$u_b(\overline{\theta},\overline{\theta},u_a(\overline{\theta},\underline{\theta},v)) > u_b(\overline{\theta},\underline{\theta},u_a(\underline{\theta},\underline{\theta},v))\ \forall v \in [0;\phi(\overline{\theta},\underline{\theta},0)],$$

obtains if

$$2\frac{\underline{c}}{\overline{c}}[\pi(\overline{c},\underline{c}) - \pi(\underline{c},\underline{c})] + \frac{K}{2}\left(\frac{1}{\underline{c}} - \frac{1}{\overline{c}}\right) > \frac{1}{\overline{\theta}}. \tag{4.9}$$

This is implied by the assumption in (4.4).

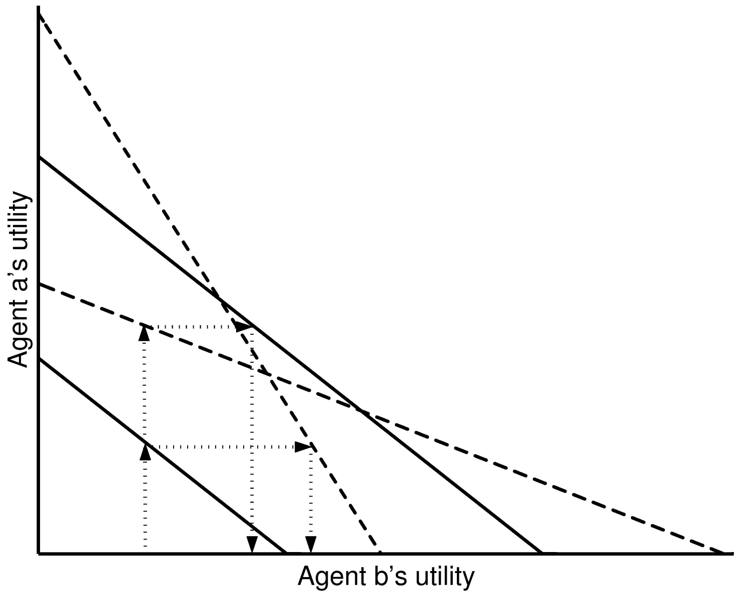

Figure 4.3: Generalized decreasing differences in the ex ante market

In this case ex ante matching is negative assortative with respect to agents' types θ_i. Then only heterogeneous coalitions are formed. The associated equilibrium in the investment subgame involves investment for the high type agents, but not for the low type agents. Figure 4.3 depicts the utility possibility frontiers of heterogeneous (dashed lines) and homogeneous coalitions (solid lines). In comparison to Figure 4.2 the utility possibility frontier of a homogenous high type coalition has contracted towards the origin, thus rendering this kind of coalition comparatively unattractive. Indeed, heterogeneous coalitions have become so attractive that generalized decreasing differences obtain as traced by the arrows. Let us sum up our findings in the following proposition.

Proposition 4.1 *Under Assumptions (4.1) - (4.4), ex ante markets induce negative assortative matching with respect to type and high type agents invest whereas low type agents do not. Ex post markets induce positive assortative matching with respect to attribute and high type agents invest whereas low type agents do not.*

To compare both regimes with respect to total surplus it suffices to compare the sums of the low and the high types' utilities in each regime. That is, ex ante markets generate higher surplus than ex post markets if

$$2\pi(\overline{c},\underline{c})-\frac{1}{\overline{\theta}}+x^{*}\left(\frac{1}{\underline{c}}-\frac{1}{\overline{c}}\right)-\frac{K}{\underline{c}}>\pi(\overline{c},\overline{c})+\pi(\underline{c},\underline{c})-\frac{1}{\overline{\theta}}+\frac{K}{2}\left(\frac{1}{\underline{c}}-\frac{1}{\overline{c}}\right) \tag{4.10}$$

where again x^* denotes the equilibrium labor share of a high type in the ex ante matching market. The condition weakens in x^*, so that a sufficient condition can be derived by setting $x^* = 0$. The so derived sufficient condition for (4.10) coincides with assumption (4.4). Hence, ex ante matching markets are more desirable in this example from a surplus point of view.

4.3.3 Discussion

In this example ex post matching generates an inefficient matching pattern and over-investment within coalitions. This is ameliorated when the matching market takes place before the investment in attributes. However, side payments in the ex ante market do not need to induce efficient cost sharing. In general, there is no reason to expect them to do so. The intuition driving the results in this example is that the cost of redundant investments can be partly internalized within a coalition in case of ex ante matching. In case of ex post matching redundant investments take place because they serve as a means of utility transfer.

Both the ex post matching market outcome and the incentive for over-investment in human capital given by better prospects in the labor markets has some empirical support in the finding of over-education. This describes a mismatch between the educational attainment requirement of a given task and the actual educational attainment of the worker employed in it (see e.g. Freeman, 1976). For the ex post market setting of our example there is overeducation in high type coalitions and undereducation in low type coalitions. This is indeed a well-documented feature of labor markets in European countries and the US (see the survey of Hartog, 2000). Cockx and Dejemeppe (2002) study the competition for jobs among the unemployed in Belgium. They find that higher education than formally required for a job significantly increases the hiring probability of an agent. Thus, overeducated individuals crowd out adequately educated individuals. Pischke and von Wachter (2005) find evidence that increasing compulsory schooling from 8 to 9 years in Germany had zero effect on earnings of those affected. It has to emphasized that professional formation plays a prominent role in Germany, in particular for those likely to leave school after 8 years, such that there existed ex ante markets. Thus a tentative explanation for these finding is that closing the ex ante markets after 8 years of schooling did not lead to the acquisition of education in the ninth year of schooling as those students affected by the change were prepared to into professional formation. It has to be emphasized that the results in our example are not at odds with empirical treatment studies finding positive private returns to additional schooling using natural experiments such as Duflo (2001). This is precisely what the example predicts. However, as Angrist (1995) noted, general equilibrium effects may overcompensate the positive returns among the treated.

4.3.4 Extensions

How is then the desired timing of markets affected by varying the extent of non-transferabilities in the market? In particular, will a decrease in the severity of market imperfections reverse the optimal market order? Let us introduce a means of transferring utility between coalition members at a one to one ratio and call it liquidity. To be precise, assume that each agent has the possibility of perfectly transferring up to b units of utility. For the ex post market the utility possibility frontier for heterogeneous matches is then given by

$$u_1(\overline{c}, \underline{c})(u_2) = \left(1 + \frac{\underline{c}}{\overline{c}}\right) \pi(\overline{c}, \underline{c}) + \left(1 - \frac{\underline{c}}{\overline{c}}\right) b - \frac{K}{\overline{c}} - \frac{\underline{c}}{\overline{c}} u_2.$$

To be exact, only a part of the utility possibility frontier is given by the equation, namely where u_2 is sufficiently low such that the utility transfer by means of liquidity has been used up, i.e. $u_2 \leq \pi(\overline{c}, \underline{c}) - b$. For $u_2 > \pi(\overline{c}, \underline{c}) - b$, the utility possibility frontier is given by $u_1(\overline{c}, \underline{c}, u_2) = 2\pi(\overline{c}, \underline{c}) - \frac{K}{\underline{c}} - u_2$. For homogeneous coalitions nothing changes. Calculating a sufficient condition for the generalized decreasing differences condition yields

$$\pi(\overline{c}, \overline{c}) - \pi(\overline{c}, \underline{c}) < \frac{\underline{c}}{\overline{c}}[\pi(\overline{c}, \underline{c}) - \pi(\underline{c}, \underline{c})] + \left(1 + \frac{\underline{c}}{\overline{c}}\right) b.$$

This condition strictly weakens in b. This means that for sufficiently transferable utility the ex post market yields the efficient matching pattern, namely negative assortative. Still high types invest and low types do not. The ex ante market yields negative assortative matching for all b and investment incentives within coalitions do not change. The only coalition where introducing liquidity has an impact is a heterogeneous match. There the utility possibility frontier changes into

$$u_1(\overline{\theta}, \underline{\theta})(u_2) = \left(1 + \frac{\underline{c}}{\overline{c}}\right) \pi(\overline{c}, \underline{c}) - \frac{1}{\overline{\theta}} - \frac{K}{\overline{c}} + \left(1 - \frac{\underline{c}}{\overline{c}}\right) b - \frac{\underline{c}}{\overline{c}} u_2.$$

Again, this expression is valid only for sufficiently low u_2, namely $u_2 \leq \pi(\overline{c}, \underline{c}) - b$. For homogeneous matches nothing changes, so that matching must remain negative assortative.

Effects of the introduction of liquidity on the desirability of ex ante versus ex post markets are ambiguous and depend on equilibrium payoffs of agents. However, whenever there is sufficient liquidity for the ex post market to yield negative assortative matching, both regimes reach efficiency in terms of investment incentives and matching pattern. The only remaining sources of inefficiencies are utility transfers by labor sharing. We already know that for sufficiently high equilibrium payoffs of the high cost agents, i.e. $u_{\underline{\theta}} > \pi(\overline{c}, \underline{c}) - b$, labor sharing within coalitions is efficient under both regimes. By strict substitutability of inputs, surplus in a coalition increases in the labor share of the low cost agent. One can show that, given b sufficiently small as to not induce first best labor sharing in all equilibria, there exist equilibria in ex ante markets where low cost agents' payoffs are smaller (and their labor share greater) than in any equilibrium with ex post markets. Hence, with equilibrium selection by Pareto dominance, ex ante markets remain more desirable than ex post markets even if both induce negative assortative matching. The intuition is simply that ex ante markets induce less inefficient compensation. For perfectly transferable utility both regimes yield first best allocations.

Suppose now again that there is no liquidity.[7] Instead of comparing two regimes one could let the agents choose whether to participate in ex ante or ex post markets. In order to avoid multiple rational expectations equilibria we assume that in the ex post market there is an influx of new agents, say measure $\epsilon > 0$ both of $\overline{c}$ and $\underline{c}$ agents. Then all rational expectations equilibria have the property that high type $\overline{\theta}$ agents invest and wait for the ex post market. This means that the existence of an ex post market where matching is positive assortative attracts the high types in this example. An application of this extension are foreign direct investments where multinationals co-operate with local firms. Cost-sharing contracts are frequently found in such co-operations. Local governments' decision whether to force investors into joint ventures reflects the timing of markets in our example. Our results suggest that intervention is indeed desirable to avoid an inefficient matching pattern, whenever investments are strategic substitutes within a coalition, e.g. when joint production faces severe coordination problems.

[7] Admitting liquidity in this setting and increasing b generates multiple equilibria as high types' payoffs in ex ante markets converge to their payoffs in ex post markets, in particular whenever b is sufficiently great to induce negative assortative matching in the ex post market.

4.4 A Differentiable NTU Matching Setup

In the remainder of this chapter we reason that ex post market equilibria with non-transferability cannot generally be expected to induce efficient ex ante investment incentives. This means there is little reason to expect that generally ex post markets will beat ex ante markets in terms of efficiency whenever utility is not fully transferable. To this end, we return to a more general setting. Agents may differ with respect to their types $(\gamma_i, \delta_i) \in \Gamma_i \times \mathcal{D}$, $i = a \in A, b \in B$, reflecting e.g. innate ability, and the degree of utility transferability which could represent wealth, for instance. For simplicity ability type spaces Γ_a and Γ_b are assumed to be compact subsets of $\mathbb{R}^2$. The transferability type of an agent $i \in K$, is $\delta_i \in \mathcal{D}^K$ where $K = A, B$. $\mathcal{D}^A$ is defined as follows.

Definition 4.3 *Define the transferability type space for market side A, $\mathcal{D}^A$, as a space of continuous, strictly decreasing functions $\delta : \mathbb{R} \times B \mapsto \mathbb{R}$ with the properties that*

- *The number of points $u \in \mathbb{R}$ where $\delta(u, b)$, $b \in B$ is not differentiable with respect to u is at most finite,*
- *$0 < \delta(u, b) \leq u$ for $u \in \mathbb{R}$ and $b \in B$, and*
- *$0 < \frac{\partial \delta(u,b)}{\partial u} \leq 1$ and $\frac{\partial \delta(u,b)}{\partial u} \geq \frac{\partial \delta(v,b)}{\partial v}$ for $u \leq v$ for all $b \in B$ and all points $u, v \in \mathbb{R}$ where $\delta(.)$ is differentiable.*

Define $\mathcal{D}^B$ likewise. Agents have the opportunity to invest in their attribute $x_i \in X$ where X is a compact subset of $\mathbb{R}_+$ at the beginning of their lives which comes at a utility cost $k(x_i, \gamma_i)$.

Assumption A 4.1 *The cost function $k(x_i, \gamma_i)$ is assumed to be convex and twice continuously differentiable.*

The agents' common utility function is assumed to be linear and to look like $u_i = g(x) + \tau(i, j) - k(x_i, \gamma_i)$ where $g(x)$ is a production function turning all team members' human capital into output, and $\tau(i, j)$ denotes a possible side payment that an agent i receives from his match j. Output of a coalition (a, b) depends on the human capital of the coalition members according to the production function $g(x)$.

Assumption A 4.2 *$g(x)$ is assumed to be twice continuously differentiable and strictly concave in each of its arguments, the Inada conditions are assumed to hold for each partial derivative, and $g(x)$ directly enters both agents' utilities.*

While the regularity conditions assumed on $g(x)$ appear to be fairly standard, the assumption that $g(x)$ represents each coalition member's utility from the coalition requires some justification. This assumption is tantamount to requiring efficient production at surplus splitting. The intuition behind the formalization of the problem is to abstract from using an explicit model of within coalition behavior, and instead to assume that outcomes of such within coalition behavior can be represented by $g(x)$ and transferability types δ_i. This means any explicit model of within coalition behavior that generates equilibrium payoffs with the regularity properties required by Definition 4.3 and efficient production at equal sharing of surplus is representable by our formalization. That is a matching game with non-transferable utility is described by $((\gamma_i, \delta_i)_{i\in A,B}, g(.), k(.))$.

Of particular interest is, of course, the link between the utility possibility frontier within a coalition and the degree of non-transferability the coalition members are subject to. Any matching (a, b) has the utility possibility frontier

$$\phi(a, b, v) = g(x_a, x_b) \left\{ \begin{array}{l} -\delta_a^{-1}(v - g(x_a, x_b), b) \text{ if } v > g(x_a, x_b) \\ +\delta_b(g(x_a, x_b) - v, a) \text{ if } v \leq g(x_a, x_b) \end{array} \right. \tag{4.11}$$

Note that the definition of $\mathcal{D}^K$ implies that all $\delta \in \mathcal{D}^K$ are invertible. As equilibrium concept we use again the one in Definition 4.2 for the matching stage. Investments have to maximize individual utility given the expected matching equilibrium. For expositional convenience we restate Definition 4.2 modified by explicitly referring to attribute investments which in turn determine utility possibility frontiers within coalitions. This is to emphasize that there are two separate issues of efficiency. Firstly, the matching pattern given attributes, i.e. the matching function $\mathcal{M}$, and, secondly, an agent's attribute investment given some expected equilibrium match.

Definition 4.4 *The allocation $((x_a, x_b)_{a \in A, b \in B}, \mathcal{M})$ where $\mathcal{M} : A \to B$ is a one-to-one mapping that assigns any $a \in A$ to some $b \in B$ and $x_a, x_b \in X$ are agents' attribute choices, and payoffs $\{u_i\}_{i \in A,B}$ is a rational expectations equilibrium of the non-transferable utility matching game $((\gamma_i, \delta_i)_{i \in A,B}, g(.), k(.))$ if*

- *$\mathcal{M}$ is measure-consistent,*
- *x_i is individually optimal given $(x_{-i}, \mathcal{M})$,*
- *$u_a = \phi(a, b, u_b)$ for all $a \in A$ and $b \in B$,*
- *there does not exist $a \in A$ and $b \neq \mathcal{M}(a) \in B$ with $u_a < \phi(a, b, v)$ and $u_b < v$.*

As a first step we are interested in determining under what circumstances ϕ preserves properties such as supermodularity of the joint production function $g(.)$. Thus it is possible to identify conditions for an efficient matching function $\mathcal{M}$ to emerge in equilibrium. We use the equilibrium concept of Definition 4.2 which exists under our assumptions (see e.g. Kaneko and Wooders, 1996). To this end we extend the work of Legros and Newman (2004a) by the following Proposition.

Proposition 4.2 *Suppose $g(x)$ has increasing differences in x and that for all $u \in \mathbb{R}$ both $\delta_a(u, i) \geq \delta_b(u, i)$ and $\delta_c(u, j) \geq \delta_d(u, j)$ for all $a, b, j \in A$ and $c, d, i \in B$ with attributes $x_a > x_b$ and $x_c > x_d$. Then positive assortative matching obtains in equilibrium and the matching allocation is efficient given attributes.*

The proof is straightforward and can be found in the appendix. It turns out that in order to obtain efficient matching, it suffices that agents with higher attributes also have weakly better means of transferring utility. The intuition of this proposition is quite simple. By making sure that an agent has better means of transferring utility the more able he is, it follows that an agent can always outbid less able agents but not more able ones. Hence, the relative ranking of any agent is preserved in the equilibrium matching. An extension of

Proposition 4.2 relating efficiency of the matching allocation to the distribution of transferability types for decreasing differences of $g(.)$ is straightforward.

In particular, Proposition 4.2 states that whenever all agents on the same market side have equal means of transferring utility within a coalition, the matching will be positive assortative for supermodular $g(.)$. This can be extended to submodular $g(.)$ that lead to negative assortative matching, so that efficient matching obtains in both cases. Moreover, this finding suggests that positive (negative) correlation of the degree of transferability with the agents' attributes is sufficient for efficient matching for supermodular (submodular) joint production functions $g(.)$.

