Perfect Matching and Search in Economic Models

JAN EECKHOUT

Thesis submitted for the degree of Doctor of Philosophy

The London School of Economics and Political Science

July 1997

Abstract

This thesis uses general matching techniques - both perfect matching and search - to study some problems in economies that are characterised by heterogeneity of their agents. Here, matching in its broadest sense is interpreted as a form of trade that is strictly limited between two partners: transactions are one-to-one, between one buyer and one seller exactly.

The first part proposes a framework that integrates two well documented strands of the existing economic literature. It is a search model that generalises the frictionless perfect matching model to a context where trade does not occur instantaneously. A general methodology with proof is given which allows us to derive the unique equilibrium allocation of agents. Though the limit case without friction reproduces the perfect matching result, with friction results deviate substantially from conclusions in both the perfect matching literature and the search literature.

The second part of the thesis concentrates entirely on frictionless matching models. First, a general class of preferences is identified that yields a unique allocation. Second, the matching model is studied when endogenous choice of characteristics is allowed and has an intuitive application to the labour market. It is shown that in the presence of job heterogeneity, too many resources are spent in order to achieve a higher ranked job. The results, including issues of turnover and distribution, are verified with some stylised facts in the empirical literature. Finally, a model that mimics a matching equilibrium and that allows for endogenous choice of characteristics is applied to the context of education in the labour market. It is shown that multiple equilibria can exist in the presence of spillovers in production.

Contents

List of Figures

Acknowledgements

I would like to thank my supervisor Professor Kevin Roberts. He has generously given me expert advice and consistently guided me in the more fruitful direction. I have been greatly influenced by his thinking. But above all, his concern and empathy have given me the feeling of security and confidence from the start until this stage of the project.

I have also benefitted from discussions and comments on different parts of this thesis from a number of people. I would like to thank in particular Steve Alpern, Tim Besley, Francis Bloch, Michele Boldrin, Kenneth Burdett, David Cass, Marco Celentani, Melvyn Coles, Leonardo Felli, Raquel Fernandez, Francisco Ferreira, Richard Freeman, Maitreesh Ghatak, Urs Haegler, Nobuhiro Kiyotaki, George Mailath, Marco Manacorda, Alan Manning, John Hardman Moore, Gustavo Nombela, Canice Prendergast, Rafael Rob, Sherwin Rosen, Lones Smith, Randall Wright.

I have received financial support in the form of a Junior Fellowship from the Royal Economic Society and a grant from the Britisch Council.

My thesis was mainly written while I was at the Centre for Economic Performance. I would like to thank Richard Jackman for encouraging me to enrol in the programme. The Centre provided a warm atmosphere and I remain with numerous friendships and good memories. I have also enjoyed the generous hospitality of Universidad Carlos III de Madrid and the European University Institute in Florence in the Summers of 95 and 96 respectively.

Several intimate and close friends have kept supporting me throughout, often longer than was good for them. I would like to dedicate this thesis to them.

Introduction

Many behavioural and market interactions studied by economists are characterised by a one-to-one transaction. In a monogamic marriage market for example, exactly one partner of one sex "trades" with one partner of the other sex. An individual's decision in the market is the choice of a partner, rather than the mere choice of the quantity consumed from whoever obtained. Associated with a partner is a consumption level. Similarly, the one-to-one transaction accurately describes job markets. Firms and workers both choose exactly one partner to engage in production. This also applies to a large number of more traditional economic choice situations. The choice of a college for education, for example, or the purchase of housing: an additional unit (an extra course or an additional square meter) cannot be bought from a separate seller. All these transactions are characterised by the same feature: exchange is between two partners only.

In this thesis, the central theme is to study several economic phenomena in the presence of markets where transactions are one-to-one. These markets will be referred to as matching markets. They deviate from the standard Walrasian market in the sense that the one-to-one trading feature involves a non-trivial allocation decision. The choice of the trading partner, i.e. the characteristics of whom to trade with, will have an effect on the equilibrium outcome. In a matching market, agents are not anonymous. This implicitly assumes some sort of heterogeneity between trading partners. If that were not the case, all agents would be entirely indifferent between different potential candidates. In the Walrasian market, arbitrage across different agents exactly cancels out ex-ante heterogeneity. A second deviating feature is that with non-trivial allocation decisions, agreement about trade is bilateral: both parties have to be willing to trade. Both features, heterogeneity and the bilateral agreement of a transaction, are present in the Walrasian market, but because of trivial allocation decisions, agents are indifferent. The indifference is entirely generated by the price mechanism that has equal unit prices across agents. In the presence of Bertrand competition, the outside option of an epsilon unit disciplines prices across agents.

Matching is of course not new. There is a long tradition in economic research dealing with matching. Two main strands of the economic literature have independently focussed on many related issues: the work on perfect matching and the search literature. Perfect Matching was introduced by Gale and Shapley [27] in the early sixties. They observed some strong behavioural regularities in the market for hospital physicians. Their pioneering work involved modelling this matching market. In subsequent research[1], it was concluded that the preference structure is of crucial importance for the equilibrium allocation, more so than the presence of transferability of utility[2]. The standard model, whether it be with or without transferable utility, features both the ingredients mentioned above: bilateral agreement and heterogeneity. The second strand of the literature, search or imperfect matching (i.e. matching with frictions), acknowledges that trading opportunities arrive at a certain cost. The implication is that the intrinsic utility from consumption is no longer a perfect indicator of value. Value in addition is determined by the cost of creating the trading opportunity. Such a cost can be interpreted as the cost of gathering information, or as the opportunity cost of waiting for the right trading opportunity. Search does not intrinsically and necessarily involve heterogeneity and bilateral agreement[3]. This

[1]For an overview, see Roth and Sotomayor [62].

[2]As in the model proposed by Becker [7].

[3]See for example the early literature on search: Mortensen [49], Diamond [17].

however implies there is no decision of choice: upon meeting, a trade *always* takes place. The subsequent search literature has incorporated both these features separately[4].

The aim of this thesis is to make contributions in two areas. In Part I, a unifying approach is proposed to both strands of the literature, perfect matching and search. The objective is to establish how the characteristics of a generalised matching model with search frictions differ from the existing results in the literature. In Part II, perfect matching models are studied. This thesis tries to make a contribution in the area of endogenous choice of characteristics of heterogeneous types. This is new in the literature and the presence of matching proves to yield results which differ substantially from the Walrasian benchmark. Throughout, the theory is employed to explain economically relevant phenomena. It is constantly argued that matching provides a good description of many phenomena, and that it highlights issues of heterogeneity which are not present in the standard neoclassical approach. It provides a theoretically rigorous framework in which thinking about heterogeneity is natural. The applications then are always to be interpreted in terms of distribution of heterogeneous types. The main underlying social agenda is to study the effect of distributional considerations on efficiency.

Part I consists of two Chapters. The first Chapter describes the model and derives the main results. The second Chapter considers some applications of the model with surprising results. A search model is proposed of the marriage market between two sets, males and females, and it incorporates the

[4]Heterogeneity has been used, amongst others, in Jovanovic [32], Diamond [18] and Pissarides [56]. Bilateral agreement is prominent in the literature on money and search (Kiyotaki and Wright [33]) where a double coincidence of wants is necessary for trade. In addition, this presupposes some particular form of heterogeneity, horizontal heterogeneity.

two main features present in perfect matching models (i.e. without search frictions). Bilateral search (i.e. search by *both* the males *and* the females) and ex-ante vertical heterogeneity (some types are preferred to others and all agents rank types identically) ensure that this model is the generalised version of the perfect matching model. It differs from traditional search models because it *jointly* incorporates both these features. As mentioned above, one approach (Kiyotaki and Wright [33]) does have both features, but their specific notion of horizontal heterogeneity (i.e. each type prefers the types nearest to her own type) implies that all agents have an identical strategy. The novelty of the results is threefold. First, a general methodology follows from the proof of existence and uniqueness of the equilibrium. The method allows for the derivation of the equilibrium allocation whatever the utility function. This includes the emergence of disconnected sets of matching for certain preferences. Second, a set of preferences is derived (i.e. satisfying multiplicative separability) that results in the partitioning of both distributions. Third, it is shown that the limit case of the search model with the friction disappearing is the perfect matching model. This is new in the sense that it deviates entirely from the existing models of search. Not only are existing models not the generalised version of the perfect matching model, they could never exhibit phenomena such as disconnected sets since heterogeneity only exists ex-post[5]. In addition, our model differs in more than one respect from research that was conducted simultaneously and that emerged after our results were found[6]. First, the proof is more general and in addition, it provides an intuitive method for solving for the equilibrium allocation. Second, the model allows us to show the equivalence with the perfect matching

[5]See for example Diamond [18] and Pissarides [56].

[6]See for example McNamara and Collins [42], Burdett and Coles [13], Bloch and Ryder [12], Sattinger [64] and Smith [69].

model. Third, it has a more general result on partitioning than Burdett and Coles [13][7]. One result that appears in the other work and that is absent in ours is the derivation of multiple endogenous distributions of singles.

In the second Chapter, the model of Chapter 1 is used to study two applications that have surprising results. First, it is shown that bachelors may exist. Bachelors are the lowest types of one of the two distributions that remain eternally unmatched. This is surprising since in the perfect matching literature[8], no one remains unmatched if both populations are of equal size (and provided a match yields more utility than being single). The result here is due to the difference in average length of search of both sexes. If one sex, say the females, on average searches longer than the males, then the lowest male types will never be matched. They involuntarily remain bachelors. This follows from the fact that matches are pairwise and that as a result, the number of agents matched per unit of time is equal in both sets. The second application abandons the assumption that matched partners are drawn from two disjoint sets and considers pairwise matching from one set. The example used to illustrate the argument is the matching of tennis players as sparring partners. The result derived is that for certain preferences, it may be the case that some types refuse to play with types identical to themselves. The reason is that when higher types are more impatient, they will be willing to accept matches with low types. Such a low type, being more patient, can afford to wait until the more preferred higher types arrive and can refuse to match with equals. This provides one reason why you may not want to be a member of the club that wants you as a member.

Part II abandons matching with frictions entirely and concentrates on

[7]Though the result on multiplicative separability has also been found by Smith [69].

[8]See amongst others Becker [7] and Roth and Sotomayor [62].

perfect matching models. In Chapter 3, an extension is made to the existing literature on perfect matching with ordinal preferences[9]. This literature has derived a large number of results on the existence of stable matchings for general preferences. The contribution of this Chapter is to identify a class of preferences for which the stable matching is unique. This is useful from a purely theoretical point of view, since quite a large share of the non-cooperative game theory is concerned with uniqueness. More importantly however is that the class of preferences identified is wide and includes the ones with the most economic relevance. Moreover, the result bears some resemblance to single-peakedness, even though this is not entirely equivalent in a model with agents from disjoint sets who have preferences over different objects (i.e. the types of the other set). This Chapter also makes a more philosophical point. It is shown that if assortative mating is defined on the preferences, then an allocation can never be negatively assorted. It is argued that assortative mating, i.e. the mating of *likes*, necessarily has to be defined on the individuals' preferences.

Chapter 4 considers the perfect matching model with transferable utility. This does not differ substantially from the model with ordinal preferences, but it allows for the derivation of pay-off functions from a joint surplus that is split between the partners of a match. The pay-off functions themselves are entirely derived from the surplus function and the allocation in equilibrium for a given surplus. Becker [7] shows that for a surplus function which has a positive cross partial derivative with respect to the types of both sets (i.e. they are strategic complements), the equilibrium exhibits positive assortative mating. The allocation is negatively assorted with a negative cross partial and agents are indifferent if the cross partial is zero. It is derived in Chapter

[9]This literature is reviewed in Roth and Sotomayor [62].

4 that in this case, the distribution of types enters the pay-off function. This follows from the argument made above: in a matching market with one-to-one transactions, the allocation is non trivial. The innovative contribution is to consider the endogenous choice of the characteristics of a type. The results that are shown have a strong bearing on the dependence of the pay-off function on the distribution of types. In this Chapter this framework is embedded in the labour market where workers match with jobs.

The central premise is that a distribution of heterogeneous jobs exists. The implication is that a worker's productivity differs in different jobs. This deviates from most of the standard analysis where it is assumed that productivity is entirely embodied in the worker. This implies that the allocation of workers to jobs is non trivial. Once matched to a job, a worker chooses the level of effort to exert. It is shown that in a repeated game, and with perfectly observable productivities, the level of effort is super optimal. This is the case if current effort affects the future type: performing well now makes a worker more productive in the future. In effect, current effort endogenously determines the future characteristic which in turn determines the future equilibrium allocation. The inefficiency result is entirely in contrast with the neoclassical model of effort choice, where the equilibrium level is optimal. It does bear some resemblance to the reputation and principle-agent literature[10], but the results stand with perfect observability! Here, the result is derived from the rank dependence of the pay-off function in a matching model. This follows from the non trivial allocation of workers to jobs. With endogenous choice of characteristics, workers play a Rank-Order Tournament: current effort increases the rank in the future distribution of types. In addition, the policy implications would be the opposite: a tax

[10]See Holmström [31], Mirrlees [46],

in the reputation model reduces the revelation of information and hence is inefficient; an income tax in the matching model, which exactly off-sets the rank effect, improves efficiency.

The matching framework with endogenous choice of characteristics (i.e. the worker's ability in the labour market) has a nice interpretation. Effort is viewed as a tool to gain promotion, i.e. to get a better job. Two further results are shown. First, inequality has an ambiguous effect on the choice of effort. Second, a higher rate of turnover increases inefficient effort. All these results are matched with a number of stylised facts from the empirical effort supply literature.

Finally, Chapter 5 maintains the endogenous choice of characteristics in a labour market environment. The characteristics are productive ability and an education technology exists which can augment the ability. The environment is not explicitly modelled as a matching model, but it has the characteristics of positive assortative mating. This is in fact an assumption, rather than an equilibrium outcome. Rather than matching, production occurs in a monopolistically competitive industry. Such an industry generates spillovers by the mere size of the industry. As a result, individual investment in education will not take into account the social return. It is shown that multiple mobility equilibria can exist. Further, sufficient conditions are derived for the distribution of types to exhibit a continuous increase in the spread.

Part I

Search

Chapter 1

Bilateral Search and Vertical Heterogeneity

This Chapter provides a framework that generalises the Perfect Matching model into a Imperfect Matching model with search frictions. The central characteristics of the perfect matching model of the marriage market are the presence of heterogeneity of types and the bilateral decision: only when there is a double coincidence of wants between a male and a female with heterogeneous preferences will they engage in marriage. Introducing a search friction certainly does not decrease the true representation of some markets. After all, in a marriage market for example, it may not be too difficult to find a spouse, but in order to do as well as possible, some longer (costly) search may be optimal. The objective is to show existence of equilibrium and to characterise the equilibrium allocation in the presence of frictions. Without frictions, the two-sided Perfect Matching model where agents are vertically heterogeneous (i.e. there exists a ranking of the types) has an allocation that exhibits positive assortative mating. The highest ranked female is matched with the highest ranked male, etc. This is due to the bilateral nature of acceptance. The second highest ranked male would like to be matched with the highest type female, but she will not accept marriage. She can do better being matched with the highest type male. With search frictions, the top female will accept males over a certain range, since waiting too long is costly.

The main contribution of this Chapter is to show that provided the distribution of singles is stationary, a (Nash) equilibrium allocation exists and is unique. This is true for any specification of the utility function. This result is surprising in two respects. First, uniqueness. The strategies of one sex are monotonic in the strategies of the other sex, so a continuum of strategies would be expected. However, an argument of iterated elimination of dominated strategies only leaves one strategy to survive. Second, existence. Whatever preferences are assumed, the allocation can always be found using the recursive elimination method. This can give rise to the existence of quite unexpected matching sets.

Given existence and uniqueness, three additional results are derived. First, the distribution of types is endogenously partitioned for preferences that are multiplicatively separable. This is unexpected since preferences are type-dependent, while by definition of endogenous partitioning, strategies are not. Second, for some preferences, matching sets are disconnected. This implies that you are rejected when you propose a match with some types, even though you are accepted by both higher and lower types. Finally, the model is shown to be robust. With frictions disappearing, the equilibrium allocation coincides with the equivalent allocation in perfect matching.

There are substantial differences with parallel results, both in the existing literature and in simultaneously conducted research. A considerable number of authors have looked at this problem for a specific utility function where the utility derived is equal to the type matched with. The papers by Mc-

Namara and Collins [42], Burdett and Coles [13], and Bloch and Ryder [12] all have these specific utiltity functions. All of them derive the partitioning result since their preferences are a limit case of multiplicative separability. The result derived here applies to a more general class of utility functions. The general partitioning result has independently been discovered by Smith [69]. Smith also looks at general preferences but provides a different proof and solution. The main novelty of the approach here is the intuitive appeal of the proof and its wide applicability to any utility specification. The fact that the equilibrium is shown to exist in a strong concept like iterated dominance provides significant behavioural foundations for both the resulting equilibrium allocation and the method or algorithm of obtaining it.

This appealing and intuitive method and solution is derived under the assumption of a stationary distribution of singles. A similar approach is adopted in McNamara and Collins [42], and Bloch and Ryder [12]. Endogenising the distribution in itself does not pose any problem (this is done in the Appendix). The problem is to find a fixed point for the equilibrium distribution and this goes at the expense of the intuitive derivation of the equilibrium allocation[1]. Burdett and Coles [13] show that due to thick market externalities, for some parameter values multiple steady state distributions can be supported in equilibrium. Our contribution is to show that given a distribution of singles, the allocation is unique for any preferences.

The generalisation of the perfect matching model to the search model is very much modelled in the tradition of the literature. The main aspect however is that both sides of the market search (i.e. there is bilateral search), and that there exists an ex ante heterogeneity of the types. In this marriage model, individuals of one sex will meet potential partners at random and

[1]Smith [69] provides a proof for a fixed point of the endogenous distribution of singles.

there are only a limited number of meetings per unit of time. Given perfect information, the type of the potential partner is observed upon meeting and can be accepted or rejected. If the type is too low, it may pay to wait until a higher type is met. Accepting, however, only implies that a match materialises provided there is a double coincidence of wants (i.e. the other party decides to accept as well). In the presence of bilateral search, the decision to form a match cannot be enforced unilaterally. The utility derived from a match is represented by any cardinal utility function that satisfies Vertical Heterogeneity. The model considered features Non-Transferable Utility.

In this framework, agents will choose strategies to accept or reject potential partners that come along in order to maximise the value function of searching. Entirely counterintuitive, the uniqueness and existence result derives from the fact that the equilibrium solution can be solved for, using an iterated strict dominance argument. The intuition is that with Vertical Heterogeneity, the top types of both sexes are most desired by all, so they can be assured to be accepted by all types. Hence, they have an iterated strict dominant strategy. Given these strategies, this argument equally applies to the next but top types, etc. In the presence of search frictions, they have to accept a range of types with positive mass, so that a finite number of iterations will suffice.

The basic model is presented in Section 1.1. Even though the marriage vocabulary prevails most dominantly in this Chapter, the model is easily generalisable to trading or labour market metaphors. In Section 1.2, the model is solved and it is shown that a unique iterated strict dominance equilibrium exists and that for multiplicatively separable utility functions the Steady State distributions are endogenously partitioned. Section 1.3 provides the intuition behind the elimination procedure and discusses several possible equilibrium outcomes in function of the pay-off specification. The equivalence of the model with the Gale-Shapley-Becker model is rigorously proved in Section 1.4. Some concluding remarks are made in Section 1.5. The Appendix provides a proof for the main Proposition and derives the endogenous distribution of singles. It is also shown that even with multiplicatively separable utility functions, strategies are type dependent out of Steady State and that the Steady State "never arrives".

1.1 The Basic Model

Consider two disjoint sets of infinitely lived individuals: females and males. They are intrinsically heterogeneous in type. Only one dimension of heterogeneity will be considered, so that their type can be represented by one variable θ. This type can be interpreted as a measure of either beauty, wealth, sexual attraction, etc. or as a composite measure of all those characteristics. Females and males are distinguished by θ_f and θ_m respectively. Both populations of singles are cumulatively distributed according to $F_i(\theta)$ over $\Theta_i = [\underline{\theta}_i, \bar{\theta}_i]$, $i \in \{f, m\}$ (with $f_i(\theta)$ the density function) and have equal measure one.

Individuals can be in two possible states. They can either be matched to a partner or they can be single. When single they are looking for a partner to be matched to. Partners of a different sex meet randomly, and upon meeting they can perfectly observe the type of the other sex. At that moment, they will decide whether to accept or reject a match with the partner met. A match is materialised only when both partners decide to accept each other. The decision is bilateral and cannot be enforced unilaterally.

Being single is a dire state. It does not yield any intrinsic utility. The state of being matched on the other hand brings all potential pleasure that exists in this world. It is modelled as an instantaneous utility derived at the moment the match is formed, i.e. when both individuals decide to accept the match the marriage is instantaneously "consum(mat)ed". The non-transferable utility to an individual of sex i characterised by type θ_i from being matched to a type θ_j is $u_i(\theta_j, \theta_i)$, with u continuous and in general no symmetry is required $u_i \neq u_j$. Preferences exhibit Vertical Heterogeneity, $u_{\theta_j} > 0$. This implies that there is a ranking of the types of the other sex on which all individuals agree. All men agree that Juliette Binoche is the most beautiful woman and all women have no doubts about who is the least endowed man. Note that utility is type dependent without any restrictions. Showing an equilibrium exists in the presence of a general utility specification is exactly the objective of this paper. Clearly, utility is cardinal, since a search model intrinsically puts a cardinal value on the time of search. The general utility specification allows for any cardinal value of the vertically heterogeneous preference orderings as long as the values are bounded: $u(\underline{\theta}_i) > 0$, $u(\bar{\theta}_i) < \infty$[2].

Typically, in a search environment it is recognised that the instantaneous utility from "consumption" does not measure the exact satisfaction, since it does not take into account the (in)direct cost incurred during search. In a search model with prices, price is no longer a perfect indicator of the value, as is the case in the neoclassical model. Search models do however use the neoclassical tools by collapsing instantaneous utility and search costs into Value functions, using some mechanistic representation of a environment with friction. Here, a constant returns to matching[3] search technology is specified

[2] In what follows, the notation $u(\cdot)$ may be used to signify $u_i(\cdot, \theta_i)$ when there is no confusion possible.

[3] Constant returns to matching implies that the number of meetings in the market is

as follows. When single, an individual bumps into someone of the other sex with probability β. This arrival rate β is distributed according to a Poisson process. Infinitely lived agents are not matched for life. With probability α, a match is dissolved[4]. For the purpose of this chapter, on-the-job search, endogenous separation and polygamy are ruled out.

Crucially, not all potential partners met will yield a match. In the first place, an individual may not be entirely satisfied with the type of the other sex and will prefer to search further until a more preferred type is met. Second, an individual may be very willing to enter a match, but the potential partner may wish to postpone the match. An individual's strategy will be determined subject to being accepted, so in the first instance, a strategy of an individual will be determined taking the strategies of all other players as given. An equilibrium will then be a rule such that an agent maximises the value function taking into account that all other agents adopt such a maximising strategy.

An individual's optimising strategy will be derived from maximising the value functions V_0 and V_1 in both possible state, the value for being single and matched respectively. They will in general be different depending on the type θ_i. The value functions of both states are written in the form of Bellman equations which give the current option value, given a positive interest rate r.

$$rV_0(\theta_i) = \beta \max E_{\theta_j}[0, u(\theta_j, \theta_i) + V_1(\theta_i) - V_0(\theta_i) \mid \text{given acceptance by } \theta_j] \tag{1.1}$$

proportional to the number of individuals searching. As a result, the number of meetings per person is constant.

[4]Modelling finitely lived agents with an exogenous inflow of new births yields the same results.

$$rV_1(\theta_i) = \alpha[V_0(\theta_i) - V_1(\theta_i)] \tag{1.2}$$

Note that being single has a positive option value associated with it even though being single does not yield any intrinsic instantaneous utility. The reason is of course that there exists the probability of being matched at some future point in time. When single, a potential partner is met with probability β. The type of the partner is randomly drawn from the pool of singles. Provided the type θ_j accepts the match, the instantaneous utility derived is $u(\theta_j, \theta_i)$. Marriage will be proposed if being matched to θ_j yields a higher utility than the value of looking further until a more preferred type is met. This is the case when $u(\theta_j, \theta_i) + V_1(\theta_i)$ is higher than the expected value of remaining single $V_0(\theta_i)$. Since separation occurs with fixed probability α, the option value of being matched is given by α times the residual value of switching from being matched to being single.

The decision of an individual of type θ_i is either to accept or reject a type θ_j that is met. We will represent this by the binary variables $\pi_i(\theta_j, \theta_i)$ which is defined as $\pi_i(\theta_j, \theta_i) = 1$, if a match with θ_j is accepted by θ_i, and $\pi_i(\theta_j, \theta_i) = 0$, if it is rejected. Clearly, acceptance does not necessarily imply that a match materialises, given the bilateral nature of the decision to form a match. A type of the other sex θ_j accepts a type θ_i if $\pi_j(\theta_i, \theta_j) = 1$. Whether a type θ_i is accepted is given by the inverse function of $\pi_j(\theta_i, \theta_j)$. Hence, once any potential trading partner is met, the match is materialised with probability $\psi_i(\theta_i)$

$$\psi_i(\theta_i) = \int_{\Theta_j} \pi_i(x, \theta_i)\pi_j(\theta_i, x)dF_j(x) \tag{1.3}$$

where $F_j(\theta_j)$ is the cumulative density function of singles in the market.

Remark 1 *Since the whole population is not single at the same time, the* *measure of people searching is not equal to the measure of the population. More importantly, since in general not all types have the same strategy (i.e. the strategy is type dependent), the distribution of singles $F_i(\theta_i)$ is not equal to the distribution of the entire population, say $H_i(\theta_i)$. In the Appendix, the relation between the H and F is derived. All our results go through with an endogenously derived distribution of singles, both in the steady state and out of the steady state.*

There are however two reasons why our results are derived under an exogenously given distribution of singles F. First, we do not have a proof for a fixed point of this distribution. Second, there is a source of multiplicity of steady state equilibria which is independent of the potential multiplicity this paper shows not to exist. Burdett and Coles [13] provide an example where separate beliefs about the steady state distribution can be supported in equilibrium. This multiplicity is due to thick market externalities in the search technology, very much as in Diamond [18]. The main contribution of Proposition 1 below is to show that given a distribution of singles (of which more than one may exist), there exists a unique equilibrium allocation. Below, it will become apparent that that in itself is a most nontrivial result.

Equation (1.1) can now be rewritten in terms of the binary variables π_i and π_j and the distribution of single males and females $F(\theta_i)$ and $F(\theta_j)$.

$$rV_0(\theta_i) = \beta \int_{\Theta_j} \pi_i(x,\theta_i)\pi_j(\theta_i,x)[u(x,\theta_i) + V_1(\theta_i) - V_0(\theta_i)]dF_j(x) \qquad (1.1')$$

An individually optimal solution to (1.1') and (1.2) for a type θ_i is a strategy of acceptance π_i such that (s)he is indifferent between remaining single and being matched. An equilibrium requires that individuals use a strategy such that they accept all matches for which the values of being matched is higher than the value of remaining single. In what follows, this will be referred to as a reservation strategy: a type θ_j is offered marriage if $u(\theta_j, \theta_i) + V_1(\theta_i) - V_0(\theta_i) \geq 0$. The reservation strategy implies that for each θ_i there must then be a critical type $\theta_j = \phi_j(\theta_i)$ which solves the equation

$$u(\phi_j) \geq V_0(\theta_i) - V_1(\theta_i) \tag{1.5}$$

Remark 2 *The reservation strategy restricts the strategy space since it rules out strategies where lower type males choose to reject a high type female because they know they will be rejected anyway. That would yield a degenerate equilibrium where everyone rejects everyone. Because we impose the strategy "accept all types for which the expected value of a match is higher than the marginal type", high types cannot be rejected strategically.*

An optimal strategy $\pi_i(\theta_j, \theta_i)$ will be determined in function of the critical type ϕ_j associated with the strategy (1.5). An Imperfect Matching Equilibrium can now be defined using the notion of Nash equilibrium. It is a list of optimising strategies taking into account that all other agents use their optimising strategy.

