Essays in
Equilibrium Search Theory

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von
Philipp Kircher
aus Plettenberg.

Tag der Promotion: 5. Juli 2006

Dekan:	Prof. Dr. Johannes Köndgen
Erstreferent:	Prof. Dr. Georg Nöldeke
Zweitreferent:	Prof. Andrew Postlewaite, Ph.D.

Tag der mündlichen Prüfung: 5. Juli 2006

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

to my parents

Acknowledgements

I am grateful to Georg Nöldeke for his advice during many discussions and his support for my plans and ambitions. His style of managing a chair, his interaction with students and scholars and his way of doing research are in many respects exemplary for me.

The interaction with Andrew Postlewaite has taught me a lot about the meaning of economic research, and I am thankful for his support throughout my studies. I learned a lot during our collaboration which led to the third chapter of this thesis. I also thank George Mailath for his involvement in many of the discussions.

Manolis Galenianos has not only been a good friend, but our joint research has been a fruitful, sometimes agonizing but usually fun and inspiring activity. Chapter One of this thesis owes to this enterprise. I enjoyed the many long debates and his great hospitality during my frequent returns to Philadelphia.

Ken Burdett and Randy Wright have been extremely supportive. Their interaction with students and their curiosity for the work of others have been a very rewarding experience for me. I also enjoyed the many discussions with Jan Eeckhout.

Stephan Lauermann's challenging questions and comments have been a stimulus for many debates. The discussions with Sandra Ludwig have been very enriching. The friendship with Jorge Gallardo-Garcia and Daniel Velasques-Nunez made Philadelphia a temporary home for me.

At the University of Bonn I thank Benny Moldovanu and Urs Schweizer for their continuing support. I acknowledge the financial support of the Bonn Graduate School of Economics and the European Recovery Program.

I am glad about the friendship with graduate students in Bonn and Philadelphia that made life enjoyable, interesting and fun. I thank Braz Camargo, Thomas Gall, Felix Marklein, Wolfram Merzyn, Timofiy Myolovanov, Arnim Speck and Ludo Visschers for proof-reading and comments.

Finally, I owe much to my family and friends, without whom I would not have gone so far, and who have put up with all the negative sides of grad student life that I levied on them.

Contents

Introduction **1**

I.1 A Brief Review of Search Theory . . . 4

I.2 Chapter Summaries . . . 8

1 Directed Search with Multiple Job Applications **13**

1.1 The Model . . . 16

1.1.1 Environment and Strategies . . . 16

1.1.2 Outcomes and Equilibrium . . . 17

1.1.3 A Preliminary Result . . . 20

1.2 A Special Case: $N = 2$ . . . 21

1.2.1 Worker Optimization . . . 21

1.2.2 Characterization of Firm Optimization . . . 25

1.2.3 Existence and Uniqueness of Equilibrium . . . 26

1.3 The General Case: $N \geq 2$ . . . 27

1.3.1 Worker Optimization . . . 28

1.3.2 Firm Optimization . . . 29

1.3.3 Existence and Uniqueness of Equilibrium . . . 30

1.4 Further Equilibrium Properties . . . 31

1.4.1 Efficiency . . . 31

1.4.2 The Empirical Distribution . . . 34

1.5 Extensions . . . 34

1.5.1 Free Entry . . . 34

1.5.2 Endogenous Number of Applications . . . 35

1.5.3 The Dynamic Version . . . 36

1.6 Directed vs. Random Search . . . 37

1.7 Conclusion of Chapter 1 39

2 Efficiency of Simultaneous Directed Search with Recall 43

2.1 The Model 47

2.1.1 Environment and Strategies 47

2.1.2 Expected Payoffs and Equilibrium Definition 49

2.2 Equilibrium Characterization 54

2.2.1 Workers' Search Decision 55

2.2.2 Firms' Wage Setting 57

2.2.3 Equilibrium Outcome 61

2.3 Efficiency 64

2.3.1 Search Efficiency 65

2.3.2 Entry efficiency 68

2.3.3 Application Efficiency 70

2.4 Generalization to $N > 2$ applications 71

2.4.1 Extended Setup and Main Result 71

2.4.2 Convergence to the Competitive Outcome 73

2.5 Discussion and Outlook 74

2.5.1 Discussion of Main Assumptions 74

2.5.2 Relation to the Literature 75

2.5.3 Conclusion of Chapter 2 78

3 Strategic Firms and Endogenous Consumer Emulation 79

3.1 The model 81

3.2 Optimal Behavior 87

3.2.1 Consumer Search 87

3.2.2 Consumers' stationary behavior 91

3.2.3 Firms' behavior 95

3.2.4 Equilibria 98

3.3 Discussion 102

3.3.1 Social Interaction 103

3.3.2 Robustness 104

3.3.3 Conclusion of Chapter 3 107

Appendix 109

A.1 Appendix to Chapter 1 109

A.2 Appendix to Chapter 2 118

A.3 Appendix to Chapter 3 135

References 145

List of Figures

1.1 Workers' application behavior. IC-H and IC-L are the workers' indifference curves for 'high' and 'low' applications. Linearity is only used for illustration. 24

1.2 Firms' equilibrium behavior. IP-H and IP-L are the isoprofit curves for high and low wage firms. 27

1.3 Equilibrium wage dispersion for $N = 15$ and $b = 1$. 28

1.4 Sum of firms for $b = 1$ and various N. 32

1.5 Wage densities with random search, for $N = 15$ and $b = 1$. 39

2.1 Equilibrium behavior. IC_1 and IC_2: Worker's indifference curve for the low and high wage, respectively. IP_1 and IP_2: Isoprofit curves for low and high wage firms. 61

2.2 Inefficiencies at a unique market wage. Areas A and B indicate mutual benefits for both workers and firms. 68

A.1 f(x) for $x \geq 0$. 116

A.2 $f(\frac{a_2}{2d}) + f(\frac{a_2}{2(1-d)})$ for $d \in [0, 1]$. 117

A.3 $f'(x)x^2$ for $x \in [0, a_2]$. 118

Introduction

Market participants often lack knowledge about product characteristics or about the actions of other market participants. This introduces frictions into the trading process, and economists have long realized that these have to be treated seriously in order to understand some important phenomena that we observe in actual market interactions.

The idea that frictions are important in real life markets goes back to the observation that seemingly identical factor inputs receive very different remuneration. Prices for goods and wages for labor of apparently identical quality vary substantially, and it has been a challenge for economists to explain the observed variation and its properties.[1] The realization that frictions are important and of substantial magnitude has prompted the question whether markets still operate (second best) efficiently given the frictions. This question has received substantial attention as it determines whether policy interventions might be warranted on the grounds of efficiency improvements. To analyze markets with frictions we face the question of how agents cope with the frictions, which is at the heart of all search models. Since dispersion of factor prices, efficiency concerns and adaptation to market frictions present central themes also in this thesis, we will briefly expand on each of these.

1. In the context of the labor market Stigler (1962) argued that labor of similar quality obtains very different salaries. Later studies support this view in the sense that a large fraction of the variation in wages remains unexplained even if one controls for observed and unobserved worker heterogeneity.[2] Competitive models imply the law of one price (or one wage, respectively), and therefore labor economists have pursued alternative models that explicitly account for market frictions in order to explain the observed variations in wages.[3]

---

[1]For evidence on prices see Stigler (1961) and the subsequent literature on price dispersion. For wages see the next paragraph.

[2]See for example Abowd, Kramatz, and Margolis (1999) and further references in Chapter 1.

[3]For example, the model by Burdett and Mortensen's (1998) has received strong popularity. See the next

2. Unemployment is a sign of frictions and one of the major issues in welfare economics. While some of it might be attributed to labor market interventions by the government or cartelized labor bargaining, early economists have already argued that there is no full employment even if economic activity is relatively unrestricted. Friedman (1968) and Phelps (1967) point out that the unemployed might neither waste time nor leisurely consume society's resources, but rather exert effort to overcome the market's frictions. Only because the unemployed undertake the productive activity of "search" do they find opportunities to become employed and are able to re-enter the workforce. Friedman and Phelps promoted the view that there exists a "natural rate" of unemployment which efficiently allocates resources in a frictional economy; and monetary policy interventions cannot improve long-run efficiency. This has spawned a long literature about the market's ability to allocate resources efficiently given the frictional forces that drive unemployment, which requires models that can explicitly account for these frictions.

3. The presence of frictions affects all agents in the economy, and they will take this into account in their decision-making. This naturally leads to equilibrium search models where both sides of the market adjust their actions optimally to the frictions and to the behavior of the other agents. If all information in the market is common knowledge, each individual's search strategy is a function of this information and his knowledge of the other agents' equilibrium strategies. If imperfect information is the source of the frictions and some agents have private information, then individual strategies will also condition on the observed behavior of others because this entails valuable information. The idea that agents learn form one another goes back at least to Nelson (1970) who argues that consumers' search activities are guided by the information obtained from others. When private information is important, then also information transmission between agents is likely to be of importance and should be accounted for in the theoretical analysis.[4,5]

section for further references. For an estimation approach see, e.g., van den Berg and Ridders (1998).

[4]Imperfect information and its transmission between economic agents has been important in the analysis of such mundane problems as the adoption of one of various competing kinds of technology as well as in the study of more fundamental problems such as bank runs and speculative attacks against currency regimes. See Chamley (2004) for an overview of the literature.

[5]We will illustrate the relevance of private information and observability in the context of a consumer product market. The labor markets that we analyze do not feature private information and information transmission is not an issue.

The following three chapters of this thesis reconsider the points just raised in the context of specific economic environments. Chapter One presents a novel model of the labor market that features wage dispersion among homogenous workers. The shape of the wage density is downward sloping, which conforms more closely with stylized facts than the predictions of "standard" models of wage dispersion.[6] The model is simple: Firms compete for labor by posting their wage, and workers then decide where to apply. The novel feature is that workers can apply to multiple firms within one period and therefore face a strategic simultaneous choice problem. Since casual observation suggests that people indeed send many applications at once at the beginning of their search, the understanding of the consequences of such behavior might be interesting in its own right.

Chapter Two considers efficiency properties in a setup similar to Chapter One. The equilibrium interaction in Chapter One as well as in some related work is inefficient, in contrast to the large literature that considers a single application for each worker. In accordance with other work, our analysis in Chapter One entails a simplifying assumption about the way that a firm can fill its vacancy. Chapter Two is concerned with the question whether the inefficiency is due to an inherent problem of the simultaneous choice among multiple alternatives, or whether it is an artifact of this simplified assumption. We naturally extend the setup to allow firms to contact all their applicants in order to fill their job (which was restricted in Chapter One) and find that this achieves an allocation which is efficient given the frictions. This confirms the hypothesis that the natural rate of unemployment is efficient, and extends positive efficiency results from models with a single application to the setup where agents can choose their intensity of search by varying the number of applications they send.[7]

While the first chapters deal only with public information, chapter three considers a private information setting in which learning from the actions of others is important. Models of social learning mainly focus on situations where each consumer has only a single decision in his life. We embed social learning into a standard repeated search environment of a consumer

[6] Burdett and Mortensen (1998) is arguably the current "standard" model in the analysis of wage dispersion. It predicts an increasing an convex shape. We will see that this is a feature that is common in partially directed search models.

[7] In contrast to Chapter One, we manage to introduce the number of applications as the choice variable for the worker and analyze the efficiency of that choice together with free entry of firms and minimization of the frictions in the market.

product market in which agents have a natural incentive to learn from the behavior of others. We allow for heterogeneity among consumers and strategic actions by the firms. This allows predictions regarding the group that will "lead" the market and regarding the way firms treat their various consumers.

All chapters share the following principles: They all consider equilibrium models in which both sides of the market act strategically. Agents act optimally given the frictions, the other agents' strategies and the observable information. Second, all chapters feature search intensity as a choice variable. The class of directed search models on which the first two chapters are based already have the appealing feature that market frictions are explicitly modeled. It therefore lends itself to the explicit introduction of search intensity, which is simply a choice about the number of applications. In the third chapter we want to allow different consumers to consume with different frequencies, and we want to incorporate the feature that more attractive products are consumed more frequently. We deploy a novel model of search intensity, which is based on a trade-off between an outside option and market consumption coupled with a taste shock that changes the attractiveness of per period consumption. Despite this feature the derivation of the optimal search strategy remains surprisingly simple. Third, all chapters allow for some choice about whom to meet as a potential trading partner, and therefore they differ from the classical random search approach in which both sides of the market meet purely randomly. In the first two chapters workers decide on the identity of the firm to which they apply based on its wage offer. In the last chapter people condition their search process on what they observe from others.

This thesis is aimed at studying the interaction between search intensity and non-random search in an equilibrium setting. We believe that these are important elements in many markets, and we hope to shed additional light on their interplay through our theoretical analysis.

I.1 A Brief Review of Search Theory

To illustrate how this thesis relates to developments in search theory, we present a short review here. It will be brief because each chapter is intended to be readable as a self-contained unit, and we want to avoid too much repetition. For a thorough review of search theoretic models see Rogerson, Shimer and Wright (2005).

The developments in search theory were spawned by observations like dispersion of prices for identical goods or factor inputs (e.g. Stigler (1961, 1962)), the observations of a correlation between unemployment and vacancies (Beveridge, 1944) and between unemployment and inflation (Phillips, 1958). While the question of efficiency was early on taken up by monetary economists like Friedman (1968) and Phelps (1967), they had little theory to back up the ideas. Models that evolved thereafter were not so much concerned with monetary influences but rather with the aim to include unemployment in equilibrium models at all and to analyze the resulting properties.

Initial models that analyzed behavior in a search environment were partial equilibrium models which only focused on the behavior of one individual agent. They considered *sequential* search, where the agent meets different options (e.g. jobs that pay different wages) and in the end takes one of the options he has met. Waiting longer, i.e. meeting more options, is costly either due to a fixed cost of waiting or due to discounting. DeGroot (1970), McCall (1965) and Weitzman (1978) develop optimal stopping rules for such environments. Stigler (1961) and Chade and Smith (2005) propose optimal algorithms for *simultaneous* choice problems.

Insights derived from the partial models received criticism (Rothschild, 1974) because they neglect the other side of the market. Therefore later models integrated the sequential search models in an equilibrium framework in which both sides of the market optimize. These models are known as *random search* models. They assume that if a worker meets a vacant firm, this firm is simply a random draw from the set of all vacant firms. Early on Diamond (1971) considered such a framework under the assumption that firms make take-it-or-leave-it wage offers, and established that firms can extract monopoly profits in such a market because switching costs provide them with local market power. In a model with identical workers his insights imply that workers receive none of the surplus even though there are many employers present that want to hire, which is known as the Diamond Paradox. Subsequent models assumed Nash bargaining between workers and firms, which implies that workers receive some rents (see e.g. Diamond (1982), Mortensen (1982a,1982b), Pissarides (1984)). This approach to labor markets has been used to investigate many questions from unemployment insurance to business cycle fluctuations. For the understanding of this thesis two important results emerge: First, these models still do not yield any dispersion of wages if workers and firms are assumed to be identical. Second, these models are generally not constrained efficient given the frictions. This has been highlighted, e.g., by Hosios (1990) and implies that fiscal or monetary interventions might have scope to improve the allocation of economic resources.

To introduce wage dispersion two roads have been successfully explored. The first and more immediate is to relax the assumption of homogeneity. Albrecht and Axell (1984) and Reinganum (1979) analyze variants of this. If one wants to stay in the paradigm of homogenous workers and homogeneous firms, attractive models of wage dispersion have been surprisingly hard to generate. Burdett and Judd (1983) and Butters (1977) clarify the crucial point to generate such dispersion: workers have to be able to see multiple employment options simultaneously. If some workers see only one option, but some see multiple options, then firms that offer low wages can only hire those workers that do not have a better option but make a high profit per hired worker. Firms that offer high wages attract many workers but make less profit per hired worker. In equilibrium this balances such that all firms make identical profit. For this logic it is crucial that not all workers have two offers, because if they did the firms would all compete in a Bertrand style. Burdett and Mortensen (1998) translate the idea of multiple options into a sequential search setup in which workers can also search on the job. Meetings are still random and firms make take-it-or-leave-it offers. Low wage offers are accepted only by unemployed workers who do not have alternative options, while high wage offers are also accepted by workers who have a job but with low pay, and firms can again make equal profits despite wage dispersion.[8] Acemoğlu and Shimer (2000) explore a setup similar to Burdett and Judd (1983) and show that the logic still applies even if firms are capacity constrained and can only hire a single worker.[9] Gautier and Moraga-González (2005) show that with capacity constraints all workers may see and apply to two options and still wage dispersion obtains. Bertrand competition will not occur since firms cannot always hire all applicants, and therefore some applicants are left with only the low option, which gives the low wage firms the chance to hire.

The models just discussed can be considered random search because the options available are drawn at random and only reflect a small subset of the available offers. This implies an increasing and convex shape of the wage distribution in all of these models, which conflicts with empirical findings.[10] These models do feature a part of non-random (or directed) search

[8]On-the-job-search overturns the Diamond Paradox.

[9]The paper also has additional insights on technology dispersion.

[10]Mortensen (2003) discusses stylized facts, which point at a distribution that is skewed with a long right

in the sense that workers can chose non-randomly between their available options. Hence, we refer to these models either as *partially random* or *partially directed* search.

The final generation of models reviewed here extends the scope of the information that is available to the searching agents by allowing workers to see all wage offers in the market. In so called *directed search* models each firm is capacity constrained (i.e. has a single job to fill) and can publicly post the wage it will pay. The workers can see all wage announcements and decide to which firm to apply. If more than one worker applies, the firm picks one of them and offers the job to him. This setup was proposed e.g. by Montgomery (1991), Peters (1991) and Burdett, Shi and Wright (2001). Two elements distinguish this environment from the previous ones: Wage competition is global because each worker can see all the offers, and there are no chance moves that introduce randomness. Thus, the frictions do not stem from randomness in the environment, but arise as a consequence of the type of equilibrium that is considered. It is assumed that workers cannot coordinate, which is modeled as a restriction to symmetric strategies. In a larger market it therefore cannot be an equilibrium that all workers apply to the same firm: If they did, all but one of them would remain unemployed. As a consequence of this, the subgame perfect symmetric equilibrium involves randomization. This implies that sometimes too many workers apply for a single job and some of them remain unemployed, while some vacancies might not receive any application and remain vacant. The restriction to symmetric strategies therefore embeds the frictions. Despite the miscoordination the optimality of the symmetric strategy implies that workers apply to higher wage offers more frequently. Hence, firms can improve their hiring probability by increasing the wage. In a large market the individual agent has no market power, and due to the competitive element these models are also called *competitive search.*[11]

Two features of directed (or competitive) search models are worth highlighting. First, Moen (1997) and others show that the equilibria in such a setup are efficient given the frictions. He can relate his result back to Hosios (1990) and show that the condition that is generally not

tail.

[11]Competitive search refers more broadly to models in which market makers set up the wage and the meeting probabilities in a competitive way, but the meeting probabilities have to satisfy the consistency conditions of a given matching function, which might be more general than the one that comes out of the game just specified. Moen (1997) uses such a setup. Shimer (1996) shows how this can be interpreted for a specific matching function in light of the physical structure presented here.

met by random search models with Nash bargaining is met in this setup. The models provide a clear extensive form for the market interaction in which unemployment can be interpreted and show that the competitive forces in this environment lead to an allocation that cannot be Pareto improved by monetary or fiscal policy (unless the frictions are eliminated). Second, these models do not feature wage dispersion when restricted to homogenous workers and homogenous firms. This is shown e.g. in Burdett, Shi and Wright (2001) and Moen (1997), and is due to the fact that the models restrict attention to a setting where each worker can only apply to a single firm in any given period.[12] The only exceptions are Delacroix and Shi (forthcoming) and Albrecht, Gautier and Vroman (forthcoming). The former model directed on-the-job search and obtain wage dispersion with a discrete number of wages. They are able to characterize the transition of a worker's wage income over time, yet the model's tractability is limited and discussion of the shape of dispersion or the efficiency of the model is omitted. The latter are the first to consider a setup with multiple applications per worker and integrate Chade and Smith (2005) into an equilibrium framework. They provide workers with bargaining power if the worker obtains two or more job offers, which leads to a single market wage.[13] Wage dispersion only arises through bargaining. They are the first to consider a directed search model that features inefficiencies beyond those that are unavoidable parts of the frictions, i.e., the competitive element does not guide the market towards efficiency.

I.2 Chapter Summaries

In this section we briefly outline the equilibrium models that we analyze in the following chapters, present our main findings and relate them to the literature mentioned in the previous section.

In **Chapter One** we develop an equilibrium directed search model of the labor market where

[12]Burdett, Shi and Wright (2001) only solve for equilibria in which firms offer one wage. In a finite economy such as theirs uniqueness results are difficult. For an appropriate limit game it can easily be shown that only a unique wage can arise. Convergence results by Peters (2000) establish that wages will be close to this unique wage when the finite market is sufficiently large. For Moen (1997) uniqueness follows out of the optimality. In Moen (1997) the model is not phrased in terms of applications, but the job arrival rates do not allow more than one choice at any given time.

[13]Since they only solve for equilibria with a single wage, there might be other equilibria as well.

workers can simultaneously apply for multiple jobs. The main result is that all equilibria exhibit wage dispersion despite the fact that workers and firms are homogeneous. Wage dispersion is driven by the simultaneity of the application choice. Risk-neutral workers apply for both 'safe' and 'risky' jobs. The former yield a high probability of a job offer, but for low pay, and act as a fallback option; the latter provide higher potential payoff, but are harder to get. Consistent with stylized facts, the density of posted wages is decreasing and high wage firms receive more applications per vacancy. Unlike most directed search models, the equilibria are not constrained efficient.

Chapter One integrates Chade and Smith (2005) into an equilibrium framework. In contrast to Albrecht, Gautier and Vroman (forthcoming) we do not assume a final stage of bargaining, which requires us to pursue the choice problem of the worker in a setting where he is really confronted with a set of different wages. The workers' choice problem provides a lot of structure and we can show that the market separates into some smaller "micro-segments" that operate similar to Burdett, Shi and Wright (2001). Within each segment one wage is offered, which implies many wages across the various market segments. Our novel results on the shape of wage dispersion is derived from two simple assumptions: Firms can compete for labor, and workers apply to several firms. We are able to show clearly why partially directed search yields an unrealistic increasing and convex shape for the wage density, whereas our directed search setup results in a decreasing distribution.

In **Chapter Two** we analyze the properties of a directed search labor market in which workers choose how many applications to send simultaneously after observing the firms' wage offers. The number of applications can be interpreted as an explicit form of search intensity. Since workers might reject some job offers in favor of better ones, we allow rejected firms to contact ("recall") other applicants by applying the deferred acceptance algorithm to the endogenous network. The equilibrium is generically unique, all workers choose to send the same number of applications, and firms offer a discrete number of wages. The equilibrium is constrained efficient given the workers' lack of coordination: entry of firms, number of applications, and number of matches are efficient. Wage dispersion is necessary for the market to achieve constrained efficiency despite homogeneity of workers and firms. For small application costs the equilibrium outcome converges to the unconstrained efficient competitive outcome.

Chapter Two reconsiders the case of multiple applications, but now the number of applications is a choice and firms can contact additional workers after a rejection. Our efficiency result reverses the negative results on efficiency obtained by Albrecht, Gautier and Vroman (forthcoming) and in Chapter One. The positive efficiency properties established by Moen (1997) are recovered. We are able to clarify the property that is important for efficiency. Of key importance is a tight link between firms' and workers' expected payoffs. This link is present in one application models and in the case of recall, but fails without recall. While other models have looked at entry of firms and sorting in the market as the margins of efficiency, we introduce a new margin by additionally analyzing the workers' choice regarding the number of applications. Our convergence result establishes the close connection between competitive search equilibria and classical Walrasian markets with perfect competition.

In **Chapter Three** we consider a search model of a consumer product market in which information is private and agents can learn from the actions of others. Consumers and firms are heterogenous and act strategically. Consumers search for high qualities among a large set of firms, and can condition their choices on observed actions of other consumers. When they can identify consumers who are more likely to have found a high quality firm, uninformed individuals will optimally emulate those consumers. One group of consumers arises endogenously as "leaders" who are being emulated. Follow-on sales induce firms to give preferential treatment to these lead consumers, which reinforces their learning.

Chapter Three extends the standard random search environment in which each agent collects information only through personal experience by introducing observability of other agents' actions. While agents naturally have an incentive to learn from other players actions, information transmission is not considered in most random search models. The models that do consider social learning are usually restricted to environments in which agents make one-time choices (see e.g. Banerjee (1992) or Banerjee and Fudenberg (2004)).[14] The main contribution to the social learning literature is the introduction of consumer heterogeneity and the firms' ability to interact strategically. Consumer heterogeneity allows us to analyze who obtains the leadership role in the market, and the firms' decisions provide an important feedback in the learning process. We show that the decisions of the firms and endogenous leadership status interact in an important manner.[15]

[14] Ellison and Fudenberg (1995) analyze repeat purchases in a boundedly rational setting. Their decision rule is such that the model can be reinterpreted as a setting in which agents only consume once when born, and newborn uninformed consumers enter the market with some private knowledge of one firm.

[15] Chapter One and two abstract from heterogeneities in order to isolate the effect of frictions on dispersion

Before moving to the main body of this treatise, a brief word on the use of the first person plural may be in order. Chapter One has been developed in collaboration with Manolis Galenianos, while Chapter Three owes to the collaboration with Andrew Postlewaite.[16] The use of the plural in the second chapter and in the Introduction are due to the well-published advice by Thomson (1999, p. 180) to use this form even for single-authored publications in economic theory.

The next three chapters each present one idea as a self-contained unit.

---

and efficiency. Chapter Three introduces heterogeneities because they importantly interact with the social learning aspect.

[16]The joint work with Manolis Galenianos has been presented and circulated under the title "Directed Search with Multiple Job Applications". The joint work with Andrew Postlewaite has been presented and circulated under the titles "Why do the rich get more for what they pay?" and "Why are the wealthy treated so well?".

Chapter 1

Directed Search with Multiple Job Applications

"What accounts for pay differentials among workers" is a classic question in economics. Empirical research has documented that a large part of wage variation cannot be explained by productivity differences. For instance, Abowd, Kramarz, and Margolis (1999) find that observable worker characteristics can explain only 30% of wage differentials. Controlling for firm characteristics or unobserved worker heterogeneity helps account for part of the other 70 percent, but a substantial residual remains, suggesting that a model with search frictions might be a useful way to think about this issue.[1]

Prominent examples of random search models that generate equilibrium wage dispersion include Burdett and Judd (1983), Albrecht and Axell (1984), and Burdett and Mortensen (1998).[2] We propose a new model of wage dispersion with homogeneous workers and firms, based not on random but on directed search, and one additional feature that we think is

[1]Abowd, Kramarz, and Margolis (1999) can account for about half of the residual variation when controlling for unobserved worker heterogeneity. Postel-Vinay and Robin (2002) estimate a model with observed and unobserved worker heterogeneity as well as productivity heterogeneity in firms and conclude that "the contribution of market imperfections to wage dispersion is typically around 50 [percent]." In a similar exercise, van den Berg and Ridder (1998) report that "search frictions explain about 20 [percent] of the variation in observable wage offers."

[2]Note that in the introduction of this thesis we classified some papers as (partially) random or (partially) directed. For the purpose of separating clearly the forces that come from any form randomness, we stick with the first convention and will omit the qualifier as all models considered here that are not completely directed but feature wage dispersion are partially random.

an important characteristic of the search process: job seekers can apply for several jobs at the same time. In random search models, workers looking for employment do not know the wages offered by different firms. In directed search they observe the wages posted by all firms before deciding where to apply. However, they do not know how many other workers apply to the same firm and, since firms have a limited number of vacancies, they may get rationed. Nevertheless, the equilibria of these models are usually constrained efficient.[3]

So far, most research in the area has focused on workers applying for one job at a time, which results in a unique equilibrium with a single wage (at least when agents are homogeneous). In this paper, workers apply for N jobs simultaneously, which yields very different results. Despite the assumption of homogeneity, all equilibria exhibit wage dispersion. Even though workers are risk neutral, they care about the probability of success of each job application because their payoffs only depend on the most attractive offer they receive. The resulting portfolio choice problem is the driving force for dispersion. Furthermore, equilibria are *not* constrained efficient. Two predictions of our model, that are consistent with stylized facts, are of particular interest since they are not typically found in random search models: the density of posted wages is declining,[4] and firms that post higher wages receive more applicants on average.[5] It is worth emphasizing that these results are not due to comlex modelling assumptions, but obtain for a simple and intuitive view of the labor market: Firms compete for labor by publicly announcing their wage, and workers then decide to which firms to apply.

The intuition behind the main result of dispersion is quite straightforward. A worker faces a portfolio choice problem when deciding where to send each of his N applications, since the probability of getting a job is different at different wage levels. This occurs because higher paying firms attract more applicants on average and hence an application to such a firm succeeds with lower probability. Loosely speaking, a worker's optimal strategy is

[3]One of the reasons why models of directed search have become more popular is that they provide a more explicit explanation of the matching process and wage determination than random search models. Rogerson, Shimer, and Wright (2005) discuss this point in their recent survey of search-theoretic labor models.

[4]In contrast, the Burdett-Mortensen model delivers a wage density that is upward sloping. While this can be fixed by extending the framework, it is often said to be a failing of the basic model (see Mortensen (2003)).

[5]Holzer, Katz, and Kruger (1991) provide evidence for this point. Note that in random search models firms cannot influence the inflow of workers since their wage is not observed until after they meet a worker.

to apply to jobs that offer different levels of risk and payoff. Some applications are sent to 'safe' wages that guarantee a high probability of getting a job, but for low pay. Since this provides insurance, it is optimal to take on more risk with the other applications. As a result, he also applies to firms where the probability of getting the job is lower but the potential payoff is high.

The willingness of workers to send each application to a separate wage level creates an incentive for firms to post different wages. It turns out that in any equilibrium exactly N wages are posted, and every worker applies once to each distinct wage. From the firms' perspective, the lower margins of high wages are balanced with a higher probability of filling a vacancy, leading to the same expected profits. It is important to reiterate, however, that this intuition fails in the single application case. The incentive for firms to post different wages arises only because every worker applies to multiple jobs.

Well-known papers on directed search include Montgomery (1991), Peters (1991), Shimer (1996, 2005), Moen (1997), Julien, Kennes, and King (2000), Burdett, Shi, and Wright (2001) and Shi (2002). Delacroix and Shi (forthcoming) develop a directed search model with on-the-job search, which shares some features with our model since employed workers can take on more risk when looking for jobs. Albrecht, Gautier, and Vroman (forthcoming) is the only other directed search paper where workers apply to multiple jobs simultaneously. The authors make different assumptions and they reach very different results as will be discussed in the conclusions.[6] Chade and Smith (2005) solve a portfolio choice problem that is similar to ours, but in a very different partial equilibrium context.

The rest of the paper is structured as follows. Section 2 presents the model, states the main theorem, and proves a straightforward initial result. Section 3 discusses the special case of two applications, which provides many of the important insights. The following section extends the results to an arbitrary (but finite) number of applications. Section 5 evaluates the efficiency of the equilibrium and the empirical distribution of wages. Section 6 introduces important but straightforward extensions such as free entry, endogenous choice of the number of applications, and a dynamic labor setting. Section 7 compares the distribution of wages that results from random and directed search and section 8 concludes.

[6] The basic difference is that, in this paper, firms commit to the wages they post, while Albrecht, Gautier, and Vroman (forthcoming) assume that firms making job offers to the same worker engage in Bertrand competition.

1.1 The Model

In this section we introduce the main features of the model, and define outcomes, payoffs, and equilibrium. At the end we state the main theorem and prove a preliminary result.

1.1.1 Environment and Strategies

There are continua of measure b workers and measure 1 firms with one vacancy each. All workers and all firms are identical, risk neutral, and they produce one unit when matched and zero otherwise. The utility of an employed worker is equal to his wage and the profits of a firm that employs a worker at wage w are given by $1 - w$. The matching process has four distinct stages. Firms start by posting (and committing to) wages. Then, workers observe all postings and send out N applications. Firms follow by making a job offer to one of the applicants they have received, if any. Last, workers that get one or more offers choose which job to accept. If a firm's chosen applicant rejects the job offer then the firm remains unmatched.[7] Firms therefore compete for workers in two separate stages: they want to attract at least one applicant in the second stage and they try to keep that applicant in the last stage; we label these 'ex ante' and 'ex post' competition, respectively.

As is common in the directed search literature, trading frictions are introduced by focusing attention on symmetric mixed strategies for workers. The assumption is that, since the market is large, workers cannot coordinate their search and hence they all use the same strategy. For simplicity, we also assume that their strategies are anonymous, i.e. all firms that post the same wage are treated identically by workers. This assumption, however, is not necessary: it is possible to let workers condition on the firms' names (say, a real number in [0,1]) but this would clutter the exposition without changing the results. Last, the firms also follow anonymous strategies, meaning that they treat all workers the same in the event that they receive multiple applicants. This is the standard environment in the directed search literature, such as Peters (1991) or Burdett, Shi, and Wright (2001), except for the innocuous assumption of the anonymity of workers' strategies, and the key difference that we allow multiple applications.

[7] Chapter 2 relaxes this assumption and allows the recall of all applicants in the case a firm's offer is rejected. Though the matching process is quite different, the unique equilibrium exhibits an N-point distribution of posted wages suggesting that the qualitative features of our model are robust. However, constrained efficiency is recovered in chapter 2. See section 1.4.1 for further discussion.

Before describing the actual strategies, observe that the last two stages of the game can be solved immediately. In the fourth stage, workers with multiple job offers choose the highest wage, and randomize with equal probabilities in the case of a tie. In the third stage, firms with many applicants choose one at random. Therefore we only need to consider the strategies for the first two stages. A strategy for the firm is a wage w that it posts in the beginning of the game. Workers observe all the wages and decide where to apply. Denote the distribution of posted wages by F and note that, due to anonymity, the workers' strategies can be summarized by the wages to which the applications are sent. Therefore, a pure strategy for a worker is an N-tuple of wages to which he applies and a mixed strategy is a randomization over different N-tuples. We denote the workers' strategy by $G(F)$, which is a mapping from the posted wages to the set of all cumulative distribution functions on $[0,1]^N$.