4.5 Attribute Investments

Now we turn towards the agents' investments in attributes. As a benchmark, (constrained) optimal investment levels given a match of agents a and b are given by

$$\max_{x_a, x_b} 2g(x_a, x_b) - \sum_{i=a,b} k(x_i, \gamma_i).$$

Assume that this is concave problem. The optimization problem has the usual first order condition

$$k'(x_i^*, \gamma_i) = 2g'(x_a^*, x_b^*), \tag{4.12}$$

for $i = a, b$ where the starred variables are the optimal solution.

Returning to the case of ex-ante investments in attributes prior to matching, agents choose their investment levels based on the outcome of the matching stage. Now suppose that the second stage matching equilibrium is representable by a matching function $\mathcal{M} : A \mapsto B$.

Assumption A 4.3 *The matching function $\mathcal{M}(a)$ determines a continuously differentiable function $x_{\mathcal{M}}(x_a) : X \mapsto X$ of the attribute of the match of an agent $a \in A$ depending on a's attribute. $x_{\mathcal{M}}(x_a)$ is assumed to be invertible.*

We postulate a function rather than a correspondence as we focus on rational expectations equilibria. The function $\tau : A \times B \mapsto \mathbb{R}$ denotes equilibrium side payments from $a \in A$ to $b \in B$ for matches (a, b) in units of utility for the receiver. An agent a's optimization problem is then given by

$$\max_{x_a} g(x_a, x_{\mathcal{M}(a)}) + \tau(x_a, x_{\mathcal{M}(a)}) - k(x_a, \gamma_a).$$

Assume for the moment that this a concave problem as well and that $\tau(.)$ is differentiable with respect to its arguments. The first order necessary condition is given by

$$\begin{aligned} \frac{\partial k(x_a^*, \gamma_a)}{\partial x_a} &= \frac{\partial g(x_a^*, x_{\mathcal{M}(a)})}{\partial x_a} + \frac{\partial g(x_a^*, x_{\mathcal{M}(a)})}{\partial x_{\mathcal{M}(a)}} \frac{\partial x_{\mathcal{M}(a)}}{\partial x_a} \\ &\quad + \frac{\partial \tau(x_a, x_{\mathcal{M}(a)})}{\partial x_a} + \frac{\partial \tau(x_a, x_{\mathcal{M}(a)})}{\partial x_{\mathcal{M}(a)}} \frac{\partial x_{\mathcal{M}(a)}}{\partial x_a}. \end{aligned} \tag{4.13}$$

Note that the derivative of the side payment will depend on the change in the ability to transfer utility of a's match $\mathcal{M}(a)$ as well. Now the problem of analyzing efficiency reduces to comparing the respective first order conditions (4.12) and (4.13). To put this more simple ex-ante investments are constrained efficient if and only if

$$\begin{aligned} \frac{\partial g(x_a^*, x_{\mathcal{M}(a)})}{\partial x_a} &= \frac{\partial g(x_a^*, x_{\mathcal{M}(a)})}{\partial x_{\mathcal{M}(a)}} \frac{\partial x_{\mathcal{M}(a)}}{\partial x_a} + \frac{\partial \tau(x_a, x_{\mathcal{M}(a)})}{\partial x_a} \\ &\quad + \frac{\partial \tau(x_a, x_{\mathcal{M}(a)})}{\partial x_{\mathcal{M}(a)}} \frac{\partial x_{\mathcal{M}(a)}}{\partial x_a}. \end{aligned} \tag{4.14}$$

It is immediate from the stability condition of a matching equilibrium that the RHS of (4.14) is positive and thus investment incentives are greater than in the usual incomplete contracts case.

4.5.1 Constant Side Payments

Let us assume for now that equilibrium side payments are constant across agents which excludes the possibility of negative assortative matching. In particular, this case includes strictly non-transferable utility. Then ex ante investments are efficient if and only if for all a

$$\frac{\partial g(x_a^*, x_{\mathcal{M}(a)})}{\partial x_a} = \frac{\partial g(x_a^*, x_{\mathcal{M}(a)})}{\partial x_{\mathcal{M}(a)}} \frac{\partial x_{\mathcal{M}(a)}}{\partial x_a}. \tag{4.15}$$

This means that the direct effect of a change in x_a on joint profit must be equal to the indirect effect via a change in the equilibrium match $\mathcal{M}(a)$. It is immediate that this condition holds for example if the matching is positive assortative, $g(.)$ has unity elasticity of substitution, and attributes are distributed equally on both market sides as in Peters and Siow (2002). Rewriting (4.15) yields

$$\frac{g'_{x_a}(x_a^*, x_{\mathcal{M}(a)})}{g'_{x_b}(x_a^*, x_{\mathcal{M}(a)})} = x'_{\mathcal{M}(a)} \;\forall a.$$

Obviously, this condition does not hold generally as the matching function $\mathcal{M}(a)$ typically depends on the distribution of types on both market sides. Some cases where (4.15) holds and efficiency can be achieved are linear matching functions and CES production functions as stated in the following proposition.

Proposition 4.3 *Suppose $\mathcal{M}$ has the property $x_{\mathcal{M}(a)} = \lambda x_a$ for all a and some $\lambda > 0$ in a rational expectations equilibrium and the joint profit function $g(x, y)$ has the property $g_x(x, y)x = g_y(x, y)y$ for $x = \lambda y$, then ex-ante investments are constrained efficient.*

Proof: Obvious.

This proposition states that for a production function with constant optimal factor input ratio investments are efficient if the matching function matches agents such that their attributes correspond exactly to this optimal factor input ratio.

4.5.2 Choosing the Utility Possibility Frontier

The key insight to matching models with non-transferable utility is that agents' attributes determine the location of the utility possibility frontier in a given coalition in the matching market. By adjusting the attribute an agent is able to shift the utility possibility frontier. In our formalization agents choose the origin of their utility possibility frontier as depicted in figure 4.4.

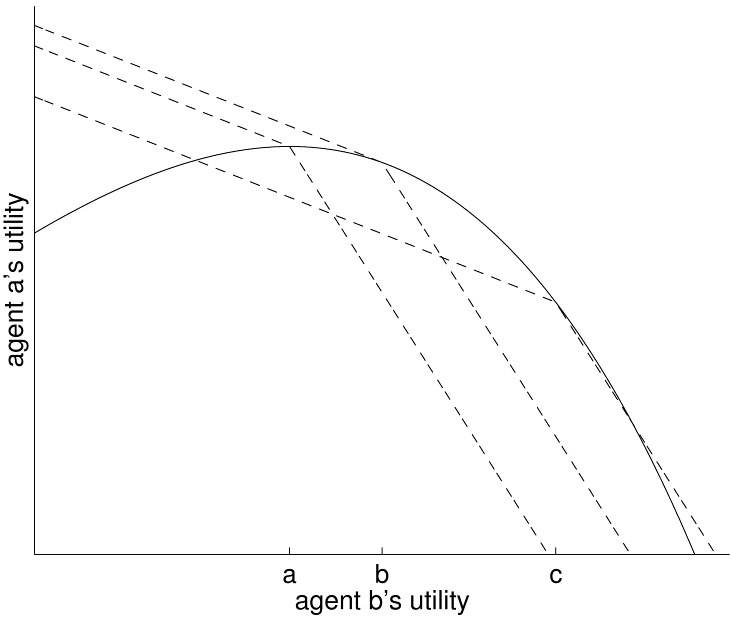

Figure 4.4: Utility possibility frontiers at the investment stage

Suppose an agent a expects to be matched with an agent b with attribute x_b. This gives rise to the set of possible utility possibility frontiers in figure 4.4. The dashed lines are instances of possible utility possibility frontiers in the matching stage associated to some points on the solid line. This in turn reflects a and b's utility depending on a's choice of attribute x_a. The attribute choice x_a corresponding to a in the graph has the property $g'(x_a, x_b) = k'(x_a)$, i.e. this is the investment choice when external effects are not realized at all. b and c are arbitrary choices of u_b indicating that for imperfect transferability of utility in general there exists no dominant utility possibility frontier in the sense of set inclusion. This means that if agent a expects to pay high side payments, the utility possibility frontier associated to $u_b = c$ dominates the one associated to $u_b = b$ and vice versa if a expects to receive high side payments.

In case of perfectly transferable utility the slopes of all attainable utility possibility frontiers are -1. Therefore, there exists a unique optimal frontier including all others, namely at u_b such that $u'_a(u_b) = -1$. However, this identity implies for the associated x_a that necessarily $k'(x_a) = 2g'(x_a, x_b)$, i.e. x_a is chosen efficiently given x_b. For non-transferable utility this does not need to hold and we will show that indeed non-transferabilities of utility generally distort investment incentives.

Let us now develop this intuition formally. An agent a's set of attainable utility possibility frontiers can be characterized by

$$\begin{aligned} \mathcal{F}(x_b) \quad = \quad & \{(u_a, u_b) : u_a = g(x_a, x_b) - k(x_a) \\ & \text{s.t. } u_b(x_a, x_b) = g(x_a, x_b) - k(x_b)\}. \end{aligned}$$

$\mathcal{F}$ corresponds to the solid line in figure 4.4 and is the locus of all origins of the utility possibility frontier of agents a and b attainable by choice of x_a when x_b is given. We now state a useful lemma characterizing an important property of $\mathcal{F}$ determining incentives for investment in attributes.

Lemma 4.1 *$\mathcal{F}(x_b)$ has the property that for $x_b > 0$ $u_a(u_b)$ is a continuously differentiable function and*

$$\frac{\partial u_a}{\partial u_b} = 1 - \frac{\frac{\partial k(x_a, \gamma_a)}{\partial x_a}}{\frac{\partial g(x_a, x_b)}{\partial x_a}}.$$

Moreover, $u_a(u_b)$ is strictly concave.

Proof: In Appendix.

The Lemma states an important link between the slope of $\mathcal{F}$ at the point of the choice of x_a to investment incentives. In general, an agent chooses the utility possibility frontier that maximizes his utility for the utility level his match needs to be given to ensure stability of

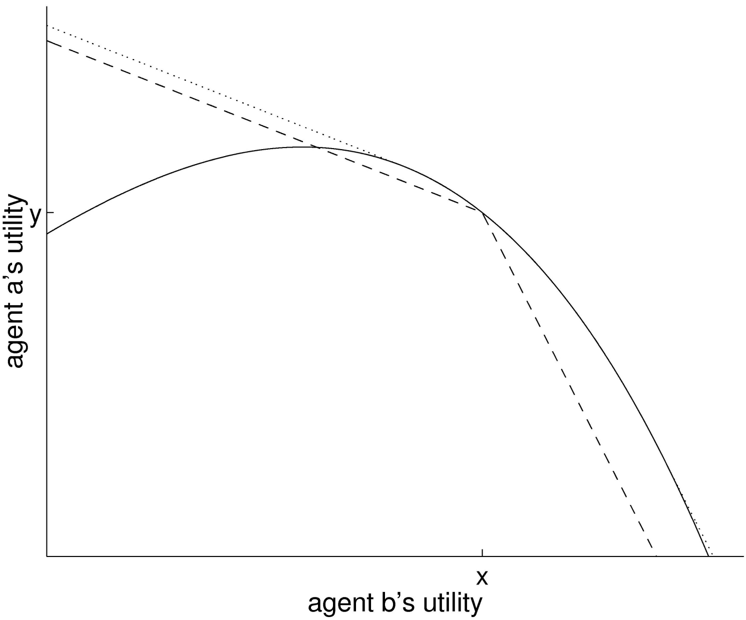

Figure 4.5: Utility possibility frontier for the linear NTU case

equilibrium. For transferable utility this utility possibility frontier is always given by the tangent to $\mathcal{F}(.)$ with slope -1.

In the case of non-transferable utility choosing the utility possibility frontier amounts to minimizing the utility cost incurred for the transfer of a given amount of utility. Figure 4.5 shows the linear case. The dashed line represents the utility possibility frontier if the attribute investment is efficient and the dotted line gives the optimal choice of investment given an agent's beliefs over his equilibrium match and payoff. In the neighborhood of the efficient investment at the point (x, y) an agent chooses to transfer utility by adjusting attribute investments where the dotted line coincides with the solid line. However, for large utility transfers a combination is optimal. For the non-transferable utility case we find that efficient investment incentives need not to be expected.

Proposition 4.4 *Suppose $\delta_i \in \mathcal{D}^K$ for all $i \in K$, $K = A, B$ and Assumptions A 1-3 hold. Then investments in attributes are constrained efficient if and only if for rational expectations equilibrium payoffs u_a and u_b and attribute investments x_a and x_b it holds that $u_a + u_b = 2g(x_a, x_b) - (k(x_a) + k(x_b))$.*

Proof: In Appendix.

To put this in other words, investment incentives will not even be constrained efficient unless equilibrium payoffs require only side payments that are not subject to the non-transferability of utility. As a consequence, for non-transferabilities, that can be represented by functions δ_i whose slope is bounded away from 1, for constrained efficient investment incentives it is necessary that side payments are not used in equilibrium. The following corollary is then immediate from Proposition 4.3 and 4.4:

Corollary 4.1 *Suppose agent $a \in A$ has δ_a such that $\delta_a(u, b) < u$ for all $u > 0$ and all $b \in B$. Then in any rational expectations equilibrium agent a's ex ante investment x_a is constrained efficient if and only if*

$$\frac{\partial g(x_a^*, x_{\mathcal{M}(a)})}{\partial x_a} = \frac{\partial g(x_a^*, x_{\mathcal{M}(a)})}{\partial x_{\mathcal{M}(a)}} \frac{\partial x_{\mathcal{M}(a)}}{\partial x_a}.$$

This means that constrained efficient investment incentives hinge on a knife-edge condition on the primitives of the model such as the distribution of agents and the functional form of $g(.)$. Hence, the matching equilibrium with non-transferable utility cannot be expected to induce efficient investment incentives.

However, for general functions δ we are able to obtain a similar result. Define $((\gamma_i, \delta_i^{TU})_{i \in A,B}, g(.), k(.))$ to denote the transferable utility version of the non-transferable utility matching game $((\gamma_i, \delta_i)_{i \in A,B}, g(.), k(.))$. That is, $\delta_i^T U(u) = u$ for all $i \in A, B$ and all finite $u \geq 0$. Since all efficient allocations belong to some equilibrium allocation of the transferable version of the matching game we are in a position to state a more general proposition. Denote by $\tau_{max}(i, j)$ the maximum amount of utility that can be transferred to i in coalition (i, j).

Proposition 4.5 *An equilibrium allocation* $((x_a, x_b)_{a,b\in A,B}, \mathcal{M})$ *of the non-transferable utility matching game* $((\gamma_i, \delta_i)_{i\in A,B}, g(.), k(.))$ *is efficient only if it coincides with some equilibrium allocation of its transferable utility version and equilibrium payoffs* u_a, u_b *are such that for all* $i \in A, B$ *and the match* j $u_i - g(x_i, x_j) - k(x_i) \leq \tau_{max}(i, j)$.

Proof: First note that the transferable version of any non-transferable utility matching game has to reach an efficient matching allocation (in the sense of aggregate utility maximization) given individual attributes by definition. Note further that all efficient allocations must be equilibria of the transferable utility version of the matching game.

To see this, suppose the contrary and let $\mathcal{M}$ maximize aggregate utility of $i \in A, B$. A blocking coalition exists if and only if there is some $a \in A$ and some $b \in B$ with $b \neq \mathcal{M}(a)$ and $2g(a, b) > g(a, \mathcal{M}(a)) + g(\mathcal{M}^{-1}(b), b)$. This condition implies, however, that $\mathcal{M}$ does not maximize aggregate utility, a contradiction.

On the other hand, for the transferable utility version it must hold that $\frac{\partial u_a(u_b)}{\partial u_b} = -1$ for all relevant u_b. Hence, in the transferable utility version agents' attribute investments are constrained efficient and the matching function induces efficient matching given attributes. This means that all efficient allocations of the non-transferable utility game $((\gamma_i, \delta_i)_{i\in A,B}, g(.), k(.))$ are in the set of equilibrium allocations of its transferable utility version. This proves necessity. ■

This last proposition provides a useful tool in assessing the properties of a specific model. Given the equilibrium of its transferable utility version, it is then easy to calculate whether the allocation will survive the introduction of market friction leading to non-transferabilities by analyzing the equilibrium side payments.

4.6 Conclusion

This chapter has analyzed the allocative role of markets with non-transferable utility in the classical hold-up problem. It finds that non-transferabilities quite generally preclude markets from setting ex ante investment incentives efficiently. Since markets that open before the investment typically do not induce efficient incentives either, there emerges a role for the timing of markets. Indeed, as we show, ex ante markets may dominate ex post markets both in terms of aggregate output and Pareto efficiency whenever non-transferabilities are sufficiently severe.

In particular, whenever investments are strategic substitutes the private return from investments on the market may exceed the social return. The reason is that transferring utility at the market stage may be more costly than increasing investment beyond the efficient level, thereby inducing over-investment. We have applied this reasoning to a model of educational attainment when contracting is limited to cost-sharing. As a result in ex post markets higher educated agents crowd out lower educated agents in the competition for jobs that require low education. This feature has been frequently cited as over-education in the empirical literature.

Moreover, this framework allows for the analysis of the evolution of market institutions. The optimal timing of markets depends, of course, on the severity and on the specific nature of the non-transferabilities. Suppose that restrictions on contract enforcement induce non-transferable utility at the market stage and a suitable technology such that indeed the ex ante market allocation dominates the ex post market allocation. However, as an economy develops more efficient instruments of enforcement, non-transferabilities can be expected to diminish. In our framework this results in a change of the optimal timing of markets, interpretable for instance by a move from contracting within corporations to spot markets, thus opening a new and interesting field for future research.