Definition 1 *For given distributions of singles F_i and F_j, an Imperfect Matching Equilibrium is a list $(\pi_i(\theta_j, \theta_i), \pi_j(\theta_i, \theta_j))$, $\forall \theta_i \in \Theta_i$, $\forall \theta_j \in \Theta_j$ satisfying:*

1. *Equations (1.1') and (1.2);*
2. *The reservation strategy (1.5).*

1.2 The Results: Existence and Uniqueness

In this Section, the main result of existence and uniqueness of the Imperfect Matching Equilibrium is derived. For that purpose, three Lemmas are shown.

Lemma 1 claims that a reservation strategy implies all types θ_j above the critical type ϕ_j are accepted and all the types below are rejected. Lemma 2 proceeds to prove that there is a unique reservation strategy holding the strategy of all other players constant. Lemma 3 shows that the reservation strategy has an upper and a lower bound. With these Lemmas, the main result in Proposition 1 can be shown.

Lemma 1 provides the relation between the strategy π_i and the reservation type ϕ_j.

Lemma 1 *A reservation strategy $\pi_i(\theta_j, \theta_i)$ satisfies:*

$\pi_i(\theta_j, \theta_i) = 1$, *if* $\theta_j \geq \phi_j(\theta_i)$;

$\pi_i(\theta_j, \theta_i) = 0$, *if* $\theta_j < \phi_j(\theta_i)$.

Proof. $\phi_j(\theta_i)$ has to satisfy the reservation strategy $u(\phi_j, \theta_j) \geq V_0(\theta_i)$. Since $u_{\theta_j} > 0$ and $\frac{\partial[V_0(\theta_i) - V_1(\theta_i)]}{\partial\theta_j} = 0$, the Lemma is always (never) satisfied for $\theta_j \geq (<)\phi_j(\theta_i)$. It follows that any $\theta_j \geq (<)\phi_j(\theta_i)$ will be accepted (rejected), so that $\pi_i(\theta_j, \theta_i) = 1 (= 0)$. ■

Remark 3 *The use of the decision variables π_i may at this stage appear cumbersome notation, since from Lemma 1, strategies are monotonic in θ_j, so that $\pi_i = 1$ always constitutes a connected set in θ_j. However, not only do we need $\pi_i(x, \theta_i)$, but also its inverse $\pi_i(\theta_j, x)$. In general, $\pi_i = 1$ is not a connected set in θ_i. As a result, with the decision variables the calculation of integrals can be made without knowing the internal boundaries of the disconnected sets.*

Lemma 2 shows that, given the strategies of all other players, the reservation strategy is unique. Using the reservation strategy (1.5) equations (1.1') and (1.2) can be rewritten

$$T_i(\phi_j) = (r + \alpha)u(\phi_j) - \beta\gamma_i(\phi_j) = 0 \tag{1.6}$$

with $\gamma_i(\phi_j) = \int_{\Theta_j} \pi_i(x,\theta_i)\mu_i(x,\theta_i)[u(x,\theta_i) - u(\phi_j)]dF_j(x)$. The first-order condition (1.6) embodies the trade-off made by every individual agent. With a reservation strategy, all types above the reservation type $\phi_j(\theta_i)$ are accepted and a match is materialised if you are accepted by these types. Given acceptance, Vertical Heterogeneity implies that the higher the reservation value, the higher the expected value of the match. The cost of increasing the reservation value though is that the probability of leaving the pool of singles decreases: being more choosy means that (utilityless) waiting times increase. In the limit, the prince(ss) of your dreams arrives with probability zero, hence the expected time of being single is infinite and utility is zero. Moreover, without a direct search cost, the opportunity cost of waiting is utility foregone while you could be matched to a partner. Solving (1.6) yields a critical type $\phi_j(\theta_i)$, $\forall\theta_i$, and hence a reservation strategy $\pi_i(\theta_j,\theta_i)$, $\forall\theta_i$. Lemma 2 shows it is unique.

Lemma 2 *Given π_j, and for $\phi'_j = \max\{\theta_j \in \Theta_j \mid \pi_j(\theta_j) = 1\}$:*

(i) ϕ_j is the unique solution to $T_i(\phi_j) = 0$;

(ii) $\phi_j \leq \phi'_j$.

Proof. First, it follows from the definitions of $\psi_i(\theta_i)$ and ϕ'_j that for $\phi_j > \phi'_j$, $\psi_i = 0$ and as a result $\gamma_i = 0$. Since $u(\theta_j) > 0$, $\forall\theta_j$ (from $u_{\theta_j} > 0$ and $u(\underline{\theta}_j) > 0$), it follows that $T_i(\phi_j) > 0$ for $\phi_j > \phi'_j$ and that there is no solution to $T_i(\phi_j) = 0$ in $(\phi'_j, \bar{\theta}_j)$. Next, $(T_i)_\phi = (r+\alpha)u_\phi - \beta\gamma_\phi > 0$, $\forall\phi_j \in \Theta_j$ since $\frac{\partial\gamma}{\partial\phi_j} = -u_{\theta_j}(\phi_j,\theta_i)\int_{\Theta_j}\pi_i(x,\theta_i)\pi_i(\theta_i,x)dF_j(x) < 0$ and $u_{\theta_j} > 0$. Given that $T_i(\phi_j) > 0$ for $\phi_j > \phi'_j$ and that $(T_i)_\phi > 0$, a solution to $T_i(\phi_j) = 0$ is in $[\underline{\theta}_j, \phi'_j]$. This establishes (ii) $\phi_j \leq \phi'_j$.

Since T_i is strictly increasing, the solution is unique. If $T_i(\underline{\theta}_j) \leq 0$, there is an interior solution. If $u(\phi_j) \geq V_0(\theta_i) - V_1(\theta_i)$ holds with strict inequality for some θ_i, there is no interior solution $T_i(\phi_j) = 0$. An optimising agent will then choose the unique ϕ_j as the minimum $\theta_j \in \Theta_j$, satisfying the reservation strategy. This maximises the expected value $V_0(\theta_i) - V_1(\theta_i)$ and the solution is a corner solution. This establishes (i) ϕ_j is the unique solution to $T_i(\phi_j) = 0$. ■

The proof of uniqueness of the reservation value is made using the fact that the value function is monotonic in the reservation value. Part (ii) of the Lemma shows that the reservation value cannot be above the highest type that is willing to accept you. On the other hand, if there is no interior solution below that, the solution is the corner solution $\underline{\theta}_j$. Together with Lemma 1, it then follows that any type above the reservation type is accepted. Like a unique optimal response in a normal form game does not imply a unique Nash equilibrium, uniqueness of the reservation strategy, given the strategies of all other players does not imply equilibrium is unique. This is illustrated in Figure 1.1. The lower graph is the reservation strategy of all types θ_i: above the reservation type, all θ_j are accepted, below they are rejected. The upper graph is the reservation value for all types θ_j: to the right of the graph, all θ_i are accepted, to the left all are rejected. Clearly, left of θ_i^* not all θ_j are willing to accept a match. Only the types θ_j below the upper graph (i.e. inverse of the reservation strategy of the types θ_j) will accept. The vertical distance between the two graphs is then the range over which matches are materialised. Measured over the distribution F_j it determines the probability of acceptance ψ_j.

To establish the nature of an equilibrium, it has to be determined how the strategy of a player changes in the presence of a change in the strategies of all other players, i.e. a change in the upper schedule. Intuitively, the lower schedule will move upwards in the presence of an upward schedule of

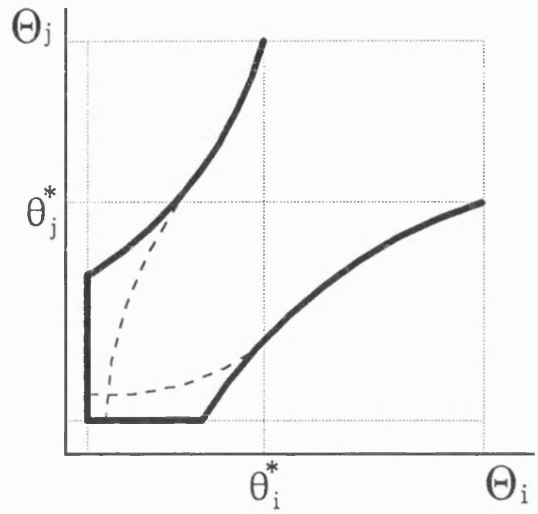

Figure 1.1: Reservation Strategies

the upper one. This is shown rigorously in Lemma 3(ii). As a result, many equilibria may be envisaged: one unique response for each strategy of the other players. However, Lemma 3(i) proves that agents accept matches with strictly positive mass (if not so, they will become matched with probability zero). This implies that, given the strategies of all other players, there is an upper bound to the reservation strategy. Part (iii) then provides the proof that if accepted by some positive mass, there is also a lower bound to the reservation strategy.

Remark 4 *For the remainder of the paper, the following notation is used.*

- $\pi_i^1(\theta_k) \geq \pi_i^2(\theta_k)$, $k \in \{i,j\}$ *means that for a given* θ_{-k}, $\pi_i^1(\theta_j,\theta_i) \geq \pi_i^2(\theta_j,\theta_i)$, $\forall\theta_k$ *and with strict inequality for some* k *with positive mass;*

- $\pi_i^1(\theta_k) = \pi_i^2(\theta_k)$ *if for a given* θ_{-k}, $\pi_i^1(\theta_j,\theta_i) = \pi_i^2(\theta_j,\theta_i)$, $\forall\theta_k$*;*

- $(\theta_i^1, \theta_j^1) < (\theta_i^2, \theta_j^2)$ *if at least one of the following two equations holds with strict inequality:* $\theta_i^1 < \theta_i^2$ *or* $\theta_j^1 < \theta_j^2$*;*

- $(\theta_i^1, \theta_j^1) = (\theta_i^2, \theta_j^2)$ *if both* $\theta_i^1 = \theta_i^2$ *and* $\theta_j^1 = \theta_j^2$*.*

Remark 5 *With every value of* $\phi_j(\theta_i)$*, there is associated a value* $\pi_i(\theta_j, \theta_i)$*. It follows that the whole schedule* $\phi_j(\theta_i)$*,* $\forall \theta_i$ *is defined by* $\pi_i(\theta_j, \theta_i)$*. In terms of notation,* $\pi_i = \tau_i(\pi_j)$ *is the reaction function* τ_i *yielding the unique solution* π_i *for a given* π_j*. That is, for a given* π_j*,* T_i *is solved* $\forall \theta_j$*.*

Lemma 3 *(i) An optimising agent will always accept matches within a range of agents with strictly positive mass;*

(ii) $\pi_j^1(\theta_j) \geq \pi_j^2(\theta_j)$ *implies* $\phi_j^1 > \phi_j^2$*;*

(iii) $\pi_j(\theta_j) \geq 0$ *implies there is an upper bound on the reservation value* ϕ_j*.*

Proof. (i) A population with zero mass implies that $\gamma_i = 0$. Since $u(\theta_j) > 0$, $T_i(\phi_j) > 0$. $T_i(\phi_j) = 0$ can only be satisfied for some $\gamma_i = 0$. This implies accepting a population with strictly positive mass. This applies to all types of both sexes since $u(\theta_j) > 0$ and $u(\theta_i) > 0$, $\forall \theta_i, \theta_j$;

(ii) $\pi_j^1(\theta_j) \geq \pi_j^2(\theta_j)$ ceteris paribus implies $\gamma_i^1 \geq \gamma_i^2$, by definition of γ_i. If ϕ_j^1 is the unique solution to $T_i(\phi_j^1 \mid \mu_i^1) = 0$, i.e. $T_i(\phi_j^1) = 0$ given π_j^1, then it follows that $T_i(\phi_j^1 \mid \pi_j^2) > 0$, since $T_\phi > 0$. The unique solution to $T_i(\phi_j^2 \mid \pi_j^2) = 0$ then satisfies $\phi_j^1 > \phi_j^2$;

(iii) For $\pi_i(\theta_j) \geq 0$, $\gamma_i > 0$. Since $T_\phi > 0$, a decreasing ϕ_j implies a decreasing $T(\phi_j)$. As a result, there will exist a value X satisfying $u(X) > 0$ such that $T(X) < 0$. No agent will choose such a reservation strategy. Hence, there is a lower bound X^* where $T(X^*) = 0$. If $X^* \notin \Theta_j$, $X^* = \{\min \theta_j \in \Theta_j \mid \theta_j > X, T(X) = 0\}$. ■

Lemma 4 *For utility functions of both sexes satisfying*

$$(r+\alpha)u_{\theta_i}(\phi_j) - \beta \int_{\Theta_j} \pi_i(x,\theta_i)\pi_j(\theta_j,x)[u_{\theta_i}(x) - u_{\theta_i}(\phi_j)]dF_j(x) = 0 \quad (1.7)$$

the equilibrium mapping $\phi_j(\theta_i)$ is type-independent, for a given π_j and F_j.

Proof. From Lemma 2, there exists a unique reservation strategy, given the strategy of all other players. Type-independence of the reservation strategy will occur when, taking π_j as given, $\frac{\partial \phi_j(\theta_i)}{\partial \theta_i} = 0$, i.e. when a lower type has the same reservation value. With $T_\phi > 0$, the implicit function theorem implies $T_{\theta_i} = 0$, or equation (1.7). ∎

Proposition 2 *For multiplicatively separable utility functions, the distributions of types are endogenously partitioned.*

Proof. Consider the general formulation of a multiplicatively separable utility function: $u(\theta_j, \theta_i) = v(\theta_j)w(\theta_i)$. $T(\phi_j) = 0$ can be rewritten as

$$(r+\alpha)v(\theta_j) - \beta \int_{\Theta_j} \pi_i(x,\theta_i)\pi_j(\theta_i,x)[v(x) - v(\phi_j)]dF_j(x) = 0 \qquad (1.8)$$

It is easily verifiable that $T_{\theta_i} = 0$, provided $\pi_j(\theta_i, \theta_j)$ is independent of θ_i, i.e. $\pi_j(\theta_i^1, \cdot) = \pi_j(\theta_i^2, \cdot)$, $\forall \theta_i^1 \neq \theta_i^2$. By requiring that u_j is multiplicative, this is automatically satisfied $\forall \theta_i$ in the same partition, provided that individuals have time invariant strategies. ∎

1.3 Discussion of the Results

Proposition 2 provides a good example in order to get some insight into the algorithm of the iterated elimination of dominated strategies which establishes the existence and uniqueness of equilibrium. Using the Steady State equilibrium strategies, the partitioning result will be used to illustrate Proposition

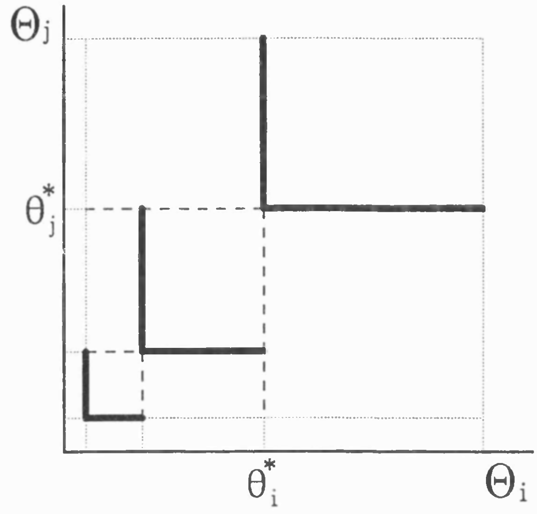

Figure 1.2: Partitioning

1. Consider Figure 1.2. If all types θ_i are accepted by all θ_j, all θ_i choose the same type-independent reservation value θ_j^*, given a multiplicative utility function. From Lemma 3(ii), it follows that all reservation strategies above this value are dominated and can thus be eliminated. Being accepted by less than all the types θ_j would certainly lower their reservation value. The same holds for all θ_j. However, if all θ_j have dominated strategies above this upper bound, these dominated strategies can be eliminated. It follows that all θ_i above θ_i^* are accepted by all with certainty, so that $\forall \theta_i \geq \theta_i^*$ their reservation strategy θ_j^* is dominant given the above elimination of dominated strategies. Likewise, the iterated strict dominance strategy for all $\theta_j \geq \theta_j^*$ is θ_i^*. This gives rise to the first set of partitions. Now given the dominant strategies of the types in the highest partitions, the same iterated dominance argument can be repeated. All $\theta_i < \theta_i^*$ are now rejected by all $\theta_j \geq \theta_j^*$. Provided they are accepted by all types below the highest partition, they will choose a

revised (type-independent) reservation strategy (the second horizontal line). Again, the iterated dominance argument implies that the second partition is formed by all types that are accepted with certainty. This goes on until all types belong to a partition.

Before the algorithm is extended to the general case, two remarks. The partitioning result is quite surprising. Though utility functions are type-dependent, the strategies are not. For a special case of multiplicatively separable utility functions with $u_{\theta_i} = 0$, the result is fairly intuitive since the utility function is type-independent[5]. It follows that the first order condition (1.6) is type independent. Utility derived and hence the opportunity cost are identical ex ante for types within one partition. Hence, they will solve for the same solution. With type dependent utility functions this is equally the case but for different reasons. Consider for example the case where higher types derive more utility from being matched with a high type of the other sex (i.e. the utility exhibits strategic complementarities). It follows that the expected value of being matched is increasing in type. On the one hand, higher types will be more choosy and have higher reservation values. On the other hand, without direct search costs, the cost of search is the opportunity cost of not being matched. As a result, the search cost is increasing in type. The higher types are more impatient and choose lower reservation values. For multiplicative utility functions, these two effects cancel out against each other and the first-order condition (1.8) is homogeneous of degree zero in the own type.

[5]The partitioning result, obtained in different frameworks, has always been derived for a special case of the type-independent utility function: $u_i = \theta_j$. See McNamara and Collins [42], Bloch and Ryder [12], Burdett and Coles [13]. The exception being Smith [69] who looks at multiplicative pay-offs and derives a similar result to the one in Proposition 2.

Note further that in case $u(\underline{\theta}_i) = u(\underline{\theta}_j) = 0$ the number of partitions goes to infinity. At the bottom it is always more lucrative to wait a bit more and not accept the lower types since they yield utility going to zero. The proof is beyond the purpose of this Chapter.

The uniqueness result is independent of both the specification of the pay-off function and of the fact whether both sexes have the same pay-off functions. The proof for existence and uniqueness uses a generalised iterated strict dominance argument as discussed for the case of multiplicatively separable utility functions. Imagine for example that the reservation value is increasing in type when accepted by all other types, as is the case in Figure 1.1 above. All strategies above this schedule are dominated. It follows that there exists a pair (θ_i^*, θ_j^*), above which all types are accepted with certainty. Hence, all types above (θ_i^*, θ_j^*) have a unique iterated strict dominant strategy. Taking these dominant strategies as given, all types below will revise their upper bound above which all strategies are dominated, so they choose a new reservation value below the dashed line. A new pair (θ_i^*, θ_j^*) exists for which there is now a iterated dominant strategy. This can then be repeated a finite number of times. The proof for the case where the reservation schedule is decreasing in type needs some additional feature. Consider Figure 1.3. In the panel on the left, the case is depicted for reservation schedules decreasing in type. This the case for example for utility functions like $u_i = \theta_j + \theta_i$. The dashed line is the (decreasing) reservation value conditional upon acceptance by all types of the other sex. All reservation values above this schedule are strictly dominated. At the intersection of the two dashed schedules, the pair (θ_i^*, θ_j^*) is defined. All types above have a dominant strategy, given by the fat line and equal to the dashed line. The types immediately below (θ_i^*, θ_j^*) are now accepted by some types above their first reservation schedule, but not

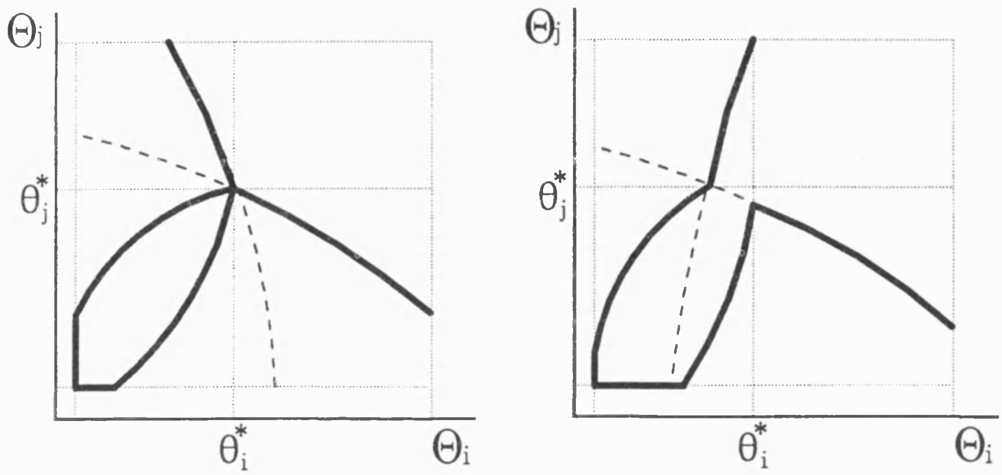

Figure 1.3: Downward Sloping Reservation Strategies

by all. Hence they will revise their upper bound downwards. However, from Lemma 3(iii), and given acceptance by some, they now also have a lower bound. This holds for both sexes. Given the lower bound of the other sex, they will revise their upper bound and given the upper bound of the other sex, they will revise the lower bound. This goes infinitely until the unique reservation schedule is determined. The panel on the right in Figure 1.3 is merely a variation on the same theme. One schedule is upward sloping, the other downward. Again, by eliminating dominated strategies starting from the top (i.e. above (θ_i^*, θ_j^*)), the whole schedule can be constructed uniquely.

Remark 6 *Figure 1.3 clearly illustrates that the slope of the schedule ϕ_j is not only a function of the utility function. It can be shown that for utility functions exhibiting log-supermodularity (i.e. $u_{12}u > u_1u_2$) the reservation value, given acceptance by all, is increasing in type and decreasing if it is log-submodular (see Smith [69]). However, the equilibrium schedule is not necessarily downward sloping over the whole range even if there is log-*

submodularity but π_j is type independent. This is for example the case at the upper part of the distribution. In Figure 1.3, even though in both cases at least one of the utility function is log-submodular, at the lower end the schedule is upward sloping. The reason is that in that range, π_j is type-dependent.

1.4 Perfect Matching Equivalence

In this Section, it is shown that the equilibrium is indeed the generalisation of the Perfect Matching model. First, the perfect matching model is defined in more detail. Second, it is shown that the Bilateral Search model with Vertical Heterogeneity yields the same outcome as the Perfect Matching model when the search friction disappears in the limit. The search friction disappears when waiting time goes to zero, i.e. when the arrival rate β goes to infinity.

The Perfect Matching model used as the benchmark here is the model by Gale and Shapley [27] and rigorously discussed in Roth and Sotomayor [62]. Originally it was formulated for a finite number of agents and for any set of preferences. Here, it will be extended to a continuum of agents and the preferences will be such that they exhibit Vertical Heterogeneity, the Beckerian aspect. In what follows, it will be referred to as the Gale-Shapley-Becker model. Consider two disjoint sets of agents Θ_i and Θ_j, both with mass one. Individuals are characterised by a type θ_i, cumulatively distributed over $F_i(\theta_i)$. Vertical Heterogeneity of preferences can be represented by any utility function $u(\theta_j, \theta_i)$ as long as $u_{\theta_j} > 0$. A matching μ is defined as a one-to-one correspondence from $\Theta_i \cup \Theta_j$ onto itself of order two (i.e. $\mu^2(\theta_i) = \theta_i$), such that $\mu(\theta_i) \in \Theta_j$ and $\mu(\theta_j) \in \Theta_i$. A matching μ is individually rational if it is not blocked by any individual agent. It is stable if it is individually rational and if it is not blocked by any pair of agents, one female and one male. Clearly, this establishes that a stable matching is a core concept and thus a cooperative equilibrium. In Chapter 3 (Corollary 4), it is shown that there exists a unique stable matching, $\mu(\theta_i) = \theta_j \Leftrightarrow F_i(\theta_i) = F_j(\theta_j)$: in equilibrium, only individuals of the same rank match.

The Equivalence between Perfect Matching and Search can now be established. Note however that there is an entirely different use of equilibrium concept: cooperative versus non-cooperative equilibrium. What will be shown is that the non-cooperative search equilibrium yields the same outcome as the cooperative stable matching without friction when the search friction is infinitely small (i.e. $\lim \beta \to \infty$). It can actually be shown that the stable matching is equivalent to the trembling hand equilibrium which rules out degenerate non-cooperative equilibria. Note also that our restriction to reservation strategies has a similar impact.

Proposition 3 ***Equivalence.*** *The Gale-Shapley-Becker Perfect Matching model is the limit case of the search model when trading opportunities arrive instantaneously (i.e.* $\lim \beta \to \infty$*).*

Proof. For $\lim \beta \to \infty$, the system of equations (1.1) and (1.2) collapses. The state of being single now coincides with the state of being matched since a match is instantaneously realised. It follows that the value of being single has to equal the expected value of being matched: $V_0(\theta_i) = E[V_1(\theta_i) \mid \pi_j = 1]$. An individual θ_i will choose a reservation value ϕ_j such as to maximise the expected value of being matched subject to being accepted. This implies

$$\max_{\phi_j} EV_1(\theta_i) = \frac{\int_{\Theta_j} \pi_i(x,\theta_i)\pi_j(\theta_j,x)u(x)dF_j(x)}{\int_{\Theta_j} \pi_i(x,\theta_i)\pi_j(\theta_j,x)dF_j(x)} \tag{1.9}$$

EV_1 is monotonically increasing in ϕ_j, provided acceptance by sex j: $\frac{\partial EV_1(\theta_i)}{\partial \phi_j} > 0$, s.t. $\pi_j = 1$ is derived from

$$\int_{\Theta_j} \pi_i(x,\theta_i)\pi_j(\theta_j,x)u(x)dF_j(x) - u(\phi_j)\int_{\Theta_j} \pi_i(x,\theta_i)\pi_j(\theta_j,x)dF_j(x) \tag{1.10}$$

s.t. $\pi_j = 1$, which is satisfied since $u_{\theta_j} > 0$. The solution to this problem is a corner solution: EV_1 is maximised when ϕ_j is maximised, s.t. $\pi_j = 1$. With $\phi'_j = \max\{\theta_j \in \Theta_j \mid \mu_i(\theta_j) = 1\}$, the optimal choice of ϕ_j is $\phi_j(\theta_i) = \phi'_j(\theta_i)$, $\forall \theta_i$. Likewise, $\phi_i(\theta_j) = \phi'_i(\theta_j)$, $\forall \theta_j$. Applying the algorithm in the proof of Proposition 1 then gives the following allocation: a type θ_i will match with θ_j if and only if $F_i(\theta_i) = F_j(\theta_j)$. This is equivalent to the stable matching $\mu(\theta_i) = \theta_j \Leftrightarrow F_i(\theta_i) = F_j(\theta_j)$. ■

1.5 Concluding Remarks

In this paper, the Perfect Matching model is extended to an Imperfect Matching Model with search frictions. A search model is proposed featuring vertical heterogeneity and bilateral search. Equilibrium in a concept as strong as iterated elimination of dominated strategies is shown to exist and is unique, irrespective of the utility specification of individuals. It is derived that its limit case without friction is the Gale-Shapley-Becker perfect matching model. Equilibrium properties can be derived in function of the pay-off function and as a result, for multiplicatively separable utility functions, equilibrium strategies are type-independent even though the utility function does depend on the type. It follows that both distributions of types are endogenously partitioned, segregating the market into classes.