1.1.2 Outcomes and Equilibrium

We define $q(w)$ to be the probability that a firm posting w receives at least one application and $\psi(w)$ to be the conditional probability that a worker who has applied to such a firm accepts a *different* job offer (i.e. the probability that the firm does *not* get the worker). Let $p(w)$ be the probability that a worker applying to wage w gets an offer and $\mathcal{W}$ be the support of the posted wages (i.e., $\mathcal{W} \equiv suppF$). When a wage is not posted by any firm ($w \notin \mathcal{W}$), we have $p(w) = 0$. Last, we define the *value* of an *individual* application to some wage w to be $p(w)\ w$. Given any N-tuple $\mathbf{w} = (w_1, w_2, ..., w_N)$ chosen by the worker, we assume without loss of generality that $w_N \geq w_{N-1} \geq ... \geq w_1$ for the remainder of the paper.

The expected profits of a firm that posts w and the expected utility of a worker who applies to $\mathbf{w}$ are given by

$$\pi(w) = q(w)\ (1 - \psi(w))\ (1 - w) \tag{1.1}$$

$$\begin{aligned} U(\mathbf{w}) &= p(w_N)\ w_N + (1 - p(w_N))\ p(w_{N-1})\ w_{N-1} \\ &\quad + ... + \prod_{i=2}^{N} (1 - p(w_i))\ p(w_1)\ w_1. \end{aligned} \tag{1.2}$$

The expected profits are equal to the probability that at least one applicant appears times the retention probability times $(1 - w)$. A worker gets utility w_N from his highest application, which is successful with probability $p(w_N)$. With the complementary probability that application fails and with probability $p(w_{N-1})$ he receives w_{N-1}. And so on.

On $\mathcal{W}$, both $p(w)$ and $q(w)$ depend on the average *queue length* at w, which is denoted by $\lambda(w)$. Intuitively, the queue length is the number of applications divided by the number of firms at a particular wage rate. Formally it is defined by the integral equation

$$\int_0^w \lambda(\tilde{w})\, dF(\tilde{w}) = b\, \hat{G}(w), \tag{1.3}$$

where $\hat{G}(w)$ is the expected number of applications that a single worker sends to wages no greater than w.[8] The right hand side of equation (1.3) gives the number of applications that are sent up to wage w by *all* workers, while the left hand side gives the number of firms that post a wage up to w multiplied by the average number of applications they receive.

When a worker applies for a wage w he randomizes over all firms offering that wage rate due to anonymity. As a result, the number of applications received by a firm posting w is random and follows a Poisson distribution with mean $\lambda(w)$.[9] Therefore the probability that a firm posting w receives at least one application is $q(w) = 1 - e^{-\lambda(w)}$ and the probability that a worker who applies to such a firm gets an offer is $p(w) = (1 - e^{-\lambda(w)})/\lambda(w)$, where $p(w) = 1$ when $\lambda(w) = 0$.[10]

In order to evaluate $\psi(w)$ for some $w \in \mathcal{W}$ we need to find the probability that, after applying to w, a worker takes a different job. Let $R_j(w_j, w_{-j})$ be the probability that a worker who applies to (w_j, w_{-j}) accepts the job posting w_j if made an offer. This occurs if the worker has no offer that is strictly higher *and* if w_j is picked in the case of a tie after randomizing. The indexes of applications can be relabeled so that higher indexes are given preference when tied. This means that $R_j(w_j, w_{-j}) = \prod_{k>j}(1 - p(w_k))$ and we can integrate over all possible wages where workers apply to.[11] Letting $Pr[j|w]$ be the conditional probability that a worker who applied to $w \in \mathcal{W}$ did so with his jth application and $G_j(w_{-j}|w)$ be the

[8]If $G_i(w)$ is the marginal distribution of w_i, then $\hat{G}(w) = \sum_{i=1}^{N} G_i(w)$.

[9]Suppose that n applications are sent at random to m firms. The number of applications received by a firm follows a binomial distribution with probability $1/m$ and sample size n. As $n, m \to \infty$ keeping $n/m = \lambda$ the distribution converges to a Poisson distribution with mean λ.

[10]Notice that the anonymity of the worker strategies is not a necessary condition for this point to hold. Symmetry and optimality clearly imply that firms with the same wage must have the same expected queue length. Poisson matching follows.

[11]The relabeling is without loss of generality since the randomization can occur before the applications are actually sent.

conditional distribution over the other applications, given that the jth application was sent to wage w, $\psi(w)$ is given by

$$\psi(w) = 1 - \sum_{j=1}^{N} \Pr[j|w] \int R_j(w, w_{-j}) \; dG_j(w_{-j}|w). \tag{1.4}$$

So far $\lambda(w)$ and $\psi(w)$ have been defined for wages on the support of F, meaning that the workers' optimization problem can be solved for a given distribution of posted wages. However, off the equilibrium path payoffs need to be evaluated in order to solve the firms' problem, and this requires that $\lambda(w)$ and $\psi(w)$ are well defined on the full domain [0,1]. That is, a firm needs to know the queue length and the retention probability it will face at *any* wage. Therefore, although no one is actually applying to wages that are not posted, the queue lengths at such wages could be positive since they represent how many workers *would* apply there *if* these wage were offered; and similarly for $\psi(w)$. The problem is that when $w \notin \mathcal{W}$, $\lambda(w)$ and $\psi(w)$ are not pinned down by equations (1.3) and (1.4), as both F and G have zero density at those wages.

To get around this issue we define λ and ψ as if 'many' firms post every wage in [0,1] so that the reaction of workers can be meaningfully evaluated. We introduce a fraction of noise firms of measure ϵ that post a wage at random from a full support distribution, $\tilde{F}$. Equivalently, one can interpret it as a mistake that firms make with probability ϵ. Given a candidate F, the distribution of posted wages becomes $F_\epsilon(w) = (1-\epsilon)\; F(w) + \epsilon\; \tilde{F}(w)$ and the game can be analyzed from the second stage onwards. Let $G(F_\epsilon)$ denote a best response of workers to F_ϵ. The outcomes λ_ϵ and ψ_ϵ can be calculated in the entire domain [0,1] using F_ϵ, $G(F_\epsilon)$, and equations (1.3) and (1.4). As $\epsilon \to 0$ the perturbed distribution converges to F, and we define $\lambda(w) = \lim_{\epsilon\to 0} \lambda_\epsilon(w)$ and $\psi(w) = \lim_{\epsilon \to 0} \psi_\epsilon(w)$ for all $w \in [0,1]$. Noise firms are simply a means to evaluate the profits a firm would obtain when deviating, and none of our results depend on the exact choice of $\tilde{F}$.[12]

[12] Two different approaches have been taken to solve the same problem in the $N = 1$ case. The market utility approach, used in Shimer (1996, 2005), Moen (1997), Acemoğlu and Shimer (1999), posits that workers respond to deviations by firms so that they are indifferent between applying anywhere. In our framework this approach yields identical result, but it is less appealing due to the complexity of specifying indifferences over sets of wages. Peters (2000) and Burdett, Shi, and Wright (2001), on the other hand, solve for the subgame perfect Nash equilibrium of the finite model and then take the limit of that equilibrium as the number of agents goes to infinity. While arguably the correct (or most reasonable) approach, with multiple applications this is intractable because the probability of success of each application is correlated (see Albrecht, Gautier, Tan, and Vroman (2004)).

Last, we should note that the refinement described above is *not* the trembling hand equilibrium refinement since we do not take the limit of equilibria of perturbed games. However, we prove at the end of the next section that the equilibria we find *are* indeed trembling hand perfect (or, the equivalent in a continuum economy). The reason we do not work directly with that, more common, refinement is that characterizing a perturbed equilibrium involves some complicated technicalities that do not add to the exposition and are hence relegated to the appendix.
We can now define an equilibrium, given a distribution with full support $\tilde{F}$.

Definition 1 *An equilibrium is a set of strategies* $\{F, G\}$ *such that*

1. $\pi(w) \geq \pi(w')$ *for all* $w \in \mathcal{W}$ *and* $w' \in [0,1]$.

2. $U(\mathbf{w}) \geq U(\mathbf{w}')$ *for all* $\mathbf{w} \in suppG(F)$ *and* $\mathbf{w}' \in [0,1]^N$.

The first condition captures the profit maximization by firms and the second one ensures that workers best respond.

We now state the main theorem of this paper.

Theorem 1.1.1 *An equilibrium always exists and it is unique when* $N = 2$. N *different wages are posted by firms and every worker sends one application to each distinct wage. The number of firms that post a given wage is decreasing with the wage. The equilibria are not constrained efficient.*

1.1.3 A Preliminary Result

The next lemma will be useful in the following sections. Let $\underline{w}$ be the lowest posted wage that receives some applications with positive probability, i.e. $\underline{w} = \inf\{w \in \mathcal{W} | \lambda(w) > 0\}$.

Lemma 1 *Given any distribution of posted wages, worker optimization implies that* $\lambda(w)$ *is continuous and strictly increasing on* $(\underline{w}, 1] \cap \mathcal{W}$.

Proof: Recall that the probability of getting a job is given by $p(w) = (1 - e^{-\lambda(w)})/\lambda(w)$ for $w \in \mathcal{W}$. If $\lambda(w)$ is not strictly increasing there exist $w, w' \in \mathcal{W}$ such that $w > w', p(w) \geq p(w')$, and $\lambda(w') > 0$. A worker who applies to w' with positive probability can profitably deviate by switching to w since the wage is higher and the probability of getting an offer is at least as high. Therefore $\lambda(w)$ is strictly increasing above any posted wage that has a positive expected queue length, and hence on $(\underline{w}, 1] \cap \mathcal{W}$. Suppose that $\lambda(w)$ is discontinuous at some $\hat{w} \in [\underline{w}, 1] \cap \mathcal{W}$. Then the probability of getting a job offer is also discontinuous at $\hat{w}$ and a worker applying in a neighborhood of that wage has an obvious profitable deviation. *QED*

The properties described in the lemma are very natural. The expected number of applicants increases with the wage that a firm posts, which also implies that the probability of getting an offer for that job is strictly decreasing. $\lambda(w)$ is continuous because the workers' best response to the offered wages 'smooths out' any discontinuities of F: even if a positive measure of firms post a particular wage, the workers respond by sending a positive measure of their applications to that wage and hence the queue length does not jump. Furthermore, these results hold for any perturbation and hence they hold for the unperturbed game as well. This means that the queue length that a firm expects is continuously increasing in the wage, regardless of whether that wage is posted or not. Last, *any* noise distribution with full support leads to monotonicity and continuity which are the main points of the lemma.

1.2 A Special Case: $N = 2$

We now look at the special case where workers send only two applications which provides many of the main insights. The case of a general N is discussed in the next section. We start by solving for the best response of workers given an arbitrary distribution of posted wages. We then characterize the wages that firms post. Finally, the existence and uniqueness of equilibrium is proved.

1.2.1 Worker Optimization

We first find the best response of workers for an arbitrary distribution of posted wages. The posted distribution could be the result of a perturbation but in that case the subscript ϵ is omitted to keep notation simple. When a worker decides where to apply he faces a menu of wage and probability pairs from which to choose. The queue length, and hence the probability of success, is determined by the distribution of posted wages, F, and the strategy that other workers use to apply for jobs, $G(F)$. The worker solves

$$\max_{(w_2,w_1)\in[0,1]^2} \; p(w_2)\; w_2 + (1 - p(w_2))\; p(w_1)\; w_1, \tag{1.5}$$

where $w_2 \geq w_1$ by convention. Differentiability of $p(w)$ is not guaranteed so the problem cannot be solved by taking the first order conditions. We show that we can separately evaluate the low wage application and then solve for the high wage application even though this is still a simultaneous choice problem. That is, the problem admits a convenient recursive solution.

The low wage application is exercised only if w_2 fails, which means that the optimal choice for w_1 has to solve

$$\max_{w \in [0,1]} \; p(w) \; w. \tag{1.6}$$

Let u_1 denote this maximum value. Given that a worker sends his low wage application to a particular w_1 that solves (1.6), his optimal choice for the high wage application is a solution to

$$\max_{w \geq w_1} \; p(w) \; w + (1 - p(w)) \; u_1. \tag{1.7}$$

Let u_2 denote the highest utility a worker can receive when sending two applications. Since every worker optimizes, all wages receiving low applications offer the same value u_1, and all pairs of wages where workers apply give the same total utility u_2.[13] Furthermore, no one applies to wages below u_1 because the value of these openings is too low.

The next step is to show that the two problems can actually be solved independently of each other. Let $\bar{w}$ be the highest wage that offers u_1, i.e. $\bar{w} = \max\{w \in \mathcal{W} | p(w) \; w = u_1\}$.[14] The first proposition follows.

Proposition 1 *Given any distribution of posted wages, workers optimize only if $w_1 \leq \bar{w} \leq w_2$ holds for every pair (w_1, w_2) where they apply.*

Proof: Suppose this is not true. Since $w_1 \leq w_2$ the only other possibilities are $\bar{w} < w_1$ or $w_2 < \bar{w}$. By construction $w_1 > \bar{w}$ implies that $p(w_1) \; w_1 < u_1$ which cannot be optimal. If $w_2 < \bar{w}$ then a worker can deviate and send his high wage application to $\bar{w}$ instead of w_2. This deviation is profitable because

$$\begin{aligned} p(\bar{w}) \; \bar{w} \quad &+ \quad (1 - p(\bar{w})) \; p(w_1) \; w_1 - [p(w_2) \; w_2 + (1 - p(w_2)) \; p(w_1) \; w_1] = \\ [p(\bar{w}) \; \bar{w} \quad &- \quad p(w_2) \; w_2] + [p(w_2) - p(\bar{w})] \; p(w_1) \; w_1 > 0. \end{aligned} \tag{1.8}$$

[13] It is not hard to see that a pair of wages is a solution to (1.5) if and only if it solves (1.6) and (1.7).

[14] The maximum is well defined since $\lambda(w)$ is continuous and $\mathcal{W}$ is a closed set.

The first term of (1.8) is non-negative since $\bar{w}$ provides the highest possible value by definition. The second term is strictly positive because $\bar{w} > w_2 \Rightarrow p(\bar{w}) < p(w_2)$, by lemma 1. *QED*

This result has several implications. The workers are indifferent about which combination of posted wages they apply to as long as they are on opposite sides of $\bar{w}$. As a result, all wages below $\bar{w}$ (and above u_1) offer the same value, u_1, since every worker sends his low application there; similarly, all wages above $\bar{w}$ offer u_2 when paired with a low wage. These results hold for any perturbed distribution of wages and hence they hold in the limit as $\epsilon \to 0$. Recalling that $\lambda(w)$ is continuous in w and that $p(w) = (1 - e^{-\lambda(w)})/\lambda(w)$, the following conditions uniquely define the queue length:[15]

$$p(w)\ w \quad = \quad u_1,\ \forall\ w \in [u_1, \bar{w}] \tag{1.9}$$

$$p(w)\ w \quad + \quad (1 - p(w))\ u_1 = u_2,\ \forall\ w \in [\bar{w}, 1]. \tag{1.10}$$

These observations are illustrated in figure 1.1. The high indifference curve, IC-H, traces the wage and queue length pairs where workers are willing to send a high wage application, while IC-L is the indifference curve for the low wage applications. The two curves intersect at $\bar{w}$ where workers are indifferent about whether they apply with a 'high' or a 'low' application. The equilibrium queue length for any wage is given by the upper envelope of the two indifference curves: if $\lambda(w)$ is any lower at some wage w, then workers at other wages could apply to w and move to a higher indifference curve (note that utility increases in the southeast direction). In other words, the queue length is 'bid up' to IC-H for $w > \bar{w}$ and IC-L for $w < \bar{w}$. Hence the dashed line is the indifference curve that firms anticipate.

Note that while the *total* utility of any pair of wages is always equal to u_2, wages that are strictly above $\bar{w}$ give value that is *strictly lower* than u_1. Workers nevertheless apply there which may appear to be counterintuitive at first sight: if a worker can apply to wages that offer value u_1, why would he choose some wage with a strictly lower individual value? The answer is that the return to failure in the high wage application is not zero: it is equal to the value that the next application brings in, as can be seen in equation (1.10). As a result, when the worker chooses where to send his high wage application he faces a tradeoff between the value that he can get from that particular application and the probability of

[15] We previously assumed that $p(w) = 0$ for $w \notin \mathcal{W}$. Here we are implicitly defining $p(w) = (1 - e^{-\lambda(w)})/\lambda(w)$ for all w, which simplifies the notation of the firm's problem and does not change the workers' behavior since workers are indifferent between wages under this specification.

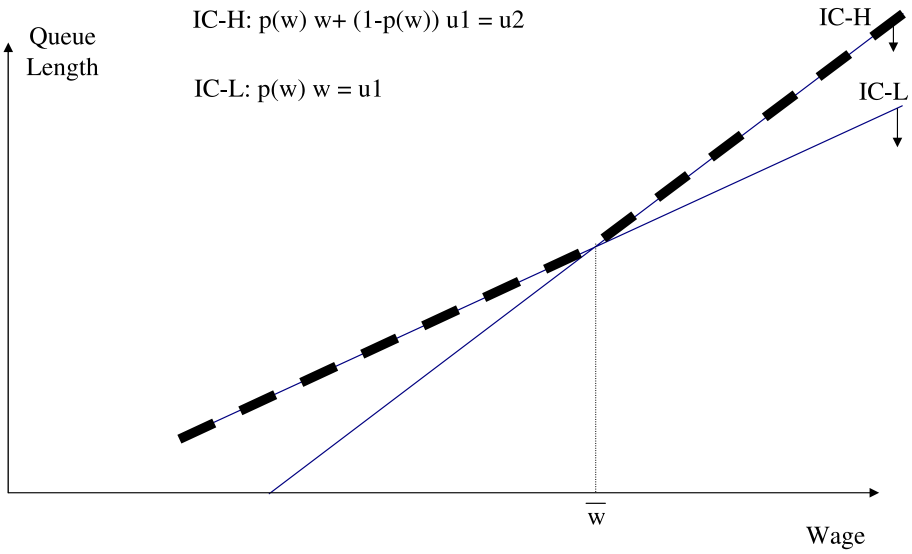

Figure 1.1: Workers' application behavior. IC-H and IC-L are the workers' indifference curves for 'high' and 'low' applications. Linearity is only used for illustration.

exercising his fallback option, i.e. the low wage application. Since the low wage provides with insurance against the possible failure of w_2, it is profitable for the worker to try a risky application that has high returns conditional on success (i.e., the wage is high) and also offers a high chance of continuing to the next application. Therefore, the low wage application goes to a relatively 'safe' region and the high application is sent to a 'risky' part of the wage distribution.[16]

The next result proves that any equilibrium exhibits wage dispersion.

Proposition 2 *There does not exist an equilibrium in which only one wage is posted.*

Proof: See the appendix. QED

The main intuition of the proof is straightforward. When a single wage is posted, workers are indifferent about which firm to work for and hence they randomize when receiving multiple job offers. A deviant firm that posts a slightly higher wage hires its preferred applicant

[16]This is an important difference between our paper and other papers on directed search with wage dispersion in which the value of sending an application is always the same for identical workers. They restrict attention to one application, which assumes away any active portfolio choice by the worker which drives our results.

for sure even when that worker receives other offers (the deviant firm also has a slightly higher expected queue length). This deviation raises profits since the increase in the hiring probability is discrete, while the increase in the wage can be arbitrarily small. Note that it is the ex post competition among firms that precludes the possibility of a single wage equilibrium.

1.2.2 Characterization of Firm Optimization

We now turn to the analysis of the first stage of the model. We prove that exactly two wages are posted in equilibrium and we characterize them.

When posting a wage, firms solve

$$\max_{w\in[0,1]} \; q(w)\; (1-\psi(w))\; (1-w), \tag{1.11}$$

taking the equilibrium objects $\{\bar{w}, u_1, u_2\}$ as given. The probability that a firm receives at least one applicant, $q(w)$, depends on the average queue length according to $q(w) = 1-e^{-\lambda(w)}$. Whether a wage is above or below the cutoff $\bar{w}$ determines the type of application it receives (high or low). This helps evaluate the probability of losing a worker after making an offer, $\psi(w)$. We label the firms that attract high (low) wage applications as high (low) wage firms. The next proposition states the result of the maximization which is proved in the appendix. A discussion follows to provide intuition about the main points.

Proposition 3 *In equilibrium, all high wage firms post $\bar{w}$ and all low wage firms post $\hat{w}_1 \in (u_1, \bar{w})$ which is derived by the first order conditions.*

Proof: See the appendix. *QED*

The reason why all firms of a particular type post the same wage is not surprising: conditional on attracting a particular type of applications, firms compete with each other in the same way as in the one application case (e.g. Burdett, Shi, and Wright (2001)), subject to some additional boundary conditions. As a result there is a unique solution to each of their profit maximization problems and two distinct wages are posted, (w_1^*, w_2^*).

To examine this in some more detail note that workers never reject an offer by a high wage firm since these firms are the applicants' best alternative. Therefore $\psi(w) = 0$ and the maximization problem of high wage firms is given by

$$\max_{w\in[\bar{w},1]} [1 \quad - \quad e^{-\lambda(w)}]\ (1-w) \tag{1.12}$$

$$\text{s.t. } \ p(w)\ w \quad + \quad (1-p(w))\ u_1 = u_2. \tag{1.13}$$

When profits are equalized across the two types of firms, the point of tangency between the isoprofit curve of the high wage firms and the high indifference curve of workers, $\hat{w}_2$, always occurs at a wage which is below $\bar{w}$, as illustrated in figure 1.2. This means that in any equilibrium the high wage firms would like to post as close to $\hat{w}_2$ as possible without moving in the low application area, and therefore they post at their lower boundary and $w_2^* = \bar{w}$.

The retention probability of low wage firms can now be calculated. When a low wage firm makes a job offer, it loses its applicant only if he is successful in his high wage application which occurs with probability $p(\bar{w})$. As a result, low wage firms solve

$$\max_{w\in[0,\bar{w}]} [1-e^{-\lambda(w)}]\ [1 \quad - \quad p(\bar{w})]\ (1-w) \tag{1.14}$$

$$\text{s.t. } p(w)\ w \quad = \quad u_1. \tag{1.15}$$

Since the retention probability enters the maximization problem as a constant, it has no marginal effect on the choice of low wage firms. Proposition (2) ensures that in equilibrium low wage firms cannot be posting $\bar{w}$. As a result their profit maximizing wage occurs at the point of tangency between their isoprofit curve and the low indifference curve of workers, i.e. $w_1^* = \hat{w}_1$. Last, note that figure 1.2 does not include the retention probability $\psi(\cdot)$ which is constant but different for the two types, and the difference in $\psi(\cdot)$ allows for equal profits even though the isoprofit curves do not intersect.

It is now easy to see that the density of posted wages is falling. Each wage receives one application per worker so $\lambda(w_i^*) = b/d_i$ where d_i is the fraction of firms posting w_i^*. $d_1 > d_2$ follows from the fact that the queue length is increasing with the wage rate. This result is driven by the fact that workers 'want' their high wage application to be risky (or, the queue length to be high). If this is not the case, a worker would be better off by not applying to the low wage and instead sending both applications to the high wage.

1.2.3 Existence and Uniqueness of Equilibrium

Turning to the existence of equilibrium, we need to find the 'correct' fraction of firms to post each wage so that profits are equalized across types of firms. More formally, an equilibrium

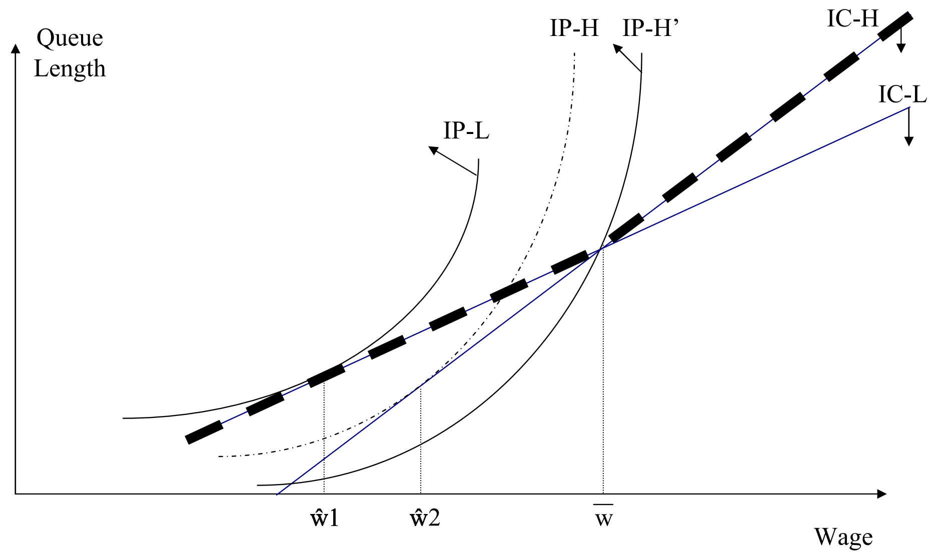

Figure 1.2: Firms' equilibrium behavior. IP-H and IP-L are the isoprofit curves for high and low wage firms.

exists if there is $\{d_1, d_2\}$ such that $d_1 + d_2 = 1$ and there is no profitable deviation when w_i^* is posted by d_i firms. Furthermore, the equilibrium is unique when there is a single pair of d_i's that satisfies the two conditions above.

Proposition 4 *An equilibrium exists and it is unique.*

Proof: See the appendix. QED

At this point it should be remarked that the full support of $\tilde{F}$ is the only property of the trembling distribution that is used in solving the model. As a result, the unique equilibrium that was constructed survives any choice of $\tilde{F}$. Furthermore, using the trebling hand refinement leads to the identical equilibrium set, as the following proposition proves.

Proposition 5 *The equilibrium is the unique trembling hand perfect equilibrium of this game.*

Proof: See the appendix. QED

1.3 The General Case: $N \geq 2$

We turn to the model with a general N. The analysis mirrors the one of section 3 and we prove that all results generalize, except for uniqueness. We provide computational evidence for uniqueness at the end of the section. Figure 1.3 illustrates the distribution of posted and received wages for an economy with equal number of workers and firms and $N = 15$. Properties of the distribution of received wages are discussed in the next section.

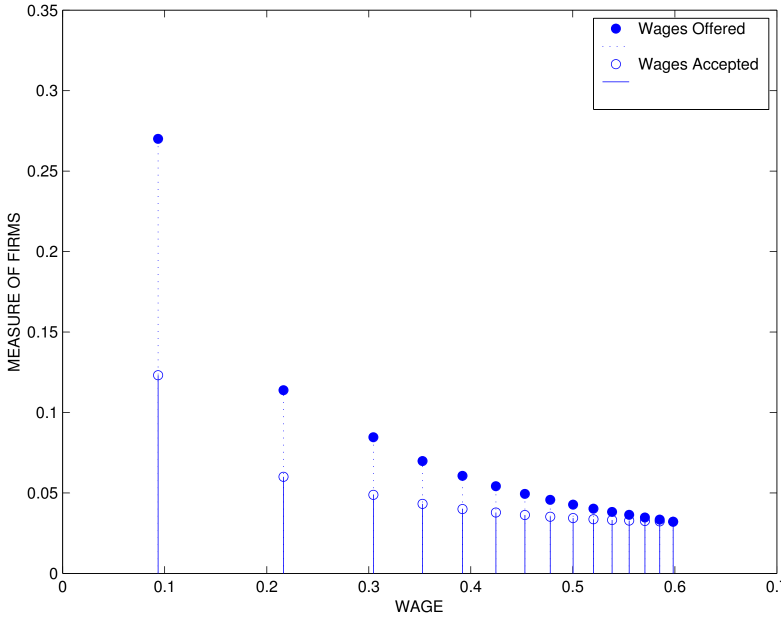

Figure 1.3: Equilibrium wage dispersion for $N = 15$ and $b = 1$.

1.3.1 Worker Optimization

Let $\mathcal{W}_i$ be the support of w_i for all i, i.e. $\mathcal{W}_i$ is the set of wages that receive the ith application of workers. As before, the utility of the lowest i applications has to be the same in any N-tuple of wages which defines the following recursive relationship

$$u_i = p(w_i)\ w_i + (1 - p(w_i))\ u_{i-1},\ \forall\ w_i \in \mathcal{W}_i,\ \ i \in \{1, 2, ...N\}, \tag{1.16}$$

where $u_0 \equiv 0$. Note that $u_i > u_{i-1}$ since $w_i \geq w_{i-1}$. Moreover, u_i is the highest possible utility a worker can get from i applications when his fallback option is u_{i-1}. Let $\bar{w}_i$ be the highest wage that provides with total utility equal to u_i when the fallback option is u_{i-1}, i.e $\bar{w}_i = \max\{w | p(w)\ w + (1 - p(w))\ u_{i-1} = u_i\}$. Also, let $\bar{w}_0$ be the lowest wage that receives applications with positive probability. We now generalize proposition (1).

Proposition 6 *When workers send N applications optimally, $w \in \mathcal{W}_i$ implies that $w \in [\bar{w}_{i-1}, \bar{w}_i]$ for $i \in \{1, 2, ...N\}$.*

Proof: The proof is by induction. It is sufficient to show that the following property holds for all i: $w < \bar{w}_i \Rightarrow w \notin \mathcal{W}_k$ for $k \geq i+1$. The initial step for $i = 1$ was proved in the previous section, where $\bar{w}_1 = \bar{w}$. We assume that the property holds for $i-1$ and show that a contradiction is reached if it does not hold for i. In other words, if $w < \bar{w}_{i-1} \Rightarrow w \in \mathcal{W}_{i-1}$ holds, then there is no $\tilde{w} \in \mathcal{W}_{i+1}$ such that $\tilde{w} < \bar{w}_i$ (if $\tilde{w} \in W_k$ for $k > i+1$ the same argument goes through). Define $v(w, u_{i-1}) = p(w)\ w + (1 - p(w))\ u_{i-1}$ to be the utility of applying to a particular wage w when the fallback option is u_{i-1}. We want to show that $v(\bar{w}_i, u_i) > v(\tilde{w}, u_i)$ for all $\tilde{w} < \bar{w}_i$. Note that

$$\begin{aligned} v(\tilde{w}, u_{i-1}) &= p(\tilde{w})\ \tilde{w} + (1 - p(\tilde{w}))\ u_{i-1} \\ &\leq p(\bar{w}_i)\ \bar{w}_i + (1 - p(\bar{w}_i))\ u_{i-1} = v(\bar{w}_i, u_{i-1}), \end{aligned}$$

since the second line is the optimal choice when u_{i-1} is the fallback option and hence it provides with the maximum level of utility. Replacing u_{i-1} with u_i in both lines above we get the terms to be compared. Since $\bar{w}_i > \tilde{w} \Rightarrow (1 - p(\bar{w}_i)) > (1 - p(\tilde{w}))$ the second term increases by more and the inequality becomes strict which proves the result. QED

An implication of the proposition is that when a worker applies to a firm of type i he receives the posted wage w if he is successful in his application or u_{i-1} if he is unsuccessful. Therefore the queue lengths facing the firms attracting the ith application are given by the following equation:

$$p(w)\ w + (1 - p(w))\ u_{i-1} = u_i, \quad \forall\ w \in [\bar{w}_{i-1}, \bar{w}_i] \tag{1.17}$$

which is a straight generalization of equations (1.9) and (1.10).

1.3.2 Firm Optimization

We now turn to the first stage of the model. For the remainder of the paper firms that receive the ith lowest application of workers are referred to as *type i firms*. The profit maximization problem of each type of firm is solved and profits are then equalized across types.

When posting a wage, firms take as given the cutoffs $\{\bar{w}_k\}_{k=0}^N$ and the equilibrium utility levels $\{u_k\}_{k=1}^N$, which determine the utility provided to workers for their lowest k applications.

A firm of type i solves the following profit maximization problem:

$$\max_{w \in [\bar{w}_{i-1}, \bar{w}_i]} \; q(w) \; [1 - \psi(w)] \; (1 - w), \tag{1.18}$$

where the queue lengths are determined by equations (1.17).

Proposition 7 *In equilibrium, all type i firms post the same wage $w_i^* = \bar{w}_{i-1}$ for $i \geq 2$. All type 1 firms post $\hat{w}_1$ which it is determined by the first order conditions.*

Proof: See the appendix. QED

The logic of the proof is similar to the one of proposition (3). The solution to the problem of type N firms is shown to be $\bar{w}_{N-1}$. This means that $\psi(w) = p(\bar{w}_{N-1})$ for type $N-1$ firms and the solution to their profit maximizing problem is $\bar{w}_{N-2}$. This implies that $\psi(w) = (1 - p(\bar{w}_{N-1}))\; (1 - p(\bar{w}_{N-2}))$ for type $N-3$ firms and so on. In general, the retention probability of a type i firms is $1 - \psi(w) = \prod_{n=i+1}^{N} (1 - p(w_n^*)) \equiv 1 - \psi_i$. Given ψ_i, the maximization problem for a type i firm becomes

$$\max_{w \in [\bar{w}_{i-1}, \bar{w}_i]} \; q(w) \; (1 \; - \; \psi_i) \; (1 - w) \tag{1.19}$$

$$\text{s.t.} \quad p(w) \; w + (1 \; - \; p(w)) \; u_{i-1} = u_i, \tag{1.20}$$

and the solution lies at the lower boundary for all $i \geq 2$. Finally, it should be noted that the density of posted wages is falling for the same reasons as in section 3.