Appendix A

Mathematical Appendix to Chapter 2

Proof of Lemma 2.4

Taking the derivative of y^*+x^* as defined in Lemma 2.1 with respect to γ yields

$$\frac{\partial(x^*+y^*)}{\partial\gamma} = \frac{\frac{1+\ln\gamma+(\ln\gamma)^2}{\gamma}w_s - (1+r)h - (2+r)w_n}{(1+r)\gamma(\ln\gamma)^2}.$$

Inspection of the derivative shows that (x^*+y^*) has a global maximum in γ, strictly increases before attaining it, and strictly decreases thereafter. But it is also the case that

$$\lim_{\gamma\to\infty}(x^*+y^*) = h + \frac{2+r}{1+r}w_n > h.$$

This implies that the derivative with respect to γ must be strictly positive whenever the function value (x^*+y^*) is smaller than its limit. Hence, given an interior solution with $(x^*+y^*) < h$, increasing γ also increases (x^*+y^*) until reaching the boundary h.

Concerning the second part of the lemma, y^* is obviously increasing in γ. To show that x^* also is, take any γ_1, γ_2 with $\gamma_1 > \gamma_2$ within the definition range. Then it can easily be checked that

$$\left.\frac{\partial U_s(x)}{\partial x}\right|_{\gamma_1} > \left.\frac{\partial U_s(x)}{\partial x}\right|_{\gamma_2} \quad \forall\, x.$$

On the other hand it holds that

$$U_s(h) = \ln(w_s) \;\; \forall\, \gamma.$$

This implies by monotonicity that

$$U_s(x)|_{\gamma_1} < U_s(x)|_{\gamma_2} \;\; \forall\, x < h. \tag{A.1}$$

This in turn implies that any tangent to $U_s(x)|_{\gamma_2}$ lies strictly above the area under $U_s(x)|_{\gamma_1}$ for all $x < h$.

Now take some point $z > U_s(x)|_{\gamma_2}$, $x < h$. Let $T_2(x)$ be any tangent to $U_s(x)|_{\gamma_2}$ with the property that it contains the point z. Denote by $T_1(x)$ a tangent to $U_s(x)|_{\gamma_1}$ with the property that it also contains z. Let the respective points of tangency t_i be implicitly defined by $T_i(t_i) = U_s(t_i)|_{\gamma_i}$, $i = 1, 2$.

Then it must hold that $T_2(t_1) > T_1(t_1)$ as long as t_1, t_2 lie to the left of h implying that that the slope of T_2 is greater than the slope of T_1. This must be true because $T_2(x) > U_s(x)|_{\gamma_2}\, \forall\, x$ due to concavity of U_s and because of inequality (A.1).

This implies that the tangent to both $U_n(x)$ at $x = x_1$ and $U_s(x)|_{\gamma_2}$ must be steeper than the tangent to both $U_n(x)$ at $x = x_2$ and $U_s(x)|_{\gamma_1}$ for all tangency points to the left of h. This follows from monotonicity of $U_n(x)$. This and strict concavity then also imply that $x_1 > x_2$. This means x^* increases in γ as long as an interior solution is the case. ■

Sustainability

Sustainability holds if h is sufficiently close to w_n

Using assumption A2.1.(i), inequality (2.5) is implied by:

$$\begin{aligned}(2+r)\gamma w_n &> \left(\frac{\ln\gamma}{(1-\alpha)(1+i)-1}-1\right)((i-r)h-(2+r)w_n)\\ \Leftrightarrow \gamma &> \left(\frac{\ln\gamma}{(1-\alpha)(1+i)-1}-1\right)\left(\frac{(1+r)(\gamma-1)}{(2+r)}\frac{h}{w_n}-1\right).\end{aligned}$$

Taking some $\gamma > 1$ with $1 < \frac{\ln\gamma}{(1-\alpha)(1+i)-1}$, this condition holds for h sufficiently small compared to w_n.

A case where sustainability fails

Looking at a rewritten inequality (2.5) we find:

$$\gamma > \left(\frac{\ln\gamma}{(1-\alpha)(1+i)-1}-1\right)\left(\frac{(1+r)\gamma h - w_s}{(2+r)w_n}\right).$$

It is now possible to find a γ sufficiently close to its lower bound given by Assumption A2.1, such that this inequality does not hold provided w_s is sufficiently close to $(1+r)h$ and $(2+r)w_n$.

Proof of Lemma 2.5

First we use the fact that every gambler weakly prefers the lottery because the convex combination of the corresponding utility levels lies above the certain utility level for gamblers. Expected utility from gambling is given by:

$$\begin{aligned}E[U(I^G(x))] &= \frac{x-x^*}{y^*}\ln((1+i)(x^*+y^*-h)+w_s)+\\ &\quad +(1-\frac{x-x^*}{y^*})\ln((1+r)x^*+(2+r)w_n).\end{aligned}$$

That means that for all $x \in (x^*, x^* + y^*)$ it must hold one of the following:

$$\begin{aligned}
&\forall x \in (x^*, f): \\
&E(U^G(x)) > \ln((1+r)x + (2+r)w_n) > \ln((1+i)(x-h) + w_s) \\
&\forall x \in [f, x^*+y^*): \\
&E(U^G(x)) > \ln((1+i)(x-h) + w_s) > \ln((1+r)x + (2+r)w_n).
\end{aligned}$$

Yet Jensen's inequality implies:

$$E(U^G(I^G(x))) \leq \ln(E[I^G(x)]).$$

So that $\forall x \in (x^*, f) : E[I^G(x)] > (1 + r)x + (2 + r)w_n$ and $\forall x \in [f, x^* + y^*) : E[I^G(x)] > (1 + i)(x - h) + w_s$. This establishes the lemma. ■

Proof of Proposition 2.3

Step 1: Note that aggregate utility levels U_{t+k} and $\hat{U}_{t+l}$, $k, l \in \mathbb{N}$ are converging sequences in $\mathbb{R}$. To see this, write down e.g. U_{t+k} as a function of endowments

$$U_{t+k} = \int_0^{\overline{x}_s} EU(I(x))dF_{t+k}(x).$$

So the dynamic behavior of U_{t+k} is entirely governed by the distribution of initial endowments among period-t-gamblers in period $t + k$. But we know that $F_{t+k}(x)$ converges. Hence, so does U_{t+k}. An analogous argument applies to $\hat{U}_{t+k}$.

Step 2: We show that if $\hat{U} \equiv \lim_{l\to\infty} \hat{U}_{t+l} > U \equiv \lim_{k\to\infty} U_{t+k}$, then there exists an $N < \infty$ such that $\hat{U}_{t+N} > U_{N+l}$. Because of convergence, the sequences $\{\hat{U}_{t+k}, k \in \mathbb{R}\}$ and $\{U_{t+k}, k \in \mathbb{R}\}$ have the properties

$$\begin{aligned}
&\forall\, \epsilon > 0\, \exists\, K : |U_{t+k} - U| < \epsilon \;\; \forall\, k \geq K \text{ and} \\
&\forall\, \epsilon > 0\, \exists\, L : |\hat{U}_{t+l} - \hat{U}| < \epsilon \;\; \forall\, l \geq L.
\end{aligned}$$

Let $2\epsilon = |U - \hat{U}|$, then it must hold for all $n \geq N \equiv \max\{K, L\}$ that $\hat{U}_{t+n} > U_{t+n}$.

Step 3: If condition (ii) in Proposition 2.2 holds, then $\hat{U} > U$ trivially, thus establishing the first part of the proposition.

Step 4: Assume condition (i) in Proposition 2.2 does not hold. Then only in the first period agents are willing to gamble both under γ and $\hat{\gamma}$. Steady state income under γ is given by:

$$\begin{aligned} I_{SS}(x) &= \int_0^g dF_t(x)[\overline{x}_n(1+r) + w_n(2+r)] \\ &\quad + \int_g^{\hat{x}^*+\hat{y}^*} dF_t(x)[(\overline{x}_s - h)(1+r) + w_s], \end{aligned}$$

and steady state income under $\hat{\gamma}$ by:

$$\begin{aligned} \hat{I}_{SS}(x) &= \left[\int_0^{\hat{x}^*} dF_t(x) + \int_{\hat{x}^*}^{\hat{x}^*+\hat{y}^*} dF_t(x) - P_t^h\right][\overline{x}_n(1+r) + w_n(2+r)] \\ &\quad + P_t^h[(\overline{x}_s - h)(1+r) + w_s], \end{aligned}$$

where P_t^h is the fraction of agents among gamblers who escape the poverty trap under $\hat{\gamma}$ defined by

$$P_t^h = \int_{\hat{x}^*}^{\hat{x}^*+\hat{y}^*} \frac{x - \hat{x}^*}{\hat{y}^*} dF_t(x).$$

Integration by parts yields

$$P_t^h = F_t(\hat{x}^* + \hat{y}^*) - \frac{1}{\hat{y}^*}\int_{\hat{x}^*}^{\hat{x}^*+\hat{y}^*} F_t(x)dx.$$

The sign of the change in steady state aggregate income after an increase of γ is given by the sign of the net fraction of agents that is placed on the high income growth path, so that $\hat{I}_{SS}(x) - I_{SS}(x)$ can be written as:

$$\begin{aligned} & P_t^h - F_t(\hat{x}^* + \hat{y}^*) + F_t(g) > 0 \\ \Leftrightarrow \quad & \hat{y}^* F_t(g) > \int_{\hat{x}^*}^{\hat{x}^*+\hat{y}^*} F_t(x)dx. \end{aligned} \tag{A.2}$$

Using the fact that $F_t(g)$ is a constant, the LHS can be rearranged:

$$\begin{aligned}\hat{y}^* F_t(g) &= (\hat{x}^* + \hat{y}^*)F_t(g) - \hat{y}^* F_t(g) \\ &= \int_{\hat{x}^*}^{\hat{x}^*+\hat{y}^*} F_t(g)dx.\end{aligned}$$

Inserting this into (A.2) yields:

$$\int_{\hat{x}^*}^{\hat{x}^*+\hat{y}^*} [F_t(g) - F_t(x)]dx > 0.$$

By definition the last inequality means that if $F_t(x)$ second order stochastically dominates any lottery $[(\hat{x}^* - c), (\hat{x}^* + \hat{y}^* + c), F_t(g), (1 - F_t(g))]$ with $c > 0$ on the interval $[\hat{x}^*, \hat{x}^* + \hat{y}^*]$, then aggregate utility from the steady state endowments is higher under $\hat{\gamma}$. ∎

Proof of Proposition 2.4

Consider an agent with endowment $0 < x_i < h$. Suppose the agent chooses a symmetric lottery (x, y) such that $x + \Delta_1 = x_i$ and $\beta y = \Delta_1 + \Delta_2$ and $x < f < x + \beta y$. This lottery is preferred to the certain endowment x_i, if

$$\frac{x_i - x}{y} U_s(x + \beta y) + \frac{x + y - x_i}{y} U_n(x) \geq U(x_i).$$

That is

$$\frac{\beta \Delta_1}{\Delta_1 + \Delta_2} [U_s(x_i + \Delta) - U_n(x_i - \Delta)] \geq U(x_i) - U_n(x_i - \Delta).$$

Let $x_i = f - \epsilon$, $\epsilon > 0$ sufficiently small, such that $U_s(f - \epsilon + \Delta_2) - U_n(f - \epsilon) \geq U_s'(f - \epsilon + \Delta_2)\Delta_2$. Using the fact that $U(.)$ is piecewise concave, a sufficient condition can then be obtained by an adequate first order approximation:

$$\beta \left(\frac{\Delta_1}{K_1} + \frac{\gamma \Delta_2}{K_2 + (1+i)\Delta_2} \right) \geq \frac{\Delta_1 + \Delta_2}{K_1 - (1+r)\Delta_1}. \tag{A.3}$$

with $K_1 = (2+r)w_n + (1+r)(f-\epsilon)$ and $K_2 = w_s + (1+i)(f-\epsilon-h)$. Closer inspection shows that for ratios

$$\frac{\Delta_2}{\Delta_1} > \frac{1-\beta}{\gamma\beta - 1},$$

$\gamma\beta > 1$, there exist $\Delta_1, \Delta_2 > 0$ sufficiently small such that (A.3) holds. By piecewise concavity and continuity of $U(.)$, for any $\Delta_2 > 0$ there exists $0 < \epsilon < \Delta_2$ such that $U_s(f-\epsilon+\Delta_2) - U_n(f-\epsilon) \geq U_s'(f-\epsilon+\Delta_2)\Delta_2$ and the approximation is valid. Hence, for any $\beta > \frac{1}{\gamma}$ there exists a neighborhood to the left of f preferring some lottery to a certain endowment x_i.

Let now $x_i = f+\epsilon$, $\Delta_1 > \epsilon > 0$ sufficiently small such that $U_s(f+\epsilon) - U_n(f+\epsilon-\Delta_1) \leq U_n'(f-\Delta_1)\Delta_1$. Such an ϵ exists due to continuity of $U(.)$. Using the fact that $U(.)$ is piecewise concave, an adequate first order approximation to obtain a sufficient condition is now

$$\beta\left(\frac{\Delta_1}{K_1} + \frac{\gamma\Delta_2}{K_2 + (1+i)\Delta_2}\right) \geq \frac{\gamma(\Delta_1+\Delta_2)}{K_1 - (1+r)\Delta_1}. \tag{A.4}$$

with $K_1 = (2+r)w_n + (1+r)f$ and $K_2 = w_s + (1+i)(f+\epsilon-h)$. Fortunately, condition (A.4) coincides with condition (A.3) so that the first part of the proposition is verified.

To prove the second part, recall the definitions of x^* and y^* from Lemmata 2.1 and 2.2. Denote the convex hull of $U(.)$ by $U^C(.)$. By definition all points $U(x)$ with $x \leq x^*$ or $x^* + y^* \leq x$ lie on the convex hull of $U(.)$. For these x it must hold that $U^{C''}(x) \leq 0$ by piecewise concavity of $U(.)$. This means that for these x there exists no convex combination $\lambda U^C(x+\Delta_2) + (1-\lambda)U^C(x-\Delta_1) > U(x)$ with $\lambda = \frac{\Delta_1}{\Delta_1+\Delta_2}$. For $\lambda = \frac{\beta\Delta_1}{\Delta_1+\Delta_2}$, $\beta < 1$, the strict inequality becomes weak. That is, for endowment levels $x_i < x^*$ and $x_i > x^* + y^*$ the certain income from x_i is preferred to any costly endowment lottery.

Conversely, for every costly lottery that is utility-maximizing for some agent it must hold that $x \geq x^*$ and $x + \beta y \leq y^*$. For any optimal costly lottery it must hold for y that

$$\frac{\partial U_s(x+\beta y)}{\partial x + \beta y} = \frac{U_s(x+\beta y) - U_n(x)}{\beta y}. \tag{A.5}$$

That is, given x, βy has to be chosen such as to maximize the slope of the lottery. $U_s'(.)$ decreases in its argument, so that an x that maximizes the corresponding βy is associated to the flattest lottery given (A.5). That is the lottery that is tangent to $U_n(.)$. This lottery is given by x^* and $x^* + y^*$. Hence, $x + \beta y \leq x^* + y^*$. To conclude the proof, suppose that $x + \beta y \leq f$. Then due to concavity of $U_n(x)$ in $x \in [0, f]$ and Jensen's inequality there exists no $x_i \in [x, x + \beta y]$ such that a costly lottery with $\beta y > 0$ is preferred to $U_n(x_i)$. ■

Appendix B

Mathematical Appendix to Chapter 3

B.1 Proofs

Proof of Lemma 3.1

(i) Let θ_1 denote asset holdings in coalition N_1 and θ_2 those in coalition N_2. Assume without loss of generality that $|\theta_1(i)| > |\theta_2(i)|$ for some agent i who is assumed to be able to finance assets using his wealth for now. Then agent i prefers θ_1 in N_1 to θ_2 in N_2 if and only if:

$$\begin{aligned} v_i(n_1, \theta_1, \omega(i), r) &> v_i(n_2, \theta_2, \omega(i), r) \\ \Leftrightarrow \pi_i(n_1, \theta_1) + (1+r)(\omega(i) - |\theta_1(i)|c) &> \pi_i(n_2, \theta_2) \\ &\quad + (1+r)(\omega(i) - |\theta_2(i)|c) \\ \Leftrightarrow \pi_i(n_1, \theta_1(i)) - \pi_i(n_2, \theta_2) &> (1+r)(|\theta_1(i)| - |\theta_2(i)|)c. \quad \text{(B.1)} \end{aligned}$$

This means that preference for more asset ownership is non-increasing in the interest rate. Firstly, note that inequality (B.1) strictly tightens in r. Secondly, there exist a unique r_0 such that (B.1) holds with equality implying preferences for ownership are well-defined with respect to r. An analogous argument applies to the case of an agent having to borrow to finance $|\theta_2(i)|$ assets. However, for endowments $|\theta_1(i)|c > \omega(i) > |\theta_2(i)|c$ inequality (B.1) changes into

$$\pi_i(n_1,\theta_1)-\pi_i(n_2,\theta_2) > (1+r)(\omega(i)-|\theta_2(i)|c)-(1+i)(\omega(i)-|\theta_1(i)|c),$$

where $(1+i) = (1+r)\gamma$. This has the same properties with respect to r as (B.1), so that the above result applies.