In other work, it has been shown that the Imperfect Matching Model for specific utility functions can result in multiple steady states. This is due to a sorting effect in the distribution of singles which has an impact on the number of singles searching and thus on the matching technology. The results here can be taken to generalise that result, incorporating the impact of search externalities from sorting in the population. The main problem with a general existence proof of a steady state equilibrium is that usual techniques do not readily generalise with non continuous strategy spaces.

Irrespective of the establishment of the existence of a steady state equilibrium, the Imperfect Matching Equilibrium and its properties prepare this framework to be applied to economically relevant environments. Certainly, the labour market is a prominent candidate. Not only is there an equilibrium rate of unemployment, there is also an equilibrium rate of idleness of jobs. In the Imperfect Matching model, the nature of it can be studied. In addition, in subsequent work (see Chapter 2) it has been established that under certain conditions, the lowest types of one set will never be matched. The reason for that is that the even the lowest types of the other set prefer to search longer, rather than accepting a match with a low type. In the labour market, the implication is that there is a threshold level of skills or ability below which there are no jobs available. In addition to voluntary unemployment, most commonly explained in the search literature, this gives rise to the existence of involuntary unemployment in equilibrium.

1.6 Appendix

Proposition 1.

Proof. Iterated elimination implies n iterations. Therefore, the following notation is introduced. First, because of the argument of iterated elimination, the variable $\mu_i(\theta_j, \theta_i) = \pi_j(\theta_i, \theta_j)$ is introduced in order to distinguish the acceptance rule by others from the strategy by other players. Clearly, in equilibrium they are the same. $\Pi_i(\theta_j, \theta_i; n) = \Pi_i(n)$, $\forall i, j$ is the schedule $\pi_i(\theta_j, \theta_i)$ calculated in iteration n, provided $\mu_i = 1$. $\phi_j(\theta_i; n)$ is the reservation value associated with $\Pi_i(\theta_j, \theta_i; n)$, provided $\mu_i = 1$. Likewise, $\mu_i(n)$ is $\mu_i(\theta_j, \theta_i)$, $\forall i, j$ in iteration n.

In each iteration n, the algorithm below will allow to determine the unique strategies for a connected set with positive mass. Given the outcome of the anterior iterations that all types $(\theta_i, \theta_j) \geq (\theta_i^*(n-1), \theta_j^*(n-1))$ have determined their unique strategy, the n-th iteration starts. It consists of 5 steps (below). It can be established that there exists a set of dominated strategies (i.e. there is a maximum reservation value) for all remaining types of both sexes. These are determined in step 1-3. These dominated strategies imply that all types of the other sex higher than the reservation value will never be rejected. In step 4, the connected set of all types is determined who will never be rejected irrespective of other players' strategies, which is the result of the other sex's dominated strategies. It follows that all these types have a unique strategy (step 5) which is the result of iterated elimination of dominated strategies within this iteration. If the connected set is empty, the unique strategy for a strictly positive connected set is determined according to Lemma 5. As a result, after the n-th iteration, all $(\theta_i, \theta_j) \geq (\theta_i^*(n), \theta_j^*(n))$ have a unique iterated strict dominant strategy.

1. After $n-1$ iterations the schedules $\Pi_i(n-1)$ and $\Pi_j(n-1)$ are uniquely determined for all $(\theta_i, \theta_j) \geq (\theta_i^*(n-1), \theta_j^*(n-1))$. It follows that in the next iteration the schedules $\mu_i(n)$ and $\mu_j(n)$ are uniquely defined in that range. For the other types, maximal acceptance (i.e. $\mu(n) = 1$) allows us to determine the dominated strategies. Hence, determine $\mu_i(n) = \Pi_j(n-1)$, if $\theta_j > \theta_j^*(n-1)$; $\mu_i(n) = 1$, otherwise. Likewise $\mu_j(n) = \Pi_i(n-1)$, if $\theta_i > \theta_i^*(n-1)$; $\mu_j(n) = 1$, otherwise. At the start of the procedure $(n = 1)$, $\mu_i(n) = 1$, $\forall \theta_j$ and $\mu_j(n) = 1$, $\forall \theta_i$;

2. $\Pi_i(n) = \tau_i(\mu_i(n))$ and $\Pi_j(n) = \tau_j(\mu_j(n))$

3. Consider all types $(\theta_i, \theta_j) < (\theta_i^*(n-1), \theta_j^*(n-1))$. Taking into account the unique strategies of all higher types and by determining $\mu_i(n)$ and $\mu_j(n)$ in terms of maximal acceptance (i.e. from step 1, there exists no $\mu_i \geq \mu_i(n)$, it follows from Lemma 2 that all reservation strategies $\theta_j > \phi_j(\theta_i; n)$ are strictly dominated for all θ_i. From Lemma 1, $\Pi_i(n) = 1$ for all $\theta_j > \phi_j(\theta_i)$. Likewise, all reservation strategies $\theta_i > \phi_i(\theta_j; n)$ are strictly dominated for all θ_j and $\Pi_j(n) = 1$ for all $\theta_i > \phi_i(\theta_j; n)$;

4. Define

$$(\theta_i^*(n), \theta_j^*(n)) = \left\{ \begin{array}{l} (\min \theta_i, \min \theta_j) \mid \Pi_i(n)\Pi_j(n) = 1 \text{ and} \\ \forall \theta_i \leq \theta_i^*(n-1) : \Pi_j(n) = 1, \forall \theta_j \leq \theta_j^*(n-1) \\ \forall \theta_j \leq \theta_j^*(n-1) : \Pi_i(n) = 1, \forall \theta_i \leq \theta_i^*(n-1) \end{array} \right\} \quad \text{(A1.1)}$$

In the first round $(n = 1)$, define $\theta_i^*(0) = \bar{\theta}_i$ and $\theta_j^*(0) = \bar{\theta}_j$. Note that it follows from equation (A1.1) that $[\theta_i^*(n), \theta_i^*(n-1)]$ and $[\theta_j^*(n), \theta_j^*(n-1)]$ are connected sets. If $(\theta_i^*(n), \theta_j^*(n)) < (\theta_i^*(n-1), \theta_j^*(n-1))$ at least one of these sets is non empty. The unique iterated strict dominant strategies are determined in step 5. Alternatively, if $(\theta_i^*(n), \theta_j^*(n)) = (\theta_i^*(n-1), \theta_j^*(n-1))$

the sets are empty and the pair $(\theta_i^*(n), \theta_j^*(n))$ and the unique iterated strict dominant strategies are determined according to Lemma 5 below.

5. From step 3 and from equation (A1.1), all types $\theta_i \in [\theta_i^*(n), \theta_i^*(n-1)]$ and $\theta_j \in [\theta_j^*(n), \theta_j^*(n-1)]$ have $\mu_i(n) = \mu_j(n) = 1$ independently of any other players strategy (because the strategies are dominated), since $\mu_i = \Pi_j$ and $\mu_j = \Pi_i$. The reservation strategy of all these types is thus independent of the strategy of any other player. By eliminating the dominated strategies, all these types have a unique iterated strict dominant strategy $\Pi_i(n)$ and $\Pi_j(n)$ respectively (from Lemma 2(i));

This iterative procedure is repeated until $\Pi_i(N) = \Pi_i(N+1)$ and $\Pi_j(N) = \Pi_j(N+1)$. Because $(\theta_i^*(n), \theta_j^*(n)) < (\theta_i^*(n-1), \theta_j^*(n-1))$ and from Lemma 3(i), every agent chooses to accept matches from a population with strictly positive mass. As a result, the populations eliminating strictly dominated strategies in every iteration have strictly positive mass. It follows that the equilibrium list $(\Pi_i(N), \Pi_j(N))$ is obtained after a finite number of N iterations. ■

Lemma 5 *If according to equation (A1.1)* $(\theta_i^*(n), \theta_j^*(n)) = (\theta_i^*(n-1), \theta_j^*(n-1))$ *a new pair can be defined such that* $(\theta_i^*(n), \theta_j^*(n)) < (\theta_i^*(n-1), \theta_j^*(n-1))$ *and such that there exists a unique iterated strict dominant strategy for all types in the interval* $([\theta_i^*(n), \theta_i^*(n-1)], [\theta_j^*(n), \theta_j^*(n-1)])$.

Proof. First, $(\theta_i^*(n), \theta_j^*(n)) = (\theta_i^*(n-1), \theta_j^*(n-1))$ implies that both $\mu_i(n) \geq 0$ for $\theta_j > \phi_j(n-1)$ and $\mu_j(n) \geq 0$ for $\theta_i > \phi_i(n-1)$. This follows from Lemma 3(i) and (ii). If it is not satisfied say for sex i, there would exist a range of dominated strategies with strictly positive mass below θ_j^* for the types of sex i. Hence, the connected set would be non-empty and the equality no longer holds. Therefore, the pair can be redefined such that the set is non-empty

$$\left(\theta_i^*(n), \theta_j^*(n)\right) = \begin{pmatrix} \min\{\theta_i \mid \mu_i(n) \geq 0, \forall \theta_j > \phi_j(n-1)\}, \\ \min\{\theta_j \mid \mu_j(n) \geq 0, \forall \theta_i > \phi_i(n-1)\} \end{pmatrix} \tag{A1.2}$$

The proof now is similar in spirit to the proof of dominance solvability of the Cournot model in Gabay and Moulin [26] and Moulin [51] and involves a Cournot-tatonnement process. First, additional notation is introduced for this stage of the elimination process only. $\Pi_i(s \mid n)$ is the subiteration s which determines the schedule $\Pi_i(n)$. Elimination of strictly dominated strategies will occur by defining an upper bound and a lower bound in every subiteration s: $\Pi_i^u(s \mid n)$ and $\Pi_i^l(s \mid n)$. $\mu_i^u(s \mid n)$ and $\mu_i^l(s \mid n)$ are analogously defined. Likewise for individuals of type j. From Lemma 3(iii) and given $\mu_i(n) \geq 0$ a lower bound on the reservation strategy exists. Given $\mu_i^l(1 \mid n) = \mu_i(n-1)$ if $\theta_j \geq \theta_j^*(n-1)$ and $\mu_i^l(1 \mid n) = 0$ otherwise, $\Pi_i^l(1 \mid n) = \tau_i(\mu_i^l(1 \mid n))$. All strategies $\pi_i \geq \Pi_i^l(1 \mid n)$ are strictly dominated. Likewise for $\Pi_j^l(1 \mid n)$. On the other hand, from Lemma 2, it can be established that for upper bounds $\Pi_i^u(1 \mid n) = \tau_i(\Pi_j^l(1 \mid n))$ and $\Pi_j^u(1 \mid n) = \tau_j(\Pi_i^l(1 \mid n))$ all strategies $\pi_i \leq \Pi_i^u(1 \mid n)$ and $\pi_j \leq \Pi_j^u(1 \mid n)$ are strictly dominated. In every following iteration, $\Pi_i^l(s \mid n) = \tau_i(\Pi_j^u(s-1 \mid n))$ and $\Pi_j^l(s \mid n) = \tau_j(\Pi_i^u(s-1 \mid n))$ are determined. From Lemma 3(iii), all strategies $\pi_i \geq \Pi_i^l(s \mid n)$ and $\pi_j \geq \Pi_j^l(s \mid n)$ are strictly dominated. Likewise, all strategies $\pi_i \leq \Pi_i^u(s \mid n) = \tau_i(\Pi_j^l(s \mid n))$ and $\pi_j \leq \Pi_j^u(s \mid n) = \tau_j(\Pi_i^l(s \mid n))$ are strictly dominated. If this procedure is repeated ad infinitum, $\Pi_i^l(\infty \mid n)$ and $\Pi_i^u(\infty \mid n)$ will converge to $\Pi_i(n)$ and $\Pi_j^l(\infty \mid n)$ and $\Pi_j^u(\infty \mid n)$ to $\Pi_j(n)$: 1. $\Pi_i^l(\infty \mid n) \geq \Pi_i(n)$ and $\Pi_i^u(\infty \mid n) \leq \Pi_i(n)$; 2. $\Pi_i^l(\infty \mid n) = \tau_i(\Pi_j^u(\infty \mid n)) = \tau_i(\tau_j(\Pi_i^l(\infty \mid n)))$, which is possible only if $\Pi_i^l(\infty \mid n) = \Pi_i(n)$; 3. Similarly for $\Pi_j^l(\infty \mid n) = \Pi_j(n)$. The same reasoning holds for $\Pi_i^u(\infty \mid n) = \Pi_i(n)$ and $\Pi_j^u(\infty \mid n) = \Pi_j(n)$.

There is a unique strategy $\Pi_i(n)$ and $\Pi_j(n)$ for all types $\theta_i \in [\theta_i^*(n), \theta_i^*(n-1)]$ and $\theta_j \in [\theta_j^*(n), \theta_j^*(n-1)]$. Hence, the pair $(\theta_i^*(n), \theta_j^*(n))$ is defined as in (A1.2). ■

Endogenous distribution of Singles

Let $H_i(\theta_i)$ be the distribution of the entire population of sex i and $H_i^s(\theta_i)$ the distribution of singles. All these distributions can be time variant. If n_i is the fraction of singles of type θ_i at a particular moment in time, the density function $h^s(\theta_i)$ of singles associated with the population density $h(\theta_i)$ is given by

$$h^s(\theta_i) = \frac{n_i}{\int_{\Theta_i} n_i dh(\theta_i)} h(\theta_i) \tag{1.4}$$

At any moment in time, the law of motion is given by $\dot{n}_i = \beta\psi_i n_i + \alpha(1-n_i)$. Clearly, out of the steady state, the distribution of singles h^s changes over time. Clearly, in a steady state $F = H_t^s$. This is also true out of steady state if players do not hold rational expectations and believe the observed distribution will not change over time[6]. If agents hold full rational expectations, they will take into account the change in the distribution over the expected duration $(\beta\psi_i)^{-1}$ of being single. The belief about the distribution of singles then satisfies $F(\theta_i) = \int_0^{(\beta\psi_i)^{-1}} H_t^s(\theta_i) dt$.

Endogenising the distribution of singles leaves the existence and uniqueness of the allocation in tact (though there is a new source of multiplicity). In addition, the characterisation of equilibria for given preferences and the perfect matching equivalence still hold. There is however a strong implication for the off the steady state characterisation of the partitioning result. This is shown below.

[6]This corresponds to what is called a Partial Rational Expectations belief in Burdett and Coles [13]

Non Steady State Type Dependence

In Section Three, it was shown that provided the distribution of types is time stationary, there is endogenous partitioning of the distribution of types. Highly appealing as this may seem, the result fails to hold out of steady state: reservation strategies are type dependent. This implies that even a steady state equilibrium can only be shown to exist if the out of steady state equilibrium exists. Since our general Proposition of existence of equilibrium out of steady state is shown with and without type dependence, it follows that the steady state partitioning result can come about from any initial condition. All the other work so far (Smith [69], Burdett and Coles [13]) could only conjecture that the steady state would come about. However, a proof for the case of type-dependent strategies as in Proposition 1 is necessary.

In Proposition 4, it is shown by example that out of Steady State, there is no partitioning even with multiplicatively separable utility functions. Consider a school ballroom matching market where dance starts at 8pm after the opening dance by the Headmaster and goes on forever. Matches go on for one song and hence dissolve after α^{-1} time on average. At the initial condition $t = 0$ (i.e. 8pm), no one is matched. Suppose a symmetric context with uniform distributions of types and (multiplicatively separable) utility functions $u_i = \theta_j$, which in the unique Steady State endogenously partitions both distributions. The Steady State distribution of singles H^s_T is not identical to the original distribution of the entire population, all of which are single at $t = 0$: $H^s_0 = H$. The distribution of singles is different from the distribution of the population because between partitions there are different probabilities of entering a match. In effect, Burdett and Coles [13] show that H^s_T is stochastically dominated by the uniform distribution H. In addition, out of Steady State, the distribution of singles H^s_t will change continuously over time $0 < t < T$. By the same argument, all distributions at any later date are stochastically dominated by the earlier ones. The Rational Expectations[7] strategy of all types at t is calculated given $F_t = \int_t^{t+(\beta\psi_i)^{-1}} H_t^s dt$, which is stochastically dominated by any F at an earlier time. By counter example, it is now shown that the out-of-Steady-State strategies are type dependent.

Proposition 4 *Out of Steady State, equilibrium strategies do not endogenously partition the distributions of types, even with multiplicatively separable utility functions.*

Proof. Consider the highest types. At $t = 0$, their reservation strategy is ϕ_0, calculated taking into account F_0. All types $\theta > \phi_0$ have the same reservation strategy. In the steady state at $t = T$, all top types have reservation strategy ϕ_T. Given stochastic dominance of H_T^s by the uniform distribution H, it follows that $\phi_T < \phi_0$: the density of singles is lower for higher types in the Steady State than at the initial condition, so they have to be less choosy in the Steady State (from (1.6)). By the same argument, at any moment in time $t \in [0, T]$, ϕ_t, $\forall \theta > \phi_0$ is lower than any earlier reservation strategy of the higher types. It follows that people entering at different times have different strategies. All types of the other sex have, provided they are accepted by everyone, a time dependent strategy. This implies that acceptance depends on the time of entry. At time t some one who entered at time $t^* < t$ will accept all $\theta > \phi_{t^*}$. Since this applies to both sexes, a type $\theta_i \in [\phi_T, \phi_0]$, faces acceptance $\pi_j = 1$, provided the type of the other sex θ_j entered at t^* such that $\theta_j < \phi_{t^*}$; $\pi_j = 0$, if entry was at t^* such that $\theta_j > \phi_{t^*}$. Since some of the high types who entered early will reject a match, the reservation strategy

[7]This result holds for any other expectations formations, including Partial Rational Expectations (see Burdett and Coles [13]) i.e. $F_t = H_t^s$, as long as strategies are time stationary.

ϕ_t of all $\theta \in [\phi_T, \phi_0]$ will be lower than ϕ_t of all $\theta > \phi_0$. Moreover, since the reservation strategy ϕ_t of the highest types is decreasing over time, the number of types that reject a type θ_i ($\pi_j = 0$) will be dependent on the type θ_i. There are more $\theta_j > \phi_{t^*}$ who entered at t^* who reject the lower the type θ_i. This implies that the strategy $\forall\theta \in [\phi_T, \phi_0]$ at time t is type dependent and increasing in type. ■

The intuition of this result is illustrated in Figure 1.4 and goes as follows. Facing the original distribution of all singles, there is a partition at $t = 0$:

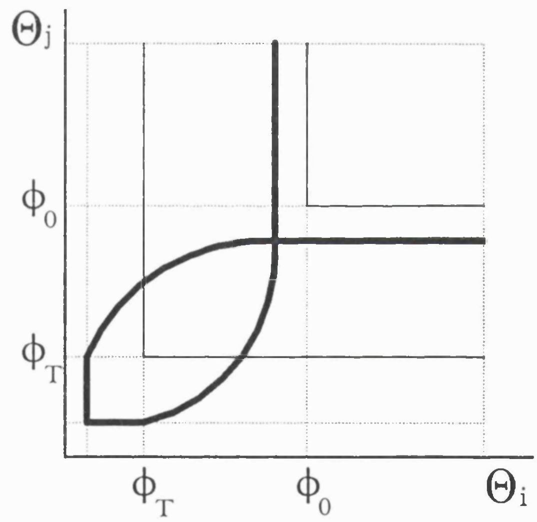

Figure 1.4: No Partitioning out off Steady State

all types above ϕ_0 have a type independent strategy ϕ_0, which is smaller than at $t = T$. However, at any time t, the new entrants in the pool of singles, will have a strategy $\phi_0 < \phi_t < \phi_T$. All types of the other sex, $\theta \in [\phi_T, \phi_t]$ are now no longer accepted by the new entrants. They are still accepted though by the ones that haven't been matched yet. As a result, they will have a strategy $\phi(\theta) < \phi_t(\bar{\theta})$. Since change is continuous and the set of individuals no longer lucky to be in the top partition grows monotonically (with monotonically changing distributions), strategies in the range $\theta \in [\phi_T, \phi_0]$ are type dependent. In Figure 1.4, the full thin line at ϕ_0 is the partition at $t = 0$. The full thin line at ϕ_T is the Steady State Partition. The thick line gives the reservation strategies at some time $t \in [0, T]$. At the top there is still a partition, since the top types are always accepted by all and thus have a type independent strategy. It is smaller than the partition at time T, because they take into account some of the top types who still have a strategy that dates from before time t: they have not been matched yet and will be more choosy. They will leave the pool of singles less quickly so that the top types can be more choosy in expected terms, since more top types search on average compared to the Steady State. Below that partition, the strategy is affected by the fact that the new entrants already have a reservation strategy of accepting you. The closer to ϕ_0, the more of the other sex that have adopted this strategy, so your reservation strategy is higher than the lower types in $[\phi_T, \phi_0]$. Hence the type dependence.

Basically, the type dependence follows from the fact that a transition path changes gradually over time. With the search friction and ex-ante strategies (i.e. strategies are determined upon entry and remain constant until matched), strategies for similar types are different at different times. The following Corollary to the Proposition 4 shows that type dependence is relevant since, starting from an out-of-Steady-State the Steady State "never arrives".

Corollary 1 *If the initial condition is out of Steady State, the Steady State only arrives after infinite time.*

Proof. Since meetings are stochastic, within finite time there will always be a positive mass of individuals who ex-post have not realised their ex-ante strategy. As a result, they will keep playing their ex-ante out of Steady State strategy. The distribution of singles H_t^s will differ from the Steady State distribution H_T^s as long as $t < T = \infty$. ■

Chapter 2

Some Applications

In this Chapter, the framework from Chapter 1 is used to discuss some particular results that follow from and are specific to the imperfect matching framework. The first Section shows that bachelors can exist. Bachelors are the lowest types of one of the two distributions who cannot find a match. This result is surprising since they exist even though the masses of both sexes are equal. In perfect matching models it is shown (Becker [7]) that no one remains unmatched if both masses are equal. The second Section deals with the possibility (in the case of pairwise matching from one distribution) of some types rejecting their equals. It provides a rationale for why you don't want to be a member of a club that wants you as a member. Again, this is counter-intuitive in a model with preferences increasing in types. Rejecting one's own type then asks the question who will accept yourself in response to a similar strategy of your equal counterparts.

Both results highlight some particular feature of the imperfect matching model. The difference between the two lies in the underlying assumption whether the pairwise matching is between types drawn from two disjoint sets or from one joint set. Clearly, bachelors cannot exist in a model with matched types drawn from one set: there always exists an equal with whom a match would generate strictly positive utility. Similarly, with matching from disjoint sets, it is impossible to reject equals.

In dealing with both issues, some interesting comparative statics features that affect the equilibrium allocation will emerge. In this search framework, relative preferences of both sets of types are crucial. This is also related to the distribution functions of the singles. The equilibrium allocation is affected by both a change in the mean of the distribution and a mean preserving spread. The model can perform well in explaining macroeconomic fluctuations and the effect on the labour market. A downturn can be interpreted as a downward shift in the mean of the distribution of jobs (i.e. the technology). Its effect would be a relative increase in the reservation strategies of firms. The effect is a decrease in the reservation strategy of the workers and would affect the existence of bachelors (which can be interpreted as involuntary unemployment in a job search context) by an increase in the set of unmatched types. These comparative statics results however are not derived rigorously in this Chapter.

For the remainder of this Chapter, the notation from the first Chapter is used, unless mentioned otherwise.

2.1 Bachelors

In a matching model, Bachelors will be referred to as individuals who are never able to engage in a match, even though they want to. It is entirely involuntary. This phenomenon is certainly distinct from individuals who have been proposed marriage but who choose to reject the offer because it pays in expected terms to wait for a better opportunity. This situation of not engaging in a match is voluntary. Our definition of Bachelors here is related to the bilateral nature of the matches. An individual would like to engage in a match, but is not accepted by any type of the other sex. Bachelors can then be defined as:

Definition 2 *Given the equilibrium* (π_i, π_j)*, let* $\theta_j^B = \min\{\phi_j(\theta_i), \forall\theta_i\}$ *and let* $B_j = [\underline{\theta}_j, \theta_j^B)$*. Bachelors are all those lowest ranked types of the distribution who never find a match:* $\theta_j \in B_j$*, provided* $\theta_j^B > \underline{\theta}_i$.

The existence of bachelors in itself is not surprising in the case of populations of different masses. In the perfect matching model without search frictions for example, Becker [7] remarks that the lowest ranked individuals of the distribution with the largest population are never matched. By construction, the perfect matching model only allows for pairwise matching from disjoint sets so that the maximum number of matches is equal to the minimum of the masses of both populations. On the other hand, it is also shown that in the case of zero outside options, all individuals are matched when both populations have equal mass.

The result that will be shown below is in sharp contrast with the perfect matching equilibrium allocation. Even though the utility derived from not being matched is zero and even though the masses of both populations are equal, some individuals will never be matched. This is not because we relax the assumption of pairwise matching. The number of individuals matched of both sexes has to be equal in all cases. The surprising result is derived from the time dimension in this search model. Before this is expanded upon, an example is developed that should help make the mechanics clearer.

The result in Proposition 5 below on the existence of Bachelors relies on an asymmetry. This will be exploited in the example.

Example 1 *Consider a model with the following parameters and functional forms:* $F \sim U(0,1]$*;* $r+\alpha = \beta = 1$*;* $u_i = \theta_j + C_i$*. Note that from Proposition 2 (Chapter 1) an equilibrium will exhibit partitioning, since the utility function is multiplicatively separable. There is an asymmetry such that both sides of the market only differ in the constant* C_i*:* $C_f = 0$*,* $C_m \geq \frac{1}{2}$*. The best response condition (1.6) given acceptance satisfies:*

$$\phi_j^2 + 2\phi_j - 1 + 2C_i = 0$$

With $C_f = 0$*,* $\phi_m(\theta_f) = \sqrt{2}-1$*, and* $\phi_f(\theta_m) \leq 0$*, provided* $C_m \geq \frac{1}{2}$*. It is easily verifiable from Proposition 1 (Chapter 1) that the equilibrium reservation strategies satisfy (this is illustrated in figure 2.1):*

$$\begin{aligned} \phi_m(\theta_f) &= \sqrt{2} - 1 \quad &, \forall \theta_f \\ \phi_f(\theta_m) &= 0 \quad &, \forall \theta_m \end{aligned}$$

The implication of these equilibrium strategies is however that $\forall \theta_m \in B_m = [0, \sqrt{2}-1]$ *never get matched and are eternal bachelors.*

The general conditions are now derived.