1.3.3 Existence and Uniqueness of Equilibrium

The next proposition establishes the existence of an equilibrium. We then provide some computational evidence for uniqueness. As in the previous section, we show that there is a sequence $\{d_1, d_2, ..., d_N\}$ such that $d_1 + d_2 + ... + d_N = 1$ and there is no profitable deviation when wage w_i^* is posted by exactly d_i firms.

Proposition 8 *An equilibrium exists for any N.*

Proof: See the appendix. QED

While existence of an equilibrium is assured, uniqueness is not. In particular, there may be more than one sequence of d_is that satisfies the above conditions. In the event of multiplicity, all equilibria have the same qualitative characteristics (e.g. $d_1 > d_2 > ... > d_N$) but the actual wages that are posted are different.

To perform the computational exercise that gives evidence of uniqueness we need some additional notation. Given an arbitrary number of type one firms d_1 and some worker-firm ratio b, we can find unique d_2, d_3,..., d_n such that profits between firms is equalized when d_i firms offer wage w_i. For consistency, these fractions have to add to one in any equilibrium. That is, the sum $S_b(d_1) := d_1 + d_2 + ... + d_N$ has to equal one. Multiplicity can only occur if there exists $d_1' \neq d_1$ with $S_b(d_1) = S_b(d_1') = 1$. Clearly, if the sum $S_b(d_1)$ is strictly increasing in d_1, such a multiplicity is not possible. What might be more surprising is the fact that $S_b(d_1)$ is strictly increasing implies also that $S_{b'}(d_1)$ is strictly increasing for any $b' > 0$, i.e. for any market tightness the equilibrium is unique. Based on this we can show

Lemma 2 *The equilibrium is unique for any $b > 0$ if $S_{b^*}(d_1)$ is strictly increasing in d_1 for some b^*.*

Proof: See the appendix. *Q.E.D.*

Figure 1.4 shows that $S_b(d_1)$ is strictly increasing for $b = 1$ under various N. Graphs for other N look similar, which suggests that the equilibrium is unique.

1.4 Further Equilibrium Properties

In this section we investigate the efficiency properties of the matching process and the empirical distribution implied by the model.

1.4.1 Efficiency

The criterion for constrained efficiency is whether the output (equivalently, the number of matches) is maximized conditional on the matching frictions, given the worker-firm ratio b. The main result is that constrained efficiency does not obtain.

To maximize the number of matches it is convenient to look at the probability that a worker finds a job.[17] It was shown in the earlier sections that in equilibrium workers send each

[17] Alternatively we could calculate the probability that a firm hires. This is somewhat more complicated though, because there are two different ways for the firm to remain vacant: it may receive no applications *or* its chosen applicant may accept a different job. This is also the reason why we cannot apply standard proofs

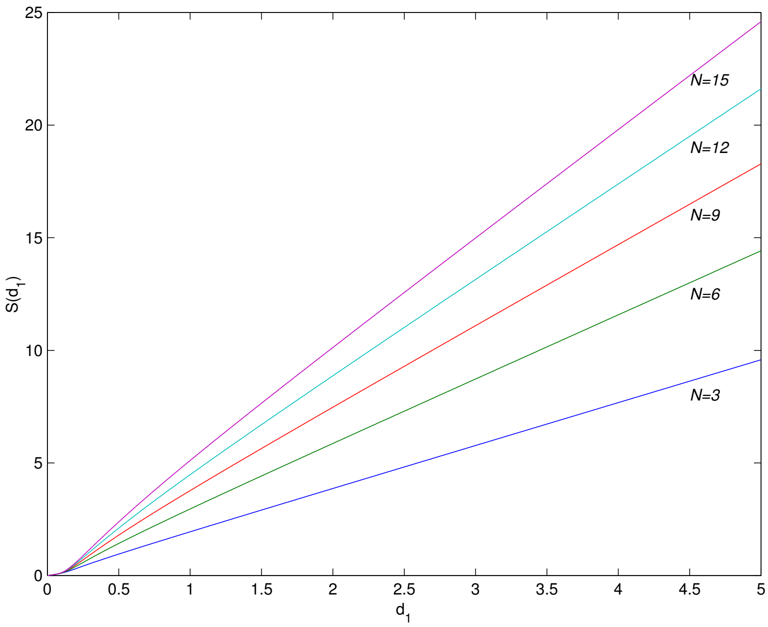

Figure 1.4: Sum of firms for $b = 1$ and various N.

of their N applications to a different group of firms, which was identified by its distinct wage. Since wages are irrelevant for efficiency purposes we label the firms posting w_i as group i. Also, we simplify notation by having $p_i = p(w_i)$ and $\lambda_i = \lambda(w_i) = b/d_i$. Letting $\mathbf{d} = (d_1, d_2, ..., d_N)$ be the vector of the fraction of firms within each group, the total number of matches is given by $b\ m(\mathbf{d})$ where $m(\mathbf{d}) \equiv 1 - \prod_{i=1}^{N} (1 - p_i)$ is the probability that a particular worker receives a job offer. The planner has to decide how many firms to allocate to each group in order to maximize output or, equivalently, to maximize $m(\mathbf{d})$.

An immediate necessary condition for optimality (which fails) is that the probability of a match cannot be increased by reallocating firms between any two groups. This condition follows directly from observing that $m(\mathbf{d}) = 1 - (1 - p_k)\ (1 - p_l)\ \prod_{i \neq k,l} (1 - p_i)$, given any two groups of firms, k and l. Therefore, an equilibrium is constrained efficient only if d_k and d_l minimize $(1 - p_k)\ (1 - p_l)$, which is equivalent to

$$\max_{d_k, d_l \geq 0} (p_k + p_l - p_k\, p_l) \tag{1.21}$$
$$\text{s.t.}\ \ d_k + d_l = 1 - \sum_{i \neq k,l} d_i.$$

that rely on minimizing the number of firms that do not receive applications as this neglects the second aspect.

This problem is identical to the case of two applications where the worker-firm ratio is given by $b/(1-\sum_{i\neq k,l} d_i)$. Therefore, we consider the $N=2$ case for an arbitrary b, letting d be the fraction of firms in the first group and $1-d$ the fraction in the second group. The planner has to decide the optimal value of d.

Proposition 9 *When $N = 2$ the number of matches is maximized only if $d = 1/2$ or $d \in \{0, 1\}$.*

Proof: See the appendix. *QED*

The proposition shows that it may be optimal for workers to send only *one* application due to congestion. In that event the planner's solution is to place all firms in one group ($d \in \{0, 1\}$). If it is optimal to send two applications, then the number of firms should be equal in both groups. As a result, all groups should have equal size when N applications are sent. However, we know that in equilibrium the number of firms posting the lower wages is larger and hence this efficiency condition is never met. Moreover, since the lack of efficiency arises from the matching process it carries over even if the number of applications is endogenized or if the ratio of workers to firms is determined by free entry subject to a fixed cost.

It is worthwhile to mention that efficiency *does* obtain in the usual directed search environment with one application. The reason is that firms can price the arrival rate (in essence, the queue length) of workers through the wages they post.[18] When workers send multiple applications firms care about the probability of retaining a worker, as well as the arrival rate of applicants. The arrival rate can still be priced using the posted wage, but the probability of retaining a worker does not depend on how many applications a firm has received: if a firm's chosen applicant has a better offer, the firm remains idle regardless of how many other workers it attracted. Therefore, the arrival rate of applicants does not change the probability of hiring at the second stage, once at least one worker has applied. Since the firm can only influence the arrival rate of workers but not the retention probability, it cannot fully price its hiring probability and hence efficiency does not obtain.

In preview of chapter 2 we already note that constrained efficiency is restored when firms can recall all the applicants they receive, in an otherwise similar model. In that environment,

[18]See Mortensen and Wright (2002) for a discussion.

the second phase of the hiring process also depends on the queue length since a firm can offer the job to all the applicants it receives, until one of them accepts (or all of them reject it). Nevertheless, if firms can only recall up to a certain (finite) number of applicants, the queue length will only partially influence the retention probability. Hence it is our conjecture that efficiency fails when recall is imperfect. Therefore, we believe that our inefficiency result can be seen as a general feature of limited recall.

1.4.2 The Empirical Distribution

As already noted, the density of posted wages is decreasing. The density of *received* wages, however, need not be decreasing as higher wages are accepted more often. The following proposition summarizes our findings.

Proposition 10 *The distribution of accepted wages is decreasing when the ratio of workers to firms is large enough.*

Proof: See the appendix. QED

1.5 Extensions

The main insights developed above carry over when we allow for free entry, for endogenous decisions concerning the number of applications, and for a dynamic labor market interaction. This section discusses each case in turn.

1.5.1 Free Entry

Consider a large number of potential firms, each of which can pay a fixed cost $K < 1$ to enter the labor market. The number of applications that each worker sends is fixed at N. Let $\Pi(b)$ denote the equilibrium profits of firms when the worker-firm ratio is b. If there are multiple equilibria, this object is a correspondence. It is easy to see that $\lim_{b\to\infty}\Pi(b) = 1, \lim_{b\to 0}\Pi(b) = 0$, and that $\Pi(b)$ is continuous in its argument (or, upper-hemicontinuous if it is a correspondence). Therefore, there is some $b^* > 0$ such that the equilibrium profits are exactly equal to K. When the equilibrium is unique, b^* is also unique.

1.5.2 Endogenous Number of Applications

We introduce a cost per application c and endogenize the number of applications that a worker sends. As earlier, attention is restricted to symmetric equilibria where every worker sends the same number of applications in expectation. Two separate issues are investigated. First, it is shown that the equilibria described in the previous sections are robust to the introduction of application costs. We then discuss the equilibria that can arise for an arbitrary value of c.

To analyze the first issue, recall that u_i is the maximum payoff a worker receives when applying i times. To determine the marginal benefit of the ith application note that in equilibrium for $i \geq 2$

$$u_i = p_i \, w_i^* + (1 - p_i) \, u_{i-1} \tag{1.22}$$

$$u_{i-1} = p_i \, w_i^* + (1 - p_i) \, u_{i-2}, \tag{1.23}$$

where the first expression holds by the definition of u_i and the second holds because $w_i^* = \bar{w}_{i-1}$ and hence $w_i^* \in \mathcal{W}_{i-1}$. Subtracting (1.23) from (1.22), the marginal benefit of the ith application is given by $u_i - u_{i-1} = (1 - p_i) \, (u_{i-1} - u_{i-2}) = \prod_{j=2}^{i}(1 - p_j) \, u_1$. Clearly, the marginal benefit of an additional application is decreasing in i and therefore $u_N - u_{N-1} > c$ is a sufficient condition for workers to send at least N applications. Moreover, since $u_N - u_{N-1}$ is strictly positive, the equilibrium does not unravel with the introduction of small costs of search.[19]

The next step is to ensure that no worker applies more than N times. It is easy to see that a worker who contemplates sending $N + 1$ applications will send his additional application to the highest wage, w_N^*. His utility from applying $N + 1$ times is therefore given by $u_{N+1} = p_N \, w_N^* + (1 - p_N) \, u_N$ which means that the marginal benefit of the extra application is $u_{N+1} - u_N = (1 - p_N) \, (u_N - u_{N-1})$. As a result, if that quantity is lower than c the worker sends at most N applications. Summing up, an equilibrium where workers apply exactly N times can be supported when the cost parameter lies in the set $[(1 - p_N) \, (u_N - u_{N-1}), u_N - u_{N-1}]$.

We now turn to the case of determining N for an arbitrary c. One possibility is that all workers send the same number of applications and they have no incentive to deviate. However, it is also possible that there is no pure strategy equilibrium in the number of applications

[19]This is not the case in other labor models, e.g. Albrecht and Axell (1984).

that workers send. In particular, a worker may prefer to send $N+1$ applications when all other workers apply N times, while preferring to send N applications when everyone applies $N+1$ times. As a result, an equilibrium in the (now endogenous) number of applications has to involve some randomization in the number of applications: some proportion of workers α applies $N+1$ times while the rest only apply N times, where α is chosen so that both types of workers receive the same ex ante utility.[20] It is worth noting that an equilibrium where workers randomize over how many times to apply looks very much like the one we have already developed. $N+1$ wages are posted and it is only the workers who send $N+1$ applications that apply to the highest wage, w^*_{N+1}. The characterization of equilibria is identical to the previous sections, except for the fact the possibility that the highest wage does not lie on the lower constraint. Numerical simulations suggest that an equilibrium can be supported for any possible cost per application, though there may be multiplicity.

1.5.3 The Dynamic Version

So far we assumed that workers and firms that are not matched remain idle. We now show that the analysis can be generalized to an infinite horizon dynamic setting where agents that remain unmatched can try to match again in the following period. The labor market opens every period with firms posting wages and workers sending N applications. We restrict attention to stationary strategies so that only N and b matter in each period. Matching proceeds as in the static model and the agents that are matched leave the market. In the beginning of every period, matches formed earlier are exogenously dissolved with probability δ and the agents reenter the market. If a job survives, the surplus is split according to the wage with which the match was consummated. Workers and firms have a common discount factor β. The total number of workers is fixed and there is free entry of firms at cost K, which determines the worker-firm ratio every period. We are interested in steady state equilibria where the number of, and value to, unmatched agents remains constant from period to period.

Let $W(w)$ be the value to a worker of being employed at wage w and let L be the value of being unemployed. Furthermore, let $J(w)$ be the value to a firm of employing a worker at w and let V be the value of a vacancy. The value of a match at w for workers and firms,

[20]It is relatively straightforward to show that the number of applications that workers send in equilibrium can only be one apart. This is due to the decreasing returns of additional applications.

respectively, is then given by

$$W(w) = w + \beta\,[(1-\delta)\,W(w) + \delta\,L] \tag{1.24}$$

$$J(w) = 1 - w + \beta\,[(1-\delta)\,J(w) + \delta\,V], \tag{1.25}$$

where it is now straightforward to solve for $W(w)$ and $J(w)$ as a function of the values of being unmatched, L and V.

To determine L and V we need to consider the maximization problem of workers and firms that are unmatched at the start of the period. This is similar to the static case, except for the fact that if they do not match they can try again in the following period. Abusing notation a little, the value to a worker of applying to $\mathbf{w}$ and to a firm of posting w is

$$\begin{aligned} L(\mathbf{w}) &= p(w_N)\,W(w_N) + (1-p(w_N))\,W(w_{N-1}) + ... \\ &\quad + \prod_{i=2}^{N}(1-p(w_i))\,p(w_1)\,W(w_1) + \prod_{i=1}^{N}(1-p(w_i))\,\beta\,L \end{aligned} \tag{1.26}$$

$$V(w) = q(w)\,(1-\psi(w))\,J(w) + [1-q(w)\,(1-\psi(w))]\,V - K, \tag{1.27}$$

where $L(\mathbf{w}) = L$ for all N-tuples $\mathbf{w}$ where workers apply and $V(w) = V$ for all w posted by firms. Note that next period's values are the same to the current ones due to stationarity. Combining the two sets of equations and noting that free entry implies $V = 0$ we get

$$\begin{aligned} L(\mathbf{w}) &= \frac{1}{\kappa}\,\{p(w_N)\,w_N + ... + \prod_{i=2}^{N}(1-p(w_i))\,p(w_1)\,w_1 \\ &\quad + [\delta + (1-\delta)\,(1-\beta)\,\prod_{i=1}^{N}(1-p(w_i))]\,\beta\,L\} \\ V(w) &= \frac{1}{\kappa}\,q(w)\,(1-\psi(w))\,(1-w), \end{aligned}$$

where $\kappa = 1 - \beta\,(1-\delta)$. This problem is the same as in the static environment, with the exception that the workers have a strictly positive outside option if all of their applications fail. It is therefore easy to prove that all the characterization results of the previous sections hold and that existence is guaranteed when taking into account the (endogenous) outside option of workers. Last, the equilibrium ratio of workers to firms in the unmatched market is pinned down by the restriction on stationary environments.

1.6 Directed vs. Random Search

We have established that the density of posted wages is downward sloping, and the density of accepted wages shares this feature under certain parameter restrictions. In this section we argue that the directedness of the application process is crucial for this result. Directed search breaks the tight link between the profit margin and the distribution of wages which is what leads to an increasing wage profile in random search models. We use a random search version of our model to illustrate this point, though a similar argument can be made for other random search models such as Burdett and Mortensen (1998), Burdett and Judd (1983) and the basic version of Acemoğlu and Shimer (2000).

Consider a version of the model where search is random rather than directed. Firms post wages but they cannot communicate them to workers. Therefore, workers apply to N firms at random, firms choose one applicant to make a job offer at the posted wage, and workers with multiple offers choose which job to accept. In other words, there is ex post but not ex ante competition among firms. This environment is examined in Gautier and Moraga-Gonzáles (2005) and it leads to wage dispersion, since posting a higher wage results in hiring a worker who has additional offers with greater probability.

Denote the probability that a worker gets an offer by $\bar{p}$, and the probability that a firm has at least one applicant by $\bar{q}$. Note that since the arrival rate of workers is independent of the posted wage, these outcomes do not depend on the wage but only on b.

Let F be the distribution of posted wages and first consider the $N = 2$ case. The probability that a firm hires a worker is given by $\bar{q}\,[1-\bar{p}+\bar{p}\,F(w)]$. The first term is the probability that at least one worker applies, while the second term is the probability that this worker has no other offer *or* his other offer is for a lower wage. Equal profits imply that the following condition has to hold for all $w \in suppF$:

$$\bar{q}\ (1-\bar{p}+\bar{p}\ F(w))\ (1-w) = \Pi$$
$$\Leftrightarrow \bar{q}\ (1-\bar{p}+\bar{p}\ F(w)) = \frac{\Pi}{1-w},$$

where $1 - w$ is the margin of the firm and Π denotes the equilibrium level of profits. Note that $\Pi/(1-w)$ is a strictly convex function of w and that $F(w)$ is the only non-constant on the left hand side of the equation above. As a result, equal profits imply that the distribution of posted wages has to be strictly convex.

The intuition behind this result is that when moving upwards on the support of F, the *percentage* decrease in the margin becomes larger at an increasing rate. For profits to remain constant, this requires an equivalent increase in the probability of hiring. In a random search environment, a higher wage firm increases its hiring probability only by getting more workers who might have different offers, i.e. only the ex post competition margin improves. This means that, when moving to the top of the distribution, a firm needs to 'overtake' an increasing number of competing firms or, in other words, the distribution of wages needs to be convex. The same reasoning holds for arbitrary N, in which case the profits are given by $\bar{q}\ (1 - \bar{p} - \bar{p}\ F(w))^{N-1}\ (1 - w)$ (see Gautier and Moraga-Gonzáles (2005)).

Figure 1.5 shows the wage density under random search for $N = 15$ and equal number of workers and firms, which allows a comparison with figure 1.3 for directed search.

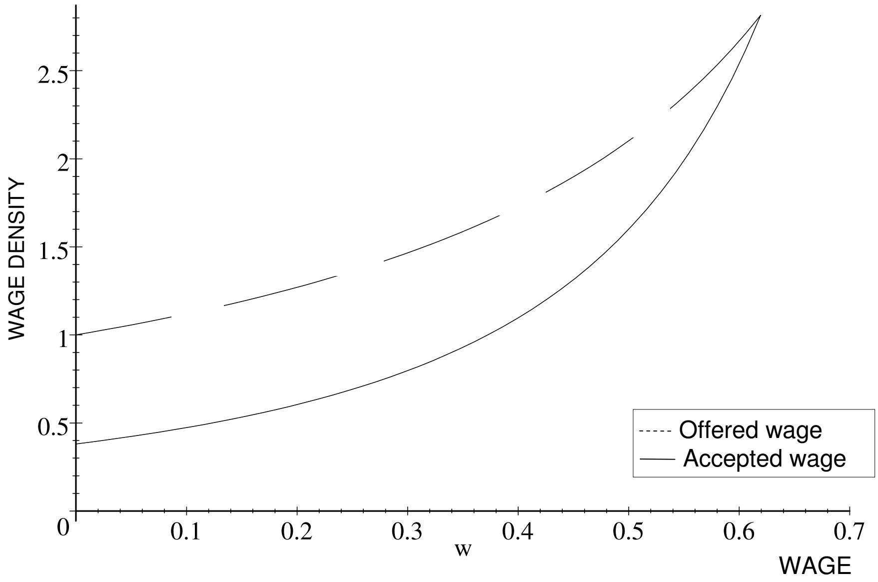

Figure 1.5: Wage densities with random search, for $N = 15$ and $b = 1$.

In a directed search environment a decreasing wage profile is possible due to the presence of ex ante competition: the ability of firms to attract more applicants by posting a higher wage gives an additional channel through which they increase the probability of hiring. Therefore, high wage firms need not 'overtake' as many of their competitors to guarantee equal profits.

1.7 Conclusion of Chapter 1

We develop a directed search model where workers apply simultaneously for N jobs. We show existence, prove uniqueness for the case of $N = 2$ and present computational evidence suggesting that the equilibrium is unique for any N. All equilibria exhibit wage dispersion, with firms posting N different wages and workers sending one application to each distinct wage. The main distinguishing feature of our model is that dispersion is driven by the portfolio choice that workers face. The matching process is a source of inefficiency because the higher paying firms enjoy higher probability of hiring a worker.

This model delivers some potentially testable predictions. In line with stylized facts, the density of posted wages is decreasing, as does the density of received wages for suitable parameter values. Firms that post high wages receive more applications per vacancy than lower-wage firms. When a firm offers its position to an applicant, the applicant does not necessarily accept, but higher wage offers are accepted with greater probability. While the aspect of wage dispersion has been repeatedly examined in the literature, the last two implications have not received much attention.

To our knowledge, the only other directed search model where workers can simultaneously apply for multiple jobs is Albrecht, Gautier, and Vroman (forthcoming). Their set-up is similar to ours, except for a crucial assumption: in their model, when two or more firms make an offer to the same applicant the potential employers engage in Bertrand competition for the worker and hence he ends up receiving the full surplus of the match. The unique equilibrium has all firms posting the reservation value of workers, with some workers receiving their marginal product due to Bertrand competition, regardless of the number of applications that workers send. The low offered wage arises because a high wage is costly if the worker only has a single offer, but it does not yield any advantage if the worker receives multiple offers. For the worker it is not that important to achieve a high paying job, but to achieve two jobs.

In our model firms commit to their wage posting. Therefore a higher wage yields an advantage even if the worker gets multiple offers, as long as the other offers are lower. Also, the worker cannot secure a high wage by applying to two low wage jobs. While ex-post bidding clearly arises in some markets, we think that our assumption of wage posting captures the operation of many real life labor markets well. In many situations Bertrand competition may not even be feasible. If alternative offers are hard to verify, workers have an incentive to invent alternative offers to raise their wage. More importantly, if firms really fear competition, they will attempt to make their offers non-verifiable to other firms to avoid this kind of Bertrand competition. We therefore pursue the assumption of posting rather than ex-post bidding,[21] and we think that the novel results warrant the analysis.

[21] In this assessment we follow the predominant opinion in the literature. See e.g. Acemoğlu and Shimer (2000), Burdett and Judd (1983), Burdett and Mortensen (1998) and Delacroix and Shi (forthcoming).

We should mention that our model can be easily extended in a number of ways. This paper shows how to incorporate entry decisions of firms and choices regarding the number of applications by workers, and discusses a discrete time infinite horizon setting, all of which retain the structure developed in the baseline model. Other potentially interesting extensions include firm and worker heterogeneity, and risk aversion. Note that the homogeneity of firms was not used when analyzing workers' optimization and, therefore, those results carry over in the case of productivity differentials among firms. The firms' optimization problem will be different, of course, and we conjecture that since more productive firms place a premium in hiring they will post higher wages. Similarly, in the case of observable worker heterogeneity each firm posts a menu of type-specific wages, and we conjecture that each type of workers has its own set of utility levels and cutoffs. Moderate risk aversion of workers can be easily accommodated in our framework by replacing w with a concave function $\nu(w)$ when specifying the worker's utility, leaving the worker's problem virtually unchanged and affecting the firms only through a modified constraint. In conclusion, we believe that this model provides a suitable structure to address further questions of economic interest.

Chapter 2

Efficiency of Simultaneous Directed Search with Recall

While unemployment is generally viewed as an undesirable phenomenon, scholars have pointed out its productive purpose in the allocation of labor in markets with frictions. The productive activity that people pursue during unemployment is usually called "search". The idea that the market achieves a natural rate of unemployment that efficiently (given the frictions) allocates the productive resources in the economy goes back at least to Friedman (1968) and Phelps (1967), and has been debated since.

Initial equilibrium models that investigated this contention are Diamond (1982), Mortensen (1982a, 1982b) and Pissarides (1984, 1985). They assumed that only a fraction of workers and firms can meet, that those meetings are random draws, and that wages are set by Nash bargaining. In general they do not support the view that the market achieves the optimal allocation given the frictions.[1] Efficiency fails essentially because cannot compete to increase their matching probability, but instead wages are determined non-competitively after meeting a worker.

The next generation of equilibrium search models allowed firms to directly compete for labor by publicly posting their wage offers.[2] Workers observe the offers and decide where to apply. Frictions arise because firms only have a single vacancy, and workers are assumed to use identical application strategies. Since it cannot be optimal for all workers to apply

[1]See also Hosios (1990). A notable exception is Lucas and Prescott (1974), who set up a very different equilibrium search model that does exhibit efficiency.

[2]For an overview of both generations of models see Rogerson, Shimer and Wright (forthcoming).

for the same job, the equilibrium requires a mixed strategy in which workers randomize and sometimes miscoordinate. This means that some jobs happen to attract many applicants, while others attract few or none. Nevertheless, higher wages induce (or "direct") workers to apply there with higher probability. In this class of directed search models, also known as competitive search models, Moen (1997), Mortensen and Wright (2002), Shi (2002) and Shimer (1996, 2005) show in various degrees of generality that the market interaction is efficient given the frictions in the market, providing theoretical support for the efficiency of the natural rate of unemployment.

These results are obtained under the restrictive assumption that each worker only sends a single application.[3] One application does not necessarily lead to employment, and workers have a natural incentive to apply to multiple firms to improve immediate employment prospects. Albrecht, Gautier and Vroman (forthcoming) and chapter 1 of this thesis introduce multiple applications per worker. Surprisingly, the competitive forces do not lead the market to constrained efficiency even under the specification of homogenous workers and firms. In the former model, entry of firms is too large compared to the level of unemployment. In the latter, wage dispersion precludes an efficient allocation of workers to firms. In both models the number of applications may be too large, i.e. decreasing the number of applications per worker would improve employment. In their main analysis both papers assume that a firm can only propose its job to one applicant, and if the applicant rejects it in favor of a better alternative, the firm remains vacant independent of the number of additional applications it received. Failure of constrained efficiency might be due to this assumption. Or it could result from an inherent externality of multiple applications that cannot be reflected in market wages and therefore distorts various margins of efficiency.

This analysis presented here has two purposes. First and most importantly, it explores an alternative assumption on the assignment of workers to firms once wages are announced and applications sent. The aim is to investigate the resulting equilibrium properties, and to analyze whether the failure of efficiency in the above models is due to the assumption about the assignment or whether it poses a deeper challenge for the efficiency in directed search

[3] In some competitive search models like Moen (1997) and Mortensen and Wright (2002) agents choose markets rather than firms. The market structure is not clearly spelled out, rather some reduced form arrival rates are assumed. Nevertheless, this implicitly limits the setup to sequential search and rules out any simultaneous choice. Shimer (1996) shows how this can be recast in a model in which agents apply to individual firms with a single application.

economies. Second, it models the number of applications as a *choice* for the worker so that the endogenous number of applications can be interpreted as a measure of search intensity, the efficiency of which can then be assessed.[4]

While the second aim requires analytical attention, the first provides a conceptual challenge. Once wages are set and applications are sent, who should work for whom? Here we interpret the applications as links in a network between workers and firms, and we assume a stable matching given the network and the announced wages. This entails that in the final matching no vacant firm has an applicant that is employed at a lower wage. This specification is based on the idea that employers can call up their applicants sequentially. If an applicant accepts, the employer is momentarily happy and stops making additional proposals. Yet if a better job is proposed to the applicant later on, he can accept the better job and reject the earlier offer. Rejected firms continue to contact ("recall") remaining applicants and propose their jobs to them. That is, we apply a version of Gale and Shapley's (1962) deferred acceptance algorithm.[5]

Apart from these additions, our model uses the standard directed search setup with homogenous workers and firms. Firms decide whether to enter the market, and if they do so they publicly post a wage commitment for their single vacancy. Workers observe these wages and decide how many applications to send and where to send them, where we retain the standard assumption of symmetric strategies that creates the market frictions. Then workers and firms are matched as explained above. In the (generically) unique equilibrium all workers choose to send the same number of applications, and the number of wages offered in equilibrium is equal to the number of applications each worker sends.

The equilibrium is constrained efficient given the workers' coordination problem. We distinguish three components of efficiency. *Search efficiency*: for a given number of applications and a given number of firms the number of matches is constrained optimal. Equilibrium wage dispersion is essential for this feature. *Entry efficiency*: the division of the match surplus is such that the constrained optimal number of firms enter. *Application efficiency*: the number of applications that workers send is constrained efficient despite the negative

[4]Both Albrecht, Gautier and Vroman (forthcoming) and the model presented in chapter 1 suggest to model search costs. Yet their is comparatively little analysis provided on it.

[5]In a finite economy the process converges in finite time and high wage firms clearly hire before lower wage firms do. For the continuum case see the appendix.

externality of an additional application on other workers. All externalities are reflected in the market wages. Finally we show that for vanishing application costs the equilibrium converges to the unconstrained efficient outcome of a frictionless Walrasian economy.

In contrast to the inefficiencies in models without recall, constrained efficiency obtains here. This is due to a commonality between workers and firms. Firms only care about applicants who do not obtain better offers. We call these applicants *effective*. Workers also only care about rival applicants that are *effective*, as the others do not compete for the job. As we will see, this commonality implies that raising the wage induces more effective applicants, and firms can "price" their applications optimally. Without recall, firms still only care about effective applicants but workers care about *all* rival applicants because any applicant who receives a job proposal precludes others from obtaining the job (even if he rejects it and in the end works at another firm). This lack of commonality between workers and firms prevents efficient pricing. A wage raise induces more applications, but potentially only from people that have an easier time getting other jobs, which can even mean less *effective* applicants. We discuss the different implications with and without recall in section 2.5.2. Note that this issue has not occurred in the prior literature because a single application allows the worker no alternative, and all applications are effective by assumption.

Our analysis also shows that wage dispersion is not merely a sign of frictions, but rather an optimal response to these frictions (even though agents are homogenous).[6] The constrained efficient allocation in the market requires different hiring probabilities among firms, as different hiring probabilities can reduce those instances in which one worker does not get a job because it is occupied by another worker while this other worker could take another job elsewhere. The market wages internalize this externality, and preferred jobs are endogenously harder get than back-up (non-preferred) jobs.[7] The paper also adds to the literature on asymptotic efficiency of search markets by showing convergence to the unconstrained

---

[6]For homogenous workers and firms such an efficiency role is novel. Models in this area include Acemoğlu and Shimer (2000); Albrecht, Gautier and Vroman (forthcoming); Burdett and Judd (1983); Burdett and Mortensen (1998); Butters (1977); Delacroix and Shi (2005); Gautier and Moraga-González (2005). See also Chapter 1 of this thesis.

[7]From the firms' point of view one can interpret wage dispersion as different prices for a differentiated good, despite the homogeneity of labor. Wage announcements "buy" applications, not labor, and higher priority applications (those where the worker sends his other applications to less attractive firms) are worth more to a firm than low priory applications (where the other applications are sent to more attractive firms).

efficient outcome in a simultaneous search environment.[8]

To my knowledge this is the first attempt to integrate the two-sided strategic considerations of a frictional search environment with stability concepts used in matching markets.[9] The paper draws on three strands of literature. We use insights from the directed search literature (e.g. Burdett, Shi and Wright (2001)) to model the frictions and information flows in the market. With multiple applications workers face a simultaneous portfolio choice. For this type of problem Chade and Smith (2004) consider an individual agent's choice and chapter 1 of this thesis derives implications in an equilibrium framework. To model recall we apply insights from the two-sided matching literature (Gale and Shapley, 1962) to the network that formed in the search process in earlier stages. Section 2.5 provides further discussion.

In the following, we first present the model. Section 3 then characterizes the equilibrium. Section 4 analyzes efficiency. Notation and exposition remain much more tractable when we consider at most two applications per worker, therefore sections 2 to 4 are restricted to this case. Section 5 lifts this restriction and, additionally, discusses convergence for vanishing application costs. Section 6 discusses the main modeling assumptions, additional literature, and concludes. Omitted proofs are gathered in the appendix.

2.1 The Model

2.1.1 Environment and Strategies

There is a measure 1 of workers and a large measure V of potential firms. The measure v of active firms is determined by free entry. Each active firm is capacity-constrained and can only employ a single worker, each worker can only work for a single firm. A vacant firm has a productivity normalized to zero, a firm that employs a worker has a productivity normalized to one but has to pay the wage bill. All agents are risk neutral. Firms maximize expected profits. Workers maximize expected wage payments.

[8]Asymptotic efficiency has been established in sequential search e.g. in Gale (1987). For an overview and quite general specifications see Mortensen and Wright (2002) and Lauermann (2005). In simultaneous search, Acemoğlu and Shimer (2000) and Albrecht, Gautier and Vroman (forthcoming) present limit results, yet converge to some constrained efficient outcomes of a still frictional economies.

[9]Gautier and Moraga-González (2005) present a three player example with a similar concept in an environment where wages are unobservable. Their main analysis of a large market assumes no recall, and also exhibits inefficiencies.

The game has three stages. First, potential entrants can become active by paying a setup cost $K < 1$, and active firms publicly post a wage. Next, workers observe all posted wages. Each worker decides on the number $i \in \{0, 1, 2\}$ of applications he wants to send at cost $c(i)$, where $c(0) = 0$ and the marginal costs $c_i = c(i) - c(i-1)$ are assumed to be weakly increasing. The worker also decides on the i active firms to which he applies. In the final stage workers and firms are matched, the posted wage is paid and matched pairs start production.