(ii) Let θ_1, θ_2, n_1, n_2 be defined as above. Now we show that a preference relation of agent i over (θ_1, n_1) and (θ_2, n_2) has a unique endowment level for which it reverses. Agent i prefers the first to the second if, depending on the degree to which $\theta_1(i)$ and $\theta_2(i)$ have to be financed by debt, one of the following inequalities holds

$$\begin{aligned}
&\pi_i(n_1,\theta_1)-\pi_i(n_2,\theta_2) > (1+i)(|\theta_1(i)|-|\theta_2(i)|)c \text{ for } \omega(i) \leq |\theta_2(i)|c, \text{ or}\\
&\qquad \omega(i)-|\theta_2(i)|c > \frac{1}{i-r}[\pi_i(n_2,\theta_2) + (1+i)(|\theta_1(i)|-|\theta_2(i)|)c]\\
&\qquad\qquad\qquad -\pi_i(n_1,\theta_1) \text{ for } \omega(i) \in (|\theta_2(i)|c, |\theta_1(i)|c), \text{ or}\\
&\pi_i(n_1,\theta_1)-\pi_i(n_2,\theta_2) > (1+r)(|\theta_1(i)|-|\theta_2(i)|)c \text{ for } \omega(i) \geq |\theta_1(i)|c.
\end{aligned} \tag{B.2}$$

Define now a wealth level ω_2 by

$$\omega_2 \equiv \frac{1}{i-r}\left[\pi_i(n_2,\theta_2) - \pi_i(n_1,\theta_1) + (1+i)(|\theta_1(i)| - |\theta_2(i)|)c\right] + |\theta_2(i)|c.$$

Clearly, ω_2 is unique given r. Note that whenever $\omega_2 < |\theta_2(i)|c$ this implies that the first inequality of (B.2) holds. On the other hand, whenever $\omega_2 > |\theta_1(i)|c$ this implies that the third inequality of (B.2) holds. To see this, let $\omega_2 > |\theta_1(i)|c$. Plugging this into the last equation yields

$$\begin{aligned}
(i-r)(|\theta_1(i)|-|\theta_2(i)|)c \;<\; &\pi_i(n_2,\theta_2) - \pi_i(n_1,\theta_1)\\
&+(1+i)(|\theta_1(i)|-|\theta_2(i)|)c.
\end{aligned}$$

This inequality can easily be transformed to yield the reverse of the third inequality of (B.2). Thus $\omega_2 > |\theta_1(i)|c$ implies that for all endowment levels (θ_2, n_2) is preferred to (θ_1, n_2) by agent i. Hence, a unique endowment cutoff value ω_0 can be defined, such that for all endowments $\omega(i) > \omega_0$, (θ_1, n_1) is preferred to (θ_2, n_2), and for all $\omega(i) < \omega_0$ the reverse is true:

$$\omega_0 = \begin{cases} \underline{\omega} & \text{if } \omega_2 < |\theta_2(i)|c \\ \overline{\omega} & \text{if } \omega_2 > |\theta_1(i)|c \\ \omega_2 & \text{otherwise.} \end{cases}$$

This means ω_0 is unique as well and agents' preferences are well-defined with respect to endowments. Moreover, it is easy to see that ω_0 is non-decreasing in r. The unique cutoff level implies that at most agents with a unique endowment level may be indifferent between any two alternatives of asset ownership and coalition size. Therefore preferences are strict almost everywhere on I.

Proof of Lemma 3.3

The first statement in the lemma can be established by noting that in any equilibrium satisfying the stability condition an owner of an $n > 1$ firm must be indifferent between any worker when hiring. If that is not the case there always exist blocking coalitions of owners not hiring their favorite workers and these workers. This follows from the fact that all owners prefer to hire the same rich workers and the utility benefit an owner has from a worker's wealth is equal for all owners that have to borrow. This immediately implies that monetary side payments only depend on the firm size.

Continuing we note that workers' transfers to owners must be decreasing in firm size for all firm sizes present in equilibrium. Suppose the contrary and let n, $n' < n$ be firm sizes that emerge in an equilibrium and $\frac{t(n)}{n} > \frac{t(n')}{n'}$. An n worker with endowment $\omega(i)$ prefers working in an n firm to an n' firm if

$$\pi_W(n) + (1{+}r)(\omega(i) - \frac{t(n)}{n}) \; > \; \pi_W(n') + (1{+}r)(\omega(i) - \frac{t(n')}{n'}) \text{ or}$$
$$\pi_W(n) + (1{+}r)\gamma(\omega(i) - \frac{t(n)}{n}) \; > \; \pi_W(n') + (1{+}r)(\omega(i) - \frac{t(n')}{n'}) \text{ or}$$
$$\pi_W(n) + (1{+}r)\gamma(\omega(i) - \frac{t(n)}{n}) \; > \; \pi_W(n') + (1{+}r)\gamma(\omega(i) - \frac{t(n')}{n'}),$$

corresponding to the cases $\omega(i) > \frac{t(n)}{n} > \frac{t(n')}{n'}$, $\frac{t(n)}{n} > \omega(i) > \frac{t(n')}{n'}$, and $\frac{t(n)}{n} > \frac{t(n')}{n'} > \omega(i)$. That is

$$\begin{aligned}
\pi_W(n) - \pi_W(n') &> (1+r)(\frac{t(n)}{n} - \frac{t(n')}{n'}) \text{ or} \\
\pi_W(n) - \pi_W(n') &> (1+r)[(\gamma\frac{t(n)}{n} - \frac{t(n')}{n'}) - (\gamma - 1)\omega(i)] \text{ or} \\
\pi_W(n) - \pi_W(n') &> (1+r)\gamma(\frac{t(n)}{n} - \frac{t(n')}{n'})
\end{aligned}$$

which leads to a contradiction in all three cases noting that the LHS of all inequality is negative. ■

Proof of Proposition 3.1

We proceed by first establishing existence and uniqueness of a matching equilibrium given some interest rate r^*. That is we use Definition 3.2 without capital market clearing. Then we show that r^* such that (3.5) holds exists and is unique.

Existence of the matching equilibrium

The proof of existence closely follows Legros and Newman (1996). Let the interest rate be given by r^*. A modified version of θ, θ^M, is needed to construct a super-additive characteristic function of the matching economy (I, θ, v_i) along the lines of Shubik and Wooders (1983). Define a modified $\theta^M(N)$, $N \in \mathcal{F}(I)$ as follows:

$$\begin{aligned}
\theta^M : \mathcal{F}(I) \to \{A^I, A^{II}, \emptyset\}^{\mathbb{R}} \times [0,1], \\
\theta^M(N) = (\theta(N), q(N)),
\end{aligned}$$

where $q(N) \neq q(O)$, with $N, O \in \mathcal{F}(I)$, whenever $N \neq O$. $q(N)$ specifies the index of the set of assets the members of coalition N own. Feasible ownership rights are defined by

$$\Theta^M(N) = (\Theta(N), q(N)).$$

Essentially, this postulates that in a coalition any member's ownership right must be assigned the same index and otherwise be feasible in the sense used above. Now let $V(O)$ with $O = \bigcup_k O_k$, where $O_k \in \mathcal{F}(I)$ are disjoint finite sets of agents, denote the characteristic function of the economy (I, θ^M, v_i):

$$\begin{aligned} V(O) \quad = \quad & \{(v_i(|O|, \theta^M, \omega(i), r^*))_{i \in O} : \\ & \text{for } i \in O_k \; \theta^M(O_k) \in \Theta^M(O_k) \forall O_i \subseteq O\}. \end{aligned}$$

$V(O)$ describes the set of agents' attainable payoff vectors in coalitions $O_k \subset O$ achievable by choosing ownership rights allocations. Note that the notation using O is equivalent to a notation using the corresponding vector of attributes $(\omega_1, \omega_2, ..., \omega_n)$ as in Kaneko and Wooders (1996). Let $o = |O|$. This notation ensures that any union of disjoint coalitions can use the same allocation as the disjoint coalitions. Then construct the comprehensive extension of $V(O)$ by defining

$$\hat{V}(O) = \{x \in \mathbb{R}^o : x \leq V(O)\}.$$

$\hat{V}(O)$ has the following properties:

$$\hat{V} \text{ is a non-empty, closed subset of } \mathbb{R}^o \; \forall \, O \in \mathcal{F}(I), \tag{B.3}$$
$$\hat{V}(O) \times \hat{V}(O') \subseteq \hat{V}(O \cup O') \; \forall \, O, O' \in \mathcal{F}(I), \tag{B.4}$$
$$\inf \sup \hat{V}(\{i\}) > -\infty, \tag{B.5}$$
$$\forall \, O \in \mathcal{F}(I), x \in \hat{V}(O) \wedge y \in \mathbb{R}^o \text{ with } y \leq x \Rightarrow y \in \hat{V}(O), \tag{B.6}$$
$$\forall \, O \in \mathcal{F}(I), \hat{V}(O) - \bigcup_{i \in O} [(\text{int } \hat{V}(\{i\})) \times \mathbb{R}^{o-1}] \text{ is non-empty and bounded.} \tag{B.7}$$

Properties B.3, B.4 and B.6 follow directly by definition. Property B.5 follows from the existence of an outside option, $V(\{i\}) \geq 0$. This and the definition of $\hat{V}$ also imply property B.7. Therefore $\hat{V}$ is a characteristic function in the sense of Kaneko and Wooders (1986).

Let us represent all agents $i_1, i_2, ..., i_p \in O$ by their respective wealth $\omega(i_k)$. Then it is straightforward that also

$$\begin{aligned}\hat{V}(\omega(i_{\rho(1)}), \omega(i_{\rho(2)}), ..., \omega(i_{\rho(p)})) &= \{(x_{\rho(1)}, x_{\rho(2)}, ..., x_{\rho(p)}) : (x_1, x_2, ..., x_p) \\ &\quad \in \hat{V}(\omega(i_1), \omega(i_2), ..., \omega(i_p))\}\end{aligned}$$

for all permutations ρ of O. Thus the conditions Comprehensiveness (property B.5), Nontriviality (implied by property B.7), and Anonymity in Kaneko and Wooders (1996) are met. We know from Lemma 3.2 that coalition sizes are bounded above by $\overline{n}$. The last condition to show is continuity of $\{x \in \mathbb{R}^p : V(\{i\}) \leq x \leq V(O)\}$ on $[0,1]^n$ for $n = 1, ..., \overline{n}$ which trivially holds.

Admitting transfers of a limited kind as defined by equation (3.2) we are in a position to apply the Theorem of Kaneko and Wooders (1996) and thus have proven existence of the f-core of the characteristic function game associated with $\hat{V}$. What remains to be shown is that an allocation in the f-core of $\hat{V}$ is also an equilibrium as in Definition 3.2 without capital market clearing. An allocation in the f-core of $\hat{V}$ for some $O \in \mathcal{F}$ gives rise to payoffs $\hat{x} \in \hat{V}(O)$ that cannot be improved upon in the sense of stability. For $x \in V(O)$ it must hold that $x \geq \hat{x}$ by construction of $\hat{V}$. Then for the equilibrium allocation neither can x be improved upon. Finally, by definition of $V(.)$ for all $x \in V(O)$ and $\hat{x} \in \hat{V}(O)$ such that $\hat{x} \leq x$ there exist disjoint subsets of O, O_k, a mapping θ^M, and side payments t such that $x_i = v_i(|O_k|, \theta^M, \omega(i), r, t(i))$. Define for all $j \in [0,1]$ $O_j = \{i \in I : \theta^M(i) = (\theta(i), j)\}$. Then the collection $(O_j, j \in [0,1] : O_j \neq \emptyset)$ defines the equilibrium coalitions in the sense of our equilibrium definition.

Uniqueness of the matching equilibrium

The proof will establish that for any matching equilibrium given r^* the equilibrium allocation of firms, that is $(\mu_n)_{n:\mu_n>0}$, is determined by side payments t which depend on firm size only. We will first proof uniqueness of side payments consistent with stability and measure consistency. Then we will show that μ_n is indeed uniquely determined by side payments almost everywhere.

Step 1: Uniqueness of side payments. As a preliminary define measures $\mu_{n_M}^{strict}$ and $\mu_{n_W}^{strict}$ as the measures of agents strictly preferring to be owner of n firms, or worker in n firms, respectively, to all other roles in firms that have positive measure in equilibrium. An equilibrium vector of side payments t induces stability and measure consistency which imply jointly

$$\mu_{n_M}^{strict} \leq \mu_n \leq \mu_{n_M} \text{ and } \mu_{n_W}^{strict} \leq n\mu_n \leq \mu_{n_W}. \tag{B.8}$$

By definitions (3.6) and (3.7) and for the strict versions accordingly, the measures μ_{n_M}, μ_{n_W} and $\mu_{n_M}^{strict}$, $\mu_{n_W}^{strict}$ are fully characterized by side payments $t(n)$. In particular, by inspection of the endowment cutoff values $\omega_M(n,.)$ and $\omega_W(n,.)$ as can be derived using the proof of Lemma 3.3, the measures μ_{n_W} and $\mu_{n_W}^{strict}$ strictly decrease in $t(n)$ for $\mu_{n_W}, \mu_{n_W}^{strict} \in (0,1)$ and the measures μ_{n_M}, $\mu_{n_M}^{strict}$ strictly increase in $t(n)$ for $\mu_{n_M}, \mu_{n_M}^{strict} \in (0,1)$. Note further that if $\mu_{n_M}|_t(n) > \mu_{n_M}^{strict}|_t(n)$ for $t'(n) \neq t(n)$, all other side payments equal, $\mu_{n_M}|_{t'(n)} = \mu_{n_M}^{strict}|_{t'(n)}$ and likewise for μ_{n_W} and $\mu_{n_W}^{strict}$. Hence, some μ_n consistent with t is not consistent with t' such that $t(j) = t'(j)$ for all $j \neq n$ with positive measure except for $t(n) \neq t'(n)$.

Let n' denote the next smaller firm size and n'' the next bigger firm size with respect to n with positive measure under t. By Lemma 3.3 we know that matching is negative assortative. This implies that μ_{n_W} and $\mu_{n_W}^{strict}$ strictly increase in $t(n')$ and $t(n''$. μ_{n_M} and $\mu_{n_M}^{strict}$ strictly decrease in $t(n')$ and $t(n'')$.

Suppose now there exist systems of side payments $t \neq t'$, both associated with a corresponding matching equilibrium such that conditions (B.8) hold.

Case $\{n : \mu_n > 0|t\} = \{n : \mu_n > 0|t'\}$: Let n, n', and n'' have positive measure under both t and t'. It follows that if $t(n) < t'(n)$ then $t'(n') > t(n')$ and $t'(n'') > t(n'')$ is necessary for μ_n, such that conditions (B.8) hold, to exist. An analogous argument applies to $t(n) > t'(n)$.

This implies that for $t' \neq t$ conditions (B.8) cannot hold if $\{n : \mu_n > 0|t\} = \{n : \mu_n > 0|t'\}$. To see this, let μ_n, $\mu_{n'}$ and $\mu_{n''}$ have positive measure each and suppose without loss of generality

$t(n) < t'(n)$. Then μ_{n_M} and its strict version increases, $\mu_{n'_M}$ and $\mu_{n''_M}$ and their strict counterparts decrease. μ_{n_W} and its strict counterpart decreases, and $\mu_{n'_W}$ and $\mu_{n''_W}$ and their strict versions all increase. Increasing $t(n')$ or $t(n'')$ to induce measure consistency again the same effect appears on the next smaller and bigger firms from a point of view of n' and n'' firms. By induction all side payments in t' must increase. Hence, for the biggest firm size in equilibrium, the measure of agents willing to own necessarily exceeds the measure of agents willing to work, thus violating (B.8). If $|\{n : \mu_n > 0|t\}| < 3$ the argument can be applied accordingly.

Case $\{n : \mu_n > 0|t\} \cap \{n : \mu_n > 0|t'\} = \emptyset$: Then there exists n' with zero measure under t but positive measure under t' and n with positive measure under t but zero measure under t' such that n' is the next bigger or larger firm size with respect to n. This is a contradiction to stability. To see this, suppose that at side payments t there does not exist $t'(n')$ such that n' is preferred to n by both workers and owners, an implication of stability of the equilibrium associated to t. At side payments $t'(n')$, however, by stability of the equilibrium associated to t' n' is weakly preferred by a positive measure of both owners and workers to n for all side payments $t(n)$.

This means that a positive measure of both owners and workers must be indifferent between n under $t(n)$ and n' under $t'(n')$. Positive measures of both owners and workers for n firms under t must induce measure consistency. However, indifference of both workers and owners for more than two consecutive firm sizes is not possible generically, as four equalities determining cutoff endowments have to hold with three degrees of freedom, an issue we will return to in step (2) of this part of the proof. Hence, measure consistency cannot be induced by an allocation with measure zero of n firms and positive measure of n' firms, a contradiction. This means $\{n : \mu_n > 0|t\} \subseteq \{n : \mu_n > 0|t'\}$ or $\{n : \mu_n > 0|t\} \supseteq \{n : \mu_n > 0|t'\}$.

Case $\{n : \mu_n > 0|t\} \supseteq \{n : \mu_n > 0|t'\}$: Suppose without loss of generality $t(n) < t'(n)$ for some n with positive measure under both t and t'. As shown previously this implies necessarily that $t(n') < t'(n')$ and $t(n'') < t'(n'')$. Suppose the next smaller or bigger firm size under t, say without loss of generality n', has zero measure under t'. To induce measure zero of n' firms t' has to be sufficiently greater than t for the labor supply in n' firms to collapse. But then agents preferring to own n' firms have to be matched in n firms and the next smaller n''' firms violating measure consistency for n firms. Restoring measure consistency for n''' and n firms requires necessarily that side payments rise in the next bigger and smaller firms. This implies by induction that for the biggest firm size under t' condition (B.8) cannot hold. This argument extends by induction to cases where more than one firm size has positive measure under t but zero measure under t'. Reversing the argument by exchanging t with t' gives the same result for $\{n : \mu_n > 0|t\} \subseteq \{n : \mu_n > 0|t'\}$.