Proposition 5 *Given the equilibrium* (π_i, π_j)*, a Bachelors* $\theta_j \in B_j$ *exists in the set* Θ_j *iff*

$$\begin{aligned} & T_j(\theta_i) > 0, \forall \theta_j \in B_j, \forall \theta_i \\ & T_i(\underline{\theta}_j) < 0, \forall \theta_i \end{aligned} \tag{2.1}$$

Proof. From the definition of B_j, it follows that $\pi_i(\theta_i, \theta_j \in B_j) = 0$. The necessary and sufficient condition is shown in two steps. First, it is shown that violating condition (2.1) for a θ_j implies there exists $\pi_i = 1$. Suppose $T_j(\theta_i) \leq 0$, then from equation (1.6) in Chapter 1, $\gamma_j > 0$. This can only be the case if there exists a $\pi_i = 1$. Hence θ_j is no Bachelor. Second, it is shown

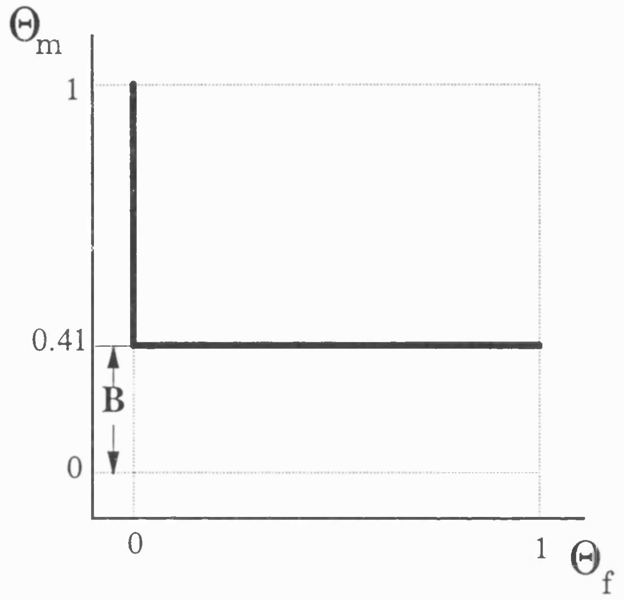

Figure 2.1: Bachelors

that if there are no bachelors, then equation (2.1) does not hold. If there are not bachelors B_j, then all θ_j are accepted by some θ_i, including $\underline{\theta}_j$. As a result, there must exist a θ_i that has $T_i(\underline{\theta}_j) \geq 0$, thus violating condition (2.1). ■

Corollary 2 *Bachelors can at most exist in one of the two disjoint sets.*

Proof. It is easily verified that condition (2.1) in Proposition 5 cannot simultaneously hold for both θ_j and θ_j. ■

Turning back to the example, it can be verified that the conditions in Proposition 5 and Corollary 2 are satisfied. Equation (1.6) in Chapter 1 for both sexes satisfies: $T_m(\theta_f) = \theta_f + C_m - \gamma_f$ and $T_f(\theta_m) = \theta_m - \gamma_m$. The implication of being Bachelor is that $\forall\theta_m \in B_m : \pi_f = 0$. Hence $\gamma_f = 0$. It is trivial now that $T_m(\theta_f) = \theta_f + C_m > 0$, $\forall\theta_m \in B_m$. For all females, it holds that $T_f(\underline{\theta}_m) = -1 < 0$.

As was mentioned above, the explanation for the surprising result that even with equal populations and pairwise matching some individuals are involuntarily unmatched lies in the duration of search. In the case of the example, males accept a match instantaneously, whatever the type of the female. Females on the other hand don't. They all prefer to reject 41% of all the potential partners and wait for a better opportunity to come. As a result, only 59% of all single females per unit of time engage in a match. If all males were always accepted, 100% of all singles would engage in a match. But it is exactly the pairwise matching restriction that prevents this from happening. As a result, exactly the same proportion of males is accepted. The remaining ones, even though willing to get matched, are rejected. The rationale then is that because of the longer search by the females, the lowest ranked males never get matched because they have to compensate for the males being too eager.

The underlying reason for this asymmetry in search strategy between males and females derives from the asymmetric utility derived from a match. Males derive a higher utility $(\theta_j + C_m)$ from each match than females $(C_f = 0)$. As a result, the opportunity cost per unit time of not being matched is higher for males, so they are more impatient. This implies they choose a lower reservation strategy. Here this means they are willing to accept any female irrespective of the type. This is not the case for the females who, knowing they are accepted by all males, prefer to wait until a higher type arrives if a low type is met.

This asymmetry only stresses the importance of the relative parameters on the equilibrium allocation. Though beyond the scope of this Chapter, it is evident from the example that the utility specification will be of fundamental importance. In addition though, it should be noted that the distribution of singles is crucial in two respects. First, the mean of the distribution. The example could equivalently be specified with identical preferences for males and females but with the distribution of females having a higher mean. The interpretation is that the "higher quality" sex will, other things equal, never have Bachelors. Second, the spread will matter. Observation of equation (1.6) shows that a higher spread will shift the reservation strategy. Whether it increases or decreases depends on whether the types matched with are above or below the mean. The likelihood of Bachelors occurring increases with an increasing spread of the opposite distribution.

2.2 One Reason Why You Don't Want to Be a Member of a Club that Wants You as a Member

Whereas the first Section of this Chapter entirely relies on the assumption of pairwise matching from two disjoint sets, here we consider pairs drawn from the same distribution. A first implication is that Bachelors can never exist (from Corollary 2), unless the population is countable and has an odd number. Several economic interpretations of pairwise matching from one set of types have been proposed. Kremer [34] for example considers the perfect matching model without search frictions. Workers, heterogeneous in their skill level need to work in pairs in order to produce. They are necessarily drawn from one distribution. But in addition, there are a number of applications in any context where partnerships are formed of equal types. Any type of contest or competition, but also all clubs where membership consists of individuals drawn from one set. Here, we will concentrate on the endogenous formation of clubs in the imperfect matching model.

Consider a group of tennis players. They are all indexed by a type θ. The distribution of singles is $F(\theta)$. As in Chapter 1, they are looking for a potential sparring partner. Partners are met at an exogenous arrival rate. Preferences are such that all players always prefer to play with a higher ranked player than with a lower ranked player. The intuition why some players are not proposed a game when they are met is that in expected terms it pays to wait for a better match. It follows that players are voluntarily inactive. Even though that is costly, it is less costly than engaging in a game with too bad a player. However, because of the search friction and the opportunity cost of not being matched, players will never wait until the "perfect match" arrives. As a result, there will be a range of types with positive mass that is accepted.

The result that is put forward in this Section highlights the possibility that a player of a certain type θ may choose not to accept a match with someone of her own type. Given symmetry, both will not make an offer. This explains the statement: "one reason why you don't want to be a member of a club that wants you as a member". Since you reject your own type, you would reject to be matched with a clone of yourself. Still, you are accepted by some other types, so you can be a member of the club.

All the results from the first chapter apply, since this can in fact be considered as a special case where $F_i = F_j = F$. Existence and uniqueness then follow from Proposition 1 (Chapter 1). Graphically, there will be complete symmetry over the 45 degree line: acceptance regions are the exact mirror image of the reservation strategies. The interest for this Section is exactly in the interpretation of the standard result when reservation strategies are decreasing in type. The next example illustrates the phenomenon.

Example 2 *Consider an economy with one set of heterogeneous types θ distributed according to $F \sim U(0,1]$ and $r + \alpha = \beta = 1$. The preferences are additively separable $u = \theta_{partner} + \theta_{own}$. For any type θ, the first order condition (1.6), Chapter 1, that solves for a reservation strategy ϕ, given acceptance by all other types, becomes*

$$\phi^2 + 2\phi - 1 + 2\theta = 0$$

Using the implicit function theorem, it follows that the reservation strategy is downward sloping:

$$\frac{\partial \phi}{\partial \theta} = -\frac{1}{\phi + 1} < 0$$

This reservation strategy is illustrated in Figure 2.2. Note that with matching from one distribution, the picture is entirely symmetric around the 45-degree line. For all $\theta > 0.5$, the reservation type is below the lowest type. For $\theta < 0.5$, it is strictly above the lowest type: $\phi > 0$. The reservation strategy locus hits the 45-degree line (i.e. $\theta = \phi$) at $\theta = \sqrt{5} - 2 \approx 0.24$. With decreasing reservation strategies, all types in the range $\theta \in (0, 0.24)$ reject a match with someone of their own type. It can easily be verified that the black solid line is the equilibrium schedule. Note that in equilibrium, not only do the lowest types reject types of their equals, but because of the downward sloping reservation schedule, they get accepted by less than they would prefer given acceptance by all.

The intuition behind this result relies on the fact that reservation strategies given acceptance by all are downward sloping. This follows from a utility

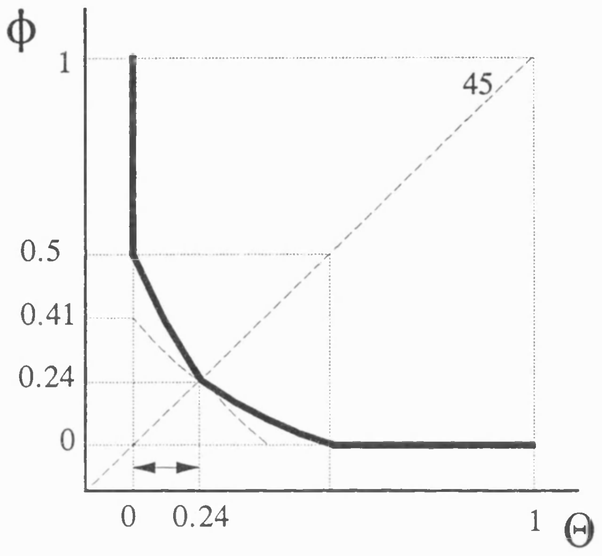

Figure 2.2: Rejection of Matches with Equals

function like the additively separable function in the example. With acceptance by all, higher types accept lower types than the lower types do. This is the result of the higher types being more impatient. Their impatience derives from the fact that with these preferences, higher types have a higher opportunity cost of not being matched. A high type in expected terms loses utility of θ per unit of time, which is higher than for a low type. In order to reduce the loss of not being matched, the high type will be prepared to engage in a match with a lower type since that minimises the time in search.

In the context of pairwise matching from one distribution, a downward sloping reservation strategy can result in types rejecting their equals. They can do so because with higher types being more impatient, they know that they will be accepted by the most desired types. This may well be in the spirit of what Groucho Marx had in mind: everyone wants to be in a club with better members than herself. Without any external effect this is impossible. On the other hand and in order for a club to exist, some external effect needs to be present. In this example, impatience drives individuals to accept partners of differing type. With some high types having a different reservation strategy, the low types can afford not to let their equals be a member of the club.

Part II

Perfect Matching

Chapter 3

Uniqueness and Negative Assortative Mating in Two-Sided Matching

In the two-sided matching literature, a large number of results have been shown on the existence of equilibrium. Typically, these are derived for general preferences, i.e. there is no restriction on the preferences of agents. The contribution in this Chapter is to extend this literature[1] by deriving a set of preferences for which the existing equilibrium is unique. This is useful for two reasons. First, understanding of uniqueness is desirable when models are used as a tool for representing economic environments[2]. Second, it turns out that the preferences for which the equilibrium is unique are of appealing economic relevance: the set is broad and it contains a number of preferences

[1]For an overview, see Roth and Sotomayor [62].

[2]A substantial part of the non-cooperative game theory has been devoted precisely to deriving uniqueness results. Milgrom and Roberts [45] even provide a method to extend uniqueness results from one model to another one with a suitable transformation of the strategy spaces.

that are commonly assumed.

The main result contains two of the preference orderings that are often used in matching and related models. First, vertical heterogeneity. The search model in Part I exhibits vertical heterogeneity: all agents have an identical ranking (i.e. there is unanimity) of the types of the other sex. As is apparent from the discussion there, quite a few pieces of related work have independently assumed those preferences[3]. In addition, the non-cooperative game theoretic literature that applies lattice theory very often makes use of preferences satisfying vertical heterogeneity. Second, horizontal heterogeneity. No two agents have identical preferences but there is a strongly systematic pattern which can best be understood in terms of "nearness". Individuals are indexed by a type on the circle. The nearer a type of the other sex, the more she is preferred. This is used amongst others in the money search models by Kiyotaki and Wright [33].

What is shown in the main Proposition is that the general set of preferences that yield uniqueness is in fact the set of convex combinations of vertical and horizontal heterogeneity. This general result shows some resemblance with the notion of single peakedness, even though it is not identical (the preferences of the two sides are over a different domain, i.e. the types of the other side, and there has to be some correspondence between the two orderings).

The uniqueness result is then used to argue for a definition of assortative mating based on preferences and not on some exogenous surplus function. The implication however is that with a preference based definition, only a very narrow class of preferences can exhibit assortative mating. It turns out

[3]See amongst others Burdett and Coles [13], Bloch and Ryder [12], McNamara and Collins [42] and Smith [69].

that only with vertical heterogeneity assortative mating is unambiguously defined. For the perfect matching model specified with preferences exhibiting non transferable utility (NTU), it follows from the main Proposition that negative assortative mating cannot occur as an equilibrium outcome. More interestingly however, the implication for models with transferable utility (TU) is that what is usually referred to as negative assortative mating in the literature[4] is drastic. The equilibrium preferences in a TU two-sided matching model exhibit vertical heterogeneity. With transferable utility, preferences will switch because the side payments are sufficiently large.

In section 3.1, the uniqueness result is derived. In section 3.2, assortative mating is defined and it turns out to be defined unambiguously only if individuals have identical preferences, i.e. there is Vertical Heterogeneity. Section 3.3 relates this to the results in assignment games and it is shown that the optimal outcome is incompatible with negative assortative mating as defined in section 3.2.

3.1 Uniqueness in a two-sided perfect matching model

The basic framework in this paper consists of a two-sided marriage model. Consider two disjoint sets $M = \{M_i\}$, $F = \{F_i\}$, $i \in I = \{1, ..., n\}$. This implies that the mass of both populations is equal: $n_M = n_F = n$. Below the analysis will be generalised to the case of a continuous distribution of types. Each female F_i has rational (i.e. complete and transitive) preferences over all males in M. Likewise for all M_i over F. Preferences will be represented using the conventional notation $F_k \succ_{M_i} F_l$, M_i strictly prefers F_k to F_l.

[4]See for example Becker [7].

Without loss of generality[5], it will be assumed that any match is preferred to being single. A matching μ can be defined as a one-to-one correspondence from $M \cup F$ onto itself of order two ($\mu^2(x) = x$) such that $\mu(M_i) \in F$ and $\mu(F_i) \in M$.

A matching μ is individually rational if it is not blocked by any individual agent. It is stable if it is individually rational and if it is not blocked by any pair of agents, one female and one male. All agents have perfect information about the preferences and there are no frictions in the market (i.e. a blocking pair can be formed at not cost). It can be shown that a stable matching exists for every marriage market (i.e. for any preferences). This existence result is put forward by Gale and Shapley [27]. In general however, the stable matching is not unique. In this section, a class of preferences is defined for which the stable matching is unique. This is the purpose of proposition 6

Proposition 6 *There exists a unique stable matching $\mu(F_i) = M_i$, $\forall i \in I$ if the preferences satisfy:*

$$\begin{aligned} &\forall M_i \in M : F_i \succ_{M_i} F_j, \forall j < i \\ &\forall F_i \in F : M_i \succ_{F_i} M_j, \forall j < i \end{aligned} \tag{3.1}$$

Proof. Suppose there exists a stable matching μ' different from μ with for some i $\mu'(F_i) = M_k$, $k \neq i$. Given the definition of a stable matching there must also exist some $j \neq k$ such that $\mu'(M_j) = F_l$, $l \neq j$. Let $\lambda = \max\{i : \mu'(F_i) = M_k, k \neq i\} = \max\{j : \mu'(M_j) = F_l, l \neq j\}$. Then $\mu'(F_\lambda) = M_k$ implies $\lambda > k$. Likewise, $\mu'(M_\lambda) = F_l$ implies $\lambda > l$. Given preferences (3.1), it follows that $M_\lambda \succ_{F_\lambda} M_k$ and $F_\lambda \succ_{M_\lambda} F_l$. F_λ and M_λ form a blocking pair and hence μ' is not a stable match. Since the blocking pair is always F_λ and M_λ, the only stable matching (there always exists a stable matching, Gale and Shapley [27]) is $\mu(F_i) = M_i$, $\forall i \in I$. ■

[5] The more general model is discussed at length in Roth and Sotomayor [62].

The intuition behind the preferences (3.1) is as follows. Suppose each individual can be given a rank i. The proof can be seen as some recursive elimination process or a specific form of the Gale-Shapley algorithm. The preferences are defined such that the for $i = I$, both the male and the female must prefer each other above anyone else. Hence, they rank each other as highest. Clearly, that pair will always be a blocking pair unless they are matched with each other. Now consider both sexes with type $i = I - 1$. Given the preferences (3.1), they must prefer type I and type $I - 1$ above all other types. It may be that they most prefer a partner of type I, but they can never be matched with type I since then both types I (i.e. each from the opposite sex) would form a blocking pair. Hence, both sexes of type $I - 1$ will be matched to each other. If they are matched to a type $j < I - 1$, together they will form a blocking pair. In general, any two partners with the same rank must prefer each other above any partner of a lower rank, given preferences (3.1). Since all types of a higher rank already form a blocking pair (from our induction argument above), they also form a blocking pair if matched with a lower type. In other words, as long as there is no type of the other sex with a lower rank which is preferred to the type with equal rank, equilibrium is unique.

Note that this does not exclude a marriage market with a unique stable matching where $\mu(F_i) = M_k$, with $i \neq k$. However, with an appropriate "relabelling" of the individuals, the same result can be obtained. This is the result of the fact that the ordering of types does not affect the outcome.

Corollary 3 *Any stable matching $\mu(F_i) = M_k$, $i \neq k$, is unique if the preferences of M_l' and F_l', $\forall l \in I$ satisfy (3.1), where $M_l' = \mu(F_i)$ and $F_l' = \mu^2(F_i)$*

Proof. Immediate from Proposition 6. ■

Corollary 4 *There is a unique stable matching $\mu(F_i) = M_i$ for preferences satisfying:*

(Vertical Heterogeneity)

$$\begin{aligned} &\forall M_i \in M : F_k \succ_{M_i} F_j, \forall k > j \\ &\forall F_i \in F : M_k \succ_{F_i} M_j, \forall k > j \end{aligned} \tag{3.2}$$

(Horizontal Heterogeneity)

$$\begin{aligned} &\forall M_i \in M : F_i \succ_{M_i} F_j, \forall j \\ &\forall F_i \in F : M_i \succ_{F_i} M_j, \forall j \end{aligned} \tag{3.3}$$

Proof. Immediate from Proposition 6. ■

In the case of Vertical Heterogeneity all individuals have identical preferences over the other types, so there is a kind of "objective" ranking of all types of both sexes.[6] In the case of horizontal heterogeneity, each individual has a different most preferred individual. This implies a "subjective" ranking of the types of the other sex. Vertical Heterogeneity and Horizontal Heterogeneity are two limit cases of the preferences (3.1). In fact, the preferences of (3.1) can be considered the set of all convex combinations of Vertical and Horizontal Heterogeneity. Note that the preferences (3.1) show some resemblance with single peakedness. However, the requirement (3.1) differs in two respects. First, it is stronger than mere single peakedness since there has to be some connection between the natural ordering of both sets. Second, in another respect it is also slightly weaker since only an order restriction is imposed, comparable in spirit to Single Crossing Conditions or Hierarchical Adherence (Roberts [57]). This is relevant when preferences are represented by utility functions.

[6]Note also that the preference restriction (3.2) is over the whole set of types thus imposing an order over the whole set. This is not the case in (3.1).

This result can be generalised to continuous preferences. Consider $M_i \in M$ and $F_i \in F$ distributed according to the cumulative distribution functions G_M and G_F. Without loss of generality, let $M = F = [0,1]$. The rank of an individual will then be given by the value of the distribution function: F_i and M_j have the same rank if and only if $G_F(F_i) = G_M(M_j)$. Proposition 6 can now be generalised.

Proposition 7 *There exists a unique stable matching* $\mu(F_i) = M_j$, *with* $G_F(F_i) = G_M(M_j)$, $\forall i,j \in [0,1]$, *if the preferences satisfy:*

$$\begin{aligned} &\forall M_j \in M : F_i \succ_{M_j} F_k, \forall k < i \\ &\forall F_i \in F : M_j \succ_{F_i} M_k, \forall k < j \end{aligned} \tag{3.4}$$

Proof. As in Proposition 6 ■

The intuition is exactly identical to the case of a discrete and countable distribution. No type of the other sex of a lower rank than your own is preferred above the type of the same rank. Clearly, the corollaries can easily be extended to the more general case. For the remainder of the paper, the more general model and notation will be used. Note that the countable discrete case is a special case of a uniform distribution: $U_F(F_i) = U_M(M_j)$ implies $i = j$.

3.2 Assortative Mating

In common language, assortative mating is loosely defined as the "mating of (un)likes". The word assortative is derived from the French *assorter*: to sort, to arrange according to similar characteristics. With such a broad definition, two possible interpretations can be pursued. One defines this assorting based on an exogenous characteristic (say colour), independent of the preferences. The other is to define "likes" endogenously, derived from the preferences of the agents. In what follows, the second interpretation is chosen. This is in line with the interpretation most adhered to in neoclassical economic thought with its foundation in utilitarianism. Revealed preference arguments however cannot entirely dismiss objectivism. Assortative mating requires some objective measure of which characteristics are "alike". Based on the individuals' preferences, some aggregate measure of "likeness" will be derived. This Section proceeds first by defining assortative mating. Then, the equilibrium outcome under that definition is derived and it is shown that no equilibrium can exhibit negative assortative mating.

All individuals of both sexes have a natural order over the types of the other sex. In the literature, assortative mating is implicitly defined by some natural order over the alternatives which is common to all types. Becker [7] for example assumes that for all females, say, the value of a match monotonically increases in the type of the male. This implies there exists a natural order of males common to all females. A similar natural order of females exists that is common to all males. Here, the common natural order in addition is obtained from the aggregation of the individuals' preferences. The formal definition reads

Definition 3 *There is an common natural order of all types of one sex provided:*

$$\forall M_i \in M : F_j \succ_{M_i} F_k, \forall j > k \qquad (3.5)$$

and

$$\forall F_l \in F : M_m \succ_{F_l} M_n, \forall m > n \qquad (3.6)$$

This definition imposes a natural order on each of the sets M and F separately. It is important to note there is no connection between the two rankings. In the general specification of the preferences (3.1), both restrictions on the preferences are connected through the type i. The intuition of the preferences (3.1) is that a type F_i does not have preferences where M_i is ranked lower than position i. Here, no such relative ranking (i.e. with cross reference to the other set) is required.[7]

The aggregate ranking (i.e. the common natural order) necessary for the notion of assortative mating has been pinned down. It is now trivial to define positive and negative assortative mating. Assortative mating necessarily makes the link to an allocation μ.

Definition 4 *Provided the preferences satisfy the common natural order restriction with F_1 and M_1 the highest ranked types, a Matching $\mu(F_i) = M_j$ exhibits:*

Positive *Assortative Mating if $G_F(F_i) = G_M(\mu(F_i))$, $\forall i \in [0,1]$;*

Negative *Assortative Mating if $G_F(F_i) = 1 - G_M(\mu(F_i))$, $\forall i \in [0,1]$.*

With the restriction of a common order, the objective is now to look at the set of equilibrium allocations within that restriction. This gives rise to the main result of this Section.

Proposition 8 ***(Impossibility of Negative Assortative Mating)*** *There can never be a stable match exhibiting Negative Assortative Mating. More-*

[7] In the definition of assortative mating, the qualifier "objective" could be added. One could argue that all individuals can also have identical preferences over the different matches μ, rather than over the individuals with whom to match. This requires preferences to satisfy horizontal heterogeneity as in equation (3.3). Such a "subjective" definition of assortative mating is tautological since all agree on what the best match is and thus no other stable matching can be the equilibrium outcome. It does not allow the individuals to be ranked, only the matches. For the remainder of this paper therefore, assortative mating will refer to objective assortative mating.

over, under the common natural order restrictions there is always a unique stable matching exhibiting Positive Assortative Mating.

Proof. The common natural order restrictions (3.5) and (3.6) jointly, and the preferences satisfying Vertical Heterogeneity (3.2), are similar but not identical. Vertical Heterogeneity is more specific since it imposes a connection between the natural order of the two sets. However, the additional condition in the definition of positive and negative assortative mating that F_1 and M_1 are the highest ranked types also imposes this connection. As a result, (3.5), (3.6) and the definition of assortative mating jointly is identical to the condition (3.2), Vertical Heterogeneity. From Proposition 7, the equilibrium is unique and satisfies: $\mu(F_i) = M_j$ with $G_F(F_i) = G_M(M_j)$, $\forall i, j \in [0, 1]$. This is Positive Assortative Mating. ■

The proposition highlights the fact that negative assortative mating is a matter of labelling the natural order. If the highest ranked types are labelled one, then there is no doubt. Suppose now the highest male is labelled zero, then the equilibrium allocation is $\mu(F_i) = M_j$ with $G_F(F_i) = 1 - G_M(M_j)$. But this is not negative assortative mating since the most preferred male is still matched with the most preferred female. The only difference is that they are labelled differently. Based on their preferences they are positively assorted.

Notice that this definition involves an extremely strong restriction on the preference structure of individuals. If it is not satisfied, assortative mating is not defined. In that sense this is a negative result, since any statement about assortative mating has to be made within a framework of these preferences. However, this does not imply that this restriction is irrelevant. Vertical Heterogeneity has intuitive appeal and is frequently assumed in economic modelling with heterogeneous agents.

So far in this Chapter, the matching problem has been considered in terms of the bare bones preferences of individuals. Introducing utilities that satisfy these preferences does not alter anything. Consider the following example. Let all males have preferences over matches with females such that $\partial u_M/\partial F_j > 0$ and similar for the females $\partial u_F/\partial M_i > 0$. Though this does not necessarily imply that utility is type independent, i.e. $\partial u_M/\partial M_i \neq 0$, it does imply that all types of one sex have identical preferences over the types of the other sex. Clearly, irrespective of whether inputs are substitutes, i.e. $\partial^2 u_F/\partial M_i \partial F_j + \partial^2 u_M/\partial M_i \partial F_j < 0$, the unique stable matching will be $\mu(M_i) = F_i$. The inference is that if the two highest types are not matched together, they can block any matching. This is also true for any pair below that is not matched to a partner of the same rank, given that all higher ranked partners are matched to a partner with the same rank.

3.3 The Assignment game

In the Sections above, the model is defined in terms of preferences. Moreover, the introduction of utilities does not alter the results, provided utility is Non Transferable (NTU). Allowing for side-payments however, will change the initial pay-offs and can change the preferences. This class of Transferable Utility (TU) models has extensively been discussed in the literature and is known as the assignment game, originally formulated by Shapley and Shubik [66]. In this Section, it is shown that with a definition of Assortative Mating derived from the preferences of individuals, Negative Assortative Mating is impossible, even in TU models .

Again, as in Section 3.2, a slightly simplified version will be discussed where the number of males equals the number of females[8]. Consider the same sets of agents F and M, and a matrix of real non-negative numbers α_{ij} associated with each partnership of (i, j) in $F \times M$. The private benefit to a female F_i from a partnership with a male M_j is f_{ij} for the female and m_{ij} for the male, so that $m_{ij} + f_{ij} = \alpha_{ij}$. A feasible assignment can be represented by a matrix $x = (x_{ij})$ of zeros and ones such that $\sum_i x_{ij} = \sum_j x_{ij} = 1$. Let $u_i = m_{ij}$ and $v_j = f_{ij}$ if $x_{ij} = 1$. A feasible outcome is stable if $u_i + v_j \geq \alpha_{ij}$, $\forall (i, j) \in S \times S$. The stability condition does not require the specification of any "out-of-equilibrium" pay-offs. The reason is that the equilibrium sought for here is cooperative and that given TU, the only relevant information is in the sum of the out-of-equilibrium pay-offs. Cooperation between a pair of a male and a female agent is needed.

Generally, results are comparable with the results from the NTU model. Most substantially, Shapley and Shubik [66] show that the set of stable outcomes coincides with the core and that it is non-empty. They also show that although side-payments are allowed between members of different partnerships, they do not occur in equilibrium. The main result deviating from NTU models is however that any stable outcome x is compatible with an optimal assignment, i.e. a stable matching maximises total output $\sum_{i,j} \alpha_{ij} x_{ij}$. Even though this optimality result does not imply uniqueness, it follows that if there exists no assignment $x' \neq x$ such that $\sum_{i,j} \alpha_{ij} x'_{ij} = \sum_{i,j} \alpha_{ij} x_{ij}$ then equilibrium is unique. Without loss of generality, this will be assumed for the remainder of the paper.