For the matching we have a large number T of subperiods in mind. In each of them a (currently) unmatched firm can make an offer. Workers accept the most attractive offer, but are able to reject offers they accepted earlier in favor of later but better offers. In the next substage those firms that got rejected have another chance to make an offer. Matches become final in period T. In the appendix we discuss properties of this matching process under the assumptions of symmetric and anonymous strategies and analyze convergence for $T \to \infty$. For the remainder of this paper we will consider the resulting limit allocation. Since workers prefer higher offers over lower offers, the stability of the limit matching implies that for almost all agents in the economy matching occurs as if higher wage firms move first. That is, we can intuitively think about the following much simpler algorithm: Out of the firms with the highest wage, select one at random. If it has at least one applicant, select one of its applicants at random, match both and remove them from the market. Otherwise only the firm is removed and stays unmatched. Repeat the process with the remaining agents in the economy. This insight will guide the specification of the relevant matching probabilities in the next subsection.[10],[11]

The notion of a large, anonymous market is captured by the assumption that agents' equilibrium strategies are symmetric and anonymous. This standard assumption of the directed search literature implies that all firms use the same entry, posting and hiring strategy, and

[10] Hiring precedence of higher wage firms yields a stable allocation on the network. It is also applied in Burlow and Levin (forthcoming). The matching resembles the process used to assign interns to hospitals in the United States. See sections 2.5.1 and 2.5.2 for details.

[11] In the matching literature agents usually have strict preferences over the other market side. We could incorporate this by assuming that agents draw at random a preference list over those agents they are linked to, and these preferences are lexicographically subordinate to money. This would remove the arbitrariness of the matchings, still the matching probabilities would be unchanged (only the order at which firms are to make proposals would now be appropriately tied to the realization of the preference lists to achieve stability).

do not condition their strategy on the identity of the worker. All workers use the same application strategies, and do not condition on the firms identity.[12] The symmetry of the workers' application strategies usually requires mixed strategies which create the market frictions: sometimes multiple workers apply for the same job, and sometimes none apply at all.

A pure strategy for a firm is its entry decision $e \in \{Enter, Out\}$ and a wage offer $w \in [0,1]$. A mixed strategy for a firm is a probability ϕ of playing $Enter$ and a cumulative distribution function F on $[0,1]$. Throughout the paper we adopt the law of large numbers convention, which for instance implies that F is also the realized distribution of wage offers and $v = \phi V$ is the realized measure of active firms. A worker observes the distribution of posted wages and decides on the number of applications $i \in \{0,1,2\}$ that he wants to send. If $i = 1$ he also decides on a wage $w \in [0,1]$ to which he applies; if $i = 2$ he decides on a wage tuple $(w_1, w_2) \in [0,1]^2$. This fully characterizes his strategy given the anonymity assumption, which implies that he randomizes equally over the firms that offer the same wage. A mixed strategy for a worker is a tuple $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)$, where γ_i is the probability of sending i applications, and a tuple $\boldsymbol{G} = (G^1, G^2)$, where G^i is a cumulative distribution function over $[0,1]^i$ which describes the way the worker sends his i applications.[13] If $i = 2$ we will throughout assume that $w_1 \leq w_2$, and will denote by $G^2_j(\tilde{w})$ the marginal distribution over w_j and by $G^2_j(\tilde{w}|w)$ the conditional distribution over w_j when $w_i = w$, $i \in \{1,2\}/\{j\}$.

2.1.2 Expected Payoffs and Equilibrium Definition

To describe the expected payoffs under mixed strategies, let $\eta(w)$ denote the probability that a firm that posts wage w hires a worker. Let $p(w)$ be the probability that an application to wage w yields an offer sometime during the matching stage. These are endogenous objects, yet once they are defined, the profit of a firm posting wage w - omitting entry costs - is

$$\pi(w) = \eta(w)(1 - w). \tag{2.1}$$

[12]These assumptions are discussed in section 6. We should note that we could purify the firms' posting strategy and relax anonymity of the workers' strategies.

[13] It will be convenient to assume that each combination in the support of the randomization of workers and firms is chosen by a continuum of agents in order to apply the law of large numbers convention. We assume the set of agents to be sufficiently large.

The profits comprise the margin $1 - w$ if a worker is hired, multiplied by the probability $\eta(w)$ of hiring.

The utility of a worker who sends no applications is $U_0 = 0$. A worker who applies with one application to wage w obtains utility $U_1(w) = p(w)w - c(1)$, i.e. the expected profit minus the cost of the application. A worker who applies to wages (w_1, w_2) with $w_1 \leq w_2$ obtains utility

$$U_2(w_1, w_2) = p(w_2)w_2 + (1 - p(w_2))p(w_1)w_1 - c(2). \tag{2.2}$$

The worker's utility is given by the wage w_2 if he is made an offer at that wage, which happens with probability $p(w_2)$. With the complementary probability $1 - p(w_2)$ he does not receive an offer at the high wage and his utility is w_1 if he gets an offer for his low wage application, which happens with probability $p(w_1)$. He always incurs the cost for the two applications.

When agents randomize over wages, their payoff is determined by appropriately averaging the payoffs at individual wages.

We will now relate $\eta(\cdot)$ and $p(\cdot)$ to the strategies (ϕ, F) and $(\boldsymbol{\gamma}, \boldsymbol{G})$. We will first determine the payoffs under the assumption that all firms and workers follow this strategy profile. We will then consider individual deviations from this profile. In the following we will talk about the offer set $\mathcal{V}$, which refers to the support of the wage offer distribution F, and about the application set $\mathcal{W}$, which refers to the support of G^1 if $\gamma_1 > 0$ joint with the union of the support of G_1^2 and G_2^2 if $\gamma_2 > 0$.

We first consider wages that are in the offer and in the application set. Let $\lambda(w)$ denote the ratio of applications per firms at wage w. It is characterized by the following mass balance:[14]

$$\gamma_1 G^1(w) + \gamma_2 G_1^2(w) + \gamma_2 G_2^2(w) = v \int_0^w \lambda(\tilde{w}) \, dF(\tilde{w}) \quad \forall w \in [0, 1]. \tag{2.3}$$

The left hand side denotes the expected mass of applications that are sent to wages up to w. It is given by the probability that workers who send one application send it below w, and the probability that workers who send two applications send either their low or their high application below w. It is the inflow of applications to wages up to w. These are dispersed

[14]We define $\lambda(\cdot)$ on $\mathbb{R}_+ \cup \{\infty\}$ to account for the case that a negligable fraction of firms might (non-optimally) receive a mass of applications.

over the firms that offer wages up to wage w. This outflow is specified on the right hand side. It is given by the ratio of applications per firm multiplied with the number of firms, aggregated over all relevant wages. We refer to λ as the gross queue length.

The crucial observation is that not all applications are "effective" in the sense that the firm can hire the applicant. The applicant cannot be hired if he receives a strictly better offer, or has already received a weakly better offer. Denote the fraction of applications that are unavailable for hiring by $\psi(w)$. Then the ratio of effective applications per firm is given by

$$\mu(w) = (1 - \psi(w))\lambda(w). \tag{2.4}$$

We call $\mu(w)$ the *effective* queue length at w.

The probability that a firm with wage w has at least one effective application is given by $1 - e^{-\mu(w)}$. This is due to the anonymity of the workers strategy, which leads to random assignment of applications to firms at a given wage. In a finite economy this implies that the number of effective applications is binomially distributed; for a large economy this is approximated by the Poisson distribution under which the probability that a firm receives no effective application is $e^{-\mu(w)}$. If the firm receives at least one effective application it will be able to fill its vacancy, because when it successively makes offers it will eventually make an offer to this application and become matched. Therefore the hiring probability for a firm is

$$\eta(w) = 1 - e^{-\mu(w)}. \tag{2.5}$$

Now consider the probability of a worker to receive an offer at wage w. His competitors for a job are only those applications that are effective, since for all others the workers decline even if they are made an offer. Each individual worker calculates his acceptance probability by considering his own application effective (because he considers the case where he is unsuccessful at better wages) but realizes that only a fraction of the other applications will be effective. Given that there are $1-e^{-\mu(w)}$ matches per firm and $\mu(w)$ effective applications per firm, the probability of an effective application to yield a match is given by[15]

$$p(w) = \frac{1 - e^{-\mu(w)}}{\mu(w)}, \tag{2.6}$$

with the convention that $p(w) = 1$ if $\mu(w) = 0$.

[15]For a careful but intuitive derivation of (2.5) and (2.6) as the limit for a finite but large economy see Burdett, Shi and Wright (2001).

Finally, consider the probability $\psi(w)$ that an offer does not lead to a match because the sender receives and accepts a different offer. $\psi(\cdot)$ is trivially zero if workers send only one application, i.e. if $\gamma_2 = 0$. Otherwise, consider some application sent to wage w, and let $\hat{G}(\tilde{w}|w)$ denote the probability that the sender had a second application and sent it to a wage weakly lower then $\tilde{w}$.[16] Similarly, let $\hat{g}(w|w)$ denote the probability that the sender sent a second application to w, i.e. $\hat{g}(w|w) = \hat{G}(w|w) - \lim_{\tilde{w} \nearrow w} \hat{G}(\tilde{w}|w)$. Then $\psi(w)$ is given by

$$\psi(w) = \int_{\tilde{w}>w}^{1} p(\tilde{w}) d\hat{G}(\tilde{w}|w) + \frac{p(w)}{2} \hat{g}(w|w). \tag{2.7}$$

That is, $\psi(w)$ is the average probability that a worker sends two applications and the other application was strictly higher and successful. If the other application was sent to the same wage, the probability that it was successful is $p(w)$, and in this case the unconditional probability for each firm to make the offer first is 1/2. The system defined by (2.4), (2.6) and (2.7) is recursive: At the highest offered wage the probabilities $p(w)$, $\eta(w)$ and $\psi(w)$ can be determined, and are then used to evaluate the corresponding terms at lower wages. To round off the specification, briefly consider the case of a wage in the offer set that is not in the application set. In this case nobody applies, and we specify $\lambda(w) = \mu(w) = 0$. We interpret the polar case of wages in the application but not in the offer set as free disposal of applications without the chance of receiving an offer. This covers all possibilities in the support of the workers and/or firms randomization, and their average ex-ante payoffs can be calculated.

Now consider deviations. For a firm, any wage in the offer set can be evaluated as described above as it arises as a possible realization of F. Yet for a deviation to wages $w \notin \mathcal{V}$ the firm has to form a belief about the workers reaction. This problem is common in the directed search literature, and the most common approach is to assume that the queue length at the deviant is exactly such that workers are indifferent between applying and not applying. If it were higher workers should adjust by applying less; if it were lower they should apply more. We will call the highest utility a worker can obtain at wages in the offer set as the Market Utility. Let $U_i^* = \sup_{\boldsymbol{w} \in \mathcal{V}^i} U_i(w)$ denote the highest utility a worker can get by sending i

[16] There are $\gamma_1 dG^1(w)$ single applications, $\gamma_2 dG_1^2(w)$ low applications and $\gamma_2 dG_2^2(w)$ high applications at w, adding to a total measure $T(w) = \gamma_1 dG^1(w) + \gamma_2 dG_1^2(w) + \gamma_2 dG_2^2(w)$. Then $\hat{G}(\tilde{w}|w) = \sum_{j=1}^{2} [\gamma_2 dG_j^2(w)/T(w)] G_{-j}^2(\tilde{w}|w)$, where $-j \in \{1,2\}/\{j\}$.

applications. Then we can define the Market Utility as $U^* = \max\{U_0, U_1^*, U_2^*\}$.

Now consider some wage $w \notin \mathcal{V}$ and assume the queue length were $\mu(w) \in [0, \infty]$, which defines $p(w)$ as in (2.6). A worker who applies there with one application can at best get $U_1(w)$. A worker who sends two applications and applies there with his low application can at best get $\hat{U}_{1,l}(w) = \sup_{\tilde{w} \in \mathcal{V}} U(w, \tilde{w})$. Applying with the high application he can get $\hat{U}_{2,h}(w) = \sup_{\tilde{w} \in \mathcal{V}} U(\tilde{w}, w)$. Let $\hat{U}(w) = \max\{U_1(w), \hat{U}_{2,l}(w), \hat{U}_{2,h}(w)\}$. We will assume

Definition 2 (Market Utility Assumption)

For any $w \notin \mathcal{V}$, $\mu(w) > 0$ *if and only if* $U^* = \hat{U}(w)$.

It states that workers are indifferent between obtaining the Market Utility and applying to the not-offered wage (possibly combined with the most attractive offered wage).[17] Only if the wage is too low it is not possible to adjust the effective queue length to obtain indifference, in which case the effective queue length is set to zero because nobody would apply there. While the arguments presented here have only intuitively appealed to "reasonable" responses of workers to deviating wage offers, papers by Burdett, Shi and Wright (2001) and Peters (1997, 2000) rigorously establish equivalence of the Market Utility Assumption and the subgame perfect response in (a limit of) finite economies in which workers send one application.

Firms also have to evaluate the profit of entering if $\phi = 0$, i.e. no other firm enters. Assume some firm enters and offers a wage. In the same spirit of subgame perfection, workers either do not apply at all because the wage is too low, or they drive the queue length down to a level where they are indifferent between applying or not applying. Anticipating this response, the firm will enter if it can find a profitable wage offer. Formally, we assume

Definition 3 (Entry Assumption)

$\phi > 0$ *if there exist* $w \in [0, 1]$ *and* $\mu(w) \in \mathbb{R}_+$ *such that* $U_1(w) = c_1$ *and* $\pi(w) > K$.

Finally, consider the deviation of an individual worker. All other workers still use their mixed strategy. The competition for each job is therefore unchanged, and he still obtains payoff $U_i(\boldsymbol{w})$ when he applies to wages in the application set. Wages $w \in \mathcal{V} \backslash \mathcal{W}$ yield an offer for sure. Other wages cannot provide profitable deviation even if they are offered by some (possibly deviating) firm when the worker's belief about the other workers' behavior corresponds to the belief summarized in the Market Utility Assumption.

[17] Inspection of the worker's problem in section 2.2.1 reveals that even if both wages are not in the offer set it is impossible to obtain a utility above U^*, given this assumption.

We define an equilibrium as follows.

Definition 4 (Equilibrium) *An equilibrium is a tuple* $\{\phi, F, \boldsymbol{\gamma}, \boldsymbol{G}\}$ *of strategies for the agents such that there exists* π^**,* U^* *and* $\mu(\cdot)$ *and*

1. *(a)* $\pi(w) = \pi^* \geq \pi(w')$ *for all* $w \in \mathcal{V}$ *and* $w' \in [0,1]$ *if* $\phi > 0$*.*

   (b) $\pi^* = K$ *if* $\phi > 0$*.*

2. *(a)* $U_i(\boldsymbol{w}) \geq U_i(\boldsymbol{w'})$ *for all* $\boldsymbol{w} \in suppG^i$ *and* $\boldsymbol{w'} \in [0,1]^i$*, if* $\gamma_i > 0$*.*

   (b) $U_i^* = U^*$ *if* $\gamma_i > 0$*.*

3. $\mu(\cdot)$ *conforms to (2.3) - (2.7) and the Market Utility and Entry Assumptions hold.*

Condition 1 *a*) and *b*) specify profit maximization and free entry.[18] Condition 2 *a*) implies that workers who send i applications send them optimally given the wage offer distribution and the behavior of other workers. Condition 2 *b*) ensures that workers send out the optimal number of applications. Condition 3 reiterates the determination of the effective queue length. The distinction between *a*) and *b*) allows the discussion of an exogenous number of applications and/or exogenous number of firms using the appropriate subset of conditions. While the exposition uses the terminology of a game, the definition resembles a competitive equilibrium with a somewhat non-standard feasibility constraint that embeds the frictions.

2.2 Equilibrium Characterization

In this section we characterize the equilibrium properties of the model and show

Summary 2.2.1 *An equilibrium exists. Generically the following holds: The equilibrium is unique; all workers send the same number of applications; the number of offered wages equals the number of applications; each worker applies with one application to each wage.*

We will proceed in three subsections: First we analyze the workers' search behavior given a distribution of wages and given the number of applications. Then we analyze the distribution of wages that firms will optimally set. Finally, we characterize equilibrium play.

[18] To ensure that entry implies zero profits it is sufficient to assume that $V > 1/K$.

2.2.1 Workers' Search Decision

To analyze the workers' search decisions, first consider a single worker who observes all wages and - given the strategy of the other workers - knows the probability of success at each wage. That is, he knows all pairs $(w, p(w))$. Equilibrium condition $2a$) implies that each wage to which workers apply has to be optimal. For a worker with $i = 1$ the application choice is trivial. An application to w' is optimal if and only if $p(w')w' = u_1 \equiv \max_{w\in[0,1]} p(w)w$, i.e. he chooses a wage with the highest expected return u_1. For a worker with $i = 2$ the analysis is slightly more involved. Let $\bar{w}$ be the highest wage out of all wages that deliver u_1, i.e. $\bar{w} = \sup\{w \in [0,1] | p(w)w = u_1\}$.

Lemma 3 *Assume that an optimal choice for a worker with $i = 2$ exists. The optimal choice involves sending one application to a wage weakly below $\bar{w}$ and one application to a wage weakly above $\bar{w}$.*

Proof: The worker maximizes

$$\max_{(w_1,w_2)\in[0,1]^2} p(w_2)w_2 + (1 - p(w_2))p(w_1)w_1. \tag{2.8}$$

Note that we have set up problem (2.8) without the restriction that $w_1 \leq w_2$. Nevertheless it is immediate that a worker who has the choice between two wages will always accept the higher over the lower. Therefore any solution to (2.8) has $w_1 \leq w_2$.[19]
Next, note that w_1 is only exercised if w_2 failed. (2.8) immediately implies that for w_1 only the expected return $p(w)w$ is important, and his optimal decision resembles that of workers with a single application. I.e. he chooses w_1 such that

$$p(w_1)w_1 = u_1. \tag{2.9}$$

Taking this into account, any high wage w_2 is optimal if it fulfills

$$p(w_2)w_2 + (1 - p(w_2))u_1 = u_2, \tag{2.10}$$

where $u_2 \equiv \sup_{w\in[0,1]} p(w)w + (1 - p(w))u_1$. Clearly any combination of w_1 and w_2 that satisfies (2.9) and (2.10) solves the maximization problem (2.8). Since we know that any

[19] Assume a worker would choose (w_1, w_2). By (2.8) he gets $U(w_1, w_2) = p(w_2)w_2 + (1 - p(w_2))p(w_1)w_1$. Now assume he reversed the order to get $U(w_2, w_1) = p(w_1)w_1 + (1-p(w_1))p(w_2)w_2$. $U(w_1, w_2) \geq U(w_2, w_1)$ if and only if $w_2 \geq w_1$.

solution to the latter problem has $w_2 \geq w_1$, it has to hold that the highest low wage associated with (2.9) has to be weakly lower than the lowest high wage associated with (2.10). The highest low wage is given by $\bar{w}$. *Q.E.D.*

At high wages the worker takes into account the possibility of obtaining a low wage offer. He is willing to accept a lower expected pay ($p(w_2)w_2 < u_1$) in return for a high upside potential if he gets a job (high w_2), because if he does not get the good job the low wage application acts as a form of insurance.[20] Since all workers face the same maximization problem we obtain

Proposition 11 *Any equilibrium with $\gamma_1 + \gamma_2 > 0$ fulfills the following conditions for the effective queue length:*

$$p(w) = 1 \;\; \forall\, w \in [0, u_1] \tag{2.11}$$

$$p(w)w = u_1 \; \forall\, w \in [u_1, \bar{w}] \tag{2.12}$$

$$p(w)w + (1 - p(w))u_1 = u_2 \; \forall\, w \in [\bar{w}, 1], \tag{2.13}$$

for some tuple $(u_1, u_2, \bar{w})$. It holds that $u_1 = \max_{w \in \mathcal{V}} p(w)w$ and

i) for $\gamma_2 > 0$, $u_2 = \max_{w \in \mathcal{V}} p(w)w + (1 - p(w))u_1$ and $\bar{w} = u_1^2/(2u_1 - u_2)$.

ii) for $\gamma_2 = 0$, if $u_1^2/(u_1 + c_2) \in (0, 1)$ then $\bar{w} = u_1^2/(u_1 + c_2)$ and $u_2 = u_1 + c_2$, otherwise $\bar{w} = 1$.

Low wages do not receive applications, wages in the intermediate range receive the low applications that workers are only willing to send if (2.9) is fulfilled, and high wages receive high applications under condition (2.10). We should note that even if workers do not send high wage applications, i.e. $\gamma_2 = 0$, the queue lengths at high wages might be determined by (2.10). If a deviant posts a high wage, the Market Utility Assumption specifies that workers are indifferent. If the second application is quite costly, indifference implies that the workers are indifferent between sending their single application to the deviant or to the offered wage. Yet if the second application is not very costly, it might be optimal to continue to send the single application to some offered wage but to send an additional application to the deviant.

[20] See Chade and Smith (2004) and also chapter 1 for more discussion of the tradeoffs under simultaneous search.

Indifference then implies that the marginal benefit of the additional application is zero, i.e. $u_2 - u_1 - c_2 = 0$, which then determines u_2 and governs the queue length at high wages.

2.2.2 Firms' Wage Setting

This subsection focuses on the nature of equilibrium wage dispersion. We first show that wage dispersion is a necessary feature of any equilibrium with $\gamma_2 > 0$. We then show that this leads to exactly two wages being offered in equilibrium.

Consider the case where $\gamma_2 > 0$, which implies that $v > 0$ as otherwise there is no reason to apply. Before we proceed, we will briefly rewrite firms' profits in a convenient way. Consider some (candidate) equilibrium characterized by u_1, u_2 and $\bar{w}$, which by proposition 11 characterizes the workers application behavior. We will call firms that end up offering wages below $\bar{w}$ as low wage firms, and those offering wage above $\bar{w}$ as high wage firms. The problem for each individual firm is to maximize $(1-e^{-\mu(w)})(1-w)$ under the constraint that $\mu(w)$ is given by (2.11), (2.12) and (2.13). This is a standard maximization problem. Writing the profit function as $\pi(w) = \mu(w)p(w)(1-w)$, we can use the constraint to substitute out the wage and write profits as a function of the effective queue length

$$\pi(\mu) = 1 - e^{-\mu} - \mu u_1 \qquad \forall\ \mu \in [\underline{\mu}, \bar{\mu}], \tag{2.14}$$

$$\pi(\mu) = (1 - e^{-\mu})(1 - u_1) - \mu(u_2 - u_1) \quad \forall\ \mu \in [\bar{\mu}, \mu(1)], \tag{2.15}$$

where $\bar{\mu} = \mu(\bar{w})$ and $\underline{\mu} = 0 = \mu(u_1)$. We can interpret this as follows. Firms that offer a wage below u_1 "buy" a queue length of zero and make zero profits. Low wage firms that offer wages in $[u_1, \bar{w}]$ "buy" a queue length according to (2.12) and obtain profits as in (2.14), while high wage firms with wages in $[\bar{w}, 1]$ "buy" a queue length as in (2.13) and obtain profits as in (2.15). Since at wages above u_1 there is a one-to-one relation between the wage and the effective queue length, we can view the individual firm's problem as simply a choice regarding the preferred effective queue length.

The profit function is continuous, but has a kink at $\bar{w}$ (respectively $\bar{\mu}$). This is due to the fact that workers trade off the effective queue length against the wage differently for high and low applications. At high wages the queue length responds stronger to a wage change since workers are more "risky" due to the fallback option at low wages. This kink implies immediately that it cannot be profitable for any individual firm to offer wage $\bar{w}$, because raising the wage induces many additional effective applicantions while reducing the wage induces only a relatively small reduction in effective applications. Therefore either it is profitable to offer a higher wage, or if that is not profitable then it is profitable to offer a lower wage.

This rules out an equilibrium in which some workers send more than one application but all firms offer the same wage, because the offered wage would coincide with the cutoff wage and individual firms would want to deviate.

Proposition 12 *There does not exist an equilibrium with $\gamma_2 > 0$ in which only one wage is offered, i.e. in which $\mathcal{V}$ is a singleton.*

We should note that the argument that rules out one-wage equilibria is different from those in most other papers on wage dispersion with homogenous workers and firms. Usually there is an appeal to a discontinuity of the following kind: If all firms offer the same wage, there is a strictly positive probability that a firms' offer is rejected because the worker accepts some other equally good offer; so if a devianting firm offers a slightly higher wage, at least as many workers apply, and all applicants accept an offer for sure. This yields a jump in profits.[21] In this model there is no discontinuity despite the fact that workers accept an offer for sure at a slightly higher wage. This positive jump in profits is offset by the fact that *fewer* workers apply to the deviant.[22] Workers internalize that only effective applications imply competition. At the market wage, only a fraction of the (other) applications are effective, while at the deviant all applications are effective. If the deviants' wage is only slightly higher, less workers apply because otherwise the competition would make an application unattractive. As a consequence profits change continuously. Nevertheless, the kink in the profit function induced by a different "risk-return"-tradeoff of workers implies wage dispersion.

Next we show that in an equilibrium in which some workers send two applications exactly two wages will be offered, one strictly below and one strictly above $\bar{w}$. This immediately implies that workers with one application send it to the low wage, and workers with two applications send one to each of the wages.

[21]This happens e.g. in Burdett and Judd (1983), Burdett and Mortensen (1998), the basic version of Acemoğlu and Shimer (2000) and in the model presented in Chapter 1.

[22]In the introduction we explained "directedness" as the ability to attract more applications when offering higher wages. By this we mean more *effective* applications. As shown here, gross applications do not need to be higher.

Proposition 13 *In any equilibrium with $\gamma_2 > 0$, exactly two distinct wages will be offered, i.e. $\mathcal{V} = \{w_1^*, w_2^*\}$. It holds that $w_1^* < \bar{w} < w_2^*$.*

Proof: Since wage dispersion implies that not all wages are zero, $u_2 > u_1 > 0$. Individual firms take u_1, u_2 and $\bar{w}$ as given. For low wage firms we can write the profits as a function of the queue length as in (2.14). The function is strictly concave on $[0, \bar{\mu}]$. Therefore all low wage firms will offer the same wage. For high wage firms profits can be written as in (2.15), which is strictly concave on $[\bar{\mu}, \mu(1)]$. Therefore all high wage firms will offer the same wage. Finally, assume one group, say low wage firms, offered the wage $\bar{w}$. Since there is wage dispersion, high wage firms will offer $w_2^* > \bar{w}$. But since their problem is strictly concave on $[\bar{w}, 1]$, they make strictly higher profits than firms at $\bar{w}$, which yields the desired contradiction. A similar argument rules out that $\bar{w}$ is offered by high wage firms. *Q.E.D.*

Given that only two wages are offered in equilibrium, we will for notational simplicity index variables referring to low wage firms by *1* and those referring to high wage firms by *2*.[23] Let d_1 be the equilibrium fraction of firms offering the low wage, and d_2 the fraction offering the high wage. Then $v_i = vd_i$ denotes the measure of firms at the respective wage, and equation (2.3) implies gross queue lengths $\lambda_2 = \gamma_2/v_2$ and $\lambda_1 = (\gamma_1 + \gamma_2)/v_1$. At high wages workers only apply strictly lower, so that $\mu_2 = \lambda_2$. At low wages, a fraction $\gamma_2/(\gamma_1 + \gamma_2)$ applies to the high wage, and so $\psi_2 = p_2\gamma_2/(\gamma_1 + \gamma_2)$, with $p_2 = (1 - e^{-\mu_2})/\mu_2$. Therefore $\mu_1 = (1 - p_2\gamma_2/(\gamma_1 + \gamma_2))\lambda_1$. With these notational simplifications we establish

Corollary 1 *In an equilibrium with $\gamma_2 > 0$ profits and wages for high and low wage firms respectively are given by*

$$\pi_1 = 1 - e^{-\mu_1} - \mu_1 e^{-\mu_1}, \tag{2.16}$$

$$w_1^* = \mu_1 e^{-\mu_1}/(1 - e^{-\mu_1}), \tag{2.17}$$

$$\pi_2 = (1 - e^{-\mu_2} - \mu_1 e^{-\mu_1})(1 - e^{-\mu_1}), \text{ and} \tag{2.18}$$

$$w_2^* = \mu_2 e^{-\mu_2}(1 - e^{-\mu_1})/(1 - e^{-\mu_2}) + e^{-\mu_1}. \tag{2.19}$$

[23]That is, let π_i be the profit, w_i the wage, λ_i the gross queue length, μ_i the effective queue length, η_i the hiring probability, p_i the probability of getting an offer when applying at a type-i firm, and ψ_i the probability that a worker accepts another offer, where $i = 1$ when we refer to low wage firms and $i = 2$ when we refer to high wage firms.

Proof: We know that neither low wage nor high wage firms are constrained, because the equilibrium wages are different from $\bar{w}$, and it is easy to see that $w_1 > 0$ (otherwise workers would not apply) and $w_2 < 1$ (otherwise high wage firms would make less profits than low wage firms). Therefore wages are given by first order conditions. For low wage firms, the first order condition of (2.14) with respect to μ leads to

$$u_1 = e^{-\mu} = e^{-\mu_1}. \tag{2.20}$$

The second equality follows since in equilibrium all low wage firms will choose the same queue length, or rather the wage associated with it. When substituted into (2.14) this leads to the expression for the profits. By (2.12) we know that $w_1^* = u_1/p_1$ and we immediately get the corresponding wage. For high wage firms, the first order condition of (2.15) implies

$$u_2 - u_1 = e^{-\mu}(1 - u_1) = e^{-\mu_2}(1 - e^{-\mu_1}). \tag{2.21}$$

Substitution back into (2.15) yields the expression for the profits. By (2.13) we know that $w_2^* = (u_2 - u_1)/p_2 + u_1$, and substitution leads to the expression for the high wage. *Q.E.D.*

In the the case of a single application ($\gamma_1 > 0$ and $\gamma_2 = 0$) the arguments above easily establish that only one wage is offered according to (2.17), yielding profits given by (2.16). Obviously in this case the probability that an offer leads to a hire is one, i.e. $\mu_1 = \lambda_1$. This is also the result obtained in Burdett, Shi and Wright (2001). The introduction of a second application essentially establishes two markets. The profits in each are given by $(1 - e^{-\mu_i} - \mu_i e^{-\mu_i})(1 - u_{i-1})$. In the low market u_0 is identical to the workers' true outside option of zero, but there is some connection to the high market induced by the strictly positive probability that an offer is rejected. In the high market the rejection probability is zero, but u_1 is greater than zero as it reflects the workers' endogenous outside option induced by the presence of the low market. Apart from these spillovers, each market operates essentially as a single one-application market.

The findings in the last sections are summarized in figure 2.1. The workers' indifference curve IC_1 for the low wage applications is given by (2.12). Low wage firms take this into account and offer a wage w_1 such that no individual firm wants to deviate, which means their isoprofit curve IP_1 is tangent to IC_1. The indifference curve IC_2 for the high application is by (2.13) steeper than for the low application because of the fallback option due to the low application. The actual queue length that firms expect is the dashed line. High wage firms

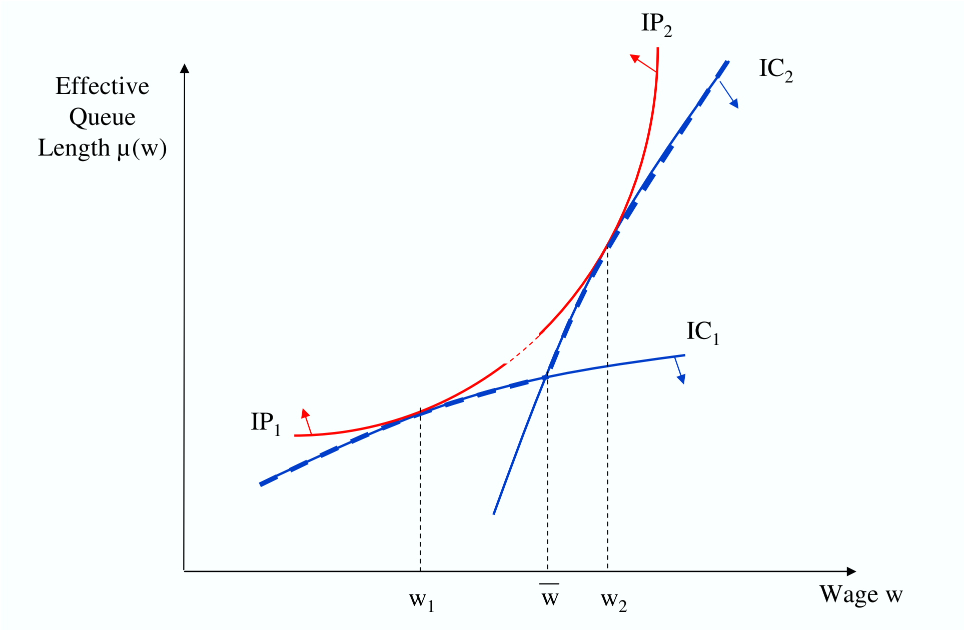

Figure 2.1: Equilibrium behavior. IC_1 and IC_2: Worker's indifference curve for the low and high wage, respectively. IP_1 and IP_2: Isoprofit curves for low and high wage firms.

take this into account and offer wage w_2 in such a way that no firm wants to deviate, i.e. such that their isoprofit curve IC_2 is tangent to IP_2. Note that a single wage $w_1 = w_2 = \bar{w}$ cannot be an equilibrium, because at the kink of the indifference curves it is impossible to place the isoprofit curve tangent, and therefore firms would want to deviate. The isoprofit curves for low and high wage firms have to coincide to provide equal profits for firms. In the next section we prove that this is possible.