Finally, the cases $t(n) = t'(n)$ for all $n \in \{n : \mu_n > 0|t\} \cap \{n : \mu_n > 0|t'\}$ either imply coincidence of side payments or can quickly be led to a contradiction to stability.

Step 2: The allocation (P, θ) is unique in t almost everywhere. Only for $|\mu_{n_M}^{strict} - \mu_{n_M}| > 0$ and $|\mu_{n_W}^{strict} - \mu_n \leq \mu_{n_W}| > 0$ for $n = n', n$, with n the next higher firm size than n' and both firm sizes with positive measure, is the statement not trivial. But then measure consistency uniquely determines μ_n and $\mu_{n'}$, since to have $|\mu_{n_M}^{strict} - \mu_{n_M}| > 0$ and $|\mu_{n_W}^{strict} - \mu_n \leq \mu_{n_W}| > 0$ for more than two consecutive firm sizes is not possible generically, as four equalities determining cutoff endowments have to hold with three degrees of freedom given by the side payments in the three firms.

Existence and uniqueness of r^*

Here only the case $w < c$ is of interest. Therefore r^* is given by conditions (3.4). For both properties it suffices to show that all three conditions are getting slacker in r. Then (3.4) postulates that r^* be the lowest non-negative interest rate such that all weak inequalities hold. The first condition is equivalent to

$$\hat{\omega} \geq \omega_U. \tag{B.9}$$

It follows immediately from the definition of $\hat{\omega}$, (3.3), that it is increasing in r. The second condition is equivalent to

$$\omega(i) \leq \frac{\gamma}{\gamma-1}\frac{t(n)}{n} - \frac{\pi_W(n)}{(1+r)(\gamma-1)} \;\forall\,\omega(i) < \omega_U \tag{B.10}$$

for all n with $\mu_n > 0$ in the matching equilibrium as defined in the last subsection. The set of conditions (B.10) is relaxing in r and in $t(n)$. Finally, no agent has a profitable deviation by forming non-equilibrium larger n' firms if

$$\begin{aligned}\overline{\omega} \quad \leq \quad & \frac{\pi_M(n) - \pi_M(n')}{(1+r)(\gamma-1)} + (n+1)c + \frac{\gamma}{\gamma-1}[(n'-n)c - \hat{t}(n')] \\ & + \frac{1}{\gamma-1}t(n) \end{aligned} \tag{B.11}$$

for at least one n given any n' with $\mu_{n'} = 0$ in equilibrium. $\hat{t}(n')$ is given by

$$\hat{t}(n') = n'[\frac{\pi_W(n')}{(1+r)\gamma} + \frac{\gamma-1}{\gamma}\omega_U].$$

Hence, the set of conditions (B.11) is relaxing in r and in $t(n)$ as well. However, the effect via side payments is second order. In particular, labor demand decreases in r. This in turn means that the binding one of conditions (B.9), (B.10), and (B.11) has relaxed. Because of continuity of the cutoff endowment levels in r and $t(.)$, the fact that $\omega(.)$ is continuously differentiable, and the possibility of agents' indifference between different roles equilibrium side payments are continuous in r as well. Given continuity in r of conditions (B.9), (B.10), and (B.11) and that $\omega(.)$ is continuously differentiable, μ_U is continuous in r as well, which is all we need to show. ■

Proof of Proposition 3.2

Starting with part (i) of the proposition, it is possible from this fact and the above expressions to derive a cutoff endowment level for agents to prefer working in an efficient K firm as opposed to working

(i) in an $n > K$ firm, in (ii) an $n' < K$ firm, or (iii) owning one asset in a symmetric firm.

$$\begin{aligned}
\omega(i) &\geq \frac{\pi_W(n) - \pi_W(K) + (1+r)\gamma[\frac{t(K)}{K} - \frac{t(n)}{n}]}{(1+r)(\gamma-1)} + \frac{t(n)}{n}, \\
\omega(i) &\leq \frac{\pi_W(K) - \pi_W(n') + (1+r)\gamma[\frac{t(n')}{n'} - \frac{t(K)}{K}]}{(1+r)(\gamma-1)} + \frac{t(K)}{K}, \\
\omega(i) &\leq \frac{\pi_W(K) - \pi_1 + (1+r)\gamma[c - \frac{t(K)}{K}]}{(1+r)(\gamma-1)} + \frac{t(K)}{K}. \qquad \text{(B.12)}
\end{aligned}$$

Conducting the same exercise for owners yields a cutoff endowment level for preferring K ownership:

$$\begin{aligned}
\omega(i) &\leq \frac{\pi_M(K)-\pi_M(n)+(1+r)\gamma[t(K)-t(n)+(n-K)c]}{(1+r)(\gamma-1)} \\
&\quad -t(K) + (K+1)c, \\
\omega(i) &\geq \frac{\pi_M(n')-\pi_M(K)+(1+r)\gamma[t(n')-t(K)+(K-n')c]}{(1+r)(\gamma-1)} \\
&\quad -t(n')+(n'+1)c, \\
\omega(i) &\geq \frac{\pi_1 - \pi_M(K) + (1+r)\gamma(Kc-t(K))}{(1+r)(\gamma-1)} + c. \qquad \text{(B.13)}
\end{aligned}$$

Assume now that endowments are scarce. We say an n firm *crowds out* an n' firm if, given equilibrium side payments $t(n)$, there exist no side payments $t(n')$ such that there is both positive measure of agents preferring to be owner in n' firms and positive measure of agents preferring to be worker in n' firms to the same role in n firms.

Lemma B.1 *A necessary condition for n firms to crowd out n' firms is*

$$\frac{t(n)}{n} - c < \frac{n'}{n'-n}\frac{\pi_W(n) - \pi_W(n')}{(1+r)\gamma} - \frac{1}{n'-n}\frac{\pi_M(n') - \pi_M(n)}{1+r}.$$

and for $n' > n > 1$ and

$$\frac{t(n)}{n} - c > \frac{n'}{n-n'}\frac{\pi_W(n') - \pi_W(n)}{1+r} - \frac{1}{n-n'}\frac{\pi_M(n) - \pi_M(n')}{(1+r)\gamma}.$$

and for $1 < n' < n$.

Proof of Lemma: Assume first that $n < n'$. Then, by Lemma 3.1 for n firms to crowd out n' firms, given $t(n)$ there must not exist $t(n')$ such that

$$\omega_M(n, n') \leq (n'+1)c - t(n') \text{ and } \omega_W(n, n') \geq \frac{t(n')}{n'},$$

where $\omega_M(n, n')$ and $\omega_W(n, n')$ is the appropriate cutoff endowment level as derived in (B.13) and (B.12). That is, $\nexists\, t(n')$ such that

$$\frac{\pi_W(n') - \pi_W(n)}{(1+r)\gamma} + \frac{t(n)}{n} - \frac{t(n')}{n'} \geq 0 \text{ and}$$
$$\frac{\pi_M(n) - \pi_M(n')}{1+r} + t(n) - t(n') + (n'-n)c \leq 0.$$

Solving for $\frac{t(n')}{n'}$ we find that this condition only holds if

$$\frac{\pi_W(n') - \pi_W(n)}{(1+r)\gamma} + \frac{t(n)}{n} < \frac{\pi_M(n) - \pi_M(n')}{1+r} + t(n) + (n'-n)c,$$
$$\Leftrightarrow \frac{t(n)}{n} - c < \frac{n'}{n'-n}\frac{\pi_W(n) - \pi_W(n')}{(1+r)\gamma} - \frac{1}{n'-n}\frac{\pi_M(n') - \pi_M(n)}{1+r}.$$

For $n' < n$ the same argument yields the necessary condition

$$\frac{t(n)}{n} - c > \frac{n'}{n-n'}\frac{\pi_W(n') - \pi_W(n)}{1+r} - \frac{1}{n-n'}\frac{\pi_M(n) - \pi_M(n')}{(1+r)\gamma}.$$

■

By the previous lemma, for efficient firms to crowd out $n > K$ firms it must hold that

$$\frac{t(K)}{K} - c \quad < \quad \frac{n}{n-K}\frac{\pi_W(K) - \pi_W(n)}{(1+r)\gamma} - \frac{\pi_M(n) - \pi_M(K)}{(n-K)(1+r)},$$

For efficient firms to crowd out $n' < K$ firms it must hold that

$$\frac{t(K)}{K} - c \quad > \quad \frac{n'}{K-n'}\frac{\pi_W(n') - \pi_W(K)}{1+r} - \frac{\pi_M(K) - \pi_M(n')}{(K-n')(1+r)\gamma},$$

Obtaining bounds on $t(K)$ by comparing the efficient firm size with symmetric $n = 1$ firms yields

$$\frac{\pi_1 - \pi_M(K)}{K(1+r)} \leq \frac{t(K)}{K} - c \leq \frac{\pi_W(K) - \pi_1}{(1+r)\gamma}.$$

Now it is possible to derive the desired necessary conditions. For all $n > K$ and $n' < K$ it must hold that

$$\begin{aligned} \frac{n}{n-K}\frac{\pi_W(K)-\pi_W(n)}{\gamma}-\frac{\pi_M(n)-\pi_M(K)}{n-K}+\frac{\pi_M(K)-\pi_1}{K} &> 0 \text{ and} \\ \gamma\frac{n'}{K-n'}[\pi_W(n')-\pi_W(K)]+\frac{\pi_M(n')-\pi_M(K)}{K-n'}+\pi_1-\pi_W(K) &< 0. \end{aligned} \tag{B.14}$$

Rewriting (B.14) yields

$$(\gamma-1)\left[\frac{\pi_M(K)}{K}-\frac{\pi_M(n)}{n}\right] > \frac{\Pi(n)}{n}-\frac{\Pi(K)}{K}+\gamma\left[\frac{n-K}{nK}\right]\frac{\Pi(1)}{2} \text{ and} \tag{B.15}$$

$$\begin{aligned} (\gamma-1)n'[\pi_W(n')-\pi_W(K)] < \; & (K+1)\frac{\Pi(K)}{K+1}-(n'+1)\frac{\Pi(n')}{n'+1} \\ & -(K-n')\frac{\Pi(1)}{2}. \end{aligned} \tag{B.16}$$

Note that condition (B.16) holds for γ sufficiently close to 1 but the LHS of (B.16) strictly increases in γ. Moreover, the LHS of (B.16) decreases in the average output of K firms, $\frac{\Pi(K)}{K+1}$, all else equal. It remains inconclusive whether the LHS of condition (B.15) increases or decreases in γ for $\gamma \gg 1$ and increases in $\frac{\Pi(K)}{K+1}$. This means for sufficiently high γ the necessary conditions fail to hold given the production function. Moreover, for given γ and production technology, increasing γ or decreasing average output of K firms cannot make conditions (B.14) hold if they did not hold before, but it can make conditions (B.14) fail if they held before.

For part (ii) of the Proposition, conditions (B.14) can be manipulated to provide us with sufficient conditions for some n and n' firms to be preferred by both some owners and some workers to K firms. To have positive measure of both n and n' in equilibrium,

(i) both conditions (B.14) must not hold. We analyzed this case in the previous part and found that sufficiently high γ and sufficiently low average output in K firms such that $\frac{\Pi(K)}{K+1} > \frac{\Pi(n)}{n+1}$ and $\frac{\Pi(K)}{K+1} > \frac{\Pi(n')}{n'+1}$ always induce conditions (B.14) to fail.

(ii) n and n' firms must not crowd out each other and therefore conditions (B.14) where K is substituted by n' and n, respectively, must fail to hold, that is

$$\begin{aligned}
0 &\geq \frac{n}{n-n'}\frac{\pi_W(n')-\pi_W(n)}{\gamma}-\frac{\pi_M(n)-\pi_M(n')}{n-n'}+\frac{\pi_M(n')-\pi_1}{n'} \text{ and} \\
0 &\leq \frac{\gamma n'}{n-n'}[\pi_W(n')-\pi_W(n)]+\frac{\pi_M(n')-\pi_M(n)}{n-n'}+\pi_1-\pi_W(n).
\end{aligned}$$

Rewriting these conditions as in inequalities (B.15) and (B.16) yields

$$\begin{aligned}
0 &\geq (\gamma-1)\left[\frac{\pi_M(n')}{n'}-\frac{\pi_M(n)}{n}\right]+\frac{\Pi(n')}{n'}-\frac{\Pi(n)}{n} \\
&\quad -\gamma\left[\frac{n-n'}{nn'}\right]\frac{\Pi(1)}{2} \text{ and} \\
0 &\leq (\gamma-1)n'[\pi_W(n')-\pi_W(n)]+(n'+1)\frac{\Pi(n')}{n'+1}-(n+1)\frac{\Pi(n)}{n+1} \\
&\quad +(n-n')\frac{\Pi(1)}{2}.
\end{aligned}$$

That is

$$\begin{aligned}
0 &\geq \gamma\left[\frac{\Pi(n')}{n'}-\frac{\Pi(n)}{n}\right]-(\gamma-1)\left[\pi_W(n')-\pi_W(n)\right] \\
&\quad -\gamma\left[\frac{n-n'}{nn'}\right]\frac{\Pi(1)}{2} \text{ and} \\
0 &\leq (\gamma-1)n'[\pi_W(n')-\pi_W(n)]+(n'+1)\frac{\Pi(n')}{n'+1}-(n+1)\frac{\Pi(n)}{n+1} \\
&\quad +(n-n')\frac{\Pi(1)}{2}.
\end{aligned}$$

Both conditions always hold for sufficiently high γ and sufficiently high average output of n firms compared to n' firms. Additionally, positive measure of n and n' firms implies that $\underline{\omega}$ is sufficiently small and $\overline{\omega}$ sufficiently large, as $\underline{\omega} < \omega_W(n,n') < \omega_M(n,n') < \overline{\omega}$.

(iii) there must not exist any other firm size n'' crowding out n or n' firms in the sense of Lemma B.1 implying the appropriate version of conditions (B.14) for n':

$$\begin{aligned} 0 &\geq \frac{n'}{n'-n''}\frac{\pi_W(n'')-\pi_W(n')}{\gamma} - \frac{\pi_M(n')-\pi_M(n'')}{n'-n''} \\ &\quad + \frac{\pi_M(n'')-\pi_1}{n''} \text{ and} \\ 0 &\leq \gamma\frac{n'}{n'''-n'}[\pi_W(n')-\pi_W(n''')] + \frac{\pi_M(n')-\pi_M(n''')}{n'''-n'} \\ &\quad +\pi_1-\pi_W(n'), \end{aligned}$$

where $n' > n''$ and $n \neq n''' > n'$. Likewise for n it must hold that

$$\begin{aligned} 0 &\geq \frac{n}{n-n''}\frac{\pi_W(n'')-\pi_W(n)}{\gamma} - \frac{\pi_M(n)-\pi_M(n'')}{n-n''} \\ &\quad + \frac{\pi_M(n'')-\pi_1}{n''} \text{ and} \\ 0 &\leq \gamma\frac{n}{n'''-n}[\pi_W(n)-\pi_W(n''')] + \frac{\pi_M(n)-\pi_M(n''')}{n'''-n} \\ &\quad +\pi_1-\pi_W(n'''), \end{aligned}$$

where $n > n'' \neq n'$ and $n' \neq n''' < n$. The previous analysis, in particular of inequalities (B.15) and (B.16), shows that if γ is sufficiently high and average output in n' and n firms is sufficiently high compared to average output in n'' and n''' firms, the set of conditions above holds. Then positive measures of both n and n' firms emerge in equilibrium. ■

Proof of Proposition 3.3

Let (P, θ, r, t) be an equilibrium associated to the endowment distribution F induced by ω and G be a redistribution from owners to workers from F induced by ω' in the sense of Definition 3.3. Then (P', θ', r', t') denotes the equilibrium associated to $\omega'(i)$.

Assume first that the interest rate remains constant, i.e. $\mu'_U = \mu_U$ or $\int_I \omega(i)di \geq c$, as aggregate endowments do not change. By assumption $\theta(i) \neq \emptyset$ for $\omega_i = \hat{\omega}$. Let $\hat{n} = |\theta(i)|$. As

$$\mu(\omega(i) \leq \omega) \leq \mu(\omega'(i) \leq \omega) \; \forall\, \omega > \hat{\omega},$$

it must hold that $\sum_{n=\hat{n}+1}^{\overline{n}} \mu_{n_M}|_{\omega,t} \geq \sum_{n=\hat{n}+1}^{\overline{n}} \mu_{n_M}|_{\omega',t}$.[1] In particular, for $n_{max} > \hat{n}$ defined as the biggest firm size with $\mu_{n_{max}}|_\omega > 0$, it must hold that $\mu_{n_{max\,M}}|_{\omega'} < \mu_{n_{max\,M}}|_\omega$. Likewise, as additionally

$$\mu(\omega(i) \leq \omega) \geq \mu(\omega'(i) \leq \omega)\ \forall\, \omega < \hat{\omega},$$

it must hold that for all $m > \hat{n}$ $\sum_{n=m}^{\overline{n}} \mu_{W_n}|_{\omega,t} \geq \sum_{n=m}^{\overline{n}} \mu_{W_n}|_{\omega',t}$ because of negative assortative matching and the fact that agents that are workers in $\hat{n}$ firms under ω, have less measure under ω' by assumption, that is $\mu_{\hat{n}_W}|_{\omega,t,r} < \mu_{\hat{n}_W}|_{\omega',t,r}$.