It now becomes nearly trivial to reproduce Becker's [7] argument: the decentralised outcome always yields the optimal outcome. Clearly, this is nothing more than the Shapley and Shubik [66] optimality result. What

[8] The notation draws heavily from Roth and Sotomayor [62].

Becker has introduced however is the notion of Assortative Mating, this by imposing some additional structure on the general assignment game. His definition of assortative mating is $\partial\alpha_{ij}/\partial M_i > 0$ and $\partial\alpha_{ij}/\partial F_j > 0$. In the $n = 2$ case for example, that implies that $\alpha_{11} > \alpha_{12} > \alpha_{22}$ and $\alpha_{11} > \alpha_{21} > \alpha_{22}$. That does not exclude the equilibrium assignment

$$x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

as long as the $\alpha_{12} + \alpha_{21} > \alpha_{11} + \alpha_{22}$. This is the case when the function α exhibits substitutability: $\partial^2\alpha_{ij}/\partial M_i \partial F_j < 0$. Stability in this assignment game requires $m_{21} + f_{12} > \alpha_{22}$ and $m_{12} + f_{21} > \alpha_{11}$.

It can now easily be shown that these conditions cannot be consistent with the definition of Assortative Mating derived from the preferences of all individuals. The reason is that given these conditions individual preferences are reversed for an equilibrium to exist. In the $n = 2$ example, it is impossible for both $m_{12} < m_{11}$ and $f_{21} < f_{11}$ (condition necessary for Assortative Mating to be defined unambiguously) since this is in contradiction with $m_{12} + f_{21} > \alpha_{11}$. In other words, the definition of Assortative Mating $\partial\alpha_{ij}/\partial M_i > 0$ and $\partial\alpha_{ij}/\partial F_j > 0$ adopted by Becker [7] is based on some exogenous characteristic, i.e. the contribution to total output, and not on the aggregation of individuals' preferences. In that sense, any outcome is possible. The fact that optimality is obtained is merely due to the property of assignment games as shown in Shapley and Shubik [66].

3.4 Concluding Remarks

The results in this Chapter are both positive and negative. The positive result is that a large class of preferences is identified for which the two-sided perfect matching model has a unique solution. Moreover, the set of preferences yielding uniqueness is the convex combination of two familiar benchmark preferences: Vertical and Horizontal Heterogeneity. The other result is negative in two respects. First, it is shown that only for a very small class of preferences (i.e. Vertical Heterogeneity) assortative mating, however often referred to in economics, is unambiguously defined. Second, if it is defined unambiguously, negative assortative mating is impossible. The implication of the first aspect is that one either accepts or rejects such restrictions on the preferences. In describing a marriage market for example, many people will have difficulties accepting there is such a thing as identical preferences over mates. Love very often is in the eye of the beholder. On the other hand, dropping "identical preferences" assumptions will leave very little economic theories entirely in tact. The implications of the second negative result are far more dramatic. If inputs are substitutes in production, negative assortative mating is socially optimal. With NTU it cannot be achieved by the market mechanism.

Chapter 4

Working for a Better Job

We like to believe that we do not merely work because there is money on our account at the end of the month. Certainly, sociologist and psychologists have a range of alternative explanations, of which status may well be a prominent one. However, even as an economist there is something to be said about the motives of a worker to provide effort. Incentives schemes and bonuses, as firm and business jargon is plenty of, is clearly relevant, but that is still related to the end-of-the-month pay check. As such, they are well understood in the traditional analysis, with or without asymmetric information and uncertainty. Nonetheless, incentive schemes cannot explain why for example a Junior Graduate at nearly all top consulting firms or merchant banks upon entry works at least 70 hours per week at a fixed annual wage and one year later earns no less than one and a half to two times the entry wage, putting in about the same number of hours. This is not a typical phenomenon related to jobs for economists. Medical doctors specialising in hospitals work legendary long hours verging on insanity. This equally applies to solicitors, barristers and architects active in partnerships. With leisure a normal good, the standard neoclassical effort supply model would predict that in all these cases the more junior employee works considerably less hours.

In the quest for an explanation of this labour market behaviour, some particular features of these markets are worth noting. First, those workers entering the firm or the labour market know that there is a hierarchy of jobs to be performed. Second, they do realise that they will not necessarily perform the same job: they have the possibility to move up in the hierarchy depending on their performance. As a result, labour market decisions do not only take into account the current wage, but also the future career plan and thus future wages. Moreover, many firms take great care in designing these career plans for their employees, conditional upon performance of course.

This Chapter would like to argue that "better-job" opportunities in a deterministic context are of significant importance to explain labour market behaviour. For that purpose, the standard neoclassical effort supply model is extended to capture the notion of "better-job" opportunities. This notion can be split up into two determinants. First, it implies that there is some hierarchy or ranking of jobs with a higher pay-off to the higher ranked jobs. In other words, the productivity of a worker is not merely a function of her ability, but also of the type of job performed. The existence of a distribution of jobs is taken as a technological characteristic: the manager's job is different from the shop floor worker's with the both jobs performed by the same person yielding different output levels. The equilibrium allocation of jobs will be such that the highest productivity workers get the highest productivity jobs. This implies that the productivity of any worker is a function of her Rank-Order.

Second, "better-job" opportunities are inherently dynamic: in a regime where workers change jobs within the hierarchy, the job performed in the next period will have some relation to the job performed in the current period.

Not only do firms tend to reward their hard-working and loyal employees with a better job, it is also efficient to allocate the higher ranked workers to the higher ranked jobs. In each period, there is an efficient allocation of workers to jobs with the workers' types determined by the effort provided in the past. Individuals will improve their type in the next period by working hard now. This affects their rank in the distribution of workers and hence the allocation to a higher productivity job. These two premises capture the notion of "better-job" opportunities and imply that labour effort is modelled as a dynamic consumption-leisure decision in a Rank-Order tournament environment.

Within this framework, it can trivially be shown that the effort supply decision of a worker does not merely take into account the trade-off between current consumption and effort, but between all future consumption that is affected by current effort. Rather than by differences in preferences, hours of work are explained by the incentive structure and the forward looking behaviour of workers. For as much as there is a real effect on the increase in ability or skill of the worker, this extra effort is socially optimal. This is nothing more than a reformulation of human capital theory. However, if the productivity of a worker is not merely a function of her ability, but also of the job performed, the implication is that there is an inefficient excess-supply of effort compared to the social optimal. This is due to the fact that every individual gets a marginal benefit from extra effort in having a higher ranked job in the next period. In equilibrium no worker will change rank, but still it is not individually rational not to put in the extra effort. The reason is that doing so prevents lower ranked types to become higher ranked. The inefficiency is the result of an external effect of the individual effort choice on the pay-off of all other workers: increasing one's effort increases one's future rank but also lowers some other workers' rank. By definition, rank represents a zero-sum game: the number of positions gained by one player equals the total of positions lost by all other players together. In fact, this Rank-Order Inefficiency feature exactly fits the definition of the rat race: "the struggle to maintain one's position in work or life". In the economic literature, this is usually referred to in a signalling context (Akerlof [3]). The main contribution of this paper is to derive the rat race result not as a signalling equilibrium (the information environment is symmetric and deterministic (but dynamic)), but as a result of the allocation of agents in the matching market.

The promotion model is used to study the effect of different labour market environments on the labour supply decision. First, the distributional impact is analysed. There are three distinct channels through which distributional considerations have an effect on the effort supply decision: the distribution of abilities, the skill premium and the job premium (i.e. the marginal return from performing a higher ranked job). Inequality in the distribution of skills decreases effort whereas inequality in the wage distribution (i.e. both a higher skill and job premium) increases effort supply. Moreover, the effect of the ability distribution and the job premium effect are inefficient. Theoretically, the net aggregate effect of inequality on both effort supply and efficiency is ambiguous. Subsequently, the rate of turnover, i.e. the number of job allocations per unit of time, is introduced in the model. It can be shown that a higher turnover increases the Rank-Order Inefficiency. It follows that effort increases with the rate of turnover. This indicates that the institutional characteristics of the labour market are of considerable importance for understanding effort decisions and efficiency.

The examples at the outset of the paper could easily be earmarked as mere anecdotal evidence. In order to support the theoretical analysis and in order to add to the credibility of the underlying assumptions, the theoretical results are matched with seven stylised facts from the empirical labour market literature. Roughly, these stylised facts can be divided into observations within labour markets which cannot be explained by the standard analysis and a comparison between labour markets.

The remainder of the Chapter is organised as follows. In the next Section, the basic model is being laid out and the paper is situated relative to the existing theoretical literature. In Section 4.2, the main results and the Rank-Order Inefficiency are derived. The first stylised facts are discussed. Section 4.3 discusses the impact of distribution on effort supply and efficiency and uses the empirical literature to sign the ambiguity. In Section 4.4, the rate of turnover is introduced. Again, the results are matched with the empirical literature. An extension to a life cycle model is proposed in Section 4.5. The Chapter is rounded up with some Concluding Remarks.

4.1 The Basic Model and Related Literature

Consider an economy with a set of heterogeneous agents. There are infinitely many individuals and the total mass of the population is one. Agents are indexed by a type $\theta \in \Theta \subset \Re^{+}$, the characteristics of the individual. The characteristics of all agents in the economy are distributed according to the cumulative distribution function $F(\theta)$, with positive density function $f(\theta)$. Characteristics are *under all circumstances* perfectly observable: information is both symmetric and complete.

Individuals have the choice between two goods, consumption C and effort (or hours worked) e. Preferences from consumption of these goods are represented by the strictly quasi-concave utility function $u(C, e) = U(C) - e$, with

$U_C > 0$ and U strictly quasi-concave. The quasi-linear preferences are chosen to simplify the analysis below. The whole analysis carries through however for any strictly monotonic and strictly quasi-concave (in both consumption and leisure, the negative of effort) functional form.

The only endowment of an agent is leisure. However, there is a technology that transforms effort into consumption. This technology is offered by atomless firms competing for the workers. This model entirely defines the standard neoclassical effort supply model. This model has been extended to the case where productivity of a worker is type-dependent. As in the optimal taxation literature (Mirrlees [46]), productive ability of the individual increases in the type θ.

For the purpose of this paper, two additional features are introduced. First, the technology transforming effort into consumption. In the introduction, the premise is proposed that there exists a distribution of jobs, each with different productivity, given an identical worker. This is interpreted as a technological feature. In an economy there are tasks of different productivity and all of them have to be performed. A job will by characterised by a type $\phi \in \Phi \subset \Re^+$ and jobs are distributed according to the cumulative distribution function $G(\phi)$ with positive density function $g(\phi)$.

Total productivity of a worker depends both on her own type and on the type of the job she performs. For a given level of effort, her productivity is given by the production function $T(\theta, \phi)$ (with $T_\theta > 0$, $T_\phi > 0$ and $T_{\theta\phi} > 0$[1]).

[1] The assumption of strategic complementarities is not vital for our framework. With strategic substitutes the highest type workers would be matched with the lowest type workers and consumption would be a decreasing function of the rank. The lowest jobs are most desirable and the Rank-Order Tournament still holds as a function of the inverse rank of the worker. The crucial part is that the cross-partial is different from zero. If that is not the case, rank does not matter in any way and the solution to the model would be

The equilibrium allocation of workers to jobs in this assignment game will determine each individual's wage. Any allocation will be referred to as a matching m. It is a one-to-one correspondence of $\Theta \cup \Phi$ onto itself of order two such that $m(\theta) \in \Phi$ and $m(\phi) \in \Theta$. Shapley and Shubik [66] show that a stable matching exists and in Becker [7] it is shown this allocation is unique. In addition, the non-cooperative Nash equilbrium with one party, say the firms holding the jobs, announcing a type of the other party coincides with this equilibrium. The unique equilibrium matching m^* exhibits positive assortative mating (Becker [7]): $\phi = m^*(\theta)$ if and only if $F(\theta) = G(\phi)$. In equilibrium, a worker is allocated to a job that has the same rank as the worker.

The surplus of this matching is split into profits and a wage: $T(\theta, \phi) = \pi(\theta, \phi) + c(\theta, \phi)$. An individual worker will at the margin be able to command what she can contribute to the surplus. Hence the first-order condition must satisfy:

$$c_\theta(\theta, \phi) = \frac{\partial T(\theta, \phi)}{\partial \theta} \tag{4.1}$$

Taking in to account the equilibrium allocation m^*, such that $\phi = G^{-1}[F(\theta)]$, the wage received by a worker of type θ is equal to

$$c(\theta) = \int \frac{\partial T\left(\theta, G^{-1}[F(\theta)]\right)}{\partial \theta} d\theta \tag{4.2}$$

The wage received is thus not only a function of the type θ, but as a result of the equilibrium allocation also of the rank. Since G is monotonically increasing and $c_\phi > 0$, it follows that $c_F > 0$. Without loss of generality, in what follows the wage of a worker is written as $c(\theta, F(\theta))$.

The same applies to the firm side. It is assumed that firms have zero mass and open one job at cost ϵ only by making a random draw from the job

identical to the standard neoclassical effort supply model.

distribution $\tilde{G}(\phi)$. With firms Bertrand[2] competing at the entry decision, expected profits $E\pi(\theta, \phi)$ are driven to zero.

Once allocated to a certain job, the total consumption is in addition a function of the amount of effort provided. Hence, $C = c(e, \theta, F(\theta))$. In order to ensure existence, functional forms will be considered that are monotonically increasing in all the arguments and further it is assumed that $c_F(\theta = 0) = 0$[3]. This model is identical to the standard neoclassical effort supply model with the additional feature of productivity a function of the job performed. This is interpreted as a deterministic Rank-Order Tournament with heterogeneous agents. In the static game, an individual strategy is a choice of effort in order to maximise utility subject to the technology. Since types, and hence the distribution of types, are exogenous, an individual's optimising strategy is independent of other players' strategies. As a result, an optimising strategy is dominant. The solution to this static optimisation problem with perfect observability satisfies all requirements for existence and uniqueness of a general competitive equilibrium. The first-order condition equalises the marginal utility of consumption with the marginal disutility of effort. From the First Theorem of Welfare Economics this solution is Pareto Efficient. Moreover, with atomless firms competing, profits are driven to zero and all the surplus accrues to the workers.

---

[2]Assuming there is a cost $\epsilon > 0$ of opening a job is necessary to avoid that firms continue to open jobs until all jobs offered are of the highest type. Note that the equilibrium distribution $G(\phi)$ is in fact the truncated distribution of $\tilde{G}(\phi)$. More jobs may be opened than the mass of workers as long as expected profits are non-negative. In equilibrium only the best jobs with a total mass one are filled. Though it is beyond the purpose of this paper, it can be shown that with $\lim \epsilon \rightarrow 0$, in equilibrium there is always a non-degenerate distribution G of jobs on offer.

[3]This assumption will ensure that a number of Pareto Dominated equilibria are eliminated (see Spence [70]).

The second additional feature captures the dynamic effect of working for a better job. The game is repeated twice with the individual type in the second period a function of the effort in the first period. With stationary utility functions, the variables will be time-indexed by $t \in \{0, 1\}$. Individuals' types in period $t = 1$, are a function of their past performance, i.e. the effort expended in period 0. $\theta_1 = g(e_0, \theta_0)$, with g concave and $g_e > 0$ and $g_\theta > 0$.

Since the type θ_1 is affected by the current choice of effort, effort today will determine future consumption from increased ability. This is nothing more than a human capital effect. Workers consider current effort as an investment in skills. Since the acquisition of skills does not depend on other players' strategies, strategies in the presence of the human capital effect remain dominant. In addition however, next period's ability also affects the distribution of types $F_1(\theta)$. With the Rank-Order Tournament which matches higher types to better jobs, productivity is a function of the rank. It follows that next period's rank and thus consumption is a function of current period's effort. Strategies now are no longer dominant, since next period's rank is also a function of all other players' strategies. Rank is a zero sum game: moving up one position implies that someone else will move down one position. It follows that an equilibrium strategy will be a solution to a fixed point problem. Note that in general, the type in period 1 is made a function of the type in period 0. This will turn out to be a necessary condition for existence, parallel to the so called Spence-Mirrlees or Single Crossing condition: higher types are more effective in transforming effort into next period's type. It is entirely reasonable to assume that if a high and a low type have expended the same effort this period, the high type will still be a high type next period. Single Crossing does not axiomatically exclude leapfrogging though since the low type can put in more effort!

Finally, a remark on the dynamic nature of the setting. It highlights the usefulness of a quasi-linear utility function. In general, consumption smoothing affects the equilibrium level of effort. By definition, in the quasi-linear case, there is no income effect. As a result, the consumption smoothing effect on effort is entirely annihilated[4].

In this Repeated Rank-Order Tournament, an Equilibrium in the decentralised system will be a decision rule which, given an optimal strategy of all other individuals, maximises the utility of the individual. This has to hold in every period. The concept is Nash Equilibrium. The objective of each individual is to maximise total discounted utility. The maximisation problem in each period $t \in \{0, 1\}$, given the discount rate $\rho < 1$, solves

$$\begin{aligned} &\max_{e_t} \sum \rho^t [U(C_t) - e_t] \\ \text{s.t.} \quad &C_t = c(e_t, \theta, F_t(\theta)) \\ &\theta_1 = g(e_0, \theta_0) \\ &F_0(\theta) \end{aligned} \tag{4.3}$$

Before explicitly deriving the solution to this problem, our approach will briefly be situated relative to the existing literature. First, the dependence of productivity on the job performed is modelled as a Rank-Order Tournament, introduced by Lazear and Rosen [36]. They are compensation schemes which pay according to an individual's ordinal rank in an organisation rather than exclusively according to his or her output level. Though the concept is by now used in a fairly general way, the initial paper and the debate in the following literature concentrates on the unobservability of effort and the efficiency of Tournament contracts. Their main results however are derived under the assumption of identical agents. Then, depending on the degree

[4]With a non-separable utility function the results in this paper will continue to hold, but they have to be disentangled from the consumption-smoothing effect.

of risk aversions, the social optimal contract can be achieved. This is no longer true when agents are heterogeneous. In this paper, the randomness of observed effort will be dropped, but agents are heterogeneous. In such a static and deterministic framework, Lazear and Rosen [36] predict a first best solution. This will also be the case in our model. The inefficiency mentioned in the Introduction is due to the repeated nature of the Tournament.

Throughout this paper, it will become apparent that many of the results are conspicuously similar to many of the standard asymmetric information (signalling, Principal-Agent) problems, even though in this context there is no asymmetry of information: types and effort are perfectly observable in all periods! Consider for example the standard Spence [70] signalling model. Implicitly, the signalling model exhibits the Rank-Order Tournament feature in the sense that agents are heterogeneous with the wage a function of the ranking of the types. There is a (partly) wasteful technology (typically education) which at some cost affects the wage received in the job. The return on education (i.e. the degree obtained) is perfectly observable, as is the effort provided in our model. The main difference in our model is that the employer does observe the types (rather than the perfectly correlated signal), and that the technology (as a function of effort) changes the types (rather than the signal). Like Spence's education technology can change the signal and thus the Rank-Order of workers in terms of the wage received, our dynamic technology can change the true types and their Rank-Order.

It is then no surprise that there is a necessary condition for equilibrium which is congruent with the Spence-Mirrlees[5] or Single Crossing condition. In

[5]Note that the model in this paper is formally related to the Mirrlees [46] Optimal Taxation framework. Though in this context, there is no information problem, it will turn out (see below) that income taxes will have welfare effects.

the signalling literature, this condition requires that higher types are more effective in "producing" the signal, i.e. the marginal cost of education is decreasing in type. If not, there is not enough informative value to the signal. In the optimal taxation literature, it requires that higher types have a lower preference for extra leisure than lower types. In our matching context, it is essential that higher types are more effective in transforming effort this period in a higher type next period. In reduced form, it will imply that higher types have a lower preference for extra leisure.

Very closely related in spirit to the Spence signalling model is Akerlof's [3] rat race model. Because of imperfect observability of effort, firms reward workers based on their output and the rank of their output. Hence, workers will put in more effort than efficient because all of the other workers do. Every worker wants to signal ability to the employer. The overlap between signalling and the deterministic rat race in the promotion context is again very conspicuous. Firms however do not use the workers' output as an indicator to be judged on, but the observable type which is a function of past effort. It highlights however an additional characteristic of the rat race: in labour market environments, the effort provided is not necessarily "wasted", as is predicted by the signalling literature. Productivity actually increases. The problem is that because of the rat race caused by the allocation mechanism according to rank, people sacrifice too much leisure. It is important to note that however conspicuous the similarity of both approaches, the source of the inefficiency here is clearly not the informational externality. The source here is the externality from allocating workers to jobs. Productivity is not merely inherent in the worker's characteristics, but also to the job performed.

The overlap between the signalling literature and this promotion analysis has been touched upon very closely by Holmström's [31] discussion of Fama

[21]. Fama argues that informational externalities are taken care of by the market in a dynamic context. Reputation effects will police moral hazard problems: in the provision of effort now, managers do not only worry about their current income, but about the income in all future periods. Hence, in a repeated moral hazard problem with past observed action, inefficiency will disappear as time goes to infinity. Holmström claims that this is true only with a linear production technology. When the production technology is non-linear, the inefficiency does not disappear. Moreover, as an example, Holmström discusses the example of matching workers to jobs, which yields a convex returns to ability function (as in Rosen's [60] superstars). Clearly, our Rank-Order feature exactly yields such a convex returns function and is nothing more than a production function with complementarities between workers matched to jobs. The main difference though is that our analysis shows that this inefficiency holds even in a full information context. What Holmström and Fama interpret as reputation does not necessarily rely on imperfect observability. Fama's claim (i.e. informational externalities disappear in the long run) still holds ground. Where Holmström believed to have rejected his claim is not due to informational externalities but to allocative externalities!

Further, an interesting investigation into the endogenous creation of social norms is of interest to our framework. Cole, Mailath and Postlewaite [14] derive different social systems in a marriage framework. In a perfect matching context, the Rank-Order feature is endogenous with complementarities between vertically heterogeneous types of both sexes. Status then becomes the Rank-Order (according to the prevailing status concept in equilibrium) of the individual. Different conceptions of what constitutes status (wealth, ancestors,...) can be supported in equilibrium. The technology affecting your rank is a capital accumulation decision and, similar to the promotion context, works intertemporally.

Finally, the Rank-Order feature generating an increase in the contribution can also be welfare improving when imbedded in a different general framework. John Morgan [47] looks into the effect of Lotteries on the financing of public goods. He shows that lotteries generate higher contributions than voluntary donations. Lotteries can be considered as a Rank-Order Tournament and the extra contribution it generates offsets the inefficient level of voluntary donations towards the provision of the public good. Clearly, in this context the Tournament is welfare improving since the free rider problem that exists in the provision of a public good results in a welfare inferior outcome to start with. In our model, the benchmark is efficient which implies that the extra effort contribution is welfare deteriorating.

4.2 The Main Results

Before discussing the solution of the decentralised economy and in order to establish a benchmark for comparison of the results, the model is first solved by the Social Planner who has a Utilitarian Social Welfare Function as an objective[6]. She chooses all individuals' level of effort in both periods such as to maximise the discounted sum of utilities. That implies that she does not take into account any strategic effect from other players who look for an opportunity to leapfrog. In fact, by optimising for all individuals jointly, the planner acts as if she signs a joint contract for all workers and from which all of them benefit. The socially optimal allocation (e_t^O, C_t^O, θ) satisfies

[6]Contrary to the asymmetric information literature where imperfect information is compared with full information, our benchmark compares the decentralised job allocation system with the centralised system.

$$\frac{dU_0}{de_0} + \rho \frac{\partial U_1}{\partial \theta_1} \frac{d\theta_1}{de_0} = 1 \text{ and } \frac{dU_1}{de_1} = 1 \tag{4.4}$$

The solution to this differential equation is unique since the initial condition $U_t(0) = 0$ yields a definite solution when there is no individual strategic interaction. In the second period, the choice of effort is identical to the choice in the static standard neoclassical framework. The efficient choice of effort in the first period is different from the static benchmark. Lemma 8 shows that, quite intuitively, effort in the presence of a human capital effect is higher than effort in a static effort supply model.

Lemma 6 *Effort in the presence of a human capital effect is higher than in the static case.*

Proof. The necessary condition for the human capital effect to exist is $c_\theta > 0$. Applying the chain rule yields that $\frac{\partial U_1}{\partial \theta_1} \frac{d\theta_1}{de_0} = U_C c_\theta g_e > 0$. With $U(C)$ strictly concave, it follows that e_1^O is strictly bigger than the effort in the static case. ■

Because the agent now takes into account the effect of current effort on next period's ability, she does not only care about current consumption but also about future consumption from becoming more skilled. The extra effort is in fact an investment in human capital and is efficient. With reference to the discussion in the introduction about bonuses and incentives schemes, note that a bonus B is entirely captured in the optimal outcome (4.4) through the "production function" $c(e) = w + B(e)$, where w is the fixed wage.

Let us turn now to the decentralised solution. It is trivial that at $t = 1$, the choice of effort will be at the socially optimal level (e_1^O, C_1^O, θ), since effort cannot get you a better job: there is no future period. For the same reason, there is no human capital effect and effort is at the level of the static labour supply model. In period 0 however, the individually rational choice of effort will incorporate the effect of current effort on the promotion. The decentralised outcome (e_0^*, C_0^*, θ) satisfies:

$$\frac{dU_0}{de_0} + \rho\frac{\partial U_1}{\partial \theta_1}\frac{d\theta_1}{de_0} + \rho\frac{\partial U_1}{\partial F_1}\frac{dF_1}{de_0} = 1 \tag{4.5}$$

Current effort still has an effect on the future skills (i.e. the human capital effect $\frac{d\theta}{de_0}$), but in addition, current effort determines which job the worker will be allocated to in the next period: $\frac{\partial F_1}{\partial e_0}$. The more effort you provide today, the higher the rank of your job tomorrow. More elaborately, the first order condition can equivalently be written as

$$\frac{dU_0}{de_0} + \rho\frac{\partial U_1}{\partial C_1}\left(\frac{\partial c}{\partial \theta_1} + \frac{\partial c}{\partial F}\frac{\partial F_1}{\partial \theta_1}\right)\frac{d\theta_1}{de_0} = 1 \tag{4.5'}$$

This highlights the fact that future consumption is affected by the effect of effort through both the increase in ability and the effect on the rank of the job performed.

The first-order condition (4.5) is a necessary condition for the existence of Equilibrium.

Proposition 9 *At $t = 0$, a unique Equilibrium (e_0^*, C_0^*, θ) exists and it is entirely defined by the first order condition (4.5), provided the following conditions are satisfied:*

1. *Single Crossing: $\frac{\partial^2 g}{\partial \theta_0 \partial e_0} > 0$;*
2. *F is continuous and has no mass points.*

Proof. See Appendix. ■

The intuition behind the Single Crossing condition is identical to the signalling and principal agent literature and is derived from the second-order condition. First, suppose that the condition holds with strict equality. It follows that the marginal effect of current effort on next period's type is the same irrespective of the current type. Hence, no one can obtain any future benefit from putting in extra effort now and the equilibrium level of effort would be equal to the level chosen by the social planner. If the condition holds with a negative sign, the provision of effort would imply that the equilibrium is not a maximum.

The second condition has no equal in the signalling literature. In the proof, it is shown technically where the condition comes from: now the effort decision in equilibrium is determined by the effect of all workers' effort on their rank, with the result that the properties of the cumulative density function will determine this rank effect. For intuitive understanding, consider the following thought experiment. Workers are distributed according to a bimodal discrete cumulative density function. Given the Single Crossing condition, in equilibrium investment by the lower types is at the socially optimal level, since they cannot leapfrog. Hence, the best response for the higher types, given optimal investment of the lower types, is to invest optimally as well. However, given optimal investment by the higher types, the best response for the lower types is to expend more effort and leapfrog to the highest rank. As a result, equilibrium does not exist in pure strategies. This is not the case with continuous distributions because, however close together, there always exists a type between the higher and the lower type which induces the higher type to invest more effort. Note though that with a discrete distribution function, equilibrium does exist in mixed strategies. The non-existence of pure strategies is equivalent to some results in auction theory or to the equilibrium strategy in a Bertrand duopoly with different marginal costs.