2.2.3 Equilibrium Outcome

In this section we derive the equilibrium outcome. Before we turn to the full equilibrium, it will be a useful first step to exogenously fix the number of applications that workers send. We will show existence and uniqueness of an (appropriately adjusted) equilibrium with and without free entry.[24]

[24]Note that the equilibrium definition does not tie down F in case $\phi = 0$ and G_i in case $\gamma_i = 0$. For the discussion of uniqueness, assume that in these cases the respective distribution takes on some unique form. As a technical detail, note that here we fix $\boldsymbol{\gamma}$ but in the Market Utility Assumption the costs still show up. To ensure consistency, assume that when $\gamma_2 = 0$ only one application plays a role (e.g. $c_1 = 0$ and $c_2 = \infty$) while when $\gamma_2 > 0$ both play a role (e.g. $c_1 = c_2 = 0$).

Lemma 4 *For given* $(\gamma_0, \gamma_1, \gamma_2)$ *with* $\gamma_1 + \gamma_2 > 0$ *it holds that*

1. *for given* $v > 0$ *there exists unique* $(F, \boldsymbol{G})$ *such that equilibrium conditions* 1a), 2a) *and* 3 *hold;*

2. *with free entry there exists unique* $(\phi, F, \boldsymbol{G})$ *such that equilibrium conditions* 1a), 1b), 2a) *and* 3 *hold.*

Proof: We will show part 2 here, part 1 is relegated to the appendix. The effective queue length μ_1 at the low market wage is according to (2.18) determined by

$$1 - e^{-\mu_1} - \mu_1 e^{-\mu_1} = K. \tag{2.22}$$

Since the left hand side of (2.22) is strictly increasing in μ_1 and is zero for $\mu_1 = 0$ and one for $\mu_1 \to \infty$, μ_1 is unique. The wage is given by (2.17).
If $\gamma_2 > 0$, the queue length at high wage firms is by (2.18) given by

$$(1 - e^{-\mu_2} - \mu_2 e^{-\mu_2})(1 - e^{-\mu_1}) = K. \tag{2.23}$$

Since μ_1 is unique, μ_2 is unique. The high wage is given by (2.19). Since $\mu_1^* = (\gamma_1 + \gamma_2 - \gamma_2 p_2)/v_1$ and $\mu_2^* = \gamma_2/v_2$, both v_1 and v_2 are uniquely determined, which characterizes the equilibrium entry and the randomization over the two wages.
Clearly firms are willing to offer these wages. The wages were determined by the appropriate first order conditions and therefore no other wage in either the high wage or the low wage region can offer a higher profit. Since (2.12) and (2.13) were used as constraints to construct the profits, they remain valid and workers are indeed willing to apply in the prescribed way. *Q.E.D.*

Knowing the equilibrium interaction for a fixed number of applications, we can turn to the analysis of the equilibrium interaction when the number of applications is endogenous. c_1 and c_2 may take any non-negative values as long as $c_1 \leq c_2$. In analogy to the free entry conditions (2.22) and (2.23) we will define the following four numbers μ_1^*, μ_2^*, u_1^* and u_2^* recursively as follows: $(1-e^{-\mu_2^*}-\mu_2^* e^{-\mu_2^*})(1-u_{i-1}^*) = K$, where $u_i^*-u_{i-1}^* = e^{-\mu_i^*}(1-u_{i-1}^*)$ and $u_0^* = 0$. These numbers are uniquely determined by the exogenous parameter K. Moreover, (2.20) established that the marginal utility of the first application in equilibrium is always u_1^* whenever at least some workers send out applications, and by (2.21) the marginal utility of the second application is $u_2^* - u_1^*$ whenever at least some workers send two applications. This is independent of the exact structure of $\boldsymbol{\gamma}$. We will establish the following proposition, which is a stronger version of summary 2.2.1 with which we started the section.

Proposition 14 *An equilibrium exists. Furthermore*

1. *For $c_1 > u_1^*$ the unique equilibrium has $v = 0$ and $\boldsymbol{\gamma} = 0$.*

2. *For $c_1 < u_1^*$ and $c_2 > u_2^* - u_1^*$, in the unique equilibrium all workers send one application and one wage is offered.*

3. *For $c_1 < u_1^*$ and $c_2 < u_2^* - u_1^*$, in the unique equilibrium all workers send two applications, two wages are offered, and each worker applies to each wage.*

Proof: In case 1 the marginal utility is too low to induce any worker to send the first application. Therefore $v = 0$ and $\boldsymbol{\gamma} = 0$. This is consistent with the Entry Assumption.
In cases 2 and 3, the marginal utility is strictly higher than the marginal cost of the first application. By the Entry Assumption there will be positive entry: At queue length μ' such that $e^{-\mu'} = c$ and wage $w' = \mu' e^{-\mu'}/(1 - e^{-\mu'})$, it holds that $U_1(w') = c$ and profits $\pi' = 1 - e^{-\mu'} - \mu' e^{-\mu'} > K$. Given positive entry, clearly each worker will send at least one application.
Now we have to analyze if it can be the case that workers only apply once, and firms only offer one wage. Assume this is the case. For $c_2 < e^{-\mu_1^*}(\mu_1^* - 1 + e^{-\mu_1^*})/\mu_1^*$ it can be shown that we get a contradiction, because a worker strictly prefers to apply twice at the unique market wage than to send only one application. Yet even if c_2 is not that small, firms might not be willing to offer only one wage - despite the fact that the wage is determined by their first order condition. At high (not offered) wages the queue length might increase fast because workers would send their high application if these high wages were offered, which is reflected in $\bar{w} < 1$ in the second part of proposition 11. Since the queue length is continuous, and the offered wage is strictly optimal on $[u_1, \bar{w}]$, a firm that is looking for a profitable deviation has to find the optimal wage in the interior of $[\bar{w}, 1]$. Since $u_2 = c_2 + u_1$ according to proposition 11, we have by (2.15) the profit $\pi(\mu) = (1 - e^{-\mu})(1 - e^{-\mu_1^*}) - \mu c_2$ for a deviant that offers a wage in $(\bar{w}, 1)$. If there is a profitable deviation, it must be profitable to deviate to $\hat{\mu}$ given by the first order condition $e^{-\hat{\mu}}(1 - e^{-\mu_1^*}) = c_2$, which implies $\hat{\mu} < \mu_2^*$ in case 2 and $\hat{\mu} > \mu_2^*$ in case 3. Substitution leads to an optimal deviation profit of

$$\pi(\hat{\mu}) = (1 - e^{-\hat{\mu}} - \hat{\mu} e^{-\hat{\mu}})(1 - e^{-\mu_1^*}). \qquad (2.24)$$

Comparing (2.24) with (2.23) establishes that $\pi(\hat{\mu})$ is strictly smaller than K in case 2, making a deviation unprofitable, and strictly larger than K in case 3, yielding a strictly profitable deviation (the wage associated with $\hat{\mu}$ is indeed above $\bar{w}$ in case 3). Therefore an equilibrium with one wage is possible in case 2 and not in case 3.

Finally, it is immediate that in case 2 an equilibrium with two wages cannot exist because by $u_2^* - u_1^* < c_2$ the marginal utility of the second application is too low, while an equilibrium with two wages can exist in case 3 since $u_2^* - u_1^* > c_2$. Therefore in case 2 everyone sends one application to the unique wage, while in case 3 every worker sends two applications, one to each of the two wages. Uniqueness is then ensured by lemma 4. *Q.E.D.*

For the case $c_2 < u_2^* - u_1^*$ it is worth emphasizing that two wages are offered because firms anticipate that workers will send an additional application when they offer a high wage (this is captured by the Market Utility Assumption). If we consider a candidate equilibrium in which all firms offer a single wage, it is this feature that leads to a high queue length for a deviant with a high wage and makes such a deviation profitable.

In the case where $c_1 = e^{-\mu_1^*}$ we have multiplicity of equilibria: for any $\gamma_1 \in [0,1]$ and $\gamma_0 = 1 - \gamma_1$ an equilibrium exists, and workers are exactly indifferent between applying once and not applying. If $c_2 = e^{-\mu_2^*}(1 - e^{-\mu_1^*})$ an equilibrium exists in which workers randomize between one and two applications, i.e. it exists for any $\gamma_2 \in [0,1]$ and $\gamma_1 = 1 - \gamma_2$.

2.3 Efficiency

To discuss the efficiency properties of the equilibria just characterized, we will follow Pissarides (2000) and others by using the following notion of constrained efficiency: An equilibrium is constrained efficient if it maximizes the output minus entry and application costs, given the frictions in the market. The frictions stem from the requirement that workers and firms use some symmetric strategies $(\boldsymbol{\gamma}, \boldsymbol{G})$ and (ϕ, F). Denoting by Υ^i the set of cumulative distribution functions over $[0,1]^i$ and by $\triangle_3$ the three-dimensional unit simplex, the strategy spaces of workers and firms are $\mathcal{G} = \triangle_3 \times \Upsilon^1 \times \Upsilon^2$ and $\mathcal{F} = [0,1] \times \Upsilon^1$. Then an equilibrium $\{\phi, F, \boldsymbol{\gamma}, \boldsymbol{G}\} \in \mathcal{F} \times \mathcal{G}$ is constrained efficient if it maximizes

$$\max_{(\phi', F') \in \mathcal{F}, (\boldsymbol{\gamma}', \boldsymbol{G}') \in \mathcal{G}} M(\phi', F', \boldsymbol{\gamma}', \boldsymbol{G}') - \phi' V K - \gamma_1' c(1) - \gamma_2' c(2), \tag{2.25}$$

where $M(\phi', F', \boldsymbol{\gamma}', \boldsymbol{G}') = \phi' V \int_0^1 \eta(w) dF'$ is the number of matches when $\eta(w)$ is determined by (2.3) - (2.7) using the relevant parameters $\{\phi', F', \boldsymbol{\gamma}', \boldsymbol{G}'\}$.

As discussed in the introduction, efficiency my fail on several dimensions: (1) Search Inefficiency: For a given vector of applications and a given number of firms the number of matches is suboptimal. (2) Entry Inefficiency: For a given vector of applications too many or too few firms enter. (3) Application Inefficiency: Workers apply too much or too little given the costs. While these inefficiencies arise without recall, we will show that with recall directed search balances all these margins.[25]

Shimer (1996) explains the efficiency property for the one-application case roughly as follows: The workers' response to a change in the wage yields an implicit price for the desired queue length, therefore firms can price the queue length of applicants exactly at its marginal cost. In this model two variables matter: The gross queue length $\lambda(\cdot)$ and the retention probability $1 - \psi(\cdot)$. Both have to be adjusted by a single-dimensional wage. This is possible because for firms *and* for workers only the combination of both matters. The workers respond to a change in the wage by changing their applications such that the *effective* queue length $\mu(\cdot)$ rises to a new level of indifference. Therefore the same logic holds here. The effective queue length is priced at marginal cost.

Proposition 15 *The equilibrium market outcome is constrained efficient.*

We will prove the proposition in the next three subsections that are dedicated to different margins. For a given number of firms and a given vector of applications, we will show that the search outcome $M(\phi, F, \boldsymbol{\gamma}, \boldsymbol{G})$ is constrained optimal. Then we show that for a given vector of applications the constrained optimal number of firms enter, given that subsequent search is optimal. And finally we will establish that workers send the constrained optimal number of applications, taking account of optimal entry and search.

2.3.1 Search Efficiency

For a given vector of applications $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)$ and firms v, we will show that the search outcome as characterized in the first part of lemma 4 is constrained efficient. As we will see, this depends on the ability of the market to generate different wages for low and high applications.

[25] See section 2.5.2 for the discussion of restricted recall.

For $\gamma_2 > 0$ we start by analyzing a narrower concept that we will call 2-*group-efficiency*. We will assume that there are two groups of firms, one preferred over the other, and all workers who send at least one application send one at random to the non-preferred group, and workers who send two applications send the second one at random to the preferred group. This setup corresponds to the equilibrium outcome. Let $d \in [0,1]$ be the fraction of firms in the preferred group. Search is two-group-efficient if d is chosen optimally given the assumptions just made.

Lemma 5 *For a given $v > 0$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)$ with $\gamma_2 > 0$, the strategy combination implied by equilibrium conditions* 1*a*)*,* 2*a*) *and* 3 *yields two-group-efficient search.*

Proof: The optimization problem is given by

$$\max_{d\in[0,1]} M(d) = vd(1-e^{-\mu_2}) + v(1-d)(1-e^{-\mu_1}), \tag{2.26}$$

where $\mu_1 = (1-\gamma_2 p_2/(\gamma_1+\gamma_2))\lambda_1$, $\mu_2 = \lambda_2 = \gamma_2/(vd)$, $p_2 = (1-e^{-\lambda_2})/\lambda_2$ and $\lambda_1 = (\gamma_1+\gamma_2)/(v(1-d))$. The first derivative is given by

$$\begin{aligned} \partial M/\partial d = v \quad \Bigg[& \quad 1 - e^{-\lambda_2} - (1-e^{-\mu_1}) + de^{-\lambda_2}[\partial\lambda_2/\partial d] \\ + & \quad \frac{e^{-\mu_1}}{v\lambda_1}\Big[-\gamma_2[\partial p_2/\partial d]\lambda_1 + (\gamma_1+\gamma_2-\gamma_2 p_2)[\partial\lambda_1/\partial d]\Big]\Bigg]. \end{aligned}$$

Noting that $\partial\lambda_2/\partial d = -\gamma_2/(d^2 v) = -\lambda_2^2 v/\gamma_2$, $\partial\lambda_1/\partial d = \lambda_1^2 v/(\gamma_1+\gamma_2)$, and then $\partial p_2/\partial d = (1-e^{-\lambda_2}-\lambda_2 e^{-\lambda_2})v/\gamma_2$ we obtain by substitution that the last expression in the first line equals $-\lambda_2 e^{-\lambda_2}$ and the second line equals $e^{-\mu_1}[-(1-e^{-\lambda_2}-\lambda_2 e^{-\lambda_2})+\mu_1]$. This yields

$$\frac{\partial M}{\partial d}\frac{1}{v} = (1-e^{-\lambda_2}-\lambda_2 e^{-\lambda_2})(1-e^{-\mu_1}) - (1-e^{-\mu_1}-\mu_1 e^{-\mu_1}) = 0, \tag{2.27}$$

where the last equality yields the first order condition. (2.27) coincides with the equal profit condition between high and low wage firms. By the proof of proposition 4 part 1 we know that this uniquely determines the measure of firms in the high and the low group. Boundary solutions cannot be optimal because one application would be waisted (global concavity follows mathematically from $\partial^2 M/\partial d^2 = -v[\lambda_2^2 e^{-\lambda_2}(1-e^{-\mu_1})/d + e^{-\mu_1}(1-e^{-\lambda_2}-\lambda_2 e^{-\lambda_2}-\mu_1)^2/(1-d)] < 0$). *Q.E.D.*

Next we show that the search outcome is constraint efficient. The proof relies on establishing that two groups are sufficient to obtain the optimal allocation.

Proposition 16 *For given $v = \phi V$ and given $\boldsymbol{\gamma}$, the search process is constrained efficient, i.e. it holds that*

$$M(\phi, F, \boldsymbol{\gamma}, \boldsymbol{G}) = \max_{F' \in \Upsilon^1, \boldsymbol{G'} \in \Upsilon^1 \times \Upsilon^2} M(\phi, F', \boldsymbol{\gamma}, \boldsymbol{G'}),$$

where $\{\boldsymbol{G}, F\}$ conform to equilibrium conditions 1*a*), 2*a*) *and* 3.

Finally, we show that two groups are indeed necessary to obtain the optimal allocation when a fraction of workers send two applications. One group of firms that all have equal hiring probability is not efficient. This implies that a unique market wage is not able to yield the optimal allocation.

Proposition 17 *For given $v = \phi V$ and given $\boldsymbol{\gamma}$ with $\gamma_2 > 0$, identical hiring probabilities for all firms cannot be constrained efficient, i.e. if $F \in \Upsilon^1$ and $\boldsymbol{G} \in \Upsilon^1 \times \Upsilon^2$ such that $\eta(w) = \bar{\eta} \;\forall\; w \in \mathcal{V}$ then $(\boldsymbol{G}, F) \notin \arg\max_{F' \in \Upsilon^1, \boldsymbol{G'} \in \Upsilon^1 \times \Upsilon^2} M(\phi, F', \boldsymbol{\gamma}, \boldsymbol{G'})$.*

In the proof we show that two groups can achieve the same hiring probabilities as a random process. But the non-preferred group is too small compared to the optimum. All workers would take a job from a firm in the preferred group. For those that end up taking jobs with firms in the non-preferred group this is their last chance to avoid unemployment. Increasing workers' matching probability in the non-preferred group at the cost of decreasing their matching probability in the preferred group improves matching for those workers for whom it is the last option to avoid unemployment at the expense of a lower matching probability for those who still might have another option.

This result is surprising because with one application different hiring probabilities for firms are only warranted when there are productivity differences.[26] Here the source for different hiring probabilities is a sorting externality. Figure 2.2 illustrates this. At a unique market wage representing a random application behavior, the indifference curve IC_1 for the low and IC_2 for the high applications cross at the same point as the firms isoprofit curve IC. Since the actual (dashed) indifference curve is kinked, it is not possible to achieve tangency with the isoprofit curve. Area A indicates mutual gains for workers and firms from sending the low applications to firms with different queue length and wage. Similar gains are indicated

[26] Shimer (2005) analyzes productivity differences when only one application is possible. If workers and firms are homogenous only one hiring probability would be efficient (similar to our case for $\gamma_1 > 0$ but $\gamma_2 = 0$), and only with heterogeneities of firms or workers different hiring probabilities are efficient.

by area B for the high applications. The exact choice of w influences which gains are more prominent, but due to the kink it is never possible to eliminate both. Figure 2.1 shows that it is possible to achieve tangency with two wages.[27]

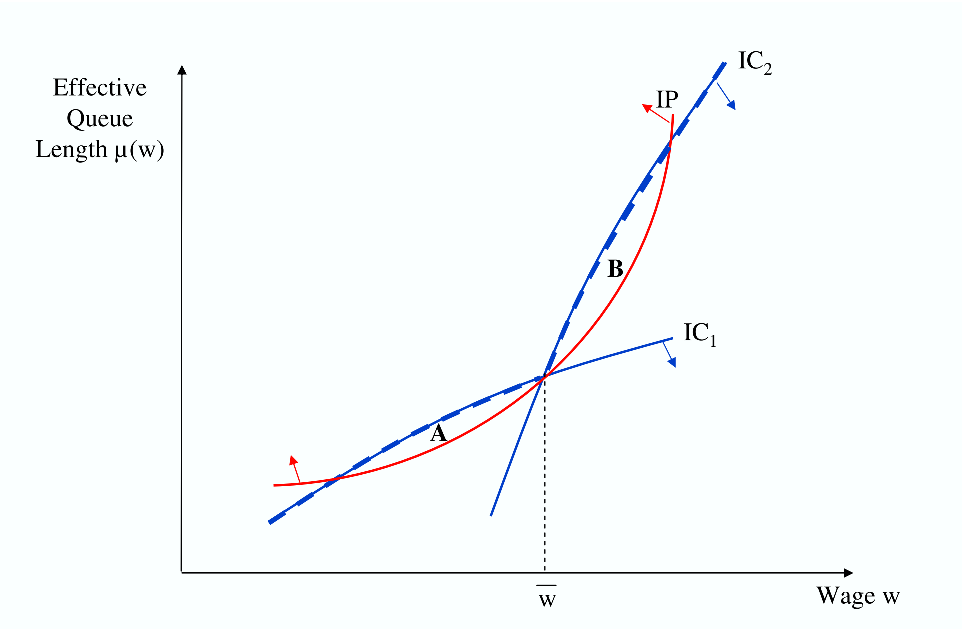

Figure 2.2: Inefficiencies at a unique market wage. Areas A and B indicate mutual benefits for both workers and firms.

2.3.2 Entry efficiency

We now turn to the efficiency of the entry decision. Let $M^*(v, \gamma)$ be the number of matches when there are γ applications and v firms and search is constrained efficient. For given γ we have determined in the second part of proposition 4 the unique entry in the (appropriately adjusted) equilibrium. This entry is constrained efficient given the application behavior.

[27] Neither the figure nor the intuition apply to the case without recall. Without recall, it is not possible to graph all relevant aspects in two dimensions, because next to the wage both the gross queue length $\lambda(\cdot)$ and the retention probability $\psi(\cdot)$ remain separately important. Regarding the intuition, without recall equal hiring probabilities for workers mean that the number of workers who have their last option in either group is the same. Only with recall the rejected positions in the low group become available again for yet unmatched workers, which induces a proportionally larger number of workers employed without another option in the non-preferred group, yielding a positive externality when placing relatively more firms in the non-preferred group.

Proposition 18 *Given $\boldsymbol{\gamma}$, entry is constraint efficient. That is*

$$M^*(v,\boldsymbol{\gamma}) - vK = \max_{v' \in [0,\infty)} M^*(v',\boldsymbol{\gamma}) - v'K, \tag{2.28}$$

where v arises when equilibrium conditions $1a)$, $1b)$, $2a)$ and 3 are fulfilled.

Proof: The number of matches is given by $M^*(v,\boldsymbol{\gamma}) = 1 - \gamma_2 \prod_{i=1}^{2}(1-p_i) - \gamma_1(1-p_1)$, where p_1 and p_2 are the probabilities of getting a job at less preferred and the more preferred group under two-group-efficient search. If $\gamma_0 = 1$, then $v = 0$ arises and is clearly optimal. If $\gamma_0 < 1$, clearly $v = 0$ is not optimal given $K < 1$. Obviously $v \to \infty$ is also not optimal. The first order condition to the problem is

$$\begin{aligned} K &= \gamma_2[\partial p_1/\partial v](1-p_2) + \gamma_2[\partial p_2/\partial v](1-p_1) + \gamma_1[\partial p_1/\partial v] \\ &= [\partial p_1/\partial v](\gamma_1+\gamma_2)(1-\psi_1) + [\partial p_2/\partial v](1-p_1). \end{aligned} \tag{2.29}$$

p_i depends on v directly since the measure $d_i v$ of firms in group i depends on v directly. It also depends on v indirectly since the two-group-efficient fraction d_i is a function of v. Yet by the envelop theorem the indirect effect is zero and we can neglect the effect on d_i. Consider the first term on the right hand side first. We can write $\partial p_1/\partial v = [\partial p_1/\partial \mu_1][\partial \mu_1/\partial v]$. One can show that $[\partial \mu_1/\partial v](\gamma_1+\gamma_2)(1-\psi_2) = -d_1\mu_1^2 - \gamma_2\mu_1[\partial p_2/\partial v]$. Noting that $[\partial p_1][\partial v] = -[(1-e^{-\mu_1}-\mu_1 e^{-\mu_1})/\mu_1^2][\partial \mu_1/\partial v] = -[(p_1 - e^{-\mu_1})/\mu_1][\partial \mu_1/\partial v]$ we obtain

$$[\partial p_1/\partial v](1-p_2) = d_1(1-e^{-\mu_1}-\mu_1 e^{-\mu_1}) + \gamma_2(p_1 - e^{-\mu_1})[\partial p_2/\partial v].$$

Then (2.29) reduces to

$$\begin{aligned} K &= d_1(1-e^{-\mu_1}-\mu_1 e^{-\mu_1}) + \gamma_2(1-e^{-\mu_1})[\partial p_2/\partial v] \\ &= d_1(1-e^{-\mu_1}-\mu_1 e^{-\mu_1}) + d_2(1-e^{-\mu_1})(1-e^{-\mu_2}-\mu_2 e^{-\mu_2}) \\ &= 1-e^{-\mu_1}-\mu_1 e^{-\mu_1}, \end{aligned}$$

where the second line follows by taking the appropriate derivative and the last line follows as a consequence of two-group-efficient search (see (2.27)). The last line also denotes the profits of low wage firms in equilibrium. Applying (2.27) again yields a condition equal to the profits of high wage firms. The first order condition is unique by the same argument that established that a unique v implies zero profits, and the entry implied by equilibrium conditions $1a)$, $1b)$, $2a)$ and 3 coincides with the entry implied by the first order condition. Since the first order condition is unique and boundary solutions are not optimal, it describes the global optimum. *Q.E.D.*

2.3.3 Application Efficiency

The number of applications that workers send in equilibrium is also constrained efficient. We will account for the associated entry of firms and the search outcome, and therefore also immediately establish the overall constrained efficiency of the equilibrium as in Proposition 15.

To gain intuition, consider the case of an individual worker who sends one application rather than none. By the above analysis his marginal benefit is $e^{-\mu_1^*} - c_1$. The benefit for society comprises the cost $-c_1$ and the additional production of one unit of output in the case that the firm did not have another effective applicant, the probability of which is $e^{-\mu_1^*}$. Therefore private and social benefits coincide. Similarly, if two wages are offered and a worker sends a second application rather than only a single one, his private marginal benefit is $e^{-\mu_2^*}(1 - e^{-\mu_1^*}) - c_2$. Additional production arises only if the high firm does not have another effective applicant but the low firm does, which has a probability $e^{-\mu_2^*}(1 - e^{-\mu_1^*})$. Again social and private benefits coincide. Note that the marginal benefit is essentially independent of $\boldsymbol{\gamma}$, and therefore the decisions of other workers summarized in $\boldsymbol{\gamma}$ provide no externality on other workers. This is due to the fact that any positive externality on firms is dissipated in free entry, and the entry compensates any negative effects on other workers.[28]

Let $v(\boldsymbol{\gamma})$ be the entry for a given vector $\boldsymbol{\gamma}$ of applications as implied by equilibrium conditions 1a), 1b), 2a) and 3. Then $M^{**}(\boldsymbol{\gamma}) = M^*(v(\boldsymbol{\gamma}), \boldsymbol{\gamma})$ denotes optimal number of matches for a given vector $\boldsymbol{\gamma}$ of applications given optimal entry and optimal search.

Proposition 19 *The equilibrium vector $\boldsymbol{\gamma}$ of applications is constrained efficient, i.e. $\boldsymbol{\gamma} \in \arg\max_{\boldsymbol{\gamma}' \in \triangle_3} M^{**}(\boldsymbol{\gamma}') - Kv(\boldsymbol{\gamma}') - \gamma_1' c(1) - \gamma_2' c(2)$.*

Proof: For a given $\boldsymbol{\gamma}$ we know that equilibrium conditions 1a), 1b), 2a) and 3 yield the optimal entry and the optimal number of matches. Moreover, under these conditions firms always receive zero profits and all surplus accrues to workers. Comparing different $\boldsymbol{\gamma}$, it is immediate that each worker always attains a marginal utility of $u_1^* - c_1$ for his first application, and $u_2^* - u_1^* - c_2$ for his second application. Clearly the equilibrium conditions in Proposition 14 specify the socially optimal entry. For the case where $c_1 = u_1^*$ (or $c_2 = u_2^* - u_1^*$) the privat and social benefits of the first (or second) application are zero, and therefore every

[28] This argument applies equilibrium conditions 1a), 1b), 2a) and 3 for different $\boldsymbol{\gamma}$, i.e. we compare partial equilibria for different $\boldsymbol{\gamma}$.

equilibrium for this case is constrained efficient. *Q.E.D.*

2.4 Generalization to $N > 2$ applications

In this section consider the case where workers can send any number $i \in \mathbb{N}$ of applications, at a cost $c(i)$. We retain the assumption that $c(0) = 0$ and that marginal costs $c_i = c(i) - c(i-1)$ are weakly increasing. We also assume $c(i) > 0$ for some $i \in \mathbb{N}$. We will establish existence, (generically) uniqueness and constrained efficiency of the equilibrium. The analysis in the preceding sections is a special case for $c(i) = \infty$ for all $i > 2$. Since many arguments are straightforward generalizations of that special case, we focus mainly on the changes that are necessary to adapt the prior setup. At the end of this section we show convergence to the outcome of a competitive economy when the costs for applications vanish.

2.4.1 Extended Setup and Main Result

The extension mainly requires adaptations of the workers' setup, while it remains essentially unchanged for firms. Define N as the largest integer such that $c(N) \leq 1$. Clearly, it is neither individually nor socially optimal to send more than N applications. Then the workers strategy is a tuple $(\boldsymbol{\gamma}, \boldsymbol{G})$, where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, ..., \gamma_N) \in \triangle_N$ and $\boldsymbol{G} = (G^1, G^2, ..., G^N) \in \times_{i=1}^{N} \Upsilon^i$. γ_i denotes the probability of sending i applications, and G^i denotes the cumulative distribution function over $[0,1]^i$ that describes to which wages the applications are sent. Let again $(w_1, ..., w_i)$ satisfy $w_1 \leq w_2 \leq ... \leq w_i$ and let G_j^i denote the marginal distribution of G^i over w_j. Then we can define $\mathcal{W}$ as the union of the support of all G_j^i with $\gamma_i > 0$. A worker who applies to $(w_1, ..., w_i)$ attains in analogy to (2.2) the utility

$$U_i(w_1, ..., w_i) = \sum_{j=1}^{i} \left[\prod_{k=j+1}^{i} (1 - p(w_k)) \right] p(w_j) w_j - c(i). \tag{2.2$'$}$$

A worker who applies nowhere attains $U_0 = 0$. Instead of (2.3) the relevant condition is now

$$\sum_{i=1}^{N} \left[\gamma_i \sum_{j=1}^{i} G_j^i(w) \right] = v \int_0^w \lambda(\tilde{w}) \, dF(\tilde{w}). \tag{2.3$'$}$$

To specify $\psi(w)$ in the extended setup, consider a firm at wage w that receives an application and let $\hat{G}(\tilde{\mathbf{w}}|w)$ denote probability that the sender applied with his other $N-1$ applications to wages weakly below $\tilde{\mathbf{w}}$. If the sender only sent $i < N-1$ other applications, then we code (only for this definition) the additional $N-1-i$ applications as going to wage -1. So

$\tilde{\mathbf{w}} = (\tilde{w}_1, ..., \tilde{w}_{N-1}) \in ([0,1] \cup \{-1\})^{N-1}$. Let $h(\tilde{\mathbf{w}}|w)$ count the number of applications sent to wage w when the worker applies to $\tilde{\mathbf{w}}$ and w. Replacing (2.7) we now specify

$$\psi(w) = \int \left[1 - \frac{1 - (1 - p(w))^{h(\tilde{\mathbf{w}}|w)}}{p(w) h(\tilde{\mathbf{w}}|w)} \prod_{\tilde{w}_j > w} [1 - p(\tilde{w}_j)] \right] d\hat{G}(\tilde{\mathbf{w}}|w). \qquad (2.7')$$

The product $\prod_{\tilde{w}_j > w}[1 - p(\tilde{w}_j)]$ describes the probability that the worker will not take a job at a strictly better wage. Its multiplier gives the probability that a worker will not turn down a job offer because of a job at another firm with the same wage, conditional on failing at higher wages (see the appendix for a derivation). Then the integrand gives the probability that the worker takes the job at a different firm, which is integrated over the relevant wages to which workers apply.

The definitions for all other variables, i.e. μ, p and η and π remain unchanged. The definition of the Market Utility Assumption now has to take into account the expanded possibilities of workers. Let $U_i^* = \sup_{\boldsymbol{w} \in \mathcal{V}^i} U_i(\boldsymbol{w})$. Then $U^* = \max_{i \in \{0,...,N\}} U_i^*$ denotes the Market Utility. Let $X_i(w) \subset [0,1]^i$ denote the set of i-tuples $(w_1, ..., w_i)$ with $w_j = w$ for some $j \in \{0, ..., i\}$ and $w_k \in \mathcal{V}$ for all $k \neq j$. Then we can define $\hat{U}_i(w) = \sup_{\boldsymbol{w} \in X_i(w)} U_i(\boldsymbol{w})$ as the optimal utility if the worker applies to wage w and to $i-1$ other offered wages. The Market Utility assumption then states that for $w \notin \mathcal{V} \bigcup \mathcal{W}$ we have $\mu(w) > 0$ if and only if $U^* = \max_{i \in \{1,...,N\}} \hat{U}_i(w)$. With these adjustments the equilibrium definition extends to this section.

We are now in the position to extend the result from the previous section. Again we recursively define μ_i^* and u_i^* as functions of the exogenous parameter K. Let $u_0^* = 0$. For all $i \in \mathbb{N}$ let $(1 - e^{-\mu_i^*} - \mu_i^* e^{-\mu_i^*})(1 - u_{i-1}^*) = K$ and $u_i^* = e^{-\mu_i^*}(1 - u_{i-1}^*) + u_{i-1}^*$. Note that $u_i^* - u_{i-1}^*$ is strictly decreasing in i, while c_i is weakly increasing. We will show

Proposition 20 *An equilibrium exists. It is constrained efficient. Generically it is unique: if $c_{i^*} < u_{i^*}^* - u_{i^*-1}^*$ and $c_{i^*+1} > u_{i^*+1}^* - u_{i^*}^*$, every workers sends i^* applications, i^* wages will be offered, and every worker applies to each wage.*

The proof relies essentially on an induction of the arguments presented in sections 2.2 and 2.3 to higher numbers of applications. The workers again partition the wages into intervals relevant to each of their applications. The equilibrium interaction in each interval corresponds to that in the one application case, again with the adjustment that the workers "outside option" incorporates the expected utility that can be obtained at lower wages, while the queue length incorporates the fact that some applicants are lost to higher wage firms. Efficiency obtains again for similar reasons, only that now i^* wages are necessary to obtain the optimal allocation in the search process.

2.4.2 Convergence to the Competitive Outcome

We will now show that the equilibrium allocation converges to the unconstrained efficient allocation of a competitive economy when application costs become small. For $K < 1$ the competitive outcome has an equal number of workers and active firms, i.e. $v = 1$, each active firm and each worker is matched, and the market wage is $1 - K$ and coincides with the utility of each worker.