Then $\sum_{n>\hat{n}} \mu_n|_{\omega,t} \geq \sum_{n>\hat{n}} \mu_n|_{\omega',t}$ if side payments and the interest rate remained constant. Note that equality only holds if ω and ω' coincide for all i with the property $|\theta(i)| \neq 1$, i.e. for all agents in $n > 1$ firms. This means both supply and demand for $n > \hat{n}$ workers weakly decrease. Equating supply and demand using side payments then gives measures $\mu_n|_{\omega',t',r}$ for which it must hold that $\sum_{n=\hat{n}+1}^{\overline{n}} \mu_n|_{\omega',t',r} \leq \sum_{n=\hat{n}+1}^{\overline{n}} \mu_n|_{\omega,t,r}$ and for n_{max} in particular that $\mu_{n_{max}}|_{\omega',t',r} < \mu_{n_{max}}|_{\omega,t,r}$. However, it holds that $\omega_U|_{\omega,t} \leq \omega_U|_{\omega',t}$. Moreover, $t \neq t'$ so that capital market clearing as defined in (3.5) may be violated. That is the interest rate may have to adjust. Given that this is only a second order effect decreasing one market side, it must hold that indeed $\sum_{n>\hat{n}} \mu_n|_{\omega,t,r} \geq \sum_{n>\hat{n}} \mu_n|_{\omega',t',r'}$ and $\mu_{n_{max}}|_{\omega,t,r} \geq \mu_{n_{max}}|_{\omega',t',r'}$ which asserts the statement in the proposition. ■

Proof of Proposition 3.4

Let us first look at a stochastic decrease inducing a uniform scarcity of workers given old side payments t and the old interest rate r. We denote measures after the decrease given t and r by primes. That is,

$$\mu_{M_n} - n\mu_{W_n} = \mu_{M_n'} - n'\mu_{W_n'} \text{ for } t, \omega', r.$$

[1]We follow the convention of writing $\mu_{M_n}|_{t,\omega}$ to indicate the measure of individuals weakly preferring to be owners of n firms given side payments t and the endowment distribution $\omega(i)$.

Then $t'(n) < t(n)$ for all n because workers are scarce. Moreover, $\omega_U \leq \omega'_U$ and capital market clearing requires $r' > r$. The cutoff endowment level to prefer owning an $n > 1$ firm to owning an $n' > n$ firm is given by

$$\begin{aligned} w_M(n,n') &= \frac{\pi_M(n)-\pi_M(n')}{(1+r')(\gamma-1)} - \frac{1}{\gamma-1}[\gamma(t(n')-(n'-n)c)-t(n)] \\ &\quad +(n+1)c, \end{aligned}$$

if $t(n) > t(n') - (n' - n)c$. For the opposite case the following holds.

$$\begin{aligned} w_M(n,n') &= \frac{\pi_M(n')-\pi_M(n)}{(1+r')(\gamma-1)} - \frac{1}{\gamma-1}[\gamma t(n)-t(n')+(n'-n)c] \\ &\quad +(n'+1)c. \end{aligned}$$

For workers the cutoff endowment level to prefer an $n > 1$ firm to an $n' > n$ firm is given by

$$w_W(n,n') = \frac{\pi_W(n') - \pi_W(n)}{(1+r')(\gamma-1)} + \frac{\gamma}{\gamma-1}\frac{t(n)}{n} - \frac{1}{\gamma-1}\frac{t(n')}{n'}$$

for $\frac{t(n')}{n'} < \underline{\omega} < \frac{t(n)}{n}$. Therefore side payments cannot decrease uniformly, i.e. $t(n) - t'(n) \neq t(n') - t'(n')$ for some $n \neq n' > 1$. Suppose the contrary. Then clearly for both owners and workers, $k = W, M$, $w_k(n,n') < w'_k(n,n')$ and $w'_k(n,n') - w_k(n,n') = w'_k(n,n'') - w_k(n,n'')$ for all $1 < n < n' < n''$. This means cutoff values increase linearly in transfers. This in turn implies that measures of owners do not decrease uniformly but those of firm sizes with higher $w_M(.,n)$, i.e. the larger firms, decrease slower. Measures of workers increase faster for firms with high per worker side payments, i.e. small firms. This means $t(n) - t'(n) < t(n') - t'(n')$ for $n < n'$ and relative side payments in small firms increase relative to those of large firms.

A similar argument applies to the case where a change in the endowment distribution leads to uniform scarcity of owners. It is straightforward that all side payments have to increase in this case. Proceeding analogously to the above reasoning it can be shown that then relative side payments of small firms decrease relative to those of large firms. ■

Proof of Proposition 3.5

The support of the firm size distribution is given by $\{n : \mu_n > 0\}$. Let the greatest firm size in the economy be denoted by $\overline{n}$. If there exist large firms in equilibrium it must hold that

$$n_{max} = \max_{n \leq n^*(r^*)} \{n : \mu_{n_W} > 0 \wedge \mu_{n_M} > 0\}.$$

To develop a necessary condition for a firm of size $n > 1$ to emerge in an equilibrium, note first that workers in and owners of equilibrium $n > 1$ firms must prefer this to being members of symmetric $n' = 1$ firms, i.e.

$$\omega_M(n,1) = \frac{\pi_1 - \pi_M(n)}{(1+r)(\gamma-1)} + c + \frac{\gamma}{\gamma-1}[nc - t(n)] \text{ and}$$
$$\omega_W(n,1) = \frac{\pi_W(n) - \pi_1}{(1+r)(\gamma-1)} + \frac{\gamma}{\gamma-1}c - \frac{1}{\gamma-1}\frac{t(n)}{n}.$$

To have an n firm in equilibrium it must necessarily hold that

$$\overline{\omega} > \omega_{n_M} > \omega_{n_W} > \omega_U.$$

Collecting terms yields

$$\overline{\omega} > \frac{\gamma n(\pi_1 - \pi_W(n)) - (\pi_M(n) - \pi_1)}{(1+r)(\gamma-1)} + \gamma n \omega_U - (\gamma n - 1)c. \quad \text{(B.17)}$$

This inequality weakens in the spread between $\overline{\omega}$ and ω_U and in the interest rate. ω_U is strictly increasing in $\underline{\omega}$ and the endowment gap $c - w$. The interest rate in turn increases in ω_U. If endowments are abundant, then $\omega_U = \underline{\omega}$ and $r = 0$, and the necessary condition reduces to a sufficiently large spread between $\underline{\omega}$ and $\overline{\omega}$. Of course, conditions analogous to condition B.17 have to hold for each $n' \neq n$ and $n > 1$. Note that the necessary conditions give implicitly an upper bound on the largest firm size in equilibrium, n_{max}. ■

Proof of Proposition 3.6

Monotonicity of incomes in endowments follows from a revealed preferences argument. Suppose agent j chooses in equilibrium to be worker, i.e. $\theta(j) = \emptyset$ in an n firm this giving rise to an upfront payment of $t(n)$. Suppose further that agent i chooses optimally some other combination of firm size and ownership rights, $(n', \theta'(i))$ leading to upfront payments greater than $t(n)$. Formally,

$$\begin{aligned} v_i(n, \theta, .) &< v_i(n', \theta', .) \text{ and} \\ v_j(n, \theta, .) &> v_j(n', \theta', .). \end{aligned} \tag{B.18}$$

Because of the credit market imperfection it must hold that

$$v_i(n, \theta, .) > v_j(n, \theta, .),$$

i.e. agent i would derive greater income from agent j's choice than agent j. On the other hand, because of the asymmetry in investment cost, by Lemma 3.1 inequalities (B.18) can only hold if

$$\omega(j) < \omega(i).$$

This argument holds for any j and any combination of firm size and ownership rights, although weak inequalities apply for comparisons between agents with equal asset ownership admitting for the case $r = 0$.

Now we turn our attention to income and endowment gaps for agents matched into n firms. Let j denote the poorest agent in the economy with endowments $\omega(j) > 0$ who is worker in an n firm. Agent j's income is then given by

$$v_j(n, \theta, .) = \pi_W(n) + (1 + r)\gamma(\omega(j) - t(n)),$$

assuming $t(n) > \omega(j)$. Let i be the richest agent with endowments $\omega(i)$ who is owner of firm n. Agent i's income is given by

$$v_i(n, \theta, .) = \pi_M(n) + (1 + r)\gamma(\omega(i) + nt(n) - (n + 1)c),$$

assuming $\omega(i) < (n + 1)c - nt(n)$. The ratio of incomes is then

$$\frac{v_i(n, \theta, .)}{v_j(n, \theta, .)} = \frac{(1+r)\gamma\omega(i)+\pi_M(n)-(1+r)\gamma((n+1)c-nt(n))}{(1+r)\gamma\omega(j)+\pi_W(n)-(1+r)\gamma t(n)}. \tag{B.19}$$

Note at this point that whenever the poorest agent with endowment $\underline{\omega} > 0$ were to remain unmatched in equilibrium, the income spread between this agent and any firm member l must exceed the endowment spread regardless of whether the firm member borrows or lends because the payoff of the firm member's equilibrium choice exceeds $(1 + r)\omega(l)$ for almost all l. A necessary and sufficient condition for

$$\frac{v_i(n, \theta, .)}{v_j(n, \theta, .)} < \frac{\omega(i)}{\omega(j)}$$

is therefore

$$\frac{\pi_M(n) - (1 + r)\gamma[(n + 1)c - nt(n)]}{\pi_W(n) - (1 + r)\gamma t(n)} < \frac{\omega(i)}{\omega(j)} \tag{B.20}$$

This immediately implies that for sufficiently low side payments such that $\pi_M(n) < (1 + r)\gamma[(n + 1)c - nt(n)]$, the above inequality holds trivially. We know that j prefers to be worker, so that

$$\pi_M(n)+(1+r)\gamma[\omega(j)-(n+1)c+nt(n)] < \pi_W(n)+(1+r)\gamma(\omega(j)-t(n)),$$

which immediately implies (B.20). This argument extends to the case where i lends and j borrows. If j lends and i borrows we can use the last inequality on an appropriately modified (B.19) yielding

$$\frac{v_i(.)}{v_j(.)} < \frac{(1 + r)\gamma\omega(i) + \pi_M - (1 + r)\gamma((n + 1)c - nt(n))}{(1 + r)\gamma\omega(j) + \pi_M - (1 + r)\gamma((n + 1)c - nt(n))} < \frac{\omega(i)}{\omega(j)}$$

for $\pi_M(n) > (1 + r)\gamma[(n + 1)c - nt(n)]$. ■

Proof of Proposition 3.7

For the model economy to have a bimodal equilibrium firm size distribution the formal requirement is that

$$\mu_i > \mu_j > 0 \text{ and } \mu_l > \mu_j \text{ with } l > j > i > 1. \tag{B.21}$$

This implies

$$\mu_{i_W} > \mu_j \text{ and } \mu_{i_M} > \mu_j$$
$$\mu_{l_W} > \mu_j \text{ and } \mu_{l_M} > \mu_j$$

Measures μ_{n_M} and μ_{n_W} are determined by cutoff endowments $\underline{\omega}_{n_W}$, $\overline{\omega}_{n_W}$ and $\underline{\omega}_{n_M}$, $\overline{\omega}_{n_M}$ as defined in Lemma 3.1. This means there must be sufficient mass of the endowment distribution within $[\underline{\omega}_{i_W}, \overline{\omega}_{i_W}]$ and $[\underline{\omega}_{i_M}, \overline{\omega}_{i_M}]$ compared to $[\underline{\omega}_{j_W}, \overline{\omega}_{j_W}]$ and $[\underline{\omega}_{j_M}, \overline{\omega}_{j_M}]$. This means the intervals for i firms must be large or the endowment distribution must have sufficient mass around the mean due to negative assortative matching.

For l firms the intervals $[\underline{\omega}_{l_W}, \overline{\omega}_{l_W}]$ and $[\underline{\omega}_{l_M}, \overline{\omega}_{l_M}]$ have to be sufficiently large compared to those for j firms as the density function is assumed to strictly decrease to the right of its mode. This means l firms have to be more attractive for both owners and workers. This means either $|K - l| < |K - j|$, in which case $|K - i| > |K - j|$ and i firms are strictly less efficient than j firms, or side payments in l firms must be sufficiently high compared to smaller firms. This means that owners of l firms have to be sufficiently scarce compared to smaller firm sizes. Sufficient relative scarcity of owners in l can only be induced by an endowment distribution that is sufficiently skewed in the sense that the left tail of the endowment distribution contains sufficient mass.

In the following we provide explicit conditions on cutoff values necessary for $\mu_j < \mu_l$ to hold from which our argumentation can be deduced. Assume first side payments $t(l)$ and $t(j)$ such that either all workers have to borrow or both side payments are negative. Then the necessary condition is

$$\mu(\underline{\omega}_{l_M} \leq \omega(i) \leq \overline{\omega}_{l_M}) > \mu(\underline{\omega}_{j_M} \leq \omega(i) \leq \overline{\omega}_{j_M}).$$

This inequality only holds if

$$(\overline{\omega}_{l_M} - \underline{\omega}_{l_M}) - (\overline{\omega}_{j_M} - \underline{\omega}_{j_M}) \tag{B.22}$$

is sufficiently great. This condition weakens considerably if $\underline{\omega}_{j_M} > \overline{\omega}_{l_M}$, as the density decreases to the right of the peak. This only happens if workers have to borrow to pay side payments in both firms, as can be seen by noticing that workers must be indifferent between firm sizes, and owning n firms must involve less investment than owning n' firms. Hence, high side payments and sufficient scarcity of owners for zero side payments facilitates the emergence of large firms. Sufficient scarcity of owners implies sufficient skewness of the endowment distribution.

Assume now side payments are negative. Then the difference (B.22) increases in $t(l)$ by the proof of Proposition 3.4. Therefore (B.21) implies that side payments in l firms have to be relatively high in comparison to those in j firms net of the different levels of profit $\pi_M(.)$.

Finally, assume that at least some agents have to borrow to afford $\frac{t(j)}{j}$, whereas $\frac{t(l)}{l} \leq \underline{\omega}$. Hence, the cutoff endowment for preferring to work in an l firm is

$$\omega_W = \frac{\pi_W(l) - \pi_W(j)}{(1+r)(\gamma-1)} + \frac{1}{\gamma-1}\left[\gamma\frac{t(j)}{j} - \frac{t(l)}{l}\right].$$

Then (B.21) implies that $\mu(i : \omega \leq \omega_W)$ has to be sufficiently large. Noting that $\omega_W > \underline{\omega} \geq 0$ and solving for $t(l)$ we obtain

$$t(l) < \gamma\frac{l}{j}t(j) - l\frac{\pi_W(l) - \pi_W(j)}{1+r}. \tag{B.23}$$

We show now that this is consistent with $\mu_{l_M} > 0$ as the cutoff endowment for owners between l and j firms is for $(j+1)c - t(j) < (l+1)c - t(l)$ given by

$$\omega_M = \frac{\pi_M(j) - \pi_M(l)}{(1+r)(\gamma-1)} + (j+1)c + \frac{\gamma}{\gamma-1}\left[(l-j)c - t(l)\right] + \frac{t(j)}{\gamma-1}. \tag{B.24}$$

To have l owners it must hold that $\omega_M < (l+1)c - t(l)$. Combining this with (B.23) yields a condition for the existence of both agents preferring to own l firms and agents preferring to work in l firms, that depends on $t(j)$ and the efficiency of l versus j firms

$$t(j) > \frac{j}{\gamma l - j}\frac{\Pi(j) - \Pi(l) + 2l\pi_W(l) - (l+j)\pi_W(j)}{1+r} + \frac{j(l-j)}{\gamma l - j}c.$$

Note that this condition is required to be sufficiently slack in order to permit a sufficiently large measure of l firms compared to j firms. This means both the general level of side payments and relative efficiency of l firms compared to j firms is required to be sufficiently great. High side payments imply sufficient skewness of the endowment distribution. This concludes the example. ■

B.2 Numerical Examples

Numerical Example for Section 3.5.3

In this subsection we provide the numerical example that generates figure 3.4. Renegotiation is assumed to be intra-firm bargaining (Stole and Zwiebel, 1996) and the payoffs are taken from Gall (2004). In equilibrium, measure consistency requires that measures of owners equate proportional measures of workers and, given transfers, there exists no blocking coalition. We restrict our attention to transfers where $t_i = t(n)$ for an owner of an N firm, $t_j = -\frac{t(n)}{n}$ for a worker in an N firm, and $t(K) = 0$ for all $i \in I$ for K firm members. A worker i has no blocking coalition if for all $n' \neq n$

$$\begin{aligned}
\pi_W(n)+(1+r)(\omega(i)-\frac{t(n)}{n}) &\geq \pi_W(n')+(1+r)(\omega(i)-\frac{t(n')}{n'}) \text{ and} \\
\pi_W(n)+(1+r)(\omega(i)-\frac{t(n)}{n} &\geq \pi_1+(1+r)\omega(i) \text{ and} \\
\pi_W(n)+(1+r)(\omega(i)-\frac{t(n)}{n} &\geq (1+r)\omega(i). \qquad \text{(B.25)}
\end{aligned}$$

For negative values of wealth net of side payments, the interest rate is given by $i = (1+r)\gamma$. For owners j it must hold that

$$\begin{aligned}
\pi_M(n)+(1+r)(\omega(j)+t(n)-(n+1)c) \geq \pi_M(n')+(1+r)(\omega(j)+ \\
t(n')-(n'+1)c) \,\forall\, n' \neq n
\end{aligned}$$

and

$$\pi_M(n)+(1+r)(\omega(j)+t(n)-(n+1)c) \geq \pi_1+(1+r)\omega(j). \qquad \text{(B.26)}$$

Again this expression is for positive net wealth and has to be modified accordingly if j has to borrow. The set of conditions (B.25) and

(B.26) combined with labor and credit market clearing, (3.9) and (3.5), provide necessary conditions that we exploit to find equilibrium side payment schemes and interest rate by means of simulation.