Given existence and uniqueness, equilibrium can be characterised and the efficiency can be compared to the benchmark. Mere observation of equations (4.4) and (4.5) reveals that the decentralised equilibrium does not achieve first best. The term $\frac{\partial U_1}{\partial F_1}\frac{dF_1}{de_0}$ will from now on be defined as the Rank-Order Inefficiency (ROI) and Proposition 10 establishes the nature of the inefficiency

Proposition 10 *The equilibrium supply of effort in the decentralised system is super-optimal:* $e_0^O(\theta) < e_0^*(\theta)$, $\forall\theta$.

Proof. The Promotion Inefficiency term is strictly positive, since $U_F > 0$, $F_e = F_g g_e > 0$ and $g_\theta > 0$. With U_C strictly concave, it follows that $e_0^O < e_0^*$. ■

The intuition is straightforward. Because of the effect of future reallocation, workers do not only take into account the marginal benefit of current consumption, but also the marginal benefit of the future position. The marginal benefit of a higher type is higher than the marginal cost of extra effort, from the single crossing condition. This is inefficient because the extra effort does not translate into the extra income that follows from the higher rank. The reason is that each individual expends too much effort. The Single Crossing condition makes sure that it is less costly for the higher types to achieve a higher rank than for the lower types. As a result, the higher types manage to avoid being leapfrogged by the lower types. The externality from the behaviour of all other agents here is clearly not an informational externality (there is full information) but it is due to the marginal effect of being allocated to a better job. Note that in this Equilibrium, output in the economy is higher than in the absence of the Rank-Order effect. Workers provide more effort and thus generate more consumption. However, the inefficiency is due to the fact that this extra consumption goes at the cost of too much effort. In this framework, GDP per capita clearly is not a good measure for utility.

Two Corollaries follow immediately from Proposition 10.

Corollary 5 *Both the Rank-Order Inefficiency and equilibrium effort are increasing in type.*

Proof. Immediate from Proposition 10 and the Single Crossing condition. ■

Even if the more productive types do not get a higher wage from being more productive, i.e. $c_\theta = 0$, they get a higher wage from performing the higher ranked job. In addition, they also marginally have to work more in order to maintain their rank. If the marginal benefit of improving the type were constant (i.e. $g_{\theta e} = 0$), there would be no inefficiency because rank would have no incentive effect.

Corollary 6 *In equilibrium, the rank is maintained.*

Proof. Immediate from Proposition 10 and the Single Crossing condition. ■

In equilibrium, there is no leapfrogging. Everyone manages to just avoid it because the higher types are marginally more effective in improving their type in the next period.

Let us now turn to the empirical verification of these first results. To match a theoretical model to empirical evidence is always a risky undertaking, if only because of the simplified nature of theoretical models and because of the multitude of different (often opposing) forces operating in the reality. The purpose of the exercise here is to show that the model does not contradict certain stylised facts. In this Section, some within labour market observations are discussed. In the next Section, the model will be extended to the analysis of turnover in the labour market to highlight some between labour market stylised facts.

1. Basic supply factors alone cannot explain differences in hours worked

The standard labour supply model predicts that, as a result of a change in earnings, there is a trade off between the income and the substitution effect. A change in the wage rate due to taxation or technology has an ambiguous effect on the hours worked. Though there are opposing claims about the sign of the aggregate effect, many authors find it to be non-significant. In an Econometrica article, Mroz [52] reviews this literature and disputes the methodology of a number of papers that find a significant effect. He concludes: "The principal finding of this analysis is that economic factors such as wage rates, taxes, and non-labour incomes have a small impact on the supply behaviour...", Mroz [52, p.795]. Not only does this piece of evidence justify the use of a quasi-linear utility function, more importantly, it sets the agenda for this paper. Given there are differences in the supply behaviour of workers, be it within or between labour markets, there must be an explanation beyond the impact of basic supply factors.

2. Human capital effects alone cannot explain wage differentials

Using firm level performance data, Medoff and Abraham [43] and [44] distinguish relative wage differentials from relative productivity (i.e. reported performance). Within one job type, experience can explain wage differentials. About 40 percent of earnings differentials associated with experience (both outside and within the company) occur within the job type. However, they do not find support for the thesis that across different job types, experience explains relative wage differentials. They conclude: "Since the fraction of the experience-earnings relationship that occurs within [job types] is substantial, [...] a substantial portion of this relationship cannot be explained by the human capital model of productivity augmenting on-the-job training" (Medoff and Abraham [43, p.735]). Though they do not suggest a theory as to why wages do not only reflect human capital accumulation effects, their findings justify our assumption that the wage as a measure for productivity depends on *both* ability and the job performed.

3. The distribution of wages is skewed to the right

It is widely accepted that the distribution of income, and in this context of labour income, is skewed to the right, the log of income approximately being symmetrically distributed. It can be argued that the underlying factor is the distribution of skills. Though it is no sinecure to measure skills, it is clear that the distribution of say education is by far not as skewed as the distribution of wages. Hence, there must be some complementarity between the skill level and the job performed. This is exactly captured by the positive dependence of labour income on the Rank-Order in the distribution. The distributional aspects of the inefficiency in the model are extensively discussed in the next Section.

4.3 Efficiency and Distribution

In the presence of a Rank-Order tournament, it is almost tautological that distributional aspects have a real effect. Mere observation of the technology which transforms effort into consumption shows how the distribution function enters the objective function. In this Section, the aim is to analyse the effect of increased inequality on the supply-of-effort decision by workers. However, not only the distribution of skills $F(\theta)$ will have an effect but also the rates at which skills are remunerated (i.e. the wage distribution) will determine effort. In our context, the wage distribution can be disentangled into the skill premium and what will be referred to as the job premium. Like the literature on wage inequality (e.g. Murphy and Welch [53]), the skill premium is equal to the marginal increase in wages from having higher ability. In our model, this is equal to c_θ. Because productivity is not only a function of ability, but also of the rank of the job performed, the wage distribution in addition is a function of the marginal increase in the rank of the job. Technically, the marginal return to jobs (i.e. the job premium) is given by c_F. Both the skill and the job premium are derived from the effect of respectively θ and $F(\theta)$ on $C = c(e, \theta, F(\theta))$. To illustrate these three distinct distributional effects, consider the first-order condition (4.5') can be written as

$$\frac{dU_0}{de_0} + \rho \frac{\partial U_1}{\partial C_1} \left(c_\theta + c_F f(\theta) \right) \frac{d\theta_1}{de_0} = 1 \tag{4.6}$$

The term in brackets distinguishes the three channels through which equilibrium effort is affected: $f(\theta)$, the skill distribution; c_F, the job premium; and c_θ, the skill premium. The skill premium determines the human capital effect whereas the job premium and the skill distribution jointly constitute the Rank-Order Inefficiency. In what follows, we will derive the impact of inequality of each of these effects both on effort and efficiency. Without jumping to conclusions, it is clear that the skill premium effect will be efficient, since it affects the efficient human capital accumulation decision. The job premium and the skill distribution effects will be inefficient (from Proposition 10), since they relate to the Rank-Order Inefficiency. However, from the same Proposition, we know that effort and inefficiency are correlated.

The **Skill Distribution Effect.** Since θ is taken to be an indicator of the level of productive ability, $F(\theta)$ can be referred to as the skill distribution. In order to unequivocally be able to assert whether one distribution is more or less equal than the other, the distribution functions that will be compared have to satisfy second order stochastic dominance (i.e. a mean preserving spread) with respect to each other. Consider two skill distributions $F(\cdot)$ and $H(\cdot)$ (with their respective density functions $f(\cdot)$ and $h(\cdot)$) and, without loss of generality, take $H(\cdot)$ to be second order stochastically dominant, i.e. $H(\cdot)$ is more unequal than $F(\cdot)$. For reasons of mathematical tractability, the mean preserving spread is modelled as a Sandmo transformation[7]. This simply implies that, while maintaining the same mean, both distributions have identical cumulative densities if the support is multiplied by a scale factor $\lambda > 1$. If μ is the common mean, then the variable γ can be defined such that $\gamma = \lambda(\theta - \mu) + \mu$ and $F(\theta) = H(\gamma(\theta))$, $\forall\theta$. It follows that $h(\cdot) = \lambda^{-1} f(\cdot)$. The Skill Distribution Effect produces the following result

Proposition 11 *(Skill Distribution Effect) Both effort and the Rank-Order Inefficiency decrease as inequality in the skill distribution increases.*

Proof. Consider the model from Section 4.1, fixing $\Delta = 1$ and with the skill distribution $H(\theta)$. Given the mean preserving spread on H from F, it follows

$$\frac{\partial U_1(H)}{\partial H}\frac{dH_1}{de_0} = \frac{\partial U_1}{\partial H_1}h_1(\theta)\frac{d\theta_1}{de_0} = \lambda^{-1}\frac{\partial U_1(F)}{\partial F}\frac{dF_1}{de_0} \tag{4.7}$$

The Rank-Order Inefficiency in equation (3) is equal to

$$\rho\frac{\partial U_1(H)}{\partial H_1}\frac{dH_1}{de_0} = \lambda^{-1}\rho\frac{\partial U_1(F)}{\partial F}\frac{dF_1}{de_0} < \rho\frac{dU_1(F)}{de_0} \tag{4.8}$$

and with $\lambda > 1$, it is decreasing in the degree of inequality in the skill distribution. ■

[7]It is our belief that the results will not be reversed using the variance definition or the Rothschild-Stiglitz [63] definition for increasing risk as a measure of inequality. Mathematical manipulation however is not trivial, especially with the second measure.

The intuition behind this result is the following. In a compact distribution, doubling your type tomorrow through the effort in the current period, cœteris paribus increases your rank considerably. In a spread out distribution, doubling your type changes your rank only marginally. In the limit case of a distribution with $h(\cdot) \to 0$, doubling your type does not change your rank at all. The return to extra effort from promotion (i.e. from increasing the rank) is higher the more equal the distribution. As a result, for the marginal benefit of effort to equal the marginal cost, it is rational to expend more effort. Basically, a more equal distribution implies tougher competition for the same jobs. Workers will expend more effort to obtain those jobs.

From an equity point of view, this is a rather negative result. Efficiency can be increased by increasing the spread of the distribution of say education. However, promotion or rank has also a distributional effect through the job premium.

The **Job Premium Effect.** In analogy with the parameter λ above, a parameter $\beta > 1$ can be introduced to represent an increase in the job premium βc_F. As β increases, the effect on consumption of working in a higher ranked job increases. Hence, the wage distribution (i.e. the distribution of C) becomes more unequal. The following proposition establishes the result related to the job premium effect.

Proposition 12 *(Job Premium Effect) Both effort and the Rank-Order Inefficiency increase as the job premium increases.*

Proof. The Rank-Order Inefficiency is equal to

$$\frac{\partial U_1}{\partial F}\frac{dF_1}{de_0} = \frac{\partial U_1}{\partial c}c_F\frac{dF_1}{de_0} \tag{4.9}$$

$\beta > 1$ implies

$$\rho\frac{\partial U_1}{\partial C}\frac{\partial c}{\partial F}\frac{dF_1}{de_0} = \rho\frac{\partial U_1}{\partial c}\beta c_F\frac{dF_1}{de_0} < \rho\frac{\partial U_1}{\partial c}c_F\frac{dF_1}{de_0} \tag{4.10}$$

so that the Rank-Order Inefficiency increases as the job premium rises. ■

Again, the intuition is straightforward. The higher the marginal benefit of being in the higher job (i.e. the higher β), the higher the incentive to over-invest in effort in the previous period. Clearly, if the marginal-return-to-jobs is zero, i.e. the job premium $c_F = 0$, the inefficiency due to wage inequality is zero because jobs are homogeneous. Of course, in that case the Rank-Order effect is completely switched off. From an efficiency viewpoint, there is clearly a welfare gain from reducing the marginal (i.e. with respect to jobs) pay-distribution.

The question remains which of the two opposing inequality effects dominates. From a theoretical point of view, the question is answered easily: $\beta\lambda^{-1} \lesseqgtr 1$ implies that the increase in inequality (jointly in the job premium and in the distribution of ability) decreases (increases) effort and thus efficiency since the job premium effect dominates (is dominated by) the skill distribution effect.

The Skill **Premium Effect**. Finally, there is also the (efficient) effect from inequality on the equilibrium effort: the effect of the skill premium on the accumulation of human and thus on effort. Introducing a parameter $\delta > 1$ allows us to study the effect of an increase in the skill premium c_θ on the provision of effort: δc_θ. Simple observation of the first-order condition (4.10) and Lemma 8 reveals that δc_θ increases the equilibrium supply of effort as $\delta > 1$.

Consider now both the human capital effect and the promotion inefficiency to work simultaneously. An increase in inequality of all sources, i.e. λ, β and $\delta > 1$, now includes the ambiguous effect of the Rank-Order Inefficiency on effort and the positive effect on effort from human capital investment. The net effect on effort remains ambiguous. Though empirically there is no work that distinguishes between the different effects, the literature does suggest an answer for the net effect.

4. The number of hours worked and aggregate inequality are correlated

A considerable problem in the empirical wage inequality literature is the notion of skill as a supply factor. Even if years of education are taken as a proxy, one could argue that there are considerable differences between the type of education within a country and between countries. The aggregate effect of wages however is more tractable. Without controlling for skill, the distribution of hourly wages captures all three effects on the labour-supply decision simultaneously. In a cross-country comparison, Bell and Freeman [9] show a significant link between the number of hours worked and (hourly) earnings inequality, indicating that the job and skill premium effects dominate the skill distribution effect.

One final remark on the inequality of the distribution of income. As the inefficiency increases, the spread in the distribution of yearly income (after effort) unambiguously increases. The benefit of the higher GDP is not only at the cost of higher disutility, it is also at the cost of equity.

4.4 Comparing Turnover Regimes

In this Section, it is analysed in greater detail how the excess supply of effort is sensitive to different institutional configurations of the labour market. The inefficiency is closely related to the potential upward mobility. The degree of mobility obviously depends on the speed at which workers can move up in the hierarchy. Therefore, labour markets are compared where these opportunities arrive at a different rate (i.e. markets are characterised by their turnover rate). The Rank-Order Inefficiency is then analysed in function of this rate of turnover. As the inefficiency is explained in function of the incentives of future earnings, different turnover rates will yield different incentive schemes. The results are then taken to explain two more stylised facts which emphasise how the model can account for distinct behaviour between labour markets.

The model is generalised to the case with infinitely-lived agents and an infinitely repeated effort-consumption decision. In addition, and in order to capture the impact of turnover on the equilibrium level of overinvestment, the discrete time model will be scaled by a factor Δ, the duration of one period. It will allow us to capture turnover in the job market. If workers provide a constant level of effort for their whole lives, their net present value of utility will be constant irrespective of Δ. As the duration of a period decreases, the turnover increases. The rate of turnover is then defined as Δ^{-1}. The interpretation is that employers allocate workers to a certain rank job based on the currently observed type, which in this linear scale factor is a function of per unit time effort. Turnover as such has no real technological effect, but it affects the incentives for future job opportunities. One final remark, in this Section the human capital effect will be made redundant, i.e. $c_\theta = 0$. Controlling for the promotion incentive effect, human capital accumulation can reasonably be assumed independent of mobility. Without loss of generality, it is assumed there is no human capital at all. The stage game becomes:

$$\begin{aligned} & \max_{e_t} \rho^{\Delta t} \Delta [U(C_t) - e_t] \\ \text{s.t.} \quad & C_t = c(e_t, F_t(\theta)) \\ & \theta_t = g(e_{t-1}, \theta_{t-1}) \\ & F_0(\theta) \end{aligned} \tag{4.11}$$

As in Section three, the social planner's problem can be reduced to finding the allocation which maximises the sum of utilities. Moreover, since $c_\theta = 0$, the solution (e_t^O, C_t^O, θ) is identical in every period and satisfies

$$\frac{dU_t}{de_t} = 1, \forall t \tag{4.12}$$

Note that the first-order condition in the presence of this linear scale factor is independent of Δ.

The decentralised Subgame Perfect Equilibrium solution now not only takes into account next period, but the discounted sum of the effect in each future period. The reason is that current effort has an effect on next period's type, and under the Single Crossing condition, next period's type has a positive effect on the type two periods ahead: $g_{\theta e} > 0$ and $g_\theta > 0$. Hence, current effort has a positive effect on the rank and the consumption two periods from now. This is true for all future periods, appropriately discounted. In the decentralised outcome the first-order condition is

$$\frac{dU_t}{de_t} + \sum_{s=1}^{\infty} \rho^{\Delta s} \frac{dU_{t+s}}{de_t} = 1, \forall t \tag{4.13}$$

The results from the former Section generalise to this model. The Rank-Order Inefficiency is still positive $\forall s > 1$, since $g_e > 0$. As a result, also in the generalised case, effort is super-optimal. Note that the extra effort is not "technologically" related to the rate of turnover but to the change in incentives. The model is specified such that, given identical provision of effort under different promotion regimes, the net present value of utility is constant. However, if incentives are affected by the Δ, effort provision differs depending on the promotion regime because incentives are affected. The way in which effort is affected is the purpose of Proposition 13.

Proposition 13 *For discount rates $0 < \rho < 1$, the (inefficient) supply of* *effort is monotonically increasing in the rate of turnover.*

Proof. From applying the implicit function theorem to the first-order condition (4.13), it follows that

$$\frac{\partial e_t}{\partial \Delta} = -\frac{\ln \rho \sum_{s=1}^{\infty} \rho^{\Delta s} s \frac{\partial U(C_{t+s})}{\partial e_t}}{U_{ee} + \sum_{s=1}^{\infty} \rho^{\Delta s} \frac{\partial U(C_{t+s})}{\partial g} g_{ee}} < 0 \tag{4.14}$$

since U and g are concave in e and with $0 < \rho < 1$. As a result e_t is decreasing in Δ, implying it is increasing in Δ^{-1}, the rate of turnover. ■

Corollary 7 *(**No Turnover Regime**) As $\Delta \to \infty$, the decentralised solution is equal to first-best.*

*(**Permanent Turnover Regime**) As $\Delta \to 0$, the inefficiency is maximised.*

Proof. Immediate from Proposition 13 and equation (4.13). ■

In order to gain some intuition behind Proposition 13, consider the following thought experiment. Compare two Turnover Regimes: Δ is one year (i.e. high turnover) versus Δ is ten years (low turnover). In the high turnover regime, extra effort is provided only during one year rather than 10 years, so the marginal cost of providing extra effort is scaled by $\Delta = 1$. Certainly, also the benefit from extra effort only lasts for one year rather than 10 years, so the benefit as well is scaled by Δ. The difference between the two regimes lies in the discounting. Providing 10 years of effort now to benefit from it only in the next period of ten years definitely has a lower net present value than in the case of one year. As a result, discounting makes the difference. This is apparent from equation (4.13), where Δ enters the first-order condition only in the discount factor $\rho^{\Delta s}$. With no discounting, the intertemporal price of utility is 1 and agents value current and future utility equally. In that case, there is no incentive effect due to the Turnover Regime. This is confirmed by equation (4.14): ln 1 = 0, implying that $\frac{\partial e_t}{\partial \Delta} = 0$. In case agents are myopic (i.e. $\rho = 0$), the Rank-Order Inefficiency (equation (4.13)) is equal to zero since it is multiplied by zero. Basically, agents are not forward looking and "better-job" incentives do not affect current effort.

It follows from Propositions 13 and Corollary 7 that as turnover increases, life-time earnings increase since more effort generates more consumption. Again, this extra consumption goes at the cost of too much effort.

In what follows, two stylised facts are discussed which indicate some differences between different labour markets.

5. American labour markets exhibit a higher turnover rate than German markets

Topel and Ward [72] for the US and Acemoglu and Pischke [1] for both the US and Germany use data on job mobility to calculate rates of turnover. They find rates that are twice as high in the US compared to Germany. Considering turnover as an institutional feature, this could be attributed to the degree of deregulation. Given that American and German labour markets exhibit a significantly different Promotion Regime, this stylised fact can be used to verify the hypothesis in Proposition 13. The next stylised fact provides the necessary information on effort provision in both countries for this verification.

6. Americans work longer hours than Germans

In a cross-country comparison, in particular between the US and Germany, Bell and Freeman [9] confirm Fact 1 above, i.e. basic supply factors cannot explain differences in labour market behaviour. In addition they claim that after controlling for demographics, preferences and average wage and income, there is still a difference of 15% in hours worked to be explained. On average, out of six days, Americans work one day extra compared to Germans. Together with Fact 5, this confirms our hypothesis that effort and turnover are positively correlated.

4.5 Extension: Life Cycle

The analysis of "better-job" opportunities so far may seem slightly odd. In equilibrium, the rank of the distribution is maintained (Corollary 6), so that effectively, there is no better jobs are ever obtained at all. At the same time however, the individual's temptation to outwit the higher types in her neighbourhood does cause the inefficiency. Because all individuals try to outwit each other and given the Single Crossing condition, there is no effective change in the rank. Trivially, imperfect observability would induce changes in the rank but still, ex-ante the distribution would maintain rank. The mere introduction of a life-cycle story with overlapping generations yields the observed effective movement between jobs.

Consider the two-period model in Section Three, but now with overlapping generations. Young individuals are born at $t = 0$, and work in this period. At $t = 1$, old individuals work and die. The difference is that at the moment a young individual enters the economy, half the population is old. The inflow distribution of young types, $F^Y(\theta)$ is exogenous. and its population has mass one half. The distribution of old individuals $F^O(\theta)$ is endogenous and is a function of the effort expended by last period's young workers, the currently old ones. As a result, the effective type distribution observed at any moment in time is $F(\theta) = \frac{1}{2}F^Y(\theta) + \frac{1}{2}F^O(\theta)$. If $g(e, \theta) > \theta$, then an old worker born of type θ is more productive than a young person born of the same type θ. It is easy to verify that effectively all workers will get a better job. Older individuals will move up the ladder. The young people that enter with the same type as the currently old ones did when they were young are now lower down the rank. However, from the results in Section Three, it follows not only that there will be over-investment in effort, but also that between people of the same age, the rank is maintained. Distributions $F_t^Y(\theta)$ and $F_{t+1}^O(\theta)$ maintain the rank. The promotion behaviour that is observed is between cohorts, but the within cohort ranking contributes to the inefficiency. This is supported by research on the explanation of inequality in the UK. Gosling, Machin and Mehir [29] find that inequality can nearly entirely be explained within rather than between cohorts.

Consider further another life-cycle consideration. The extended model in Section Four gives some additional insights into the effort decision of finitely lived agents. Recall from Section Two that in the two-period model, there is no investment in the last period. In general, for a fixed period model (T periods), effort is decreasing as time increases. In the last period, effort is at the optimal level.

Proposition 14 *For a finitely lived agent, the excess supply of effort decreases over time*

Proof. The first-order condition for the T period model becomes

$$\frac{dU_t}{de_t} + \sum_{s=1}^{T-t} \rho^{\Delta s} \frac{dU_{t+s}}{de_t} = 1, \forall t \tag{4.15}$$

It follows directly from this equation that as t increases, the number of positive terms in the Rank-Order Inefficiency decreases. From Proposition 10, effort is positively related to the Rank-Order Inefficiency. As a result, effort will decrease over time. In the last period, the Rank-Order Inefficiency term is equal to zero, thus yielding the first-order condition for the social optimum. ■

This result can explain the following stylised fact within labour markets.

7. Hours decrease with age

Controlling now for differences between professions (e.g. self employed versus public sector) and skill, workers at a later stage of their career work less hours. In measuring productivity of different age groups, Medoff and Abraham [43] find that within job types, young workers produce more output than older workers of similar ability. This indicates that, given ability is related within the job type, younger workers supply more labour than older workers.

4.6 Concluding Remarks

In an attempt to explain the effort supply decision of workers, it has been shown that the future career opportunities can have considerable effects on the hours of work. In part, this is efficient in a human capital context: the extra effort is an investment in the ability of the worker. However, if the premise is accepted that there exists a distribution of jobs with different productivity, the productivity of a worker is not merely depending on ability, but also on allocation of the worker to the job. Since in equilbrium the allocation depends on the rank of the worker, effort becomes a tool to achieve higher ranked jobs. It is shown that the extra supply of effort is inefficient since rank is a zero sum game. Moreover, the inefficiency increases as turnover increases. On the other hand, the theoretical effect of inequality on effort is ambiguous. A more unequal distribution of skills decreases both effort and inefficiency, whereas a rise in the job premium increases them. A rise in the skill premium increases effort but is efficient. The theoretical results are supported by seven stylised facts from the empirical labour-supply literature.

In the presence of a Rank-Order effect, there are some important welfare implications. One implication relates to the importance of GDP as an indicator of welfare. Simple observation of a general equilibrium economy (Edgeworth box) shows that maximising GDP is equivalent to maximising the utilitarian social welfare function if and only if the marginal utilities of all individuals are equal in equilibrium. However, redistribution (whether it is through lump sum taxes or marginal taxes) involves a Pareto deterioration of the richer individuals. In the inefficient outcome of our equilibrium, output is higher than in the social optimum. The inefficiency is due to the over-supply of hours. Hence, the economy maximises GDP but not total utilitarian welfare. Contrary to the Edgeworth economy, in this production economy there exists a welfare improving allocation which is Pareto Superior. There exists a Pigouvian tax rate for all workers, equal to the Rank-Order Inefficiency, that would improve overall welfare and no individual would be worse off. GDP on the contrary would decline. The relevant measure for comparison between countries is then not yearly GDP, but hourly GDP, since that corrects for the differences in effort. Provided of course that the preferences between different countries are identical. Bell and Freeman [9] report that according to such a purely utilitarian criterion, Germans comfortably lead over Americans, even though GDP per capita in the US is higher. Further, in the presence of this Pareto Superior tax schedule, Mirrlees' optimal taxation problem is not as dramatic as it seems. Even if the government cannot monitor the types of the workers, there exists one implementable tax schedule (i.e. the Pigouvian tax schedule) which is first best. Finally, one cannot argue that our analysis does not take into account the effect of more hours on the productivity of the firm, since this is embodied in the production function. Under perfect competition, all the firms' potential benefits accrue to the workers.

To conclude, one remark on our approach of modelling the labour market in a dynamic framework and in the presence of a matching game. Combined with vertical heterogeneity, it yields a result that is analogous to the results in the signalling literature. By the same token, this approach can be extended to the whole static imperfect information literature. Efficiency wage models for example, are often used to explain unemployment. However, rather than the fear of being caught shirking, this fear may be interpreted as the fear of losing the job. Even with a wage rate that is flat within one period, workers put in effort, because they know that next period's wage (unemployment versus the same wage) depends on the current period's effort. In this dynamic context there is no asymmetry in information between worker and firm. The comforting result of the analogy is that, as shown in this paper, the results of several imperfect information models can be extended to full information models with Repeated Tournament features. Those seemingly equivalent results however are derived from allocation externalities, rather than information externalities. As a result, the understanding of the mechanisms generating those outcomes differs dramatically.