We will consider a sequence of cost functions such that the marginal cost of the i'th application converges to zero for all $i \in \mathbb{N}$. Rather than looking at these functions directly, it will be convenient to simply consider the associated equilibrium number i^* of applications that each worker sends.[29] Vanishing costs amounts to $i^* \to \infty$. Let $v(i^*)$ denote the equilibrium measure of active firms, $\eta(i^*) = \int \eta(w)dF$ and $\varrho(i^*) = v(i^*)\eta(i^*)$ the average probability of being matched for a firm and a worker in the economy. Let $w(i^*)$ denote the average wage conditional on being matched and $U^*(i^*) = u^*_{i^*} - c^{i^*}(i^*)$ the equilibrium utility when i^* applications are sent, where $c^{i^*}(\cdot)$ denotes some cost function that supports an equilibrium with i^* applications per worker. We will show that

Proposition 21 *The equilibrium outcome converges to the competitive outcome, i.e.* $\lim_{i^* \to \infty} v(i^*) = 1$, $\lim_{i^* \to \infty} \eta(i^*) = \lim_{i^* \to \infty} \varrho(i^*) = 1$ *and* $\lim_{i^* \to \infty} w(i^*) = \lim_{i^* \to \infty} U^*(i^*) = 1 - K$.

The structure of the proof uses the intuition for the competitive economy: For a given measure v of active firms the competitive economy implies that (only) the long side of the market gets rationed and the short side appropriates all surplus. We will show that for small frictions (i^* large) this still holds approximately. Then it trivially follows that $v(i^*) \to 1$ because otherwise the firms either generate to much or too little profits to cover entry. Since nearly all agents get matched, zero profits imply a wage of $1 - K$.

[29] For the case of multiple equilibria, consider for simplicity the case where all workers send the same number of applications.

2.5 Discussion and Outlook

2.5.1 Discussion of Main Assumptions

The paper builds on a micro-foundation of frictional markets based on coordination problems between agents. One underlying assumption is that firms treat similar workers alike and do not condition on applicants' names, which seems plausible in larger markets. The other assumption is that workers cannot coordinate to each apply to a different firm. This is modeled by the requirement that in equilibrium workers use symmetric strategies, which has the advantage that a worker does not need to know his "role" in the application process and can deploy the same strategy as everybody else. That such coordination problems actually arise even in small groups has been shown experimentally by Ochs (1990) and Cason and Noussair (2003). A deeper discussion can be found e.g. in Shimer (2005). It is worth mentioning that anonymity of the worker's strategy is not a crucial restriction. If firms do not condition their hiring on workers' names and workers use symmetric strategies, than all firms with the same wage have to have the same effective queue length (otherwise workers could get a higher utility by applying to those firms with the lower queue lengths). This arises in essentially all papers in this field, and the anonymity assumption saves some notational complexities in making the point precise.[30]

When we allow workers to choose their search intensity by sending multiple applications, new modeling choices arise that are absent in one-application models. After the application stage every worker is "linked" to the multiple firms to which he applied. A firm might be "linked" to multiple workers who applied to it (others might only have one worker or no worker at all). Since each firm can only hire one worker, and each worker can only work for one firm, any multiple application model has to specify how matches between firms and workers are formed given those links. It also has to specify the division of surplus for a given match.

For the allocation the main novelty in this paper is to allow firms to contact additional applicants when their offer is rejected. The analysis of this recall process remains tractable because of the assumption that workers can reconsider their options when they receive a better offer. This is the case for example when workers receive a job contract and have

[30] See e.g. Shimer (2005) for the infinite economy case. Burdett, Shi and Wright (2001) demonstrate this property nicely in a finite economy.

a certain time period to sign and return that contract. If they get a better offer during this time period, they can switch their future employer essentially costlessly. This has the convenient feature that it takes most strategic considerations out of the worker's acceptance decision in the extensive form matching process.

For the division of the surplus this paper assumes wage commitments. That is, the wage paid in the final period is the posted wage, and firms cannot counter the proposals of other firms. This again has the advantage of keeping the matching process tractable.[31] It might be a good approximation in environments in which firms are able to make their offers non-verifiable to other firms in order to avoid counter-proposals. It might also be reasonable if individual company (or university) rules do not allow more than the allocated budget for hiring decisions. A final case for this assumption arises in environments in which market rules require binding job and wage descriptions ahead of the final matching. One example of such an environment is the market for hospital interns in the United States. Roth (1984) shows that the algorithm used to match interns with hospitals coincides with the Deferred Acceptance Algorithm by Gale and Shapley (1962), on which the allocation in this paper is based. He points out that participating hospitals have to specify wages and job descriptions way in advance of the actual matching. Similar to the assumption in this model, the algorithm only matches those hospitals and interns that have established contact in a preceding application and interview process.

2.5.2 Relation to the Literature

Equilibrium directed search models resemble competitive economies in their assumption that prices (or wages) are observable to everybody. Yet instead of a Walrasian Auctioneer that facilitates trade, the agents have to individually try to find a trading partner. This leads to frictions if agents cannot coordinate their strategies. Coordination frictions were introduced by Montgommery (1991) and Peters (1991) through the assumption that workers use symmetric application strategies. The symmetry assumption creates coexistence of unemployment and unfilled vacancies; the wage offers direct more applications to higher offers. Most models are restricted to one application per worker. For this case Montgommery already provides an argument for constrained efficiency induced by the wage announcements,

[31]Burlow and Levin (forthcoming) make similar assumptions. See section 2.5.2.

which has been substantiated in subsequent contributions.[32] Burdett, Shi and Wright (2001) provide a detailed derivation of the equilibrium properties in a finite economy with homogenous workers and firms. With one application recall is not an issue because every job offer generates a match due to the workers' lack of alternatives. When costs in the model presented here are such that workers apply only once, our equilibrium reduces to the (limit) equilibrium in Burdett, Shi and Wright. Even with multiple applications the equilibrium interaction within each segment induced by the workers' response resembles the interaction in the one application model, with some adjustments for the spillovers of one segment onto the other (see section 2.2.2).

The only other directed search models that allow for multiple applications are Albrecht, Gautier and Vroman (forthcoming) and the model presented in Chapter 1. Both models consider a fixed number of applications per worker. The main difference arises in the allocation of workers to firms on the given network. Both models restrict recall. If a firm has at least one applicant, it makes exactly one job offer to one applicant. If that applicant takes another job the firm remains vacant.[33] This has the immediate feature that in both models too many applications lead to congestion because too often several firms offer a job to the same worker and one of the firms remains vacant. Therefore these models cannot converge to the (unconstrained efficient) competitive outcome with increasing numbers of applications.

Albrecht, Gautier and Vroman also differ in their assumption regarding the division of surplus. They assume that firms pay at least the posted wage, but engage in Bertrand competition if two of them offer a job to the same worker. They consider equilibria with a single posted wage, and show that the equilibrium wage offer equals the workers' outside option. This arises because a higher posted wage does not yield any advantage if the worker gets two offers. While paid wages are in part higher due to the Betrand competition, the low offered wage nevertheless leads to excessive entry. This remains even in an extension in which firms can recall one additional worker.

The model in Chapter 1 is closer to the model considered here in assuming commitment to the offered wage. Firms make a single offer, and workers take the highest one. Despite the difference in recall, the equilibrium interaction also features wage dispersion with the

[32]See the introduction.

[33]No recall corresponds to $T = 1$ in the matching stage of this paper.

number of wages corresponding to the number of applications. While the worker's problem is quite different in the way that strategies of other agents translate into the relevant hiring probabilities, the structural trade-offs for each worker are in fact similar, implying the separation property (lemma 3) in both models. This suggests a robustness of the equilibrium structure to the specifics of the recall process. The search process in Chapter 1 is inefficient because wage dispersion leads to different hiring probabilities, but efficient search would equalize the hiring probabilities. That is, one wage would be optimal.[34]

This paper incorporates search costs from the outstart, and introduces a tractable recall process for models with multiple applications. Constrained efficiency obtains because firms can "price" their productive input, which is the queue of effective applications. As explained in section 2.3, this arises because workers also care about the effective queue length and adjust it in response to the wage announcements.

Without recall, firms and workers care about different things. Firms are interested in effective applications. But workers now care about all applications, no matter if these are effective or non-effective: If another worker applies and gets an offer, the job is lost even if that worker takes a better position. If a firm raises the wage (keeping the other wages constant), workers adjust their applications such that the gross queue length reaches a new level of indifference. Therefore firms can only "price" the gross queue length, but not the effective queue length that they really care about.[35] The model considered here shows that a more efficient allocation on the given network translates into efficient market interaction in earlier stages of entry, wage setting and applying.

The recall process that we specify is a limit version of the deferred acceptance algorithm, introduced by Gale and Shapley (1962) to obtain stable matchings in the marriage market. Bulow and Levin (forthcoming) also study non-cooperative wage commitments prior to non-transferable utility matching. Like most papers on the marriage market, they allow for heterogeneities but neglect limited "links". Exceptions are Roth and Perason (1999) and Immorlica and Mahdian (2005), who consider a random network of links. This study neglects heterogeneities other than through wages, but treats link formation as an active

[34]See footnote 27 for the difference in terms of search efficiency.

[35]This is likely to arise with any finite recall, i.e. finite T: In the final period T, the firm cares about effective applicants while the workers' care about all applicants. Since only few firms make offers in the final period as T becomes large, this effect disappears in the limit.

choice. The equilibrium wage dispersion leads to a structure on the network that strictly improves the number of matches over a fully random network.

2.5.3 Conclusion of Chapter 2

This paper incorporates a micro-foundation for search intensity into a directed search framework. Directed search can here be interpreted as strategic but frictional link formation between workers and firms. Search intensity can be viewed as a choice on the number of links that the worker wants to obtain. We consider a stable allocation on the network arising from a process in which firms contact ("recall") additional applicants if their offer gets rejected. Firms' wage announcements price the network efficiently, given the workers' coordination problem. Equilibrium wage dispersion turns out to be the optimal response of the market to the presence of frictions, since it allows for a network structure that minimizes coordination failure. While other work has shown that multiple applications lead to inefficiencies in a directed search setting when recall is restricted, we show that constrained efficiency prevails in the presence of strategic search intensity when recall is allowed.

While we focused our attention on efficiency properties, the relative ease with which search intensity and recall can be incorporated suggests that the model can be applied to answer wider questions. In a first step we considered the connection between productivity and search intensity: Lower search costs (or equivalently higher productivity) imply more search and in the limit the equilibrium outcome approximates the unconstrained efficient competitive allocation. Additional interesting questions concern the interaction between simultaneous and sequential search in a repeated labor market interaction.[36] We expect simultaneous search to dominate in markets with long time-frames between applications and final hiring decisions. The introduction of heterogeneity is also left for future research. If only firms are heterogeneous in terms of productivity, it seems likely that higher productivity firms will offer higher wages, and we expect firms' wage offers to be clustered with the number of clusters equal to the number of applications (similar to the mass points in this analysis). For two-sided heterogeneity the matching process will have to be adapted to account for the firms preferences over different workers.

[36]By an approach similar to the extension in Chapter 1 section 1.5.3 one can show that with exogenous separations the steady-state of a repeated interaction looks similar to the one-shot interaction analyzed here.

Chapter 3

Strategic Firms and Endogenous Consumer Emulation

"One very clear impression I had of all the Beautiful People was their prudence. It may be that they paid for their own airline tickets but they paid for little else."

James Brady, Press Secretary to Ronald Reagan

From *Superchic*, Little, Brown 1974

There is a large literature on social learning in which consumers make inferences about the quality of a good by observing what other consumers have done in the past. Banerjee (1992) and Bikhchandani, Hirshleifer and Welch (1992) analyze the case in which a sequence of identical individuals consume once and prices are fixed. Bose, Orosel, Ottaviani and Vesterland (1992a,b) extend the analysis to the case in which a monopoly seller facing a sequence of identical one-time buyers behaves strategically in setting his price.[1][2] In this paper we analyze an infinite horizon problem in which a large number of firms with differing qualities face a large number of repeat buyers who make inferences about the quality of

[1]Other work on social learning that is less related to the current paper includes Smallwood and Conlisk (1979) and Ellison and Fudenberg (1995) who consider information transmission of consumption choices or consumption outcomes, respectively, between boundedly rational consumers. Banerjee and Fudenberg (2004) extend this to word-of-mouth communication among rational agents.

[2]There is research in other fields on the degree to which consumer choice is influenced by other people. See Rogers (1995) for an overview over the marketing literature and for contributions to consumer research and reference groups see e.g. Bearden and Etzel (1982), Bearden et al. (1989) and Burnkrant and Cousineau (1975). In psychology see e.g. Cohen and Golden (1972) and Pincus and Waters (1977).

firms that other individuals have patronized in the past. Buyers are heterogeneous, and correct inference of an uninformed individual depends on which buyers are better informed, which in turn depends on the (endogenous) frequencies of purchase decisions.

We consider a market environment in which consumers are initially uninformed about firms' qualities, but will know a firm's quality perfectly after purchasing once. Some individuals consume more frequently than others, hence are on average better informed about quality. When uninformed consumers can identify those who are likely to purchase more frequently, they will optimally emulate a frequent purchaser.[3] Because of the follow-on business of these frequent purchasers, they will (in equilibrium) be rewarded by firms they patronize. This will happen even though they do not pay more than others, and the price they pay does not cover the costs of the preferential treatment. Sales to these more frequent purchasers are in a sense "loss-leaders".

In our model consumers are heterogeneous with respect to income, which is assumed observable. We assume that the good of unknown quality is normal, so the relatively wealthy consume more frequently, all else equal, and acquire information faster. Each consumer observes the choices of some other customers in the market. Individuals who have not found a product of acceptable quality have an incentive to buy from the same firms as the customers they observe, in the hope that those customers made informed decisions. When wealthier customers acquire information faster, other customers will follow their decisions rather than the decisions of poorer customers, if they observe both. Wealthier customers will then attract additional business to the firm they patronize. If the firm can enhance the buying experience of a customer by providing costly service, it will provide the service to their customers who purchase more frequently, both to prevent them from switching to a competitor or to induce them to consume more frequently. While typically wealthier customers receive preferential treatment, there may be additional equilibria in which special service is given to less affluent consumers. This arises because service reinforces the learning process, and exclusive service to the poorer consumers can lead them to purchase more frequently than the wealthy. This can only arise if service is sufficiently important.

---

[3]The idea that quality might only be varified through purchases and subsequent consumption goes back at least to Nelson's (1970) concept of experience goods. He suggests that the pattern of an individual's repeated purchases might not be random, but incorporate the information of others, a process he terms guided sampling. We formalize the idea that this guiding might evolve endogenously with firms strategically engaged in the guided sampling.

In the analysis we focus on differences in income as the only source of heterogeneity between consumers. We will show that in a simple setting this translates into different opportunity costs in the search process. It will become clear in the analysis that the difference in the opportunity cost drives the results. Thus, whenever there is consumer heterogeneity that leads to one group consuming more frequently, uninformed consumers will follow the choices of members of the frequently-consuming group when possible. If the difference is observable, our model can be interpreted more broadly as an analysis of the interplay between consumer search and firm competition in a market with two-sided heterogeneity and service as the competitive strategic variable.[4]

We focus on income or wealth as the basis of heterogeneity for two reasons. First, for normal goods higher income is associated with higher consumption, which is the driving force of information acquisition in our model. Second, while income or wealth might be difficult to observe and usually have to be inferred by secondary characteristics, this might still be easier than to observe somebody's taste preferences. Observability is crucial in this setup since it is essential to a consumer's decision about whom to follow.[5]

The paper is structured as follows. The next section introduces the model. Section 3 derives the equilibrium. To reduce complexity of the analysis, we postpone interesting but straightforward discussions of the role of conspicuous consumption, visibility in the market and relative position as well as some robustness checks to section 4. Omitted proofs and derivations are provided in the appendix.

3.1 The model

This section sets out the model and the equilibrium concept, followed by a discussion of the main modelling choices. There is a countably infinite number of periods and a continuum I of consumers and a continuum J of firms. Consumers are heterogeneous with respect to their per-period income. Each consumer $i \in I$ has a type $\theta_i \in \{p, w\}$, indicating whether he

[4]In such a world service can in fact be reinterpreted as a price reduction, leading to price as the competitive variable. Note also that heterogeneity of opportunity costs in our model is identical to heterogeneity in valuations for the good.

[5]We discuss in the last section how conspicuous consumption may arise when wealth must be inferred rather than being directly observed.

is poor or wealthy. The proportion of wealthy consumers with income y_w is $\alpha \in (0,1)$. The other consumers have income $y_p < y_w$.

Firms are infinitely lived and heterogenous with respect to the quality $q \in \{q_l, q_h\}$, $q_l < q_h$, of the otherwise identical good that they produce. We denote the proportion of firms with quality q_h by $\lambda \in (0,1)$.

Consumers' types are public information. A firm's type is initially known only to the firm, and is fully revealed to a consumer after consumption of the firm's output. Consumers die each period with probability $(1-\delta)$; when a consumer dies, a new consumer of the same type is born; new agents know only the proportion of high quality firms.

The firm's problem

Each firm $j \in J$ supplies an indivisible good, the quality of which is exogenous and unchanging over time. The market price of the good, $P > 0$, is exogenously given and identical for all firms.[6] Our focus is on firms' interest in attracting customers, and for simplicity we assume that the good can be produced costlessly. We assume that the firm chooses whether to provide service to a given customer, and denote the level of service by $s \in \{0, \bar{s}\}$, $\bar{s} > 0$, where 0 denotes no service. At the time of service provision, the customer is already locked in and cannot switch to a competitor for the current period. The cost $c(\bar{s})$ of providing service is c and is incurred in the period in which the service is provided. We assume $c > P$. There is no cost to the firm if service is not provided, i.e. $c(0) = 0$.

Firms can commit to any current customer to give service the next time he returns. More specifically, we model firm j's choice $s^t_{j,i}$ in period t for consumer i as representing the firm's one-period-ahead service commitment. $s^t_{j,i}$ is the promise to provide this service level in the first period $\tau > t$ that the customer returns. First period service is zero.

Let $I^t_j(\theta)$ be the set of consumers of type θ consuming products of firm j in period t, and $\hat{s}^t_{j,i}$ firm $j's$ service provision to consumer i. Then firm $j's$ per period profit is

$$\pi^t_j = \int_{i \in I^t_j(w)} P - c(\hat{s}^t_{j,i}) di + \int_{i \in I^t_j(p)} P - c(\hat{s}^t_{j,i}) di.$$

Firms maximize the discounted present value of per period profits, $\sum_{t=0}^{\infty} \beta^t E \pi^t_j$.

The cost of the provision of service is shown in the per period profit expression above while the benefits are indirect. A firm that promises service to an individual consumer may deter the consumer from switching to a competitor or may increase the frequency with which he

[6] We argue in the discussion section that endogenizing the price would not qualitatively affect our results.

patronizes the firm. Furthermore, the consumer's choice may affect the future choices of others. These (potential) benefits to a firm that provides service are reflected in the size of the set of consumers who consume at the firm in the future. As a tie-breaking rule we assume that firms offer service when indifferent.

The consumer's problem

Consumers are heterogeneous with respect to income, but all can afford the product. That is, $y_w > y_p > P$. Income is non-storable. In each period $t \in T$ each consumer $i \in I$ has the two choices: Enter the market or not, and if entering the market, from which firm to consume. If he does not enter, that is, does not purchase the good, he spends his income on a numeraire good. Since income is non-storable and the price of the numeraire is normalized to one, the consumer obtains y units of the numeraire in the case that he does not consume in the market, and $y - P$ units if he does.

At the beginning of each period, before the consumption decision is made, a taste shock ρ is realized for each consumer that affects the degree to which he enjoys consuming the indivisible good in that period. We assume this shock is an i.i.d. draw from distribution F with density f and full support on $[\underline{\rho}, \bar{\rho}]$, where $-\infty \leq \underline{\rho} < \bar{\rho} \leq \infty$. If the consumer decides to enter the market and consume from firm j in period t his utility in that period is

$$U^t = q_j + \hat{s}_j^t + \rho^t + u(y - P),$$

where q_j is the quality of firm j, $\hat{s}_j^t$ is the service that he receives and ρ^t is the current period taste shock. $u(.)$ denotes the utility derived from the numeraire, which is assumed to be increasing and strictly concave.

If the consumer is uninformed and chooses a firm randomly, his expected utility is

$$EU^t = E_j[q_j] + \rho^t + u(y - P).$$

If the consumer decides not to consume, his utility for that period is $U^t = u(y)$. Consumers maximize the expected discounted utility $\sum_{t=0}^{\infty} \delta^t EU^t$, where $\delta \in (0, 1)$ is the probability of survival.

We assume that observing other consumers' behavior partially substitutes for an individual's initial lack of information about product qualities. After the first time a consumer purchases the indivisible good, he can costlessly observe at which firm a random wealthy consumer and a random poor consumer consumed in the previous period.[7] This implies that only players who participated in the market in the previous period are observable. For ease of exposition we assume that if a consumer is indifferent between following other participants' choices observed at different periods, he follows the most recent observation.

Equilibrium

In the spirit of Markov perfection we are interested in equilibria in which firms and consumers base their decisions only on information that is relevant for their future payoffs. We also restrict attention to anonymous strategies which condition only on the type (i.e., wealthy or poor) but not on the name of other players. For firms, the minimal payoff relevant information is the type of the consumer. We therefore consider equilibrium strategies $s(\theta)$ in which the service commitment of a firm is a function of the consumer's type. We will consider symmetric pure strategies, that is, $s(\theta)$ is deterministic and the same for firms with the same type. We denote by $s_l(\cdot)$ the strategy for low and by $s_h(\cdot)$ the strategy for high quality firms.

For a consumer, the relevant information is the combination of quality and service he can obtain; the name of the firm(s) from which he can receive that combination is not important. Consequently, a consumer conditions his actions on $D \subseteq \{q_l, q_h\} \times \{0, \bar{s}\}$, where D denotes the various quality-service combinations he has experienced. If the consumer has not yet purchased, $D = \emptyset$. A strategy for a consumer of type $\theta \in \{p, w\}$ is then a tuple $(\hat{\rho}^\theta(D), \sigma^\theta(D))$ for all D. The term $\hat{\rho}^\theta(D)$ denotes a reservation value for the taste shock: If the taste shock is above $\hat{\rho}^\theta(D)$, the consumer buys the product, otherwise he does not.[8] If he chooses to buy, $\sigma^\theta(D)$ specifies the choice of firm from which he purchases the good. If $D \neq \emptyset$, then the consumer can return to a firm with quality-service combination $(q, s) \in D$, he can follow the choice of either a wealthy or a poor player observed in the previous period, or he can search randomly for a new firm. If $D = \emptyset$ only the last option is available, as by

[7] In section 3.3.2 we discuss alternative signal technologies. Observing more than one player of any type does not alter any results. Observing only a random selection of N players' choices each period should not alter any qualitative results as long as N is sufficiently large.

The specification that newborn players in the market have no information about other players' choices prevents a situation in which everybody is following other market participants and no player is searching randomly. The specification that only players who have consumed at least once in the market can observe other players simplifies the analysis.

[8] In the appendix we show that reservation strategies are indeed optimal in our environment.

assumption the consumer must search on his own in the first period of consumption.

Let $n^{\theta,t}(D)$ denote the measure of type θ consumers with information D at time t, where the law of motion is determined by the strategies of consumers and firms. With this we can define a stationary equilibrium encompassing stationary distributions (1), consumer optimality (2) and firm optimality (3):

Definition 5 (Stationary Equilibrium) *A stationary equilibrium is a tuple* $S = (s_l, s_h, (\hat{\rho}^w, \sigma^w), (\hat{\rho}^p, \sigma^p), n^w, n^p)$ *such that*

1. $n^{\theta,t}(D) = n^{\theta}(D)\ \forall t\ \forall D\ \forall \theta.$
2. *For each consumer of type* θ*, strategy* $(\hat{\rho}^{\theta}, \sigma^{\theta})$ *is optimal in the continuation game for all* D*, when the consumer takes as given the strategies and fractions of the other players as summarized in* S*.*
3. *For each high (low) quality firm* s_h (s_l) *is optimal given* S*.*

Discussion of modelling assumptions

We will briefly discuss the modelling choices. A discussion of alternative assumptions and their impact on our results is relegated so section 3.3.2.

Firms. We allow one-period-ahead commitment in order to eliminate implausible equilibria. In particular, without commitment there is always an equilibrium in which a firm does not provide service because the firm cannot convince the customer that he will also get service in the future. In that case, giving service today is costly while it does not change the future behavior of the customer, which implies that it is unprofitable to provide service today. Therefore, even if the firm would like to give service to deter the customer from switching to a competitor or to encourage the customer to consume more frequently, it omits service because it cannot affect the customer's belief about future service through current period action. This arises no matter how profitable the business with this customer is. We restrict attention to time-independent firm strategies to focus on the effects of information transmission and to abstract from repeated game effects. Without commitment, this implicitly restricts the beliefs of the consumer to be unaffected by current period actions, rendering current period service an ineffective tool to change consumer behavior. Limited commitment power, i.e., commitment only for the next time in which the customer returns, provides a tool by which a customer's beliefs about future service can be altered. This allows firms to provide service if the business with the consumer is sufficiently profitable. In the last section we argue that the equilibrium with commitment will also be an equilibrium of the game without commitment when we allow for non-Markovian strategies.

The assumption that no service is given during the first period of consumption at a firm is also due to our restriction to Markovian strategies. If firms have a choice to give more service than promised, they would not do so since they cannot influence the consumer's future behavior. Since there is no commitment to service in the first period, they would not provide any service.

Taking prices as exogenously given allows us to focus on private information that is not fully revealed through prices. That prices do not reveal all relevant information about products is widely accepted. Strong restrictions on pricing behavior are therefore common in models of this sort to preclude revelation of too much information (see, for example, Wolinsky (1990)). On the other hand we can, with slight modification, interpret the service as a price reduction. In this sense we do allow for price competition. We discuss the possibility of price competition further in section 3.3.2.

Consumers. The combination of the numeraire good as an alternative to market consumption and the taste shock together explicitly capture the idea that the good under consideration is a normal good. The opportunity cost of going into the market is

$$u_0 := u(y) - u(y - P),$$

that is, the opportunity cost of foregone consumption of the numeraire good. Denote this opportunity cost u_0^w for the wealthy and u_0^p for the poor. The strict concavity of $u(\cdot)$ then implies that $u_0^w < u_0^p$, that is, the wealthy have a lower opportunity cost of consumption than the poor.[9] Without service the wealthy will therefore consume at lower values of the taste shock than the poor. Thus, on average the wealthy consume more often, which establishes our version of the normal goods assumption.[10]

[9] Players get in each period $y - P$ units of the numeraire for sure, independent of their current period choice. Therefore wealthy players consumption-independent level of the numeraire is higher. Only the additional amount that they might get, i.e. their opportunity cost, is lower. The term $u(y - P)$ in the utility function will be dropped for all subsequent calculations as it only reflects a constant.

[10] Heterogeneity in the opportunity costs of consumption (rather than the heterogeneity in terms of income) can be taken as primitive to allow for more general interpretations of the model. See section 3.3.1 for a discussion.

The taste shock also allows firms to encourage the customer to consume more frequently. The timing of when to consume in the market is not exogenously fixed, but rather depends on the current period taste shock and the utility of consumption. By promising service, the firms can raise the utility of consumption and can thus encourage a consumer to consume more frequently.

Our results are most interesting when $c > P$. In this case no consumer will receive service only because of his own consumption at a firm. Firms will only provide service because a consumer brings in additional customers who follow his lead. Consequently, this case clearly highlights the effects of information transmission in the market.

Equilibrium. The equilibrium concept requires optimality in the continuation after a deviation to ensure that the firms' service promises affect behavior.

3.2 Optimal Behavior

As a first step to characterizing the stationary equilibria of this game, in the following subsection we describe the optimal consumption behavior of a single consumer. Toward this end, we consider a partial problem in which the service provision by firms and the information in the market is exogenous. In the next subsection we endogenize the market information, still under the assumption that firms service provision is exogenously given. The subsequent subsection analyzes the behavior of the firms. Finally, we integrate the parts in the analysis of the equilibria.

3.2.1 Consumer Search

Consider a consumer with opportunity cost u_0 who has entered the market at least once, that is, he will observe other participants in the market. Suppose all high quality firms offer identical service s (either $\bar{s}$ or zero) to this consumer in every period, and low quality firms do not offer more service than high quality firms.[11] If the consumer has not found a high quality firm and chooses to purchase from a firm that he hasn't previously frequented, there is a probability which we denote γ that that firm will be of high quality. We take for now as exogenous the process by which this consumer chooses a new firm, and hence γ.

[11] We will later show that it is always profitable for a high-quality firm to provide service if it is profitable for a low-quality firm to provide service. Therefore this specification will be the relevant case.

We are interested in the optimal decision rule of the consumer. The problem is a standard search problem with one exception – the consumer does not search in every period. Rather, the choice to enter the market is endogenous and depends on the taste shock, where high taste shocks imply that he enjoys consumption of the uncertain good relatively more than when the taste shock is low. It also depends on the qualities and service promises of the firms he has encountered, as well as on his beliefs about the service he will be promised by firms he has not yet encountered. The decisions about how frequently to search and how to choose firms given the frequency of search are interlinked: a higher frequency of search effectively implies a higher discount factor in the choice of firms. Higher quality or higher service imply a higher value of consumption, thus increasing the frequency of consumption.

Using standard techniques, appropriately modified to this setting, we derive cutoff levels $\hat{\rho}_l$ and $\hat{\rho}_h$ for the taste shock as functions of the quality of the best firm encountered so far, for which the consumer is indifferent between consuming in the market and not consuming. When the consumer has only experienced low quality firms, he will enter the market only if his taste shock ρ exceeds the cutoff $\hat{\rho}_l$, and his frequency of consumption is thus $(1 - F(\hat{\rho}_l))$. If the customer has experienced at least one high quality firm, his relevant cutoff is $\hat{\rho}_h$ and his frequency becomes $(1 - F(\hat{\rho}_h))$. Both $\hat{\rho}_l$ and $\hat{\rho}_h$ depend on the service strategies of the firms.

To ensure that a cutoff exists that leads to this indifference, we make the following assumption on the support of the taste shock that we will retain throughout: $(\underline{\rho}, \bar{\rho}) \supset [u_0^w - q_h - \bar{s}, u_0^p - q_l]$. This implies that $\underline{\rho} + q_h + \bar{s} < u_0^w$, i.e., even in the most advantageous situation of high quality and high service, there are taste shocks sufficiently low such that not consuming is preferable. Similarly, it implies that $q_l + \bar{\rho} > u_0^p$, so that even in the most disadvantageous situation some taste shocks still induce the consumer to enter the market. Requiring the support to be sufficiently large avoids the discussion of boundary cases. Note that we consider different ranges for $\bar{s}$ in some of our statements. If a range conflicts the first inequality, it is vacuous by assumption. Yet for any shock distribution that has unbounded lower support, i.e. $\underline{\rho} = -\infty$, the inequality above does not restrict $\bar{s}$ in any way and all ranges that we consider are possible.

Before stating the results on the consumer's search behavior, it should be noted that the consumer might not search for high quality firms if all firms offer service since the consumer does not receive service in his first period of consumption at a firm. If he returns to a low quality firm, he will receive service immediately, while he will not receive any service if he continues searching for a high quality firm until he consumes at least twice at such a firm. However, if service is not too important, or the consumer is sufficiently patient, he will search for high quality firms rather than remaining with a low quality firm. We summarize this in the following lemma, which is proven in the appendix.

Lemma 6 *If all firms offer service, there exists $\underline{\delta} \in [0, 1)$ such that for $\delta \geq \underline{\delta}$ the consumer searches for high quality. If $\bar{s} < \gamma(q_h - q_l)$, consumers always search for high quality firms, i.e. $\underline{\delta} = 0$.*

The following lemma describes the optimal service strategy for the consumer. Recall that s denotes the service promised by high quality firms to this consumer, which is assumed weakly larger than the service promised by low quality firms. Let

$$\hat{\rho}_l = u_0 - E_\gamma(q) - \frac{\delta\gamma}{1-\delta} \int_{u_0 - q_h - s}^{\hat{\rho}_l} [1 - F(\rho)] d\rho \tag{3.1}$$

where $E_\gamma(q) = \gamma q_l + (1 - \gamma) q_h$. Let the state variable q be the best quality yet encountered. Then we obtain:

Lemma 7 *If $\delta \in (\underline{\delta}, 1)$ and high wage firms offer weakly higher service, then the consumer's optimal decision rule has the following structure:*
If $q = q_l$, sample a new firm if current period shock $\rho \geq \hat{\rho}_l$, otherwise don't consume. If $q = q_h$, then return to the firm with high quality if the current period shock $\rho \geq \hat{\rho}_h = u_0 - q_h - s$, otherwise don't consume.

Intuitively these cutoffs are easy to understand. At high qualities the consumer either gets u_0 or $q_h + s + \rho$, and he chooses the higher one. At low qualities the trade-off is similar, except that consumption yields only the average quality $E_\gamma(q)$. On the other hand the frequency of consumption in the future will be higher, because of the possibility of finding a high quality firm. This is reflected in the increment in terms of frequency in the last term on the right hand side of (3.1).
The primary interest in the lemma stems from its implications for the behavior of wealthy consumers, i.e., those with the lower opportunity cost, relative to the behavior of poorer consumers. To compare the two groups, consider a setting where high quality firms promise service $s_h(p)$ to poor consumers and $s_h(w)$ to wealthy consumers, and low quality firms offer

at most this level of service: $s_h(\cdot) \geq s_l(\cdot)$. Let $\hat{\rho}_l^p$ and $\hat{\rho}_h^p$ be the threshold levels for a poor consumer, and $\hat{\rho}_l^w$ and $\hat{\rho}_h^w$ be the threshold levels for a wealthy consumer. The following lemma compares these cutoff values:

Lemma 8 *Let $\delta \in (\underline{\delta}, 1)$ and $s_h(\cdot) \geq s_l(\cdot)$.*

1) *If $s_h(w) \geq s_h(p)$, then the wealthy consume more frequently than the poor: $\hat{\rho}_l^w \leq \hat{\rho}_l^p$ and $\hat{\rho}_h^w \leq \hat{\rho}_h^p$.*

2 a) *If $s_h(p) - s_h(w) = \bar{s} < u_0^p - u_0^w$, still $\hat{\rho}_l^w < \hat{\rho}_l^p$ and $\hat{\rho}_h^w < \hat{\rho}_h^p$.*

2 b) *If $s_h(p) - s_h(w) = \bar{s} > u_0^p - u_0^w$, then $\hat{\rho}_h^w > \hat{\rho}_h^p$.*

There exists unique $\xi_\gamma > u_0^p - u_0^w$ such that $\hat{\rho}_l^w < \hat{\rho}_l^p$ if $\bar{s} < \xi_\gamma$ and $\hat{\rho}_l^w > \hat{\rho}_l^p$ if $\bar{s} > \xi_\gamma$.