The algorithm takes as inputs side payments in $n = 2$ firms and an interest rate guess. Then all other side payments are calculated as to make the poorest agent indifferent between working in any $n > 1$ firm. We proceed by clearing the labor markets associated to each firm size sequentially, starting with the one with lowest side payments, i.e. the largest firm size. A market n is cleared by varying side payments in all $n' < n$ markets such that workers in those markets remain indifferent until market n clears. Thereby NAM in equilibrium (as under part (i) of Proposition 3.2) can be exploited as clearing market n does not affect markets $n' > n$[2]. Special attention is required by the algorithm to the fact that excess demand for labor for a given firm size may be set valued as agents may be indifferent between two firm sizes. In this case the measure of indifferent agents has to be assigned to both markets associated to the firm sizes such that the market where side payments are lower possibly clears.

This procedure yields market clearing side payments for all but the smallest firm size and thus an aggregate excess demand for labor for any value of $n = 2$ side payments. Repeating this step and varying $n = 2$ side payments appropriately approximates the matching equilibrium given an interest rate. For the credit market to clear this procedure has to be repeated varying the interest rate appropriately until all conditions in (3.9) and (3.5) are satisfied.

The parameters of the simulation of a drop of capital on the economy are as follows. The endowment density function $\phi(x)$ is parameterized as

$$\phi(x) = \frac{1}{G}\left[\frac{1}{2}x(gx^{g-l}\exp(-gx) + (1 - \frac{1}{2}x))(\beta\lambda^{\beta}x^{\beta-1}\exp(-(\lambda x)^{\beta}))\right]$$

[2]This is the reason why admitting partial ownership generates certain difficulties for numerical simulation. Typically, clearing markets sequentially causes circles as owners of asset II tend to sit in the middle of the endowment distribution. This leaves a nonlinear system of equations in excess demand correspondences which can be quite tedious to solve.

where G is defined such that $\int_{\underline{\omega}}^{\overline{\omega}} f(x) = 1$ given all other parameters. The increase in endowments corresponds to a change in parameters as reported in Table B.1.

The production function depending on the number of workers in a firm, n, is assumed to be

$$f(n) = (hnh)^{0.68 - 0.06\sqrt{|n-K|}}.$$

with $h = 0.90$ and $K = 3$. Cost of an asset is $c = 0.1500$ and the degree of credit market imperfection is $\gamma = 3.00$. Then we obtain the results in Table B.1.

	before	after	Results	before	after
c	0.1500	0.1500	mean endowment	0.1479	0.1498
γ	3.0000	3.0000	r	0.1900	0.1715
$\underline{\omega}$	0.0000	0.0125	$n = 1$ firms	0.0826	0.0701
$\overline{\omega}$	1.0000	1.0125	$n = 2$ firms	0.0673	0.0618
β	1.9000	1.3000	$K = 3$ firms	0.0546	0.0439
λ	10.4500	13.0000	$n = 4$ firms	0.0284	0.0271
g	2.9000	2.2000	$n = 5$ firms	0.0432	0.0603
l	3.6000	2.5000	mean output	0.4216	0.4204

Table B.1: Simulation Details for Figure 3.4

Numerical Example for Section 3.6

In order to provide some evidence of robustness of our findings to Assumption 3.2, we conduct a numerical example that generates figure 3.5. Renegotiation is again assumed to be intra-firm bargaining (Stole and Zwiebel, 1996) and the payoffs are taken from Gall (2004). However, we replace Assumption 3.2 by the following assumption on technology

Assumption A B.1 *Let $f(A_n, h_n) = 0$ for $n > 3$ and $K = 2$. Moreover, $\frac{f(A_3,h_3)}{4} < \frac{f(A_1,h_1)}{2}$. Joint ownership is not feasible.*

Thus we can limit our attention to $n < 4$ productive coalition yet admitting for all ownership constellations except for joint ownership. Table B.2 lists the possible combinations of ownership rights:

Coalition Size n	Ownership Rights	Identifier
1	$(\{A^I\}, \{A^{II}\})$	A
2	$(\{A^I, A^{II}, A^{II}\})$	B
2	$(\{A^I\}, \{A^{II}, A^{II}\})$	C
3	$(\{A^I, A^{II}, A^{II}, A^{II}\})$	D
3	$(\{A^I\}, \{A^{II}, A^{II}, A^{II}\})$	E
3	$(\{A^I, A^{II}\}, \{A^{II}, A^{II}\})$	F

Table B.2: Ownership Structures

Coalitions of type F will not emerge in equilibrium as they are dominated by type A coalitions under our assumptions. To see this let side payments in an F coalition be denoted by $t_W(F)$ for the workers' payments and $t_O(F)$ for the payments of the owner of two assets A^{II}, henceforth denoted by O. Note first that in order to have both workers and managers in F coalitions, the following condition on side payments, derived analogously to Proposition 3.2, has to hold:

$$\begin{aligned} & 2c - \frac{\pi_M(F) - \pi_W(F)}{1+r} - \frac{3}{2} t_W(F) \leq t_O(F) \\ \Leftrightarrow \quad & t_O(F) \leq 2c - \frac{\pi_M(F) - \pi_W(F)}{(1+r)\gamma} - \frac{3}{2} t_W(F). \qquad \text{(B.27)} \end{aligned}$$

To have both workers and owners of two assets A^{II} the following is required to hold

$$\frac{t_W(F)}{2} - 2c + \frac{\pi_O(F)}{(1+r)\gamma} \leq t_O(F) \leq \frac{t_W(F)}{2} - 2c + \frac{\pi_O(F)}{1+r}. \qquad \text{(B.28)}$$

Combining (B.27) and (B.28), a necessary condition for workers, managers, and owners to emerge in equilibrium is given by:

$$c - \frac{\frac{\Pi(F)}{4} - \pi_W(F)}{1+r} \leq \frac{t_W(F)}{2} \leq c - \frac{\frac{\Pi(F)}{4} - \pi_W(F)}{(1+r)\gamma}. \qquad \text{(B.29)}$$

Finally, we must have that in equilibrium some agents prefer to be workers in F coalitions to being members of $n = 1$ coalitions. This implies

$$\frac{t_W(F)}{2} \leq c - \frac{\pi_2 - \pi_W(F)}{1+r}.$$

Plugging this into (B.29) and some algebra yield a simple necessary condition for emergence of type F coalitions:

$$\frac{\Pi(A)}{2} = \pi_2 \leq \frac{\Pi(F)}{4}.$$

This condition is violated by Assumption B.1. Hence we do not need to consider F coalitions in equilibrium.

Moreover, it can be shown that investments for managers and owners in coalition types C and E must be identical, as their payoffs in the renegotiation game are identical as well. This leaves us with a vector of just four side payments $(t_W(B), t_W(C), t_W(D), t_W(E))$ that need to support the labor market equilibrium. This allows us to use the above simulation approach.

We use the same endowment distribution as in the case above. The production function depending on the number of workers in a firm, n, is assumed to be

$$f(n) = (hnh)^{0.6}.$$

with $h = 1.25$ implying $K = 2$. Cost of an asset is $c = 0.3000$ and the degree of credit market imperfection is $\gamma = 3.00$. The results are reported in Table B.3.

Identifier	A	B	C	D	E	unempl.
Measure	0.0135	0.0328	0.0139	0.0580	0.0000	0.6009

Table B.3: Simulation Results for Figure 3.5

Appendix C

Mathematical Appendix to Chapter 4

Proof of Proposition 4.2

As a first step, the following lemma translates the GID condition into our notation.

Lemma C.1 *$\phi(.)$ has GID if $g(.)$ has (weakly) increasing differences and for all $a, b \in A$ and $c, b \in B$ with $x_a > x_b$ and $x_c > x_d$ if and only if all of the following conditions hold:*

(i) for $v > g(x_a, x_d)$:

$$\begin{aligned} g(x_a, x_c) - g(x_b, x_c) \quad \geq \quad & \delta_b(g(x_b, x_c) - g(x_b, x_d) \\ & +\delta_b^{-1}(v - g(x_b, x_d), d), c) \\ & -\delta_a(g(x_a, x_c) - g(x_a, x_d) \\ & +\delta_a^{-1}(v - g(x_a, x_d), d), c). \end{aligned}$$

(ii) for $g(x_b, x_d) < v < g(x_a, x_d)$:

$$\begin{aligned} g(x_a, x_c) - g(x_b, x_c) \quad \geq \quad & \max\{-\delta_a(g(x_a, x_c) - \bar{\phi}, c), \\ & \delta_c^{-1}(\bar{\phi} - g(x_a, x_c), a)\} \\ & +\delta_b(g(x_b, x_c) - g(x_b, x_d) + \\ & \delta_b^{-1}(v - g(x_b, x_d), d), c). \end{aligned}$$

where $\bar{\phi} = g(x_a, x_d) + \delta_d(g(x_a, x_d) - v, a)$.

(iii) for $v < g(x_b, x_d)$*:*

$$\begin{aligned} g(x_a, x_c) - g(x_b, x_c) \quad \geq \quad & \max\{-\delta_a(g(x_a, x_c) - \bar{\phi}, c), \\ & \delta_c^{-1}(\bar{\phi} - g(x_a, x_c), a)\} \\ & - \max\{-\delta_b(g(x_b, x_c) - \bar{\bar{\phi}}, c), \\ & \delta_c^{-1}(\bar{\bar{\phi}} - g(x_b, x_c), b)\} \end{aligned}$$

where $\bar{\bar{\phi}} = g(x_b, x_d) + \delta_d(g(x_b, x_d) - v, b)$.

Proof of Lemma: Let $v > g(x_a, x_d)$ and apply (4.11) to the definition of GID. Then GID is equivalent to

$$\begin{aligned} & g(x_a, x_c) + \delta_a(g(x_a, x_c) - g(x_a, x_d) + \delta_a^{-1}(v - g(x_a, x_d), d), c) \\ & \geq g(x_b, x_c) + \delta_b(g(x_b, x_c) - g(x_b, x_d) + \delta_b^{-1}(v - g(x_b, x_d), d), c). \end{aligned}$$

Rearranging yields statement (i) in the lemma. Now choose $v \in [g(x_b, x_d), g(x_a, x_d)]$. GID holds if and only if

$$\begin{aligned} & g(x_a, x_c) + \delta_a(g(x_a, x_c) - g(x_a, x_d) - \delta_d(g(x_a, x_d), a) - v, c) \\ & \geq g(x_b, x_c) + \delta_b(g(x_b, x_c) - g(x_b, x_d) + \delta_b^{-1}(v - g(x_b, x_d), d), c), \end{aligned}$$

provided $g(x_a, x_c) - g(x_a, x_d) < \delta_d(g(x_a, x_d) - v, d)$. Note that from $g(x_b, x_d) < v < g(x_a, x_d)$ it cannot be concluded whether this is true. Otherwise the equivalent condition is given by

$$\begin{aligned} & g(x_a, x_c) - \delta_c^{-1}(g(x_a, x_d) - g(x_a, x_c) + \delta_d(g(x_a, x_d) - v, a), a) \\ & \geq g(x_b, x_c) + \delta_b(g(x_b, x_c) - g(x_b, x_d) + \delta_b^{-1}(v - g(x_b, x_d), d), c). \end{aligned}$$

Rearranging all these conditions gives statement (ii) in the lemma. Finally, if $v < g(x_b, x_d)$, there are four cases depending on whether

$$\begin{aligned} g(x_a, x_d) + \delta_d(g(x_a, x_d) - v, a) \quad > \quad & g(x_a, x_c) \text{ and/or} \\ g(x_b, x_d) + \delta_d(g(x_b, x_d) - v, b) \quad > \quad & g(x_b, x_c). \end{aligned}$$

Let

$$\begin{aligned} g(x_a, x_d) + \delta_d(g(x_a, x_d) - v, a) \quad < \quad & g(x_a, x_c) \text{ and} \\ g(x_b, x_d) + \delta_d(g(x_b, x_d) - v, b) \quad > \quad & g(x_b, x_c). \end{aligned}$$

Then

$$-\delta_a(g(x_a,x_c)-g(x_a,x_d)-\delta_d(g(x_a,x_d)-v,a),c)$$
$$-\delta_c^{-1}(g(x_b,x_d)+\delta_d(g(x_b,x_d)-v,b)-g(x_b,x_c),b)<0.$$

But $g(x_a,x_c)>g(x_b,x_c)$ and therefore in this case GID always holds. The other three cases are ambiguous and must be included in the lemma. ■

PAM is clearly efficient if $g(x_a,x_b)$ has (weakly) increasing differences as the TU matching equilibrium coincides with solution to social planner's problem and exhibits PAM. A sufficient condition for PAM in equilibrium is GID of $\phi(.)$ as in Legros and Newman (2004a). By inspection of the conditions in Lemma C.1 it can be checked by simple calculation, using the respective condition on v in some cases, that all hold given the assumption on δ_i. As an example we look at the case (ii) of Lemma C.1. Let v be such that the decisive condition is given by

$$\begin{aligned} g(x_a,x_c)-g(x_b,x_c) \quad &\geq \quad \delta_c^{-1}(g(x_a,x_d)+\delta_d(g(x_a,x_d)-v,a) \\ &\quad -g(x_a,x_c),a)+\delta_b(g(x_b,x_c)-g(x_b,x_d) \\ &\quad +\delta_b^{-1}(v-g(x_b,x_d),d),c). \end{aligned}$$

By monotonicity of δ_i we know that the right hand side is smaller than

$$\delta_c^{-1}(\delta_d(g(x_a,x_d)-v,a),a)+\delta_b(\delta_b^{-1}(v-g(x_b,x_d),d),c).$$

Using the assumption $\delta_c(u,i)>\delta_d(u,i)$ for all $u\in\mathbb{R}$ and $i\in A$ this in turn must be less than or equal to

$$g(x_a,x_d)-g(x_b,x_d)<g(x_a,x_c)-g(x_b,x_c).$$

This last inequality follows from the ID property of $g(.)$. Lemma C.1 gives a sufficient condition for GID of $\phi(.)$ when $g(.)$ has WID and this verifies the proposition. ■

Proof of Lemma 4.1

First note that u_a can be expressed in terms of u_b as follows.

$$u_a = u_b + k(x_b) - k(x_a).$$

However, x_a depends on u_b by the relation

$$u_b + k(x_b) = g(x_a, x_b). \tag{C.1}$$

Let $x_b > 0$, then $g(x_a, .)$ is continuous and monotone. Hence, $g(x_a, .)$ is a one-to-one mapping of the form $g : X \mapsto Y$ where

$$Y = [g(\inf_X x_a, x_b), g(\sup_X x_a, x_b)].$$

It is a well-known fact that then its inverse $g^{-1} : Y \mapsto X$ exists and is a continuous function. Furthermore we know that under this condition g^{-1} is differentiable at all $x \in X$, $g'(x) \neq 0$ which must hold due to the Inada conditions. Its derivative is then given by

$$\frac{\partial g^{-1}(x)}{\partial x} = \frac{1}{\frac{\partial g(x_a, x_b)}{\partial x_a}}.$$

This implies the first part of the lemma as

$$\begin{aligned} \frac{\partial u_a(u_b)}{\partial u_b} &= 1 - \frac{\partial k(g^{-1}(u_b + k(x_b)))}{\partial g^{-1}(u_b + k(x_b)))} \frac{\partial g^{-1}(u_b + k(x_b))}{\partial u_b} \\ &= 1 - \frac{\partial k(x_a)}{\partial x_a} \left(\frac{\partial g(x_a, x_b)}{\partial x_a} \right)^{-1}, \end{aligned} \tag{C.2}$$

where we used the identity $g^{-1}(u_b + k(x_b)) = x_a$. Moreover, $u_a(u_b)$ is clearly continuously differentiable. Differentiating (C.2) with respect to u_b and exploiting convexity of $k(.)$ as well as differentiating (C.1) yields

$$\frac{\partial^2 u_a}{\partial u_b^2} - \frac{g'(x_a, x_b) \frac{\partial^2 k(x_a)}{\partial x_a^2} \frac{\partial x_a}{\partial u_b} - k'(x_a) \frac{\partial^2 g(x_a, x_b)}{\partial x_a^2} \frac{\partial x_a}{\partial u_b}}{(g'(x_a, x_b))^2} < 0.$$

This establishes the second part of the lemma. ■

Proof of Proposition 4.4

Assume some agent $a \in A$ rationally expects to be matched with agent $b \in B$ with attribute $\bar{x}_b$ in equilibrium. Assume further agent b needs to be given at least utility $\bar{u}_b$ to accept agent a. By Definition 3, δ_i, $i = a, b$, is piecewise differentiable with $\delta_i'(u, j) \leq 1$ whenever it is differentiable. Define u_b^* implicitly by

$$\frac{\partial u_a(u_b^*)}{\partial u_b} = -1. \tag{C.3}$$

By concavity of $u_a(u_b)$ from Lemma 4.1, u_b^* is a singleton, suppose $u_{max} \geq \bar{u}_b > u_b^*$ where u_{max} is defined by $0 = u_a(u_{max})$. If x_a is chosen efficiently, generating utility $u_a(u_b^*)$ for a at zero side payments, and b is given utility $\bar{u}_b$, agent a's utility is given by

$$u_a(u_b^*) - \delta_a^{-1}(\bar{u}_b - u_b^*, b).$$

If x_a is not chosen efficiently, generating utility $u_a(u_b')$ at zero side payments with $u_b' > u_b^*$ and giving b utility $\bar{u}_b$, then agent a's utility is given by

$$\begin{aligned} &u_a(u_b') - \delta_a^{-1}(\bar{u}_b - u_b', b) \text{ for } \bar{u}_b \geq u_b', \\ &u_a(u_b') + \delta_b(u_b' - \bar{u}_b, a) \text{ for } \bar{u}_b < u_b'. \end{aligned}$$