4.7 Appendix

Proposition 9

Proof. The first-order condition (4.5) is a necessary condition for the equilibrium to exist. In addition, the second-order condition needs to be verified. The proof of the first condition of our Proposition (i.e. the sufficiency of Single Crossing) is based upon Spence [70, Proposition 1]. The second derivative has to satisfy,

$$\frac{d^2U_0}{de_0^2} + \rho\frac{d^2U_1}{de_0^2} < 0 \tag{A4.1}$$

In a Subgame Perfect equilibrium (i.e. taking the optimal strategy of all other players as given), there is the indirect effect of e_0 on the rank $F(\theta_1)$ of current period's type θ_0. Differentiating (4.5) then yields:

$$\frac{d^2U_0}{de_0^2} + \rho\frac{d^2U_1}{de_0^2} + \rho\frac{\partial C_1}{\partial F_1}f_1\frac{\partial^2 g}{\partial\theta_0\partial e_0}\frac{g_{e_0}}{g_{\theta_0}} = 0 \tag{A4.2}$$

For the second-order condition (A1) to be satisfied, it is sufficient that

$$\frac{\partial C_1}{\partial F_1}f_1\frac{\partial^2 g}{\partial\theta_0\partial e_0}\frac{g_{e_0}}{g_{\theta_0}} > 0 \tag{A4.3}$$

Given the restrictions on the functional forms above, this implies that the sufficient condition requires Single Crossing to be satisfied: $\frac{\partial^2 g}{\partial\theta_0\partial e_0} > 0$. Observation of (A4.3) in addition shows that continuity of F is sufficient (condition 2 in the Proposition). If this is not satisfied, the density function f may not exist or be zero, thus violating (A4.3).

In general, like the signalling equilibrium with continuous agents (Spence [70]), this equilibrium is not unique: integrating (4.5) gives a solution subject to a constant. However, each of the multiple equilibria exists provided that *for all* agents there exists a constant such that the solution to (4.5) is feasible. That is, as long as the marginal benefit from promotion exceeds the marginal cost. Because it is assumed that for the lowest types the marginal benefit is zero, i.e. $c_F(\theta = 0) = 0$, at least some types have only one solution that is feasible (i.e. where the constant is zero). Since it is no longer the case that *all agents* have a feasible pay-off with a non-zero constant, all other equilibria do not exist. ∎

Chapter 5

Educational Mobility: The Effect on Efficiency and Distribution

The whole literature following the emergence of endogenous growth theory has argued that externalities in the accumulation process are of considerable importance to explain different growth experiences. Romer [59] discusses the effect on the accumulation of physical capital and Lucas [38] considers externalities in the acquisition of human capital. The idea in all of this work is that the individual returns to investment are not necessarily equal to the social returns. Aghion and Howitt [2] and Grossman and Helpman [30] have followed this up by looking more in detail at the innovation process of technology. They provide micro foundations for the externality that arises in the upgrading and advance of technology. Parallel to this strand of the literature, work has been done on the externalities that arise in the accumulation process of human capital. Most recently, Bénabou [11] looks at spillovers in the education technology and its effect on the incentives to acquire human capital. With agents characterised by heterogeneous wealth holdings, it is argued that the education technology is not only characterised by the individual level of investment but also by the social level of investment, i.e. the aggregate of all individual levels. With agents heterogeneous in wealth levels, other approaches have been taken to study the impact of accumulation of human (and physical) capital in the presence of capital market imperfections: Banerjee and Newman [6], Galor and Zeira [28] and Piketty [55]. The initial wealth level can impose a constraint on the level of optimal investment in the presence of imperfect information, thus leading to the persistence in inequalities. Though the level of wealth certainly is an important determinant in investment opportunities, human capital accumulation and education is definitely also affected by the productive ability of agents in the labour market.

This paper proposes a line of research that starts from the premise that agents are heterogeneous in their ability or skill level rather than in their wealth holdings. Certainly, over the life cycle, ability and wealth may well be correlated, but the mechanism through which ability affects the accumulation of human capital is of a different nature. This has also been recognised by Arnott and Rowse [4]. Their approach is similar to that of Bénabou [11] in the sense that they study at the external effect of mixing heterogeneous agents in the education technology. However, rather than in wealth, they consider heterogeneity in ability. The second feature in which our approach differs from the literature above is that it is recognised that the external effect does not arise as such in the process of education, but in the production in the presence of heterogeneous agents. In that respect, our approach is in line with Eicher [20] who looks at the interaction between the production technology (with externalities) and the incentives to invest in education.

Human capital is endogenous and the interaction is modelled as affecting technological change. In that respect, there is no ex ante heterogeneity of agents so that it is impossible to study how individual heterogeneous agents behave. This is exactly the objective of providing micro foundations for the rationale of economy wide accumulation of human capital.

In this paper, a labour market with agents of different ability is modelled explicitly. They produce vertically differentiated goods (in quality) in separate production industries of the economy. The value of the output is a function of the ability of the workers. The production externality arises from the fact that these industries are monopolistically competitive. Industries are modelled in the fashion of Markusen [40] and Krugman and Venables [35]. The vertically differentiated goods are imperfect substitutes in consumption. As in the utility of consumers in "The Economics of Superstars" (Rosen [60]), individuals are indifferent between the consumption of different quality goods as long as they are appropriately compensated in quantity. Consumption can then be expressed in terms of efficiency units. Within this framework, the presence of a constant returns to scale education technology allows for mobility of individuals between different skill classes. The private incentives for investment however do not take into account the external industry wide effect. If a group of workers augment their skills, the size distribution of skills changes the initial exogenous distribution. With a monopolistically competitive production technology, the output per head is higher in a larger industry. This entirely captures the positive externality that exists from mobility of low skilled workers into higher ability jobs.

In the next Section, the economy and its population are described in detail and the general equilibrium outcome without mobility is derived. Modelling this micro behaviour in detail involves some inevitable algebra. Section 5.2

discusses the impact of mobility. It is shown that multiple equilibria can exist for a certain range of skill differentials. Apart from a coordination issue, there is also a problem of free riding which can result in Pareto improvements by subsidising education of the lower skilled people. Without subsidy, the cost of investment is higher than the private return, so that no investment occurs. However, the high types benefit from their mobility, so they would be better off if they can induce low types to invest. In Section 5.3, the game is played repeatedly. This allows us to consider the effect of growth. Though this paper does not pretend to provide a theory of growth, with endogenous accumulation of human capital, the growth rate will depend on the distribution of skills and thus on education. The relation between equity and efficiency will be identified. In the tradition of the literature on ergodic distributions in the presence of capital market imperfections, the limiting distribution of both wealth and ability can be related to the growth of the economy. A sufficient condition is derived for which there is polarisation, i.e. ever increasing inequality. Some Concluding Remarks are made in Section 5.4.

5.1 The Basic Model

The economy is populated with heterogeneous individuals characterised by a type q. There are a finite number of n types indexed by q_i, $i = 1...n$. Initially, they are distributed according to the densisty $\phi(q_i)$ and the size of the whole population is normalised to one, $\sum \phi_i = 1$. Individuals are both producers and consumers and the type of an agent is interpreted as her productive ability or level of skill as a worker. The higher q_i, the more productive.

Workers produce in order to derive utility from consumption. To formalise the notion of less than perfect substitutability of labour, it will be assumed that workers of a certain type will only work in the same industry as workers of the same type. An industry can then be characterised by the type of its characteristic worker q_i with production in n parallel industries. In terms of the quantity produced in the different industries, technology is identical for industries of the same size. However, workers with higher skill levels will produce higher quality goods. Though the quantity produced in two industries may be identical, the value will differ depending on the level of skills of its workers.

With technology of the quantity produced identical for all industries, we can specify the technology of a generic industry given its size ϕ. As mentioned in the introduction, a monopolistically competitive technology is assumed which will incorporate spillover effects at the production level. The production process is modelled following Markusen [40], allowing for free entry and a zero profit condition. Each industry is characterised by a sector of diversified input goods and an output sector. A worker of a certain industry can either work in the input or the output sector.

The input sector exhibits increasing returns and the output sector constant returns. Hence, the industry as a whole has an increasing returns technology. In any industry, $L \in [0, \phi]$ of the workers will work in the output sector. The other $\phi - L$ will work in the increasing returns to scale input sector, which consists of a number of equally sized firms, producing some variety r of the input good, in a monopolistically competitive environment. The more varieties, the greater the quantity of the composite input X good which is used in production of the output. This technology is as in Dixit-Stiglitz [19]:

$$X = \left(\sum_r x_r^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{5.1}$$

where $\sigma(> 1)$ is the (constant) elasticity of substitution between the different variety inputs. Every x_r, the quantity of variety r of the input, is produced with increasing returns to scale: $s_r = \alpha + \beta x_r$, where s_r is the amount of labour used in that firm. It now follows that the amount of the composite input produced is convex in the number of workers in the input sector, $\phi - L$.

The output sector exhibits constant returns, the technology of which is: $Y = L^\theta X^{1-\theta}$, where Y is the quantity produced. It follows that the output Y of the industry as a whole is convex in the number of workers ϕ.

We solve the problem for a given industry of size ϕ:

$$\frac{X}{L} = \frac{1-\theta}{\theta} \frac{\omega}{p_X} \tag{5.2}$$

$$L^\theta X^{1-\theta} - \omega L - p_X X = 0 \tag{5.3}$$

$$p_r = \frac{\sigma}{\sigma - 1} \beta \omega \tag{5.4}$$

$$\frac{\alpha(\sigma - 1)}{\beta} = x_r \tag{5.5}$$

Equation (5.2) is the profit maximisation condition in the output sector, with p_X the price of the composite input and ω is the wage (in terms of units of production) in the industry. In the presence of free entry, the zero profit condition is given by (5.3). In the input sector, each of the diversified firms r maximises profits (equation (5.4)) with (5.5) the zero profit condition. This allows us to calculate the number of firms m in the sector, each producing the same amount x_r:

$$m = \frac{\phi - L}{\alpha \sigma} \tag{5.6}$$

Because of the increasing returns in the production of the composite input and typically for this Dixit-Stiglitz type of technology, the price p_r will for

$m > 1$ be higher than the price index faced by the output producers[1]:

$$\begin{aligned} p_X &= m^{\frac{1}{1-\sigma}} p_r \\ &= \left(\frac{\phi - L}{\alpha\sigma}\right)^{\frac{1}{1-\sigma}} \frac{\sigma}{\sigma-1} \beta\omega \end{aligned} \tag{5.7}$$

after substitution for (5.4) and (5.6). Finally, market clearing implies:

$$\omega\phi = L^{\theta} X^{1-\theta} \tag{5.8}$$

Equations (5.2), (5.3), (5.7) and (5.8) are independent and contain four unknowns: ω, p_X, L and X.

The preferences of individuals over the goods produced in each individual industry depend on the quality of the good produced and hence on the level of skills of the workers. The quality of a good will simply be indexed by q. As a result, n types of goods will be produced with quality q_i, with $q_1 < q_2 < ... < q_n$. Agents all have identical preferences independent of their type. Utility is increasing in both quality q and the quantity consumed Y_q. They perfectly substitute quantity for quality. An equalising difference unit can then be defined as $c_q = qY_q$. This specification is as in Rosen [60]. This implies that one unit of the higher quality good is preferred above one unit of the lower quality good. It also follows that consumers are indifferent between one unit of a quality q good and two units of a quality $q/2$ good. As a result, even though different quality goods are not perfect substitutes, equalising difference units are. The total consumption C can thus be written as the sum of all c_q. The utility derived from consumption of C is $U = C$. Individuals are risk neutral.

[1]This transition mechanism through the price is the trick of the Dixit-Stiglitz [19] model. Since there is a 'taste' for variety in inputs, the nominal value p_r of an input is higher than the real value p_X. The more variety (i.e. the higher m), the lower the real value of the input.

From the substitutability of the equalising difference units, it follows that individuals will value one unit of a good q twice as high as one unit of a good $q/2$. In general, consumers will equate the marginal rate of substitution, q_i/q_j to the price ratio of different quality goods. If good $q_i = 1$ is taken as the numeraire good, prices satisfy

$$p_q = q \tag{5.9}$$

From equation (5.9), we know that the value of the output for industries with the same number of workers will be in exact proportion to the relative quality levels. As a result, the value of the output of an industry q can be written as qY_q. In addition, the wages between different industries can be compared. Each worker in an industry receives a wage equal to the value of the quantity $\omega(\phi)$ of that industry's goods. Hence, with different values of goods (equation (5.9)), the monetary wages will be given by

$$w_q(\phi) = q\omega(\phi) \tag{5.10}$$

The advantage of the specified preferences and the production technology is that, though wages depend on the size of the industry, the quality or skill impact is separable.

Within this framework, the actions of an individual are the choice of the level of education. Costly education will enable her to increase her level of human capital, given an initial endowment of skills. This is beneficial since that allows her to produce a higher quality good and hence she will receive a higher price for her labour. The cost of education is increasing in the level at which skills are augmented and it depends on the initial skill level q. In general, the cost of education is given by the separable cost function $F(\Delta, q) = f(\Delta)q$, with

$$\Delta = \frac{\Delta q_i}{q_i} \tag{5.11}$$

To ensure existence, the following restrictions are imposed: $f(\Delta)$ is strictly convex ($f_{\Delta\Delta} > 0$), $f_{\Delta} \geq 0$, $f_{\Delta}(0) = 0$, $f_{\Delta}(\infty) = \infty$ and $f(0) = 0$. Investment in general depends on the initial level of skills, but the returns to skill are assumed constant. Individuals will choose the level of education Δ in order to maximise utility, taking the strategy of all other players as given. An equilibrium will then be a rule such that each individual chooses an optimal strategy taking into account the optimal strategy of all other players.

5.2 The Results

With the production sector specified above, it is crucial how the wage is affected by the size of the workforce in the industry. With a monopolistically competitive production industry, there will certainly be a positive effect of ϕ on the quantity of output produced. Simultaneously, from the fact that there is free entry, profits are driven to zero and the entire output accrues to the workers in this general competitive equilibrium context. In addition, not only is the quantity produced increasing in the size of the industry, the value of the wage is increasing. This is shown in the following Lemma.

Lemma 7 *The wage in any industry is increasing in the size of the industry* ϕ

$$\frac{\partial w}{\partial \phi} > 0 \tag{5.12}$$

Proof. See Appendix. ■

This entirely captures the notion of external effects in the production technology. Though the education technology will not inherently exhibit any external effect, the returns to education are affected by the wage. Hence, since the wage depends on the degree of mobility (which is a synonym for education in this context), the decision of an individual to invest will depend on the level of investment of the others.

Consider the simplest possible case of a two period game with two types q_1 (low) and q_2 (high). In the first period of the game, all ϕ_1 (in the two industry case, set $\phi_2 = \phi$, hence $\phi_1 = 1 - \phi$) types q_1 can choose between two strategies: **I.** Invest in education at cost $F(\Delta, q_i)$[2] in order to achieve a level of human capital q_2; In the second period, the q_1 type will be able to work in the q_2 quality industry and receive w_2; **NI.** No Investment and work in the low quality production industry for the second period.

Note that in (5.11), Δ is defined as the gap between the quality (or skill) levels of the two industries. Since the cost of investment function is increasing in its argument Δ, the return on investment must be a function of the gap Δ. The return on investment behaves as in figure 5.1[3].

The return on investment function V_q can be defined as $V_q = q(1 + \Delta)\omega(\cdot) - RF(\Delta, q)$[4]. For the simplest case, where a q_1 type decides not to invest, i.e. play strategy **NI**, the return on investment function is independent of Δ because there is no cost. In the figure, this is represented by the horizontal line **NI**. The wage received is the current wage in industry 1, $w_1 = q_1\omega(1 - \phi)$. If a low type decides to play **I**, the return on investment depends on the behaviour of the other low types. The reason is that, as a result of the increasing returns, the wage is an increasing function of the size of the industry (Lemma 7). This implies a positive effect on the wage in

[2]Capital markets are perfect, so investment can be paid for by borrowed money and will be repaid at rate $R = 1+$ interest rate.

[3]Figure 1 is a mapping of Δ into the return on investment function for given densities $\{\phi_1, \phi_2\}$ (the figure is drawn for $\phi_1 < \phi_2$). As is defined as the relative gap between q_2 and q_1, the function indicates what happens in case of changing inequality.

[4]Note that $q_1(1 + \Delta)\omega(\cdot) = q_2\omega(\cdot)$.

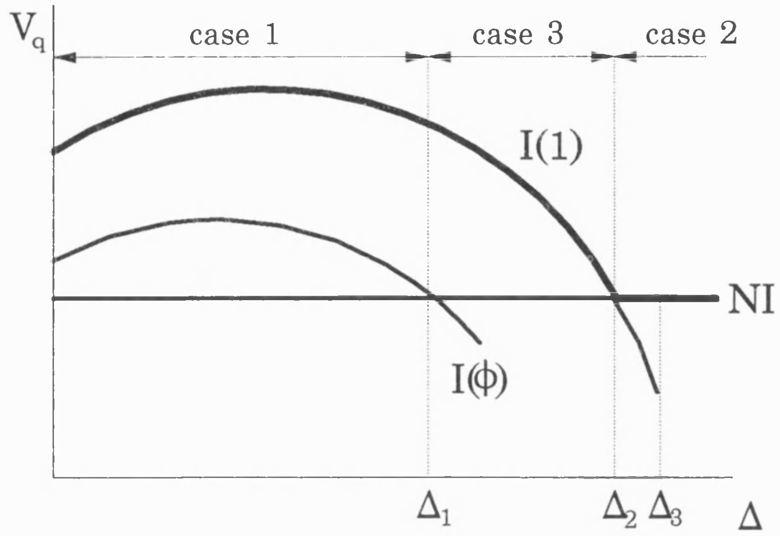

Figure 5.1: Cases with Different Equilibria

industry 2 when q_1 types decide to invest, i.e. the externality from mobility. Now, two extreme situations can be considered. First, all the q_1 types decide to invest, so the size of the new q_2 industry is $\phi_1 + \phi_2 = 1$. The return on investment is $\mathbf{I(1)}$: $q_2\omega(1) - RF(\Delta, q)$. Second, if none of the other low types decides to invest, the size of the q_2 industry does not change and the return on investment for a low type is $\mathbf{I}(\phi)$: $q_2\omega(\phi) - RF(\Delta, q)$, which is always lower than in the first situation.

The equilibrium strategies chosen by the q_1 types can now be analysed given an initial distribution, i.e. given $\{q_1, q_2\}$ and $\{1 - \phi, \phi\}$. Given the returns V_q for the strategies $\mathbf{I}(\phi)$, $\mathbf{I(1)}$ and $\mathbf{NI}$ and depending on the gap between the two types, there are three different cases:

CASE 1. $q_2\omega(\phi) - RF(\Delta, q_1) > q_1\omega(1 - \phi)$: unique Dominant Strategy equilibrium

The equilibrium strategy for all $1-\phi$ types will be **I**. The motives are obvious: the disparity between the human capital levels - and hence the cost of investment F - is so low that they are better off when earning the high wage, irrespective of the strategies of the other q_2 types.

CASE 2. $q_2\omega(1) - RF(\Delta, q_1) < q_1\omega(0)$: unique Dominant Strategy equilibrium

The equilibrium strategy for all q_1 types will be **NI**. They do not invest, because the wage in the q_1 industry is higher than the wage in q_2 net of the cost of investment, even if all other low types would invest. Since the disparity between the different levels of human capital is so high, the cost of investment cannot be compensated by the gain in wage.

CASE 3. $q_2\omega(\phi) - RF(\Delta, q_1) < q_1\omega(1-\phi)$ and $q_1\omega(0) < q_2\omega(1) - RF(\Delta, q_1)$ multiple Nash equilibria

For Δ in this region there are: 1. a pure strategy equilibrium **I** for all types q_1; 2. a pure strategy equilibrium **NI** for all types q_1; and 3. a mixed strategy equilibrium[5] where all types q_1 are indifferent between **I** and **NI**. Each worker plays **I** with probability $\rho = \rho^*$, where

$$\rho^* \in \{\rho \in (0,1) : q_2\omega\left(\phi + \rho(1-\phi)\right) - RF(\Delta, q_1) = q_1\omega\left((1-\rho)(1-\phi)\right) \tag{5.13}$$

The mixed strategy equilibrium is unstable because the slightest deviation leads to either one of the stable pure strategy equilibria (see also figure 5.2). Ex ante, the equilibria are Pareto ranked as follows: 1. all choose **I**; 2. all

[5]This mixed strategy equilibrium can analogously be interpreted as a pure strategy equilbrium where a fraction ρ decide to invest. At that point, a q_1 type is indifferent between I and NI.

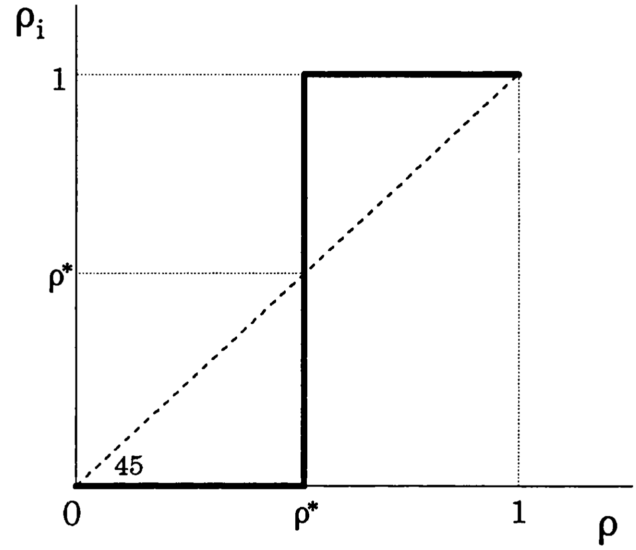

Figure 5.2: Multiple Nash Equilibria

choose NI; 3. mixed strategy equilibrium. The mixed strategy equilibrium is inferior because $w_1 = q_1\omega\left((1-\rho)(1-\phi)\right)$ is increasing in $1-\phi$: as the industry becomes smaller, the wage drops (Lemma 7), i.e. if only a fraction of the workers increase their human capital, there is a negative externality on the workers remaining immobile[6].

Let us now consider some of the welfare implications. Case 3 illustrates that there is a serious problem of coordination failure. All workers choosing I is Pareto optimal, but the emergence of this equilibrium depends on the beliefs of all the other workers. There is a role for the government to improve coordination. Legislation that makes education mandatory up to a certain age can be interpreted as one such an example of coordinating action.

[6]Ex post, I(1) dominates both the other equilibria. The equilibria all NI and the mixed strategy can not be Pareto ranked however, because in the mixed strategy equilbrium, a fraction ρ will be better off than in the all NI while a fraction $1-\rho$ are worse off.

Remark 7 ***Perfect coordination within one industry.*** *For the remainder of the paper, we will concentrate on the more interesting equilibria where no such coordination failure within an industry exists. Workers in different industries have different objectives. However, the interests within industries are identical for all workers. One way to think about this is that all workers are represented by a guild and decision are made collectively. As a result, there is no coordination failure within groups of workers of the same type. In terms of figure 5.1, this implies that only the highest curve,* $q_2\omega(1) - RF(\Delta, q_1)$*, is taken into consideration. For what concerns the equilibrium strategy, the upper envelope of both the latter and the* **NI** *curve will be chosen.*

After abstracting from the problems of coordination failure, one very substantial welfare issue remains. The game is designed such that only the lower human capital types choose a strategy. The higher types remain idle in the first stage. However, because of the externality due to increasing returns, mobility of the low types will have an effect on the wage of the high types and thus on their utility. Closer inspection of the externality shows that in fact, the low, mobile types do not receive the marginal product of their entry into the higher quality industry, but the average product, which is lower than the marginal product. Since the higher types receive the average product as well, they benefit unequivocally from entry by the low types. There will be a case for a Pareto improving subsidy S. Two conditions have to be satisfied: 1. case 2 must apply (i.e. $q_2\omega(1) - RF(\Delta, q_1) < q_1\omega(0)$ so that without subsidy, **NI** is the equilibrium strategy. This condition is fulfilled for $\Delta > \Delta_2$. As a result, there will be no positive externality on the high types; 2. the subsidy must be large enough to induce the low types to make the investment: $q_2\omega(1) - RF(\Delta, q_1) + S(1-\phi)^{-1} = q_1\omega(0)$. This in fact implies that the high types give the minimum subsidy S, necessary to achieve investment. There will obviously be an upper bound to the amount the high types are willing to subsidise. S_{max} is defined as the subsidy which makes the higher types indifferent between subsidising education with the induced mobility (and hence with the resulting higher wage) and not subsidising

$$S_{\max} = \phi \left(q_2 \omega(1) - q_2 \omega(\phi) \right) \tag{5.14}$$

In fact, S_{max} is the total value of the pure economic rent. It follows that only for $S \in [0, S_{max}]$, Pareto improvements are possible. In terms of Δ, we can establish that Pareto improvements are possible for $\Delta \in [\Delta_2, \Delta_3]$, where Δ_3 is defined as

$$\Delta_3 \in \left\{ \Delta \in \Re^+ : q_2\omega(1) - RF(\Delta, q_1) + S_{\max}(1-\phi)^{-1} = q_1\omega(0) \right\} \tag{5.15}$$

Since the external effect has an industry wide impact, it is crucial that the effort of the high types to provide additional incentives to the low types is coordinated. There is indeed a serious free rider problem which is not necessarily ruled out under the assumption of coordination within industry, but which cannot avoid deviation when there are incentives to do so. As a result, there will be a role for the government to impose a tax on the q_2 types and to subsidize education of the q_1 types if Δ is in the relevant interval.

5.3 The repeated game

The point of interest in this section is how the distribution evolves over time and how mobility can have an impact on both the distribution and efficiency. In other words, an underlying theory of growth is needed. However, because this paper does not claim to be able to explain growth, the analysis will be limited to the case of an economy growing at a constant rate. Education will be the sole motor behind growth. It is therefore essential that now all types, also the highest types, have access to this investment in education technology.

Consider the two period game from section 5.1 (for now still with two types: $n = 2$), which is repeated in a successive generations model where every parent gives birth to one child at the end of the second period. Monetary bequests are left out of the analysis, because capital markets are assumed to be perfect and as a result there will be no effect of the distribution of wealth on the investment opportunities. Children inherit the human capital which the parent has accumulated in the second period of her generation. It follows that human capital is inheritable and accumulatable and as a result, the distribution of human capital at the end of one generation is reproduced at the beginning of the next generation.

In this framework with growth, both high types and low types will invest in education, irrespective of the mobility issue. Abstracting for the moment from the possibility of mobility, the problem for a type q will be to choose the amount of investment in education such that: $\Delta^* \in \arg\max\{w_{q,t+1}(\phi(q), \Delta) - RF(\Delta, q)\} = \{q(1+\Delta)\omega(\phi(q)) - RF(\Delta, q)\}$. In general, the optimal amount of investment is given:

$$\Delta^* = F_\Delta^{-1}\left(\frac{1}{R}\omega(\phi)q\right) \tag{5.16}$$

With F separable and with the cost constant with respect to skills ($F(\Delta, q) = f(\Delta)q$), Δ^* will be independent of q

$$\Delta^* = f_\Delta^{-1}\left(\frac{1}{R}\omega(\phi)\right) \tag{5.17}$$

The intuition behind this investment technology is the following. As Δ is a measure for the gap between two levels of human capital - q after investment in education and q at the beginning of period one, i.e. the initial level of human capital - Δ^* is the optimal amount by which to augment the level of skill

q. At the same time, Δ (see equation (5.11)) is a measure for a percentage increase in the initial human capital and hence the outcome after investment depends upon the initial level of human capital q. In other words, though the optimal amount of investment is independent of q (constant returns to skill), the level of human capital after investment is the initial level q augmented with Δ^*. Because of Remark 7, there is only one optimal level of investment Δ^*. If there were no perfect coordination, any level of investment could be supported in equilibrium. If all agents choose a level different from Δ^*, the choice of this level will yield higher utility than choosing Δ^*, since no one else chooses Δ^*. With perfect coordination within the industry, this multiplicity is ruled out.