Proof: The result for the cutoff $\hat{\rho}_h$ follows directly from $\hat{\rho}_h^\theta = u_0^\theta - q_h - s_h(\theta)$, $\theta \in \{p, w\}$. For $\hat{\rho}_l$, rewrite (3.1) as $\hat{\rho}_l^\theta - u_0^\theta + E_\gamma(q) = -\frac{\delta\gamma}{1-\delta}\int_{u_0^\theta - q_h - s_h(\theta)}^{\hat{\rho}_l^\theta}[1-F(\rho)]d\rho$ and observe that the left hand side is increasing in $\hat{\rho}_l^\theta$ and decreasing in u_0^θ and the right hand side is decreasing in $\hat{\rho}_l^\theta$ and increasing in $u_0^\theta - s_h(\theta)$. For 1) and $2a$) the wealthy have lower u_0^θ and $u_0^\theta - s_h(\theta)$, therefore their cutoff $\hat{\rho}_l^w$ must be lower for the equality to hold. For $2b$), if $u_0^p - s_h(p) \approx u_0^w - s_h(w)$, then $\hat{\rho}_l^w < \hat{\rho}_l^p$ since $u_0^w < u_0^p$. Since $\hat{\rho}_l^p$ is by (3.1) strictly increasing and unbounded in $\bar{s}$ when $s_h(p) - s_h(w) = \bar{s}$ but $\hat{\rho}_l^h$ is constant, there exists a unique ξ_γ such that $\hat{\rho}_l^w = \hat{\rho}_l^p$ if $\bar{s} = \xi_\gamma$. *Q.E.D.*

If wealthy and poor consumers are treated equally by firms, this result simply restates our formulation of the normal goods assumption. The wealthy consume more frequently both in the search phase and after they have found a high quality firm. The explicit formulation allows us to compare the frequency of consumption even in the cases where consumers are treated differently by firms. As long as the service benefit does not outweigh the differences in the opportunity costs of consumption, wealthy consumers still consume more frequently even if poor consumers are treated preferentially. If the impact of service outweighs the difference in the opportunity costs, the frequency of consumption at high quality firms becomes larger for the poor than for the wealthy. This does not necessarily translate into a higher frequency of consumption at low quality firms. As long as consumers are searching they do not obtain service, and without service the wealthy have a stronger incentive to consume. Only if the service at high quality firms is sufficiently attractive, poor consumers search more frequently for high qualities than the wealthy. Otherwise it still means that wealthy consumers find high quality firms relatively faster, even if service outweighs the exogenous difference in opportunity costs.

We turn next to endogenizing the probability γ.

3.2.2 Consumers' stationary behavior

In equilibrium, both high and low quality firms decide whether to provide service to each of the two types of consumers. Before examining the full model, we take the firms' choices regarding service as fixed and examine consumers' behavior in steady state as they choose optimally, given firms' choices. We will again consider the case where all high quality firms offer weakly higher service than low quality firms.

In this case all consumers, wealthy and poor, will sample in a way that gives the highest probability of identifying a high quality firm. Thus, both the uninformed wealthy consumers and the uninformed poor follow the same group. The probability of drawing a high quality firm, γ, when following consumers of this group (which we will call "leaders") is now endogenous. γ depends on the particular stationary equilibrium we are looking at.

We note that following *any* consumer, wealthy or poor, is preferable to searching randomly. At worst, that consumer who is followed has not found a high quality firm yet, in which case the firm he or she purchased from is as likely to be high quality as a randomly sampled firm. In addition, there is positive probability that the consumer who is being mimicked has found a high quality firm and purchases only from that firm. Hence, it is strictly better to follow *any* other consumer than to sample randomly; thus, in any equilibrium $\gamma \geq \lambda$. Due to this inequality $\underline{\delta}$ in lemma 6 can be established independent of the exact value of γ, and long-lived consumers indeed search for high qualities.

An individual who follows the leaders sees only those who have consumed in the previous period. The probability that this individual will find a high quality firm is the proportion γ of the leaders who have identified a high quality firm and who consumed in the previous period. When the wealthy identify high quality firms with probability γ when following the leaders, a fraction γ^w of the wealthy who consume in any given period will purchase from a high quality firm. We will first show how γ^w is determined. If the wealthy themselves are the leaders, we then have to show that $\gamma = \gamma^w$, i.e. that a fixed point exists. Given also that the poor follow the leaders, a fraction γ^p of the poor who purchase in any given period will do so from a high quality firm. If the poor are the leaders, the fixed point will be $\gamma = \gamma^p$.

The goal of this section is to establish conditions under which $\gamma^w > \gamma^p$ and vice versa. This will determine which group is being followed, since an uninformed consumer will follow the group with the higher γ^θ.

To calculate γ^θ, $\theta \in \{p, w\}$, we will consider each group individually. We will focus on the wealthy, but the derivations for the poor are analogous when replacing w with p. In the first period in which a wealthy person consumes, he samples a firm randomly and has probability λ of drawing a high quality. In every period thereafter he draws a high quality firm with probability γ whenever he searches. To derive the stationary distribution we must keep track of the proportion of wealthy consumers whose best quality encountered so far is q_l, q_h or $\emptyset$ respectively, where $\emptyset$ stands for those who have not yet consumed. In period t denote these by $n_l^{w,t}$, $n_h^{w,t}$ and $n_\emptyset^{w,t}$.

The relevant flow equations can then be constructed as follows: In period $t+1$ consumers who have not tried any product include all newborns and those consumers who had not yet consumed at the beginning of period t, did not consume a product in period t and survived:

$$n_\emptyset^{w,t+1} = (1-\delta) + \delta F(\hat{\rho}_\emptyset^w) n_\emptyset^{w,t}, \tag{3.2}$$

where $F(\hat{\rho}_\emptyset^w) > 0$ is the probability that a wealthy person who does not yet observe other market participants does not consume in the market. The cutoff $\hat{\rho}_\emptyset^w$ is analytically complicated,[12] but our specification that in addition to the newborn, all consumers prior to their first purchase lack information about other market participants, eliminates $F(\hat{\rho}_\emptyset^w)$ in the derivation of γ^w.

Wealthy consumers in $t+1$ who have state variable q_l include those without information at the beginning of period t who consumed a product in t, drew quality q_l and survived; those who had not found a satisfactory quality before t, drew quality q_l in t and survived; and those survivors from the prior period who did not consume a product and had experienced

[12]For a given γ, the taste shock $\hat{\rho}_\emptyset^w$ is characterized by the indifference of the customer between going into the market and sampling a random firm vs. taking his outside option. If he goes into the market, his continuation payoff $EV_{\rho'}(q, \rho')$ is given in the appendix in (A.44) and (A.46). Let $X = \lambda\left[(1-\delta)q_h + \delta EV_{\rho'}(q_h, \rho')\right] + (1-\lambda)\left[(1-\delta)q_l + \delta EV_{\rho'}(q_l, \rho')\right]$, then $\hat{\rho}_\emptyset^w \in (\underline{\rho}, \bar{\rho})$ is characterized by $[1-\delta F(\hat{\rho}_\emptyset^w)][(1-\delta)\hat{\rho}_\emptyset^w + X] = (1-\delta)u_0^w + \delta\left[(1-F(\hat{\rho}_\emptyset^w))X + \int_{\hat{\rho}_\emptyset^w}^{\bar{\rho}} \rho dF(\rho)\right]$.

quality q_l before:

$$\begin{aligned} n_l^{w,t+1} &= \delta\left[1-F(\hat{\rho}_\emptyset^w)\right](1-\lambda)n_\emptyset^{w,t} \\ &+\delta\left[1-F(\hat{\rho}_l^w)\right](1-\gamma)n_l^{w,t} \\ &+\delta F(\hat{\rho}_l^w)n_l^{w,t} \end{aligned} \tag{3.3}$$

The cutoff $\hat{\rho}_l^w$ is given by (3.1) under the opportunity cost u_0^w.

Finally, the people who have state variable q_h constitute

$$\begin{aligned} n_h^{w,t+1} &= \delta\left[1-F(\hat{\rho}_\emptyset^w)\right]\lambda n_\emptyset^{w,t} \\ &+\delta\left[1-F(\hat{\rho}_l^w)\right]\gamma n_l^{w,t} \\ &+\delta n_h^{w,t} \end{aligned} \tag{3.4}$$

which is similar to (3.3) except for the last line: Since at q_h people return to the same firm, their state variable is q_h regardless of whether they consumed last period.

Stationarity is characterized by $n_\omega^{w,t'} = n_\omega^{w,t} = n_\omega^w$ for all t and t' and $\omega \in \{\emptyset, l, h\}$. We can use equations (3.2) to (3.4) to get

$$n_\emptyset^w = \frac{(1-\delta)}{1-\delta+\delta A^w}, \tag{3.5}$$

$$n_l^w = \frac{\delta n_\emptyset A^w(1-\lambda)}{1-\delta+\delta\gamma B^w}, \tag{3.6}$$

and

$$n_h^w = \frac{1}{1-\delta}\frac{\delta n_\emptyset A^w[(1-\delta)\lambda+\delta\gamma B^w]}{1-\delta+\delta\gamma B^w}, \tag{3.7}$$

where $A^w \equiv [1-F(\hat{\rho}_\emptyset^w)]$ represents the frequency of consumption for a wealthy consumer who has never consumed the product before, and $B^w \equiv 1-F(\hat{\rho}_l^w)$ represents the frequency of consumption for a wealthy consumer who has only experienced low quality firms. Since γ^w represents the fraction of wealthy consumer who consumed in a period who have found a high quality firm, we must find the measure of consumers who actually consume in any given period. Denote by φ_l^w (φ_h^w) the measure of wealthy consumers who consume at low quality (high quality) firms in any given period in the steady state.

In any period the consumers who consume at q_l are all the uninformed players $n_\emptyset^w$ who consume with probability $A^w = [1-F(\hat{\rho}_\emptyset^w)]$ and draw a low quality with probability $(1-\lambda)$, plus all informed players who have not found a high quality firm, n_l^w, who consume with frequency $B^w = [1 - F(\hat{\rho}_l^w)]$ and draw a low quality firm with probability $(1-\gamma)$. Thus we have

$$\varphi_l^w = n_\emptyset^w A^w(1-\lambda) + n_l^w B^w(1-\gamma). \tag{3.8}$$

For φ_h^w, we have similar terms for the uniformed and unsatisfied players, plus an additional term reflecting those people who had already found a firm of satisfactory quality in the past, n_h^w, times their frequency of consumption $C^w \equiv [1 - F(u_0^w - q_h - s_h(w))]$. Therefore

$$\varphi_h^w = n_\emptyset^w A^w \lambda + n_l^w B^w \gamma + n_h^w C^w.$$

Since $\gamma^w = \frac{\varphi_h^w}{\varphi_l^w + \varphi_h^w}$ we get

$$1 - \gamma^w = \frac{\varphi_l^w}{\varphi_l^w + \varphi_h^w} = \frac{n_\emptyset^w A^w(1-\lambda) + n_l^w B^w(1-\gamma)}{n_\emptyset^w A^w + n_l^w B^w + n_h^w C^w}, \tag{3.9}$$

where again A^w, B^w and C^w represent the frequencies of consumption for those wealthy consumers who have not consumed from any firm before, those who have only experienced low quality firms and those who have previously experienced a high quality firm, respectively. Substituting the value for $n_\emptyset^w$, n_l^w and n_h^w from equations (3.5) to (3.7) and rearranging yields an expression for γ^w that is independent of the initial frequency A^w, i.e.

$$\gamma^w = 1 - \frac{[1 - \delta + \delta B^w]\,[1-\lambda]}{1 - \delta + \delta(1-\lambda+\gamma)B^w + \left[\delta\lambda + \frac{\delta^2}{1-\delta} B^w \gamma\right] C^w} \tag{3.10}$$

for a given $\hat{\rho}_l^w$, which is implicitly defined in equation (3.1) in lemma 7.

The proof of the following lemma, which is left to the appendix, shows that there is a fixed point $\gamma^w = \gamma$.

Lemma 9 *There exists a fixed point $\gamma^w = \gamma$, $\gamma \in (\lambda, 1)$, in equation (3.10) such that equation (3.1) is also satisfied.*

We argued above that it is always better to follow some group than to sample randomly. Therefore in a stationary equilibrium, $\gamma = \gamma^w$ or $\gamma = \gamma^p$, where γ^p is constructed analogously. While an uninformed consumer does better by following any other consumer than searching randomly, one should expect that it is better to follow a wealthy consumer than a poor consumer. Suppose that the level of service offered to wealthy consumers is at least as large as the service offered to poor consumers. Then while uninformed, the wealthy will consume more frequently than the poor, and hence, will discover a high quality firm more quickly.

While this argument is appealing, it is not trivial. The wealthy search more often at low quality firms in any period, diluting the visibility of the wealthy who have identified a high quality firm. Partial differentiation of (3.10) reveals that $\partial\gamma^w/\partial B^w > 0$ if $\gamma > \lambda$, and therefore the dilution effect does not outweigh the greater frequency with which they consume. If the frequency of consumption while searching for a high quality firm increases for either group, this unambiguously increases the quality of their signal. Nevertheless, the visibility induced through the frequency of consumption is important: If the poor is the group that receives service (which we will show is only possible if service is sufficiently valuable), then it is possible that the poor consume less frequently prior to identifying a high quality firm than the wealthy, but it is nevertheless better to follow the poor. The reason is that the poor consume much more frequently at high quality firms, and therefore those who have identified a high quality firm are a greater proportion than the proportion of wealthy who have identified a high quality firm.
We summarize this in the following proposition the proof of which is left to the appendix. Recall that ξ_γ was introduced in lemma 8 as the threshold on service below which the poor search less frequently at low qualities even if they are the ones that receive service.

Proposition 22 *Let $\delta \in (\underline{\delta}, 1)$ and $\gamma \in (\lambda, 1)$ Consider a candidate stationary equilibrium with $s_h(\cdot) \geq s_l(\cdot)$. Then there exists $\hat{\xi}_\gamma \in (u_0^p - u_0^w, \xi_\gamma)$ such that*

1) *If $s_h(w) \geq s_h(p)$, then $\gamma^w > \gamma^p > \lambda$.*

2 a) *If $s_h(p) - s_h(w) = \bar{s} < \hat{\xi}_\gamma$, still $\gamma^w > \gamma^p > \lambda$.*

2 b) *If $s_h(p) - s_h(w) = \bar{s} > \hat{\xi}_\gamma$, then $\gamma^p > \gamma^w > \lambda$.*

Parts 1) and 2a) of the proposition provide conditions on the service level that imply that all consumers will want to follow wealthy consumers. Only if these conditions do not obtain could it be possible that it is preferable to follow the poor.

3.2.3 Firms' behavior

We next analyze service provision by the firms. We establish that it is profitable for the firm to provide service to consumers in a group if they are followed by sufficiently many consumers of the other type. We also prove that high quality firms will always provide at least as high a service level as low quality firms, which we have so far taken as exogenous.

A key observation for this result is that in any stationary equilibrium, if a high-quality firm promises in any period service $\bar{s}$ to a consumer, then the consumer will return to this firm the next period he enters the market, regardless of his expectations about future service. The intuition behind this observation is that accepting the optimal per-period outcome of high quality and high service and searching thereafter for a new firm yields higher utility than searching immediately. Therefore, high-quality firms can always ensure the return of a consumer by promising him service. The question in this section is when this will be profitable.

Our first result concerns the equilibrium profit contribution of a customer. In a stationary equilibrium, a consumer who purchases from a firm has the same expected number of followers in every period. Thus, the benefit to a firm of a single visit of a particular customer is the following: He pays price P, potentially receives service at cost $c > P$, and induces the expected discounted lifetime equilibrium profit that the firm generates from his next-period followers. Call this benefit Π. Π is an equilibrium object that depends on the strategies of the firm in question as well as the strategies of other firms and consumers. If a firm deviates and promises s' instead of the equilibrium promise s, the benefit of the next return visit is $\Pi - (c(s') - c(s))$. Since in a stationary equilibrium after a one shot deviation the continuation game is unchanged once the customer returns, only the immediate cost of service changes from $c(s)$ to $c(s')$. In particular, the behavior of the customer once he returns as well as the behavior of the followers is unchanged.[13] For example, promising no service instead of service in any period changes the benefit of the next visit of the consumer from Π to $\Pi + c$, as the firm saves the service costs next time. Yet it might delay the consumer's return, as now consumption is less valuable compared to the opportunity cost of consumption. If the consumer switches to a competitor, $\Pi + c$ will in fact never be realized.

Consequently, if a particular consumer generates sufficient indirect profit it will pay a firm to promise that consumer service to induce him to purchase more frequently. In addition, Proposition 22 provides conditions under which it is optimal to follow consumers of type θ.

[13] This is a consequence of the assumption of markov perfection which is embedded in the requirement that the equilibrium service strategy does not depend on past histories. In the situations that we analyze the consumer is not indifferent between his choices of searching for a new firm or returning to a previous firm. Furthermore, the decision situation is effectively unchanged compared to on the equilibrium path play once he returns after a one-shot deviation. Taken together, this implies that his continuation strategy once he returns will be his on-the-equilibrium-path strategy.

If the fraction of this leading group is not too large, the spillover benefits of the followers will make it profitable for high quality firms to provide service to the leaders. We state this formally in the next proposition.

Moreover, we establish that it is indeed optimal for high quality firms to outbid low quality firms in their pursuit of valuable customers. We show that if it is profitable for a low wage firm to provide service to a consumer, it is also profitable for a high wage firm to provide service in order to keep the business of this consumer. In Lemma 6 we showed that if consumers are sufficiently long-lived, that is, $\delta \in (\underline{\delta}, 1)$, they will continue to search for a high quality firm rather than return to a low quality firm. This was derived under the assumption that high quality firms offer at least weakly more service. The result here establishes that this assumption is indeed fulfilled. We summarize this in the following proposition, the proof of which is relegated to the appendix. For more compact notation of the proposition, let $\alpha^w = \alpha$ and $\alpha^p = 1 - \alpha$.

Proposition 23 *Let $\delta \in (\underline{\delta}, 1)$.*

i) *Suppose all uninformed consumers follow consumers of type θ. Then there exists $\bar{\alpha} > 0$ such that for $\alpha^\theta < \bar{\alpha}$, in any stationary equilibrium $s_h(\theta) = \bar{s}$.*

ii) *In any stationary equilibrium, either $s_h(\theta) \geq s_l(\theta)$ for $\theta \in \{p, w\}$; or $s_h(\theta) < s_l(\theta)$ but type θ consumers nevertheless do not return to low quality firms.*

We should point out that two forces can lead to high-quality firms' willingness to provide service. One is competitive pressure. If other high-quality firms offer service, then if a single firm does not provide service, the consumer might not return, preferring to search for another firm that provides service. The second is the encouragement effect, that is, the consumer returns more frequently when he is offered service. In our model the encouragement effect is important to sustain a high-service equilibrium. That is due to the Markovian assumption on the firms' strategies which implies that service provision only depends on the consumer type, and not on the firm's own or other players' past actions. In an equilibrium in which the high-quality firms are supposed to promise service, the firm's strategy specifies offering service again in the continuation game even if it deviated for a period and did not offer service. After a deviation a consumer will therefore still expect to get a service promise the next time he consumes there (even though he will not get any actual service provided in that period) and it is better for him to return than to search for a new firm in which he also will not receive any service in the first period of consumption, but might draw a low quality. Therefore, the consumer would return even if service is not provided for one period. The motivation for high quality firms to offer service then comes from the encouragement effect. If we impose less stringent assumptions on the equilibrium (in particular, dropping the restriction to Markov strategies), the competitive forces can be very important. If consumers believe that they will not be promised service by a firm ever again if it does not promise service in a given period, then service provision can be sustained by the competitive threat that consumers search for new firms that do offer service, which seems intuitively appealing. With these results on optimal strategies for consumers and firms and the steady state derivations we now turn to the stationary equilibria of the game.

3.2.4 Equilibria

We first provide a necessary condition for equilibria when the value of service is not too large. In any such equilibrium, the poor follow the wealthy, and if service is provided, it is provided only to the wealthy. This is driven by the fact that the wealthy accumulate information faster than the poor.

Proposition 24 *Let $\delta \in (\underline{\delta}, 1)$. There exists $\xi > u_0^p - u_0^w$ such that for $\bar{s} < \xi$ in any stationary equilibrium all uninformed consumers follow wealthy consumers after their initial purchase. If service is provided in equilibrium, it is only provided to the wealthy.*

Proof: By proposition 23 higher quality firms provide weakly higher service. In combination with lemma 6 we know that all consumers search for high quality firms. That is, all consumers, wealthy and poor, will in equilibrium follow the distribution that places the highest weight on high quality firms. By Proposition 22 $\bar{s} < \hat{\xi}_\gamma$ ensures that all consumers will follow the wealthy, even if the poor receive the service. Since $P < c$ service is only provided (i.e., promised and then delivered) to players who have followers. Therefore, in equilibrium only the wealthy can receive service. While $\hat{\xi}_\gamma$ depends on γ, one can show that it is bounded away from $u_0^p - u_0^w$ for all $\gamma \geq \lambda$. *Q.E.D.*

Whether stationary equilibria with or without service exist depends on the profit generated by the followers that a consumer has. Since $P < c$ service will only be provided when the profit generated by one's followers is sufficiently large. We establish existence and uniqueness separately for the cases when there are many or few poor.[14] With many poor people, the wealthy arise unambiguously as the leaders and will be provided with service to further encourage their search externality. With few poor people there still is always an equilibrium in which everybody follows the wealthy, but no service is provided. Only if service sufficiently outweighs the difference in the opportunity costs can the poor receive service and be the leaders.

When analyzing the equilibria, we will discuss cases where a consumer is followed mainly by consumers of his own type. It is therefore important to understand whether this induces service or not. Consider the case where there is only one group of consumers. Consumers without followers do not receive service, since the price of the good is lower than the cost of providing service. If consumers do not die, i.e., $\delta = 1$, there cannot be a steady state in which a non-trivial fraction of consumers has not found a high quality firm. Therefore in any steady state equilibrium, consumers do not search for new firms, and thus consumers do not have followers. Therefore no firms will find it profitable to provide service to any consumer, independently of the service provided by other firms. By continuity, there exists $\delta^* < 1$ such that this holds for all $\delta \in (\delta^*, 1)$, where δ^* is a function of the price P and the cost c.[15] That is, there exists a value δ^* such that for survival rates greater than δ^*, no firm would find it profitable to provide service to consumers who are only followed by consumers of their own type, independent of the service promises of the other firms.

The following proposition establishes the existence of equilibria in an environment when the ratio of poor to wealthy consumers is sufficiently large. All equilibria exhibit service only for the wealthy customers if the survival rate is sufficiently high or service is lower than the threshold ξ in Proposition 24.

Proposition 25 *Fix $\delta \in (\underline{\delta}, 1)$ and $\bar{s} > 0$. There exists $\alpha^* \in (0,1)$ such that for $\alpha \leq \alpha^*$ the following holds:*

[14] We have not shown that the fixed point distribution when the rich follow themselves is unique. Also, service by low quality firms might make no difference to the consumers' search. Therefore equilibria might not be unique, but all exhibit the properties we want to establish.

[15] This can also be seen by considering equation (A.51) with $N_p = 0$ and $N_r = 1 - \gamma$, where γ is the fixed point to equation (3.10). Since $1 - \gamma$ converges to 1 and $\frac{1-\gamma}{1-\delta}$ to $\frac{1}{1-F(u_0^r - q_h - \bar{s})}$ for δ converging to 1, $\Pi^{rh} \approx (P - c) < 0$ for δ large. The argument can be extended to low quality firms and poor consumers, and to the case where other firms do not provide service. It remains valid even when $\beta = \delta$, i.e., the firms' discount factor is also large.

1) *There exist stationary equilibria in which all consumers follow wealthy consumers while searching. Consumers stop searching only when they find a high quality firm, and high quality firms offer service to the wealthy and not to the poor. Low quality firms may offer service but do not attract repeat business.*

2) *All stationary equilibria are of this form if* $\bar{s} < \xi$ *or* $\delta > \delta^*$ *and* α^* *sufficiently small.*

Proof: Assume all players follow the wealthy when searching. Then the poor will never be promised service by any firm that expects repeat business, as by $P < c$ the firm would make a loss. The wealthy will be promised service by all high quality firms. These firms can induce the consumer to return by offering service. For α^* small enough Proposition 23 establishes that this will be the only choice that does not have a profitable deviation. By lemma 6 low quality firms are never repeatedly visited by a wealthy player. It is immediate that all players have an incentive to follow the wealthy: since $s_h(w) \geq s_h(p)$ by Proposition 22 $\gamma^w > \gamma^p > \lambda$, and following the wealthy is better than following the poor or sampling randomly.

For $\bar{s} < \xi$ no other equilibria exist, as by Proposition 24 all players follow the wealthy and the assumption of the prior paragraph is fulfilled. If $\delta > \delta^*$, then it is not profitable to provide service to the poor if they are followed only by other poor consumers. If α^* is sufficiently small, then each poor consumer can only have a negligible number of wealthy followers, and providing service to the poor remains unprofitable even if all consumers follow the poor. If the poor do not receive service, they prefer to follow the wealthy, and again the assumption of the prior paragraph is fulfilled. *Q.E.D.*

The proposition shows that firms indeed support the learning process when the service can be concentrated on sufficiently few wealthy people who achieve a high visibility in the market. For all consumers the outcome is clearly preferred to a world in which service is absent. Wealthy consumers benefit directly from the service and indirectly because they obtain high qualities faster. Poor customers benefit also, but only indirectly through the improved search externality provided by the wealthy. High quality firms benefit, because consumers find high quality firms faster. Yet their cost of providing service might outweigh this benefit. Low quality firms unambiguously loose compared to a world without firms' ability to interference with the consumers search process. Service increases the informational externality between consumers, and a newborn consumer tries on average fewer low quality firms before finding high quality.

As a comparison we analyze the case in which the ratio of wealthy to poor players is reversed. If there are few poor people, there is still always an equilibrium in which everybody follows the wealthy. Only if service is sufficiently important is there also a second equilibrium in which everybody follows the poor.

Proposition 26 *Fix* $\delta > \max\{\underline{\delta}, \delta^*\}$ *and* $\bar{s} > 0$. *There exists* $\alpha^{**} \in (0,1)$ *such that for* $\alpha \geq \alpha^{**}$ *the following holds:*

1) *There exist stationary equilibria in which all consumers follow wealthy consumers while searching. Consumers stop searching only when they find a high quality firm. High quality firms do not offer service to any consumer. Low quality firms may offer service but do not attract repeat business.*

2) *If* $\bar{s} < \xi$ *these are the only equilibria. There is* $\xi' > u_0^p - u_0^w$ *such that there also exist equilibria in which all consumers follow poor consumers while searching. Consumers stop searching only when they find a high quality firm. High quality firms offer service to the poor and not to the wealthy. Low quality firms may offer service but do not attract repeat business.*

3) *There do not exist stationary equilibria with other properties.*

Proof: Assume all consumers follow the wealthy. Since $\delta > \delta^*$ the wealthy do not receive service due to wealthy followers, and α^{**} small enough assures there will not be sufficient poor followers to warrant service.[16] Also the poor do not get service. Proposition 22 then establishes $\gamma^w > \gamma^p > \lambda$, and indeed everybody wants to follow the poor. By Proposition 24, for $u_0^p > u_0^w$ and $\bar{s} < \xi$ there cannot be any other stationary equilibria in which the wealthy are not being followed.

Consider a stationary equilibrium in which the poor do not follow the wealthy. It must then be the case that the wealthy follow the poor. If the wealthy did not follow the poor, the poor would not receive service, and everybody would follow the wealthy as in the previous paragraph. If the wealthy follow the poor, then by lemma 23 high quality firms would indeed want to provide service to the poor. Yet the wealthy will only follow the poor if $\gamma^p \geq \gamma^w$. By Proposition 24 this only happens for $\bar{s} \geq \hat{\xi}_{\gamma^p}$. Since γ^p is bounded away from one for all

[16]The contribution by the poor is $\delta \frac{\varphi_l^p}{\varphi_l^w + \varphi_h^w} \frac{1-\alpha}{\alpha} P \left[1 + \frac{\delta\beta}{1-\delta\beta}(1 - F(u_0^p - q^h))\right]$. Since for a given δ the term $\frac{\varphi_l^p}{\varphi_l^w + \varphi_h^w}$ is bounded from above and independent of α, the expression converges to zero as α converges to one.

$\bar{s}$ since some newborns are always searching, it is easy to see that $\hat{\xi}_{\gamma^p}$ is bounded. Therefore there exists ξ' such that $\bar{s} \geq \xi'$ implies $\gamma^p \geq \gamma^w$. *Q.E.D.*

Proposition 26 reveals the natural advantage that the wealthy possess in information gathering. Following the wealthy is always an equilibrium, as in the absence of service it is best for everybody to follow them. Only if service is very attractive will the poor search enough such that following them can be worthwhile for the wealthy. As discussed in the context of Proposition 22, service has to sufficiently outweigh the opportunity cost, but does not need to be so high that poor consumers actually find high qualities faster than wealthy consumers. It is worth noticing that in the case where both types of equilibria coexist, consumers are all better off in the equilibrium where consumers follow the poor and the poor obtain service.

Propositions 25 and 26 establish that it is the information that is revealed in the choices of the wealthier players that makes them valuable to other players and, by extension, to firms. If there are sufficiently many consumers who value this information, the wealthy are in a unique position to profit from this if service is not too valuable. Poor consumers are not substitutes for the wealthy as their actions reveal less information than those of the wealthy, even if the visibility of the poor is much better when there are fewer of them. Note that we have effectively ruled out trigger strategies in the analysis.[17] Hence, firms' decisions are primarily influenced by the per period contribution of a customer. Thus, it is not the frequency of consumption per se that allows wealthier consumers to command service, but rather the induced information that is valued by other consumers, and in turn by the firms.

3.3 Discussion

The mechanics of our model are sufficiently transparent to allow the discussion of additional social components such as conspicuous consumption, visibility and the importance of relative position in society. We discuss these in the next subsection. We discuss the robustness of our results to various changes in the model assumptions in the following subsection and then conclude.

[17] These would have allowed richer customers to impose harsher punishment on firms, as their overall lifetime consumption is higher and their effective discount factor is higher due to more frequent consumption.

3.3.1 Social Interaction

Our interpretation of the different opportunity costs u_0^w and u_0^p has been derived from differences in income that affect the consumers' budget constraints. For the analysis, u_0^w and u_0^p could be taken as primitives that result from heterogeneity with respect to something other than income. They could, for example, reflect differences in tastes. If you look for a Swedish restaurant, Swedes might have a greater preference than the average consumer, that is, have lower u_0. For running shoes, runners will consume more, and good jazz places are likely most likely discovered by following jazz enthusiasts. While our analysis can easily handle exogenous differences, our focus on income differences stems from two observations. For normal goods income differences will induce higher consumption for the wealthy. More importantly, in many situations income differences might be easier to infer than differences in taste. If taste heterogeneity is similar for different income categories but only income differences are observable, then the firms' treatment decisions and the consumers' decisions on whom to follow will be based on the observable characteristic.

The ability to distinguish between wealthy and poor is important in this context. Typically this must be inferred from some attribute, for example from the suit one wears or the car one drives, suggesting a rational basis for conspicuous consumption.[18] A standard signalling argument would explain why those who would like to consume more frequently would rationally choose to spend the money for a Rolex watch if it lead to greater service while less frequent purchasers would not. It is interesting to note that the inefficiencies associated with the excess spending on such items is at least partially offset by the increased efficiency in the search process made possible by the conspicuous consumption.

Our results also highlight the importance of visibility in the marketplace. Given our signal technology, a consumer of the group that is relatively small is most visible. Therefore it is the small group that can receive service, as service is tied to a sufficient number of followers. This can obviously be extended to a setting in which consumers in the same income category have different visibility in the market.[19] Again those with higher visibility are more likely to receive service.

[18] Fang (2001) analyses conspicuous consumption as a way to mediate informational free-riding in a labor market context; Cole, Mailath and Postlewaite (1995) investigate this in a matching setting.

[19] Assume e.g. two subgroups of wealthy players, and each consumer sees a member of the first subgroup for sure but a member of the second only with probability smaller than one.

As a final point it should be noted that our concept of wealthy vs. poor is one of relative comparison. Being materially better off then others is important, as this results in a "leader" status. The absolute level is not crucial for this. Thus, even in market settings relative comparisons can be important.[20] This provides an understanding of how some groups can enjoy leadership status even when there is only slight heterogeneity in society. The actual market outcome in terms of consumption might be quite different even though the underlying heterogeneity is small, because firms may interact with the search process in ways that magnify intrinsic heterogeneity.