Agent a optimally chooses x_a efficiently if and only if

$$\begin{aligned} &u_a(u_b^*) - u_a(u_b') \geq \delta_a^{-1}(\bar{u}_b - u_b^*, b) - \delta_a^{-1}(\bar{u}_b - u_b', b), \\ &u_a(u_b^*) - u_a(u_b') \geq \delta_a^{-1}(\bar{u}_b - u_b^*, b) + \delta_b(u_b' - \bar{u}_b, a) \text{ resp.} \end{aligned} \tag{C.4}$$

for all $u_b^* < u' \leq u_{max}$. Define $\Delta u_b = u_b' - u_b^*$ and rewrite (C.4) to get this necessary condition

$$\frac{u_a(u_b^*) - u_a(u_b^* + \Delta u_b)}{\Delta u_b} \geq \frac{\delta_a^{-1}(\bar{u}_b - u_b^*, b) - \delta_a^{-1}(\bar{u}_b - u_b^* - \Delta u_b, b)}{\Delta u_b}$$

for all $0 < \Delta u_b < u_{max} - u_b^*$. As the proof focuses on Δu_b sufficiently small to derive a contradiction the case $u' > \bar{u}_b$ is omitted. Note that by Lemma 4.1

$$\lim_{\Delta u_b \to 0} \frac{u_a(u_b^*) - u_a(u_b^* + \Delta u_b)}{\Delta u_b} = 1.$$

This and concavity of $u_a(u_b)$ imply that there exists Δu_b sufficiently small such that for $\epsilon > 0$

$$\frac{u_a(u_b^*) - u_a(u_b^* + \Delta u_b)}{\Delta u_b} = 1 + \epsilon.$$

By Definition of $\mathcal{D}^k$, $k = A, B$, δ has weakly decreasing differences so that

$$\frac{\delta_a^{-1}(\bar{u}_b - u_b^*, b) - \delta_a^{-1}(\bar{u}_b - u_b^* - \Delta u_b, b)}{\Delta u_b} \geq 1.$$

If $\delta_a^{-1}(\bar{u}_b - u_b^*, b) - \delta_a^{-1}(\bar{u}_b - u_b^* - \Delta u_b, b) > \Delta u_b$ we can approximate the expression using the first derivative of δ_a at a point $\hat{u} \leq \bar{u}_b - u_b^*$ where δ_a is differentiable:

$$\frac{\delta_a^{-1}(\bar{u}_b - u_b^*, b) - \delta_a^{-1}(\bar{u}_b - u_b^* - \Delta u_b, b)}{\Delta u_b} \geq \frac{\partial \delta_a^{-1}(\hat{u})}{\partial u}$$

Choose now

$$0 < \epsilon < \frac{\partial \delta_a^{-1}(\hat{u})}{\partial u} - 1.$$

This yields a contradiction to (C.4). Conversely, for $\delta_a^{-1}(\bar{u}_b - u_b^*, b) - \delta_a^{-1}(\bar{u}_b - u_b^* - \Delta u_b, b) = \Delta u_b$ with Δu_b sufficiently large, inequality (C.4) always holds for both $u_b' > \bar{u}_b$ and $u_b' \leq \bar{u}_b$. The case $0 \leq u_b < u_b^*$ can be treated simultaneously although δ_b does not need not to be inverted and the inequality signs need to be reversed. ■

Bibliography

Aghion, P. and P. Bolton: 1997, 'A Theory of Trickle-Down Growth and Development'. *Review of Economic Studies* **64**, 151–172.

Akerlof, G. A.: 1970, 'The Market for "Lemons": Quality Uncertainty and the Market Mechanis'. *Quarterly Journal of Economics* **84**(3), 488–500.

Angrist, J. D.: 1995, 'The Economic Returns to Schooling in the West Bank and Gaza Strip'. *American Economic Review* **85**(3), 1065–1087.

Banerjee, A. V.: 2002, 'Contracting Constraints, Credit Markets and Economic Development'. *MIT Dept. of Economics Working Paper* **02-17**.

Banerjee, A. V. and E. Duflo: 2004, 'Growth Theory Through the Lens of Development Economics'. *forthcoming in Handbook of Development Economics.*

Banerjee, A. V. and K. Munshi: 2004, 'How Efficiently is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur'. *Review of Economic Studies* **71**(1), 19–42.

Banerjee, A. V. and A. F. Newman: 1993, 'Occupational Choice and the Process of Development'. *Journal of Political Economy* **101**(2), 274–298.

Barro, R. J. and J.-W. Lee: 1993, 'International Comparisons of Educational Attainment'. *Journal of Monetary Economics* **32**(3), 363–394.

Becker, G. S.: 1973, 'A Theory of Marriage: Part I'. *Journal of Political Economics* **81**(4), 813–846.

Benhabib, J. and M. M. Spiegel: 1994, 'The Role of Human Capital in Economic Development: Eveidence from Aggregate Cross-Country Data'. *Journal of Monetary Economics* **34**(2), 143–174.

Besley, T.: 1995, 'Property Rights and Investment Incentives: Theory and Evidence from Ghana'. *Journal of Political Economy* **103**(5), 903–937.

Besley, T., S. Coate, and G. C. Loury: 1993, 'The Economics of Rotating Savings and Credit Associations'. *American Economic Review* **83**(4), 792–810.

Biggs, T., M. Shah, and P. Srivastava: 1995, 'Technological Capabilities and Learning in African Enterprises'. *World Bank Technical Paper* **288**.

Binmore, K., A. Rubinstein, and A. Wolinsky: 1986, 'The Nash Bargaining Solution in Economic Modelling'. *RAND Journal of Economics* **17**(2), 176–188.

Bolton, P. and M. D. Whinston: 1993, 'Incomplete Contracts, Vertical Integration, and Supply Assurance'. *Review of Economic Studies* **60**, 121–148.

Caballero, R. J. and M. L. Hammour: 1996, 'The Macroeconomics of Specifity'. *NBER Working Paper Series* **5757**.

Cabral, L. M. B. and J. Mata: 2003, 'On the Evolution of the Firm Size Distribution: Facts and Theory'. *American Economic Review* **93**(4), 1075–1090.

Chowdhury, T.: 1994, *Small Firms Informally Financed*, Chapt. 2, pp. 11–18. World Bank Discussion Paper 253 (eds. R. Islam, J. D. von Pischke, and J. M. de Waard).

Coase, R. H.: 1960, 'The Problem of Social Cost'. *Journal of Law and Economics* **3**, 1–44.

Cockx, B. and M. Dejemeppe: 2002, 'Doe the Higher Educated Unemployed Crowd Out the Lower Educated Ones in a Competition for Jobs?'. *IZA Discussion Paper* **541**.

Cole, H. L., G. J. Mailath, and A. Postlewaite: 2001, 'Efficient Noncontractible Investments in Large Economies'. *Journal of Economic Theory* **101**, 333–373.

Cooley, T., R. Marimon, and V. Quadrini: 2004, 'Aggregate Consequences of Limited Contract Enforceability'. *CEPR Discussion Paper Series* **4173**.

Cullen, J. B. and S. D. Levitt: 1999, 'Crime, Urban Flight, and the Consequences for Cities'. *Review of Economics and Statistics* **81**(2), 333–373.

de Meza, D. and B. Lockwood: 1998, 'Does Asset Ownership always Motivate Managers? Outside Options and the Property Rights Theory of the Firm'. *Quarterly Journal of Economics* **113**(2), 361–386.

Duflo, E.: 2001, 'Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment'. *American Economic Review* **91**(4), 795–813.

Evans, D. S. and B. Jovanovic: 1989, 'An Estimated Model of Entrepreneurial Choice under Liquidity Constraints'. *Journal of Political Economy* **97**(4), 808–827.

Felli, L. and K. Roberts: 2002, 'Does Competition Solve the Hold-up Problem?'. *CEPR Discussion Paper Series* **3535**.

Field, E.: 2003, 'Entitled to Work: Urban Property Rights and Labor Supply in Peru'. *Mimeo Harvard University.*

Freeman, R. B.: 1976, *The Overeducated American.* Academic Press, New York.

Friedman, M.: 1953, 'Choice, Chance, and the Personal Distribution of Income'. *Journal of Political Economy* **61**(4), 277–290.

Gall, T.: 2004, 'Intra-firm Bargaining Outcomes with Property Rights on Assets and Human Capital Investment'. *mimeo, University of Mannheim.*

Gall, T., P. Legros, and A. F. Newman: 2005, 'Timing Education'. *Unpublished Working Paper.*

Galor, O. and J. Zeira: 1993, 'Income Distribution and Macroeconomics'. *Review of Economic Studies* **60**, 35–52.

Garratt, R. and J. M. Marshall: 1994, 'Public Finance of Private Goods: The Case of College Education'. *Journal of Political Economy* **102**(3), 566–582.

Garrett, T. A.: 2001, 'An International Comparison and Analysis of Lotteries and the Distribution of Lottery Expenditures'. *International Review of Applied Economics* **15**(2), 213–227.

Ghatak, M., M. Morelli, and T. Sjöström: 2001, 'Occupational Choice and Dynamic Incentives'. *Review of Economic Studies* **68**, 781–810.

Grossman, S. and O. Hart: 1986, 'The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration'. *Journal of Political Economy* **94**(5), 691–719.

Grüner, H. P. and R. Schils: 2002, 'Capital Redistribution and the Market Allocation of Firm Ownership'. *CEPR Discussion Paper Series* **3130**.

Hallward-Driemeier, M., G. Iarossi, and K. L. Sokoloff: 2002, 'Exports and Manufacturing Productivity in East Asia: A Comparative Analysis with Firm-level Data'. *NBER Working Paper Series* **8894**.

Hart, O.: 1995, *Firms, Contracts, and Financial Structure.* Oxford University Press, Oxford.

Hart, O. and J. Moore: 1990, 'Property Rights and the Nature of the Firm'. *Journal of Political Economy* **98**(6), 1119–1157.

Hartog, J.: 2000, 'Over-education and Earnings: Where Are We, Where should We Go?'. *Economics of Education Review* **19**(2), 131–147.

Hirshleifer, J.: 1989, 'Conflict and Rent-seeking Success Functions: Ratio vs. Difference Models of Relative Success'. *Public Choice* **63**, 101–112.

Hirshleifer, J.: 2000, 'The Macrotechnology of Conflict'. *The Journal of Conflict Resolution* **44**(6), 773–792.

Hopenhayn, H. A.: 1992, 'Entry, Exit, and Firm Dynamics in Long Run Equilibrium'. *Econometrica* **60**(5), 1127–1150.

Jovanovic, B.: 1982, 'Selection and the Evolution of Industry'. *Econometrica* **50**(3), 649–670.

Kaneko, M. and M. H. Wooders: 1986, 'The Core of a Game with a Continuum of Players and Finite Coalitions: the Model and Some Results'. *Mathematical Social Sciences* **12**, 105–137.

Kaneko, M. and M. H. Wooders: 1996, 'The Nonemptiness of the f-Core of a Game Without Side Payments'. *International Journal of Game Theory* **25**, 245–258.

Kihlstrom, R. E. and J.-J. Laffont: 1979, 'A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion'. *Journal of Political Economy* **87**(4), 719–748.

Klein, B., R. G. Crawford, and A. A. Alchian: 1978, 'Vertical Integration, Appropriable Rents and the Competitive Contracting Process'. *Journal of Law and Economics* **21**(2), 297–326.

Krueger, A. B. and M. Lindahl: 2001, 'Education for Growth: Why and For Whom?'. *Journal of Economic Literature* **39**, 1101–1136.

Legros, P. and A. F. Newman: 1996, 'Wealth Effects, Distribution, and the Theory of Organization'. *Journal of Economic Theory* **70**, 312–341.

Legros, P. and A. F. Newman: 2002, 'Monotone Matching in Perfect and Imperfect Worlds'. *Review of Economic Studies* **69**(4), 925–942.

Legros, P. and A. F. Newman: 2004a, 'Beauty is a Beast, Frog is a Prince'. *Working Paper.*

Legros, P. and A. F. Newman: 2004b, 'Competing for Ownership'. *Working Paper.*

Little, I. M. D., D. Mazumdar, and J. John M. Page: 1988, *Small Manufacturing Enterprises: A Comparative Analysis of India and Other Economies.* Oxford University Press, Oxford.

Lucas, R. E.: 1978, 'On the Size Distribution of Business Firms'. *BELL Journal of Economics* **9**(2), 508–523.

Ludwig, J., G. J. Duncan, and P. Hirschfield: 2001, 'Urban Poverty and Juvenile Crime: Evindence from a Randomized Housing-mobility Experiment'. *Quarterly Journal of Economics* **166**(2), 655–680.

Mailath, G. J. and A. Postlewaite: 2004, 'Social Assets'. *PIER Working Paper Series* **04-025**.

Mailath, G. J., A. Postlewaite, and L. Samuelson: 2004, 'Sunk Investments Lead to Unpredictable Prices'. *PIER Working Paper Series* **04-007**.

Marshall, J. M.: 1984, 'Gambles and the Shadow Price of Death'. *American Economic Review* **74**(1), 73–86.

Matsuyama, K.: 2000, 'Endogenous Inequality'. *Review of Economic Studies* **67**, 743–759.

Neary, H. M.: 1997, 'Equilibrium Structure in an Economic Model of Conflict'. *Economic Inquiry* **35**(3), 480–494.

Ng, Y. K.: 1965, 'Why Do People Buy Lottery Tickets? Choices Involving Risk and the Indivisibility of Expenditure'. *Journal of Political Economy* **73**(5), 530–535.

Peters, M. and A. Siow: 2002, 'Competing Pre-marital Investments'. *Journal of Political Economy* **110**, 592–608.

Piketty, T.: 1997, 'The Dynamics of the Wealth Distribution and the Interest Rate with Credit Rationing'. *Review of Economic Studies* **64**, 173–189.

Pischke, J.-S. and T. von Wachter: 2005, 'Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation'. *CEPR Discaussion Paper Series* **5105**.

Rajan, R. G. and L. Zingales: 1998, 'Power in a Theory of the Firm'. *Quarterly Journal of Economics* **113**(2), 387–432.

Scott, F. and J. Garen: 1993, 'Probability of Purchase, Amount of Purchase, and the Demographic Incidence of the Lottery Tax'. *Journal of Public Economics* **54**, 121–143.

Shubik, M. and M. H. Wooders: 1983, 'Approximate Cores of Replica Games and Economies. Part II: Setup-Cost and Firm Formation in Coalition Production Economies'. *Mathematical Social Sciences* **6**, 285–306.

Simon, H. A. and C. P. Bonini: 1958, 'The Size Distribution of Business Firms'. *American Economic Review* **48**(4), 607–617.

Skaperdas, S.: 1996, 'Contest Success Functions'. *Economic Theory* **7**, 283–290.

Stiglitz, J. E. and A. Weiss: 1981, 'Credit Rationing in Markets with Imperfect Information'. *American Economic Review* **71**(3), 393–410.

Stole, L. A. and J. Zwiebel: 1996, 'Intra-firm Bargaining under Nonbinding Contracts'. *Review of Economic Studies* **63**, 375–410.

Tybout, J. R.: 2000, 'Manufacturing Firms in Developing Countries: How Well They Do, and Why?'. *Journal of Economic Literature* **38**(2), 11–44.

van Smoorenburg, M. S. M. and R. K. D. van der Velden: 2000, 'The Training of School-leavers Complementarity or Substitution?'. *Economics of Education Review* **19**(2), 207–217.

Westermark, A.: 2003, 'Bargaining, Binding Contracts, and Competitive Wages'. *Games and Economic Behavior* **43**, 296–311.

Williamson, O. E.: 1975, *Markets and Hierarchies: Analysis and Antitrust Implications.* The Free Press MacMillan, New York.

Williamson, O. E.: 2000, 'The New Institutional Economics: Taking Stock, Looking Ahead'. *Journal of Economic Literature* **38**, 595–613.

Worthington, A. C.: 2001, 'Implicit Finance in Gambling Expenditures: Australian Evidence on Socioeconomic and Demographic Tax Incidence'. *Public Finance Review* **29**(4).

Zohir, S. C.: 1994, *Small Firms Informally Financed*, Chapt. 8, pp. 83–97. World Bank Discussion Paper 253 (eds. R. Islam, J. D. von Pischke, and J. M. de Waard).

Hiermit erkläre ich, daß ich die Dissertation selbständig angefertigt und mich anderer als der in ihr angebenen Hilfsmittel nicht bedient habe, insbesondere, daß aus anderen Schriften Entlehnungen, soweit sie in der Dissertation nicht ausdrücklich als solche gekennzeichnet und mit Quellenangaben versehen sind, nicht stattgefunden haben.

Mannheim, den 19.5.2005 Thomas Gall

Curriculum Vitae

Personal Details

Full Name	Thomas Eduard Gall
Born	18.06.1977 in Gräfelfing, Germany

Education

1987-1996	Max-Planck-Gymnasium, Munich, Germany
1997-2001	Undergraduate Studies in Economics at Ludwig-Maximilians-Universität, Munich, Germany
1999-2000	Erasmus Grant in Economics at Universitat Pompeu Fabra, Barcelona, Spain
2001-2003	Graduate Studies in Economics at CDSEM, University of Mannheim, Germany
2003-2004	ENTER Fellow and Teaching Assistant in Economics at University College London, UK
2004-2005	Teaching Assistant at the Chair of Economic Policy, University of Mannheim, Germany