The optimal investment decision is illustrated in figure 5.3 for different

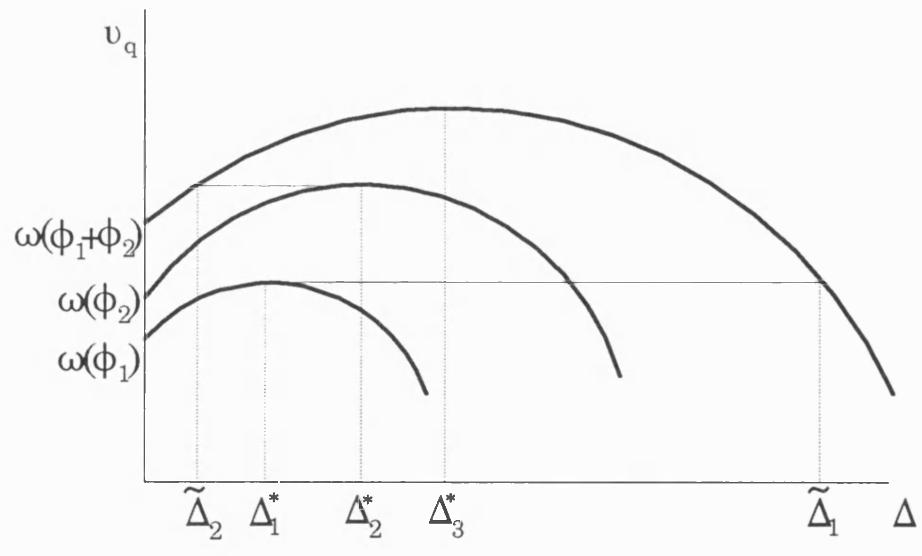

Figure 5.3: Optimal Investment

sized industries. Consider one curve, representing the return on investment of an industry for a given ϕ. Because of constant returns to skill and from equation (5.17), the return on investment, $V_q = q(1+\Delta)\omega(\phi) - Rf(\Delta)q$ can be decomposed into a component ν, independent of the initial human capital q, and a component q: $V_q = \nu(\Delta)q$. Hence, $\nu(\Delta) = q(1+\Delta)\omega(\phi) - Rf(\Delta)$. Since the cost of investment function $f(\cdot)$ is strictly convex in Δ and the gains from investment $(1+\Delta)\omega(\cdot)$ are linear in Δ, the return on investment function $\nu(\Delta)$ is strictly concave, for $\Delta \in \Re^{+}$. As a result, there will be a unique solution for Δ^{*}. The solution always exists because $f_{\Delta}(0) = 0$ and $f(0) = 0$: infinitesimally small amounts of investment have infinitely large returns which makes some investment always attractive. This is the mechanism which results in a strictly positive growth rate of an industry.

Given this technology, it follows that (still abstracting from the possibility of two groups merging) the chosen amount of investment of each individual within an industry will be the optimal amount Δ^{*}. We can now compare the growth rates of different industries, which by definition coincide with the amount of investment Δ. Because of the assumption of constant returns to skill, growth rates will, other things equal, be identical across different quality industries. However, in proposition 15 it is shown that the growth rate depends on the size of the industry.

Proposition 15 *The growth rate of an industry is increasing in the density of the workers with that level of human capital:*

$$\frac{\partial \Delta^{*}}{\partial \phi} > 0$$

Proof. See appendix. ■

The intuition of proposition 15 is shown in figure 5.3. Since $\omega(\cdot)$ is strictly increasing, the return on investment function ν is strictly higher for a higher density ϕ. Figure 5.3 gives v for three different densities, where $\phi_1 < \phi_2 <$

$\phi_1 + \phi_2$. Proposition 15 is also illustrated graphically with $\Delta_1^* < \Delta_2^*$. From proposition 15, it also follows that an industry will have the maximal growth rate when all workers in the economy work in the same industry, i.e. when $\phi = 1$. So far, we have concentrated on the optimal investment when no mobility between industries was possible. However, in line with the results from Section 5.2, the low types may be willing to invest more than Δ^*, if they can join a higher quality industry and thus benefit from the externalities from a larger work force. As we have shown, the high types too benefit from the externality. Translated to the repeated game with investment by both types, this means that the high types may be willing to invest less than optimal. However, the willingness to over/under invest is bounded by the outside option, i.e. the return when industries do not merge. The maximum/minimum individuals are willing to invest has to make them at least as well off as in the case of no industries merging. Hence the following definition which is also illustrated in figure 5.3.

Definition 5

$$\begin{aligned}\tilde{\Delta}_1 = \{\Delta \in \Re^+ : \ & \omega(\phi_1)(1+\Delta_1^*) - Rf(\Delta_1^*) = \\ & \omega(\phi_1+\phi_2)(1+\Delta) - Rf(\Delta),\ \Delta > \Delta_1^*\}\end{aligned}$$

(maximal investment by q_1 over and above Δ_1^* which makes her indifferent between the return in the merged (large) industry and the optimal return in the separate industry)

$$\begin{aligned}\tilde{\Delta}_1 = \max\{0, \Delta \in \Re^+ : \ & \omega(\phi_2)(1+\Delta_2^*) - Rf(\Delta_2^*) = \\ & \omega(\phi_1+\phi_2)(1+\Delta) - Rf(\Delta),\ \Delta < \Delta_2^*\}\end{aligned}$$

(minimal investment by q_2 below Δ_2^* which makes her indifferent between the return in the merged (large) industry and the optimal return in the separate industry)

We can now establish how the distribution will evolve over time and derive a sufficient condition for the limiting distribution.

Proposition 16 *A two industry economy with constant returns to skills will remain polarised into the two industries with a continuously decreasing ratio of human capital* $\frac{q_1}{q_2}$*, tending to zero at infinity, if both the following conditions hold:*

1. $\phi(q_1) < \phi(q_2)$ *(necessary condition);*
2. $q_1(1 + \tilde{\Delta}_1) < q_2(1 + \tilde{\Delta}_2)$ *(sufficient condition).*

Proof. See Appendix. ∎

The intuition behind Proposition 16 is the following. If industry two is larger (condition 1), it will grow faster than industry 1 (from Proposition 15). At the same time, the initial disparity between the two industries is so large that even after the q_1 types have invested maximally and the q_2 types minimally, they are still not near enough to merge. In that case, the decentralised economy will not merge. The next period, the gap between the two industries is even bigger, because the larger industry grows faster because of condition 1. It follows that the gap increases over time.

Corollary 8 *Given the conditions of Proposition 16. If the lower types do not merge with the higher types now, they will never do so.*

This follows from the proof of Proposition 16. It is shown that condition 2 holds even stronger in the next period, hence no mobility will occur. This applies for all consecutive periods, so mobility will never occur. Proposition 16 provides us with a sufficient condition for no mobility in a decentralised system. The result of the Proposition can easily be generalised to an n industry economy, as long as Proposition 16 holds between every of the neighbouring industries. Hence, we can formulate a general definition for a steady state.

Definition 6 *An n industry economy is in a steady state if either:*

1. Proposition 16 applies $n-1$ times, between every industry i and $i+1$, $i = 1...n-1$, $n > 1$; or

2. $n = 1$.

Once the steady state does occur, there will be no changes in the growth rates of the different industries. Since Proposition 16 requires the higher quality industries to be larger, they will grow at a faster rate (from Proposition 15). This gives rise to Proposition 17 about the inequality of the evolving distribution.

Proposition 17 *In the steady state of an n industry economy ($n > 1$), the normalised distribution of skills of the current generation stochastically dominates the distribution of next generation.*

Proof. See Appendix. ■

This means that there is an unambiguous increase in inequality over time, in the sense that the Lorenz curves of any two consecutive generations do not intersect. Note that it cannot be shown that the Generalised Lorenz Curves (Shorrocks [67]) do not intersect, because the GLC takes into account both first and second order stochastic dominance. The distribution with the higher mean (next generation's distribution) can never be dominated by the distribution with the lower mean (current generation), whatever the variance.

Corollary 9 *There is an unambiguous increase in the inequality of income.*

Net income is given by $V_q = q\nu(\Delta^*)$. From condition 1 in proposition 16, ϕ has to be increasing in q in the steady state and from Proposition 15, ν will be increasing in q. As a result, if the distribution of q becomes more unequal, the distribution of $q\nu(\Delta^*)$ will become even more unequal.

5.4 Conclusion

In this Chapter, a general equilibrium framework has been developed which is used to study mobility through education. The aim was to study equilibria in the presence of a monopolistically competitive production technology. Generally, it can be concluded that mobility not only leads to a higher income, but also gets the economy on a higher growth path.

In the static game, coordination problems exist. Legislation on compulsory schooling can be considered as a way of coordinating strategies of agents. Even in the case of perfect coordination however, there is no mobility for a certain range of skill in the decentralised system, even though Pareto improvements can be achieved. This is because the higher types experience a positive externality from mobility. There is a free rider problem which cannot be solved merely by inducing coordination.

When the basic game is repeated and the education technology is embedded in a constant growth economy, it emerges that the larger industries will grow faster. A sufficient condition for polarisation due to lack of mobility is derived. As the growth rate of an industry is increasing in the size of the industry, this steady state - i.e. no mobility - exhibits a low growth path. Moreover, over time, inequality in the economy increases unequivocally. The underlying reason for increasing inequality is the spread in skills. Mobility is too costly and the emergence of a poverty trap is possible. A higher growth path - which increases intertemporal social welfare - can be achieved, but it requires an intertemporal redistribution from the high types in the current generation to all types in the future generations. Since the cost of adjustment is a function of the spread between the skill levels and since inequality increases over time, adjustment becomes more costly the longer it is postponed.

Though it is dangerous to derive policy implications from a rigid model providing a simplified representation of reality, some general guidelines may be useful. In the presence of some externality in production, i.e. when a compact distribution is more efficient, the government has a role to encourage maximal mobility. Of a long term concern is the fact that no action now may cause irretrievable damage later. This is a particularly difficult dilemma because it involves intertemporal redistribution between the generations which is extremely costly for the currently skilled workers and thus an unpopular measure to impose.

5.5 Appendix

Proof of Lemma 7

The system of equations can be simplified and yields an explicit solution for ω as a function of ϕ as follows. Define:

$$A = \left(\frac{\phi - L}{\alpha\sigma}\right)^{\frac{1}{1-\sigma}} \frac{\sigma}{\sigma - 1}\beta \tag{A5.1}$$

Using equation (5.8), we can rewrite (5.2) and (5.3) respectively as:

$$X = \frac{1-\theta}{\theta}\frac{1}{A}L \tag{A5.2}$$

$$L^{\theta}X^{1-\theta} - \omega L - A\omega X = 0 \tag{A5.3}$$

(A5.2) and (A5.3) then yield:

$$\left(\frac{1-\theta}{\theta}\right)^{1-\theta} A^{\theta-1} \tag{A5.4}$$

(5.9) and (A5.2) can be written as:

$$\omega\phi = L\left(\frac{1-\theta}{\theta}\right)^{1-\theta} A^{\theta-1} \tag{A5.5}$$

(A5.4) and (A5.5) give us a very simple expression for L: $L = \phi\theta$ (A5.6). Using this and substituting in (A5.4) - i.e. rewriting the expression for A -, we find the following explicit solution for ω as a function of ϕ

$$\omega = \theta^{\theta}(1-\theta)^{1-\theta}\left(\frac{\phi(1-\theta)}{\alpha\sigma}\right)^{\frac{\theta-1}{1-\sigma}}\left(\frac{\sigma}{\sigma-1}\beta\right)^{\theta-1} \tag{A5.7}$$

Taking the partial derivative yields

$$\frac{\partial\omega}{\partial\phi} = \theta^{\theta}(1-\theta)^{1-\theta}\left(\frac{\sigma}{\sigma-1}\beta\right)^{\theta-1}\frac{\theta-1}{1-\sigma}\left(\frac{\phi(1-\theta)}{\alpha\sigma}\right)^{\frac{\theta-1}{1-\sigma}-1}\frac{(1-\theta)}{\alpha\sigma} \tag{A5.8}$$

Given α, β, $\phi > 0$; $0 < \theta < 1$; $\sigma > 1$; expression (A5.8) is positive. Since $w = \omega q$ and with $\partial q / \partial \phi = 0$, it follows that

$$\frac{\partial w}{\partial \phi} > 0 \tag{A5.9}$$

■

Proof of Proposition 15

Equation (5.19) is derived from:

$$\Delta^* = f_{\Delta}^{-1}\left(\frac{1}{R}\omega(\phi) q h(q)^{-1}\right) \tag{A5.10}$$

Derivation with respect to ϕ gives:

$$\frac{\partial \Delta^*}{\partial \phi} = \frac{1}{f_{\Delta\Delta}} \frac{\partial \omega(\phi)}{\partial \phi} \frac{q}{R} h(q)^{-1} \tag{A5.11}$$

From Lemma 7,

$$\frac{\partial \omega(\phi)}{\partial \phi} > 0 \tag{A5.12}$$

and with $h(q)$ and $f_{\Delta\Delta}$ positive, it follows that

$$\frac{\partial \Delta^*}{\partial \phi} > 0 \tag{A5.13}$$

■

Proof of Proposition 16

It follows from proposition 15 that condition 1 is necessary: the industry with the higher density has the higher growth rate. The condition is necessary because if violated, industry 1 will grow at least as fast and thus there will be no decrease in the proportion of human capital. If the density at q_1 is strictly bigger than at q_2, the proportion q_1/q_2 will increase and eventually, the two industries will merge together.

Condition 1 is not sufficient since there is the possibility that lower type individuals will bridge the gap between their levels of human capital and start producing the q_2 quality good. Condition 2 refers to the case where the gap between the two levels of human capital is so high that no investment will be made to bridge the gap. In terms of the extended model of section two - i.e. the dynamic version with investment by all agents even without social mobility -, case two applies.

Since the cost of investment function is strictly convex in Δ and the gains from investment $\omega(\cdot)(1+\Delta)$ are linear in Δ, the investment function $V_q = \nu q$ is strictly concave for Δ in $\Re^+$. Moreover, $\omega(\cdot)$ is strictly increasing, so that $\tilde{\Delta}_1$ and $\tilde{\Delta}_2$ are uniquely defined. From condition two, it follows that type q_1, when investing the maximal individually rational amount, will never be able to reach a level of human capital equal to that of type q_2, when the latter is investing the minimal individually rational amount.

Combination of the two conditions provides a sufficient condition for q_1/q_2 to be lower at the end of stage two compared to the beginning of stage one. In the next generation, the distribution is exactly reproduced as it was at the end of stage 2, so that condition one remains unchanged and condition two will hold even stronger because: 1. proposition 1 implies that q_2 types will invest more then q_1 types, which will drive down the q_1/q_2 ratio; 2. since $\tilde{\Delta}$ is independent of the level of human capital $\frac{1+\tilde{\Delta}_2}{1+\tilde{\Delta}_1}$, will remain unchanged. This scenario will be repeated, and q_1/q_2 will continue to decrease over time. Over an infinite number of future generations, q_1/q_2 will tend to zero, since

$$\lim_{t\to\infty} \frac{q_1(1+\Delta_1^t)}{q_2(1+\Delta_2^t)} \tag{A5.14}$$

■

Proof of Proposition 17

The underlying social welfare function according to which distributions are ranked is utilitarian. It will be shown that the Lorenz curve of this period's distribution is not lower than next period's for all q. Note further that the usual assumptions about the underlying welfare function and individual utilities apply. In terms of notation: next period's variables will be marked by '.

Starting from the following observations:

1. $\Delta_j^* > \Delta_i^*$, $\forall j > i$: by definition of steady state and from proposition 1;
2. $\phi(q_i) = \phi'(q_i)$: by definition of steady state;
3. $q_i' = q_i(1 + \Delta_i^*)$;

we can show that the share of total income is never larger in the next period:

$$\frac{\sum_1^k q_i\phi(q_i)}{\sum_1^n q_i\phi(q_i)} \geq \frac{\sum_1^k q_i'\phi'(q_i)}{\sum_1^n q_i'\phi'(q_i)}, \forall k = 1...n \tag{A5.15}$$

Inverting (A5.9) and using observations 2 and 3 gives

$$1 + \frac{\sum_{k+1}^n q_i\phi(q_i)}{\sum_1^k q_i\phi(q_i)} \leq 1 + \frac{\sum_{k+1}^n q_i'\phi'(q_i)}{\sum_1^k q_i'\phi'(q_i)}, \forall k = 1...n \tag{A5.16}$$

Dividing the numerator and denominator in the RHS through by $1 + \Delta_{k+1}^*$ gives

$$1 + \frac{\sum_{k+1}^n q_i\phi(q_i)}{\sum_1^k q_i\phi(q_i)} \leq 1 + \frac{\sum_{k+1}^n q_i\left(\frac{1+\Delta_i^*}{1+\Delta_{k+1}^*}\right)\phi'(q_i)}{\sum_1^k q_i\left(\frac{1+\Delta_i^*}{1+\Delta_{k+1}^*}\right)\phi'(q_i)}, \forall k = 1...n \tag{A5.17}$$

From observation 1, it follows that $1 + \Delta_i^* > 1 + \Delta_{k+1}^*$, $\forall i > k + 1$, so that the numerator of the RHS is higher than the one on the LHS. Similarly, $1 + \Delta_i^* < 1 + \Delta_{k+1}^*$, $\forall i < k + 1$, resulting in the denominator on the RHS being smaller than the one on the LHS. Hence, the RHS is bigger than the LHS, which proves (A5.15). ∎

Bibliography

[1] D. ACEMOGLU AND J. PISCHKE, "Why Do Firms Train? Theory and Evidence", *NBER Working Paper* **5605** (June 1996).

[2] P. AGHION AND P. HOWITT, "A Model of Growth through Creative Destruction", *Econometrica* **60** (1992), 323-351.

[3] G. AKERLOF, "On the Economics of Caste, the Rat Race and other Woeful Tales", *Quarterly Journal of Economics* **90(4)** (1976), 559-617.

[4] R. ARNOTT AND J. ROWSE, "Peer Group Effects and Educational Attainment", *Journal of Public Economics* **32** (1987), 287-305.

[5] A. ATKINSON, "On the Measurement of Inequality", *Journal of Economic Theory* **2** (1970), 244-263.

[6] A. BANERJEE AND A. NEWMAN, "Occupational Choice and the Process of Development", *Journal of Political Economy* **101** (1993), 274-298.

[7] G. BECKER, "A Theory of Marriage: Part I", *Journal of Political Economy* **81** (1973), 813-846.

[8] G. BECKER, "A Theory of Marriage: Part II", *Journal of Political Economy* **82(2 pt.2)** (1974), S11-S26.

[9] L. BELL AND R. FREEMAN, "Why do Americans and Germans Work Different Hours?", *NBER Working paper* **4808** (July 1994).

[10] R. BÉNABOU, "Workings of a City: Location, Education, and Production", *Quarterly Journal of Economics* **108** (1993), 619-652.

[11] R. BÉNABOU, "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance", *American Economic Review* **86(3)** (June 1996), 584-609.

[12] F. BLOCH AND H. RYDER, "Two-Sided Search, Marriages and Matchmakers", HEC mimeo, September 1995.

[13] K. BURDETT AND M. COLES, "Marriage and Class", University of Essex mimeo, November 1995.

[14] H. COLE, G. MAILATH AND A. POSTLEWAITE, "Social Norms, Savings Behaviour, and Growth", *Journal of Political Economy* **100** (1992), 1092-1125.

[15] H. COLE, G. MAILATH AND A. POSTLEWAITE, "Bargaining and Investment in a Two-Sided Matching Model", University of Pennsylvania mimeo (1995).

[16] R. COOPER AND A. JOHN, "Coordinating Coordination Failures in Keynesian Models", *Quarterly Journal of Economics* **103** (August 1988), 441-464.

[17] P. DIAMOND, "A Model of Price Adjustment", *Journal of Economic Theory* **3** (1971), 156-168.

[18] P. DIAMOND, "Aggregate Demand Management in Search Equilibrium", *Journal of Political Economy* **90** (1982), 881-894.

[19] A. DIXIT AND J. STIGLITZ, "Monopolistic Competition and Optimum Product Diversity", *American Economic Review* **67(3)**, June 1977, 297-308.

[20] T. EICHER, "Interaction between Endogenous Human Capital and Technological Change", *Review of Economic Studies* **63** (1996), 127-140.

[21] E. FAMA, "Agency Problems and the Theory of the Firm", *Journal of Political Economy* 88 (2) (1980), 288-307.

[22] R. FERNANDEZ AND R. ROGERSON, "Income Distribution, Communities, and the Quality of Public Education", *Quarterly Journal of Economics* **111(1)** (1996), 135-164.

[23] R. FERNANDEZ AND J. GALI, "To Each According To...? Markets, Tournaments, an the Matching Problem with Borrowing Constraints", New York University mimeo, (1997).

[24] F. FERREIRA, "Roads to Equality: Wealth Distribution Dynamics with Public-Private Capital Complementarity", 1995, STICERD (LSE) Discussion Paper.

[25] R. FRANK, *Choosing the Right Pond: Human Behaviour and the Quest for Status* (1985), New York: Oxford University Press.

[26] D. GABAY AND H. MOULIN, "On the Uniqueness of Nash Equilibrium in Noncooperative Games", in Bensoussan, Kleindorfer and Tapien (ed.), *Applied Stochastic Control in Econometric and Management Science*, (1980), North Holland.

[27] D. GALE AND L. SHAPLEY, "College Admission and the Stability of Marriage", *American Mathematical Monthly* **69** (1962), 9-15.

[28] O. GALOR, AND J. ZEIRA, "Income Distribution and Macroeconomics", *Review of Economic Studies* **60** (1993), 35-52.

[29] A. GOSLING, S. MACHIN AND C. MEHIR, "The Changing Distribution of Male Wages in UK, 1966-1992", mimeo, University College London.

[30] G. GROSSMAN AND E. HELPMAN, *Innovation and Growth in the Global Economy* (1991), MIT Press, Cambridge MA.

[31] B. HOLMSTRÖM, "Managerial Incentives Problems - A Dynamic Perspective", in *Essays in Economics and Management in Honour of Lars Wahlbeck* (1982) Helsinki, Swedish School of Economics.

[32] B. JOVANOVIC, "Firm-Specific Capital and Turnover", *Journal of Political Economy* **87** (1979), 972-990.

[33] N. KIYOTAKI AND R. WRIGHT, "A Search Theoretic Approach to Monetary Economics", *American Economic Review* **83(1)** (1993), 63-77.

[34] M. KREMER, "The O-Ring Theory of Economic Development", *The Quarterly Journal of Economics* **108** (1993), 551-575.

[35] P. KRUGMAN AND A. VENABLES, "Integration, Specialisation and Adjustment", (October 1992), CEP (LSE) Discussion Paper, 172.

[36] E. LAZEAR AND S. ROSEN, "Rank-Order Tournaments as Optimum Labour Contracts", *Journal of Political Economy* **89** (1981), 841-864.

[37] H. LEIBENSTEIN, "Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand", *Quarterly Journal of Economics* **80** (1965), 183-207.

[38] R. LUCAS, "On the Mechanics of Economic Development", *Journal of Monetary Economics* **22** (1988), 2-42.

[39] A. MANNING, "Endogenous Labour Market Segmentation in a Matching Model", Centre for Economic Performance (LSE) Discussion Paper No.126, (March 1993).

[40] J. MARKUSEN, "Trade in Producer Services and in Other Specialised Intermediate Inputs", *American Economic Review* **79(1)** (1989), 85-95.

[41] E. MASKIN AND J. RILEY, "Monopoly with Incomplete Information", *Rand Journal of Economics* **15(2)** (1984), 171-196.

[42] J. MCNAMARA AND E. COLLINS, "The Job Search Problem as an Employer-candidate Game", *Journal of Applied Probability* **28** (1990), 815-827.

[43] J. MEDOFF AND K. ABRAHAM, "Experience, Performance and Earnings", *Quarterly Journal of Economics* **95** (1980), 703-736.

[44] J. MEDOFF AND K. ABRAHAM, "Are Those Paid More Really More Productive? The Case of Experience", *Journal of Human Resources* **16(2)** (1981), 186-216.

[45] P. MILGROM AND J. ROBERTS, "Comparing Equilibria", *American Economic Review* **84(3)** (1994), 441-459.

[46] J. MIRRLEES, "An Exploration in the Theory of Optimal Income Taxation", *Review of Economic Studies* **38** (1971), 479-501.

[47] J. MORGAN, "Financing Public Goods by Means of Lotteries", mimeo (1995), Pennsylvania State University.

[48] O. MORGENSTERN, "Demand Theory Reconsidered", *Quarterly Journal of Economics* **63** (1948), 165-201.

[49] D. MORTENSEN, "A Theory of Wage and Employment Dynamics", in Phelps et al. (eds), *Microeconomic Foundations of Employment and Inflation Theory* (1970), Norton, New York.

[50] D. MORTENSEN, "Matching: Finding a Partner for Life or Otherwise", *American Journal of Sociology* **94** Supplement (1988), S215-S240.

[51] H. MOULIN, "Dominance Solvability and Cournot Stability", *Mathematical Social Sciences* **7** (1984), 83-102.

[52] T. MROZ, "The Sensitivity of an Empirical Model of Married Women's Hours of Work of Economic and Statistical Assumptions", *Econometrica* **55**(4) (July 1987), 765-799.

[53] K. MURPHY AND F. WELCH, "The Structure of Wages", *Quarterly Journal of Economics* **107** (1992), 285-326.

[54] M. MUSSA AND S. ROSEN, "Monopoly and Product Quality", *Journal of Economic Theory* **18** (1978), 301-317.

[55] T. PIKETTY, "Imperfect Capital Markets and Persistence of Initial Wealth Inequalities", 1992, STICERD (LSE) Discussion Paper, 92/255.

[56] C. PISSARIDES, *Equilibrium Unemployment Theory* (1990), Blackwell, London.

[57] K. ROBERTS, "Voting over Income Tax Schedules", *Journal of Public Economics* **8** (1977), 329-340.

[58] K. ROBERTS, "Welfare Considerations of Nonlinear Pricing", *Economic Journal* **89** (1979), 66-83.

[59] P. ROMER, "Increasing Returns and Long-Run Growth", *Journal of Political Economy* **94(5)** (1986), 1002-1037.

[60] S. ROSEN, "The Economics of Superstars", *American Economic Review* **71(5)** (1981), 845-858.

[61] S. ROSEN, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", *Journal of Political Economy* **71(5)** (1974), 34-55.

[62] A. ROTH AND M. SOTOMAYOR, *Two-sided Matching. A Study in Game-theoretic Modelling and Analysis* (1990), Cambridge University Press, Cambridge.

[63] M. ROTHSCHILD AND J. STIGLITZ, "Increasing Risk: I. A Definition", *Journal of Economic Theory* **2** (1970), 225-243.

[64] M. SATTINGER, "Search and Efficient Assignment of Workers to Jobs", *International Economic Review* **36(2)** (1995), 283-302.

[65] A. SHAKED AND J. SUTTON, "Natural Oligopolies", *Econometrica* **51(5)** (1983), 1469-1483.

[66] L. SHAPLEY AND M. SHUBIK, "The Assignment Game I: the Core", *International Journal of Game Theory* **1** (1972), 111-130.

[67] A. SHORROCKS, "Ranking Income Distributions", *Economica*, **50**, (1983), 3-17.

[68] A. SMITH, *The Theory of Moral Sentiments*, D. Raphael and A. Macfie (eds.), (1976) Oxford.

[69] L. SMITH, "The Marriage Model With Search Frictions", Massachusetts Institute of Technology mimeo, (February 1996).

[70] M. SPENCE, "Competitive and Optimal Responses to Signals: An Analysis of Efficiency and Distribution", *Journal of Economic Theory* **7** (1974), 296-332.

[71] C. TIEBOUT, "A Pure Theory of Local Expenditure", *Journal of Political Economy* **64** (1956), 416-424.

[72] R. TOPEL AND M. WARD, "Job Mobility and the Careers of Young Men", *Quarterly Journal of Economics* **107** (1992), 439-480.

[73] T. VEBLEN, *The Theory of the Leisure Class*, George Allen (1949) London.

[74] R. WILSON, *Nonlinear Pricing*, Oxford University Press, (1993), New York.