3.3.2 Robustness

We will discuss some of the modelling choices we have made and the robustness of our results to changes in these assumptions. One of the features of our model that deserves discussion is the validity of the commitment. We assumed that firms promise some service level and always honor their promise when the customer returns. This commitment assumption is a shorthand way of introducing reputational considerations that allow for service provision even in Markov-perfect equilibria, i.e., it allows us to rule out equilibria in which service is not provided because firms cannot convince consumers that they will get service in the future. In the analysis we have not considered whether firms would want to renege on their promise (as we assumed that this is not possible). Commitment is not necessary to support this equilibrium in a world without commitment if we allow non-Markovian (trigger) strategies: incentive compatibility of the commitments can be ensured. Equilibria without service as in Proposition 26 can trivially be supported by out of equilibrium beliefs that no firm will provide service in the future, even if it provided service in the current period. Providing service in any period thus only induces costs to the firm without altering future benefits. Equilibria with service such in Propositions 25 and 26 can be supported if consumers believe to never receive service again from a firm that deviates from equilibrium service provision. If players are sufficiently long-lived and patient, the loss of the consumers business or the slowdown of his visits still warrant service provision. Obviously, uniqueness claims do not apply to such a non-restricted environment.

---

[20]Cole, Mailath and Postlewaite (1992) and Mailath and Postlewaite (forthcoming) provide a rationale for relative comparisons in a model where benefits of higher relative standing arise from more attractive matching opportunities. Samuelson (2004) attributes relative comparisons to evolutionary pressure.

In the equilibrium analysis we have established existence for small and large fractions α of wealthy consumers. Due to our restriction to pure markovian strategies we cannot ensure existence for intermediate α for arbitrary parameters c, $\bar{s}$ and $F(\cdot)$. If service is provided, the signal is better and consumers search more efficiently, which reduces the number of followers. Therefore with service the number of followers might be too small to warrant service, while without service the number of followers might be large and service is profitable. In these cases mixed or non-markovian strategies would be necessary for existence.

We also limit attention to two quality levels. This assumption is not entirely innocuous. If there were three qualities $q_l < q_m < q_h$ with associated population fractions λ_l, λ_m and λ_h, it is possible that wealthy consumers search only for q_h firms, while poor consumers stop if they found a medium quality firm as their lower frequency of consumption acts similarly to a lower discount factor. Following the wealthy then implies a high probability of drawing a high quality firm, but also a relatively high probability of finding a low quality firm since the wealthy do not settle on a medium quality firm should they find one, and hence may search for a long time. Following the poor reduces the risk of finding a low quality firm if they searching when they identify either a medium or high quality firm. This can lead to the wealthy following the wealthy to obtain high quality, and the poor following the poor to find medium or high quality firms. Nevertheless, modified versions of our results hold if the survival rate δ is sufficiently large, since in that case both poor and wealthy will continue searching until they find a high quality firm.

We also restricted attention to only two levels of service. This simplifies the analysis, but the model could accommodate multiple levels of service $s_1, ..., s_n$ at costs $c_1, ..., c_n$. In Proposition 25 the level of service to the poor would be small even if they are followed, because the number of followers is small. Therefore they would nevertheless search less frequently, only the wealthy are followed and substantial service is only given to the wealthy. In Proposition 26 the equilibrium in which the wealthy are followed continues to exist, because without followers the poor receive little service and the wealthy consume more often. In this case all consumers receive little service. There will also be an equilibrium in which everybody follows the poor, as they then have many followers. Many followers will induce firms to provide top service, and high quality firms outbid low quality firms. This also happens if service is a continuous choice $s \in [0, \bar{s}]$.[21] $c_n > P$ or $c(\bar{s}) > P$ would

[21]If the service is unbounded, there are parameter constellations for which low quality firms provide service

again insure that maximum service is not simply given due to individual consumption. Yet individual consumption could warrant a service level above zero.

We have assumed a simple signal structure in which each consumer can observe one wealthy and one poor consumer in every period. All results hold if a consumer can observe multiple wealthy and multiple poor consumers each period. Since there is a continuum of firms, the probability of observing two or more consumers who choose the same firm is zero. Therefore several consumers of the same type are as informative as a single consumer, and an individual simply decides to follow either one of the wealthy or one of the poor. On the other hand we could have assumed that each consumer simply sees the choices of N randomly chosen consumers who purchased last period. This assumption is closer to models such as those analyzed by Ellison and Fudenberg (1995) and Banerjee and Fudenberg (2003). While analytically much more complicated, we do not think that it changes the flavor of the results as long as N is sufficiently large given the fraction of wealthy people. The reason is that a consumer only cares about observing at least one poor or at least one wealthy person, depending on whom he wants to follow. For sufficiently large N, the probability is virtually one that one poor and one wealthy consumer is in the observed set.

Some of the social learning literature assumes some private information by agents. If we assume that agents receive in each period a signal that indicates a firm which is with probability ψ of good and with $1 - \psi$ of average quality, this would change the probability of finding a high quality firm when sampling independently to $\lambda' = \psi + (1 - \psi)\lambda$, rather than simply λ, leaving the results qualitatively unchanged.

We took the price as being exogenously set, and identical across firms regardless of quality. We argued above that it seems unrealistic that even if prices differed across firms, they would perfectly convey the quality of firms, and there would remain the possibility that social learning of the sort in our model would still play a role. It is worth discussing what the equilibria of a model such as we have laid out would look like if prices were a strategic variable rather than exogenously set. Suppose that there were a symmetric equilibrium in which all low quality firms set one price and all high quality firms set a possibly different

so high as to prevent the leaders from continued search for high qualities and a pure strategy equilibrium may not exist. The reason is that low quality firms are willing to give up the full surplus to retain the customer, while high quality firms are only driven by the encouragement effect of more frequent consumption due to the markovian restriction.

price. If the difference in quality between the high and low quality firms is small, there may be a separating equilibrium in which the prices of the two types of firms are not very different, and wealthy people go to high quality firms while the poor go to cheaper, low quality firms. Suppose, however, that there was no value to the low quality firm; that is, even if the price were zero, all consumers would prefer the high quality firm. There clearly cannot be a separating equilibrium then since low quality firms could profitably charge the same price as high quality firms. If all firms charge the same price, whether any single firm has an incentive to deviate depends on consumers' beliefs when they see an out-of-equilibrium price. Trivially, beliefs that it is a low quality firm that deviates will support equal pricing.[22] Hence, if our model were extended so that pricing was endogenized, one would get the equal pricing that we assumed.

3.3.3 Conclusion of Chapter 3

We have presented an equilibrium model of social learning in which heterogenous consumers search for an experience good of high quality. Consumers also value service. Information can be obtained through personal consumption or through the observation of other players' choices. One group arises endogenously as the leading group whose actions are emulated by other consumers in the market, leading to the possibility that firms can manipulate the learning process through differential service provision. High quality firms outbid low quality firms in their pursuit of customers, and therefore even in the presence of service, consumers will not be induced to consume at low quality firms. If service is not too valuable, wealthy consumers arise unambiguously as the leading group because they consume more frequently and on average gather more information. If they have sufficiently many followers, they receive service and capture some of the benefits of the search externality they generate. As service induces them to search more frequently, and consequently learn more quickly, it strictly increases the welfare of all consumers. If service is sufficiently important, the poor can arise endogenously as the leading group because service can induce them to consume

[22]For low quality firms only new customers are important. These firms want to pool on any price above marginal cost if they otherwise get no new customers. High quality firms obtain profits from new customers, but also from existing customers. Analogous to the Diamond Paradox, existing customers face switching costs and a high quality firm can extract rents from them. In equilibrium high quality firms must charge a price such that a deviation does not increase the profits from existing customers beyond the loss of sales due to absence of new customers.

sufficiently frequently that they learn more quickly than the wealthy. While service to the poor must sufficiently outweigh the difference in opportunity costs of search, it is not necessary that the poor identify a high quality firm more quickly. If service induces them to consume much more frequently at high qualities than the rich, their informed customers are more visible, which can outweigh the slower rate of information accumulation. Nevertheless, even in this case there exists another equilibrium in which everybody follows the wealthy. The mechanics of the model are sufficiently transparent to shed light on social components such as the role of relative position in society and the role of conspicuous consumption.

Appendix

A.1 Appendix to Chapter 1

Proof of Proposition 2.

Assume an equilibrium exists such that all firms post the same wage w^*. The expected profits are given by $\pi(w^*) = q(w^*)\ (1-w^*)\ (1-\psi(w^*))$, where $\psi(w^*) > 0$ since a worker turns down a firm with positive probability in the case of multiple offers. Proposition 1 implies that $w^* = \bar{w}$ when trembles are sufficiently small, since otherwise all workers send an application to one of arbitrarily few noise firms. Therefore $\psi(w) = 0$ in the limit for all $w > w^*$. Since the queue length (and $q(w)$) is increasing in w, the profits of a firm that posts a wage just above w^* are equal to $\lim_{w \searrow w^*} \pi(w) = q(w^*)\ (1-w^*) > q(w^*)\ (1-w^*)\ (1-\psi(w^*)) = \pi(w^*)$. Therefore offering a wage just above w^* is a profitable deviation.

If all firms post $w^* = 1$ they make zero expected profits. Equation (1.9) implies that there is some $\tilde{w}$ close enough to w^* with strictly positive queue length in the unperturbed game. A firm posting $\tilde{w}$ could then hire a worker with positive probability and receive strictly positive expected profits. This offers a profitable deviation. Last, if $w^* = 0$ workers receive zero expected utility and so for any positive trembles they send both applications to positive wages. As the trembles become smaller the hiring probability of a firm with a positive wage converges to one and since $q(0)\ (1 - \psi(0)) < 1$ posting a wage slightly above zero increases the firm's profits. *QED*

Proof of Proposition 3.

The proposition is proved in two stages. The problem of the high wage firms is solved first and that of the low wage firms follows. As shown in section 3, the maximization problem of the high wage firms is given by

$$\max_{w\in[\bar{w},1]} (1 \quad - \quad e^{-\lambda(w)})\ (1-w) \tag{A.1}$$
$$\text{s.t.} \;\; p(w)\ w \quad + \quad (1-p(w))\ u_1 = u_2. \tag{A.2}$$

Using the constraint we can solve for $w = (u_2 - u_1)\ \lambda/(1-e^{-\lambda}) + u_1$ and substitute that expression into the objective function. The maximization problem can be rewritten with respect to λ as $\max_{\lambda\geq\bar{\lambda}}\ 1-u_1-\lambda\ (u_2-u_1)-e^{-\lambda}\ (1-u_1)$ where $\bar{\lambda} = \lambda(\bar{w})$. This problem is strictly concave in λ since $u_1 < 1$ and hence it has a unique solution λ_2^*, which corresponds to some w_2^*. Note that we proceeded as if $\psi(\bar{w}) = 0$ which is not necessarily the case. However, if $w_2^* > \bar{w}$ then the value of $\psi(\bar{w})$ is irrelevant; if $w_2^* = \bar{w}$ then proposition (2) shows that low wage firms cannot post $\bar{w}$ in equilibrium and hence $\psi(\bar{w}) = 0$. Therefore the maximization problem is specified correctly.

There are two candidate solutions for w_2^*. If the constraint does not bind, the wage is determined by the first order conditions of the problem, $\hat{w}_2$. If the constraint does bind then $w_2^* = \bar{w}$. We show that high wage firms enjoy strictly higher profits than low wage firms when $w_2^* = \hat{w}_2$. Setting the derivative of the problem to zero yields $u_2 - u_1 = e^{-\lambda_2^*}\ (1-u_1)$. Substituting this expression back into the profit function and rearranging gives the following:

$$\pi(\hat{w}_2) = (1-e^{-\lambda_2^*})\ (1-u_1)\ (1-\frac{\lambda_2^*\ e^{-\lambda_2^*}}{1-e^{-\lambda_2^*}}). \tag{A.3}$$

The profits of a low wage firm that posts w_1 and has expected queue length $\lambda_1 = \lambda(w_1)$ are given by

$$\pi(w_1) = (1-e^{-\lambda_1})\ (1-w_1)\ (1-\frac{1-e^{-\lambda_2^*}}{\lambda_2^*}), \tag{A.4}$$

where the first term is the probability of getting at least one applicant, the second term is the margin of the firm, and the last term is the probability that the chosen applicant does not have an offer from a high wage firm.

Comparing the two equations term by term it is easy to see that the profits of high wage firms are strictly higher: firms offering a higher wage have longer queues, so $\lambda_2^* > \lambda_1$ which means that $1-e^{-\lambda_2^*} > 1-e^{-\lambda_1}$; $u_1 = p(w_1)\ w_1$ which implies that $u_1 \leq w_1$; to prove the the third term we need to show that $\lambda\ e^{-\lambda}/(1-e^{-\lambda}) < (1-e^{-\lambda})/\lambda$ for any $\lambda > 0$. This expression can be rearranged as $\lambda^2\ e^{-\lambda} < (1-e^{-\lambda})^2$ or $\lambda^2\ e^{\lambda} - e^{2\ \lambda} + 2\ e^{\lambda} - 1 < 0$. Denote the left hand side by $H_1(\lambda)$ and note that $H_1(0) = 0$. If $H_1'(\lambda) < 0$ for all $\lambda > 0$ we have our result. But, $H_1'(\lambda) = (2\ \lambda + \lambda^2 + 2 - 2\ e^{\lambda})\ e^{\lambda}$ and $H_1'(0) = 0$. Call the term in the bracket

$H_2(\lambda)$ and note that $H_2(0) = 0$. Then $H_2'(\lambda) = 2\ (1 + \lambda - e^{\lambda})$ which is negative for $\lambda > 0$. Therefore, $w_2^* = \bar{w}$ is a necessary condition for any equilibrium.

Turning to low wage firms, they solve

$$\max_{w}\ (1 - e^{-\lambda(w)})\ (1 \quad - \quad \psi(w))\ (1 - w) \tag{A.5}$$

$$\text{s.t. } p(w)\ w \quad = \quad u_1. \tag{A.6}$$

As argued above, $\psi(w) = p(w_2^*)$ for $w \in [0, \bar{w}) \cap \mathcal{W}$, i.e. for all wage levels that are actually posted. We solve the maximization problem as if $\psi(w)$ is the same for all w, whether posted or not, which is the case when, for instance, workers randomize independently inside each of the two intervals. We then show that this simplification is innocuous. Using equation $(A.6)$, we can solve for $w = u_1\ \lambda/(1 - e^{-\lambda})$ and substitute it into the profit function to get $\max_\lambda\ (1 - e^{-\lambda} - \lambda\ u_1)\ (1 - p(w_2^*))$. The term in the second bracket has no marginal effect so the problem is strictly concave and therefore it has a unique solution λ_1^*. The first order conditions imply $u_1 = e^{-\lambda_1^*}$ and hence $w_1^* = \lambda_1^*\ e^{-\lambda_1^*}/(1 - e^{-\lambda_1^*})$.

We now consider the case where the worker strategies are such that $\psi(w)$ takes different values in $[0, \bar{w})$. An example of why this could happen is the following. Suppose that one of the pairs of wages that the workers randomize over in response to every perturbed distribution is $(\tilde{w}_1, \tilde{w}_2)$ where $\tilde{w}_2 \approx 1$. If workers applying to $\tilde{w}_1$ send their high wage application to $\tilde{w}_2$ only, then the retention probability at $\tilde{w}_1$ is very high since $\tilde{w}_2$ being close to one implies that $p(\tilde{w}_2)$ has to be very low. As the trembles become smaller, the probability that this particular pair is chosen converges to zero, however $\psi_\epsilon(\tilde{w}_1)$ remains equal to $p(\tilde{w}_2)$ and so it converges to a relatively high value. This would be troublesome if a different equilibrium could be supported in the way described. Suppose that there is such an equilibrium in which low wage firms post some $\tilde{w} \neq \hat{w}_1$. For $\tilde{w}$ to be posted it needs to provide the highest possible profits, implying in particular that $\pi(\tilde{w}) \geq \pi(\hat{w}_1)$. The last inequality can only hold if $\psi(\hat{w}_1) > \psi(\tilde{w})$ since $\{\hat{w}_1\} = \text{argmax}(1 - e^{-\lambda(w)})\ (1 - w)$. However, the fact that $\tilde{w}$ is actually posted means that $\psi(\tilde{w}) = p(w_2^*)$. Moreover, $w_2^* = \bar{w}$ implies that $p(w) \leq p(w_2^*)$ for all wages w that high firms can post and hence $\psi(\tilde{w}) = p(w_2^*) \geq \psi(\hat{w}_1)$, yielding a contradiction. Therefore no other equilibrium can be supported.

This completes the proof of proposition 3. *QED*

Proof of Propositions 4 *and* 8.

We show that there is a sequence $\{d_i\}_{i=1}^N$ such that when w_i^* is posted by d_i firms (call these type i firms), there is no profitable deviation for any type of firm. Afterwards, uniqueness is proven for the $N = 2$ case.

First, consider deviations within the same type. Since $w_1^* = \hat{w}_1$ it is immediate that type 1 firms cannot profitably deviate within their type. For type $i \geq 2$ firms, $w_i^* = \bar{w}_{i-1}$ is a necessary condition for equilibrium. $\bar{w}_{i-1}$ is the profit maximizing wage within type i only if $\bar{w}_{i-1} > \hat{w}_i$, i.e. when then wage derived from the first order condition is not feasible. We show that profits can be equalized across types only if the above condition holds. The previous proposition proved that if type i firms post $\hat{w}_i$ then they necessarily make higher profits than type $i-1$ firms, or $\pi(\hat{w}_i) > \pi(w_{i-1}^*)$. If $\bar{w}_{i-1} < \hat{w}_i$, and if all type i firms post $\bar{w}_{i-1}$ they make higher profits than if they all posted $\hat{w}_i$. This happens because they receive the same number of applications (one per worker) but pay them less (however, each firm could individually increase its profits even more by posting $\hat{w}_i$). As a result, $\pi(\bar{w}_{i-1}) > \pi(\hat{w}_i)$ and profits cannot be equalized across types i and $i-1$. If, on the other hand, $\bar{w}_{i-1} > \hat{w}_i$, then $\pi(\bar{w}_{i-1}) < \pi(\hat{w}_i)$. Therefore, if profits can be equalized across types, then $\bar{w}_{i-1}$ is the profit maximizing wage of type i firms.

The next step is to prove that profits can be equalized across types of firms. To simplify notation let $\pi_i = \pi(w_i^*)$, $p_i = p(w_i^*)$, $\tilde{\pi}_i = \pi_i/(1-\psi_i)$, and $\lambda_i^* = b/d_i$. For equal profits across types it is sufficient to show that $\pi_i = \pi_{i-1}$ for all i, which is the same as $\tilde{\pi}_i = (1-p_i)\ \tilde{\pi}_{i-1}$ since the term $(1-\psi_i)$ is common to both sides. We show that given a d_{i-1} we can find a d_i in $(0, d_{i-1})$ such that $\Delta\pi_i(d_i|d_{i-1}) \equiv \tilde{\pi}_{i-1} - \tilde{\pi}_i/(1-p_i) = 0$. This allows us to construct a sequence of d_is such that all firms make the same profits for an arbitrary initial d_1. We then show that the d_i's sum up to one.

It is useful to recall the following two equations (for $i \geq 2$).

$$u_{i-1} = p_{i-1}\ w_{i-1}^* + (1-p_{i-1})\ u_{i-2} \tag{A.7}$$

$$u_{i-1} = p_i\ w_i^* + (1-p_i)\ u_{i-2}. \tag{A.8}$$

Equation $(A.7)$ holds by the definition of u_{i-1}. Equation $(A.8)$ holds because $w_i^* = \bar{w}_{i-1}$ and hence the i firm has to provide the same utility as w_{i-1}^* if it is used for the $i-1$ lowest application.

Note that the queue lengths are the same when $d_i = d_{i-1}$, which means that $p_{i-1} = p_i$, $w_{i-1}^* = w_i^*$, and $\tilde{\pi}_{i-1} = \tilde{\pi}_i$ leading to $\Delta\pi_i(d_{i-1}; d_{i-1}) < 0$. On the other hand, $\lambda_i \approx \infty$ when $d_i \approx 0$ which means that $p_i \approx 0$ and therefore equation $(A.8)$ requires a very large w_i^* leading to $\tilde{\pi}_i < 0$ (this occurs because the firm is assumed to post $\bar{w}_{i-1}$). $\Delta\pi_i(d_i|d_{i-1}) > 0$

when $d_i \approx 0$, and there is a $d_i(d_{i-1})$ such that type i and $i-1$ firms make the same profits. Moreover, the solution $d_i(d_{i-1})$ is unique because

$$\frac{\partial \Delta \pi_i}{\partial d_i} = -\tilde{\pi}_i \; \frac{\partial (1/(1-p_i))}{\partial d_i} - \frac{1}{1-p_i} \; \frac{\partial \tilde{\pi}_i}{\partial d_i} < 0. \tag{A.9}$$

When d_i increases the queue length decreases and hence the probability of getting a job increases. Therefore the first partial is positive and the first term as a whole is strictly negative. The second partial is non-positive since $\partial \tilde{\pi}_i / \partial \lambda_i \leq 0$. Recall that when $i = 1$ the first order conditions are equal to zero because $w_1^* = \hat{w}_1$. Furthermore, when $i \geq 2$ the firm would like to post a lower wage when profits are equalized (i.e., $w_i^* = \bar{w}_{i-1} > \hat{w}_i$) which implies that the first order conditions with respect to λ are strictly negative. This proves that equation $(A.9)$ is strictly negative.

Therefore, for a given d_1 the rest of the sequence $d_2(d_1), d_3(d_1)...d_N(d_1)$ can be uniquely constructed such that all types of firms make the same profits. To find the sequence whose elements sum up to one define $S(d_1) \equiv \sum_{i=1}^{N} d_i(d_1)$ and note that it is continuous since all of its components vary continuously with d_1. Moreover, $S(1/N) < 1$ since $d_i(d_{i-1}) < d_{i-1}$ and $S(1) > 1$ so there is some d_1^* such that $S(d_1^*) = 1$ and an equilibrium exists for any N. To prove the uniqueness of equilibrium when $N = 2$ we show that d_1 and $d_2(d_1)$ are positively related along the isoprofit curve, and hence there is a unique pair that sums up to one. Implicit differentiation of d_2 with respect to d_1 while keeping profits equal yields $\partial d_2 / \partial d_1 = -(\partial \Delta \pi_2 / \partial d_1)/(\partial \Delta \pi_2 / \partial d_2)$. The denominator is positive by (A.9). A little algebra shows that the numerator is given by $\partial \Delta \pi_2 / \partial d_1 = (\partial \lambda_1^* / \partial d_1) \; e^{-\lambda_1^*} \; (\lambda_1^* - \lambda_2^*/(1-p_2))$, which is positive since the queue length is inversely related to the number of firms and $\lambda_1^* < \lambda_2^*$. This proves that the equilibrium is unique when $N = 2$. *QED*

Proof of Proposition 5

We first define an equilibrium of the perturbed game and then show that, as the trembles disappear, there is a sequence of such equilibria in the perturbed games that converges to our equilibrium of the non-perturbed game. Furthermore, there is no sequence that converges to a different limit. Let $\tilde{F}$ be distribution from which noise firms post. Let F_ϵ^{NT} denote the distribution that non-trembling firms post, which is a response to both noise and equilibrium firms. The distribution of posted wages is then defined by $F_\epsilon(w) = (1-\epsilon) \; F_\epsilon^{NT}(w) + \epsilon \; \tilde{F}(w)$. The best response of workers is given by $G(F_\epsilon)$. The outcomes can be determined as in section 2 using F_ϵ and $G(F_\epsilon)$. Given $\tilde{F}$ we can define an equilibrium.

Definition 6 *An equilibrium of the perturbed game is F_ϵ^{NT} and $G(F_\epsilon)$ such that*

1. $\pi(w) \geq \pi(w')$ *for all* $w \in suppF_\epsilon^{NT}$ *and* $w' \in [0,1]$.

2. $U(\mathbf{w}) \geq U(\mathbf{w}')$ *for all* $\mathbf{w} \in suppG(F_\epsilon)$ *and* $\mathbf{w}' \in [0,1]^N$.

We now prove that there is an equilibrium in the trembling game that converges to our candidate equilibrium. Then we show that there is no sequence that converges to a different equilibrium.

Note that, under trembles, the worker's problem is identical to the one described in section 1.2.1. When (non-trembling) firms choose which wage to post, they take into account the presence of the noise firms. High wage firms, however, solve exactly the same problem as in section 1.2.2 since they face $\psi(w) = 0$. This implies that, for small enough trembles, they all post $\bar{w}$. Low wage firms have to consider the possibility that they may lose a worker to a noise high wage firm. Suppose all workers send their applications independently to the two intervals of firms. This is obviously an equilibrium strategy. Then, the retention probability of a low wage firm does not depend on the actual wage that it posts. As a result, the low wage firms' optimization problem corresponds to that in section 1.2.2 and they post a unique wage that is derived by their first order condition. Existence can then be proved in the same way as in section 1.2.3. These trembling equilibria clearly converge to our candidate equilibrium. Finally, consider the possibility that workers follow different (equilibrium) strategies. The same logic that we used at the end of the proof of proposition 3 shows that no different equilibrium can be supported. QED

Proof of Proposition 7.

To generalize (3) to any N it is sufficient to show that unless type $i \geq 2$ firms post $\bar{w}_{i-1}$ they make strictly higher profits than firms of type $i-1$. After using the constraint to solve for the wage, and taking the first order conditions, the profits of a type i firm are

$$\pi(\hat{w}_i) = (1 - e^{-\lambda_i^*})\ (1 - u_{i-1})\ (1 - \frac{\lambda_i^*\ e^{-\lambda_i^*}}{1 - e^{-\lambda_i^*}})\ (1 - \psi_i). \tag{A.10}$$

The profits of a type $i-1$ firm are given by

$$\pi(w_{i-1}) = (1 - e^{-\lambda_{i-1}})\ (1 - w_{i-1})\ (1 - \frac{1 - e^{-\lambda_i^*}}{\lambda_i^*})\ (1 - \psi_i), \tag{A.11}$$

and they are lower for the same reasons as before. QED

Proof of Proposition 2.

For given worker-firm ratio b and given fraction d_1 of type one firms, we can find a unique sequence $\mathbf{d}_b(d_1) = \{d_1, d_2, d_3...d_N\}$ such that there is no profitable deviation for firms when wage w_i^* is posted by exactly d_i firms (see proof of proposition 8). Let $S_b(d_1) = d_1 + d_2 + ... + d_N$ for $d_i \in \mathbf{d}_b(d_1)$. Let $D(b)$ be the set of all sequences $\mathbf{d}_b(d_1)$ that sum up to one. Given any b and b' let $d_i' = (d_i\ b')/b$, $\lambda_i = b/d_i$, and $\lambda_i' = b'/d_i'$. Then $\lambda_i = \lambda_i'$ and therefore $\mathbf{d}_b(d_1) \in D(b)$ if and only if $\mathbf{d}_{b'}(d_1') \in D(b')$. Furthermore, $S_b(d_1) = 1$ if and only if $S_{b'}(d_1') = 1$. If $S_{b^*}(d_1)$ is strictly increasing for some b^*, then there is a unique d_1^* such that $S_{b^*}(d_1^*) = 1$ and hence there is a unique equilibrium. This means that the equilibrium is unique for any b. *QED*

Proof of Proposition 9.

The planner solves the following problem: $\max_{d \in [0,1]}\ m(d) = p_1 + p_2 - p_1\ p_2$. If the problem has an interior solution, the first order conditions yield

$$\frac{\partial p_2}{\partial d_1}(1 - p_1) + \frac{\partial p_1}{\partial d_1}(1 - p_2) = 0. \tag{A.12}$$

Recalling that $\lambda_1 = b/(1-d)$ and $\lambda_2 = b/d$ it is easy to see that $\partial p_i/\partial d = -\partial \lambda_i/\partial d\ (1 - e^{\lambda_i} - \lambda_i\ e^{-\lambda_i})/\lambda_i^2$, $\partial \lambda_1/\partial d = b/(1-d)^2 = \lambda_1^2/b$, and $\partial \lambda_2/\partial d = -b/d^2 = -\lambda_2^2/b$, so equation (A.12) can be rewritten as

$$(1 - e^{-\lambda_2} - \lambda_2 e^{-\lambda_2})(1 - \frac{1 - e^{-\lambda_1}}{\lambda_1}) = (1 - e^{-\lambda_1} - \lambda_1 e^{-\lambda_1})(1 - \frac{1 - e^{-\lambda_2}}{\lambda_2}). \tag{A.13}$$

It is immediate that one extremum occurs when $\lambda_1 = \lambda_2$, or $d = 1/2$. The second derivative is given by

$$\begin{aligned}\frac{\partial^2 m}{\partial d^2} = \quad & \tfrac{1}{b^2}(1 - e^{-\lambda_2} - \lambda_2 e^{-\lambda_2})(1 - e^{-\lambda_1} - \lambda_1 e^{-\lambda_1}) - \tfrac{1}{b^2}\lambda_2^3 e^{-\lambda_2}(1 - p_1) \\ & \tfrac{1}{b^2}(1 - e^{-\lambda_2} - \lambda_2 e^{-\lambda_2})(1 - e^{-\lambda_1} - \lambda_1 e^{-\lambda_1}) - \tfrac{1}{b^2}\lambda_1^3 e^{-\lambda_1}(1 - p_2).\end{aligned} \tag{A.14}$$

Substitution of (A.13) and dividing by $(1-p_1)(1-p_2)/b^2$ establishes that at any candidate extreme point the sign of the second derivative is given by $sign(\partial^2 m/\partial d^2) = sign(f(\lambda_2) + f(\lambda_1))$, where

$$f(\lambda) = \frac{(1 - e^{-\lambda} - \lambda e^{-\lambda})^2}{(1 - (1 - e^{\lambda})/\lambda)^2} - \frac{\lambda^3 e^{-\lambda}}{1 - (1 - e^{\lambda})/\lambda}. \tag{A.15}$$

Therefore, we want to show that there is no $b > 0$ such that there exists $d \in (1/2, 1)$ where (A.13) holds and

$$f(\frac{b}{d}) + f(\frac{b}{1-d}) \leq 0. \tag{A.16}$$

Figure A.1 shows $f(\lambda)$ for $\lambda \geq 0$. The function is strictly decreasing on $(0, a_1)$, strictly increasing on (a_1, a_4), again strictly decreasing on (a_4, ∞) and converges to 1 for $\lambda \to \infty$. The only roots of the function are 0 and a_2. We will discuss this function in order to establish the result. Note that for any b, the specific value of d defines $\lambda_1 = b/d$ and $\lambda_2 = b/(1-d)$. Note that for $\lambda_2 > a_3$ it is not possible to fulfill (A.16), where a_3 is such that $f(a_3) = -f(a_1)$. Therefore we will restrict the discussion to $\lambda_2 < a_3$. This also implies that we do not have to discuss any b where $2b > a_3$. For $d = 1/2$ we know that $\lambda_1 = \lambda_2$, and therefore the first order condition holds and $sign(\partial^2 m/\partial d^2) = sign f(2b)$.

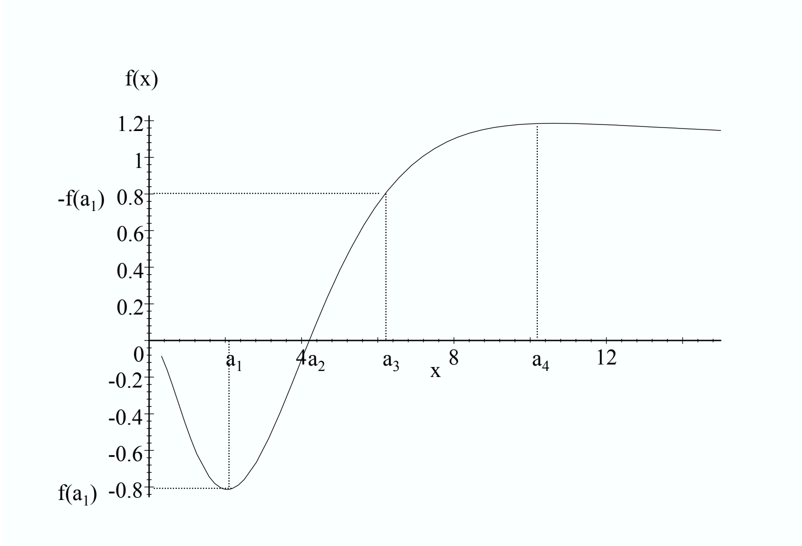

Figure A.1: f(x) for $x \geq 0$.

CASE 1: $b \geq a_2/2$. Then at $d = 1/2$ we have $2f(2b) \geq 0$. Starting from $d = 1/2$, i.e. $\lambda_1 = \lambda_2$, we will increase d and thus spread λ_1 and λ_2 apart. We will show that there does not exist $d > 1/2$ such that (A.16) holds. Assume that (A.16) holds for the given b at some $d > 1/2$. Then for any $b' \in [a_2/2, b)$ there exists a $d' > 1/2$ such that (A.16) holds. This is easy to see if there exists $d' > 1/2$ such that $\lambda_1 = b/d = b'/d' = \lambda_1'$. Then $f(\lambda_1) = f(\lambda_1')$. Since $\lambda_2 = b/1-d) > b'/(1-d') = \lambda_2'$, $f(\lambda_2) > f(\lambda_2')$. But then $f(\lambda_1) + f(\lambda_2) \leq 0$ implies $f(\lambda_1') + f(\lambda_2') < 0$. If for some $b' \in [a_2/2, b)$ no such $d' > 1/2$ exists, we reach a contradiction: There is some $b'' \in [b', b)$ such that at $d'' = 1/2$ it holds that $\lambda_1 = b/d = b''/d'' = \lambda_1'$. By the prior argument $f(\lambda_1'') + f(\lambda_2'') < 0$, but this violates $2f(2b) = f(\lambda_1'') + f(\lambda_2'') \geq 0$. Therefore, if we know that (A.16) does not hold at $\tilde{b} = a_2/2$, then we know that (A.16) does not hold for any $b > a_2/2$. Figure A.2 shows $f(a_2/2d) + f(a_2/(2(1-d)))$ for all $d \geq 1/2$, which is strictly positive for all $d > 1/2$. Therefore, (A.16) does not hold for any $b \geq a_2/2$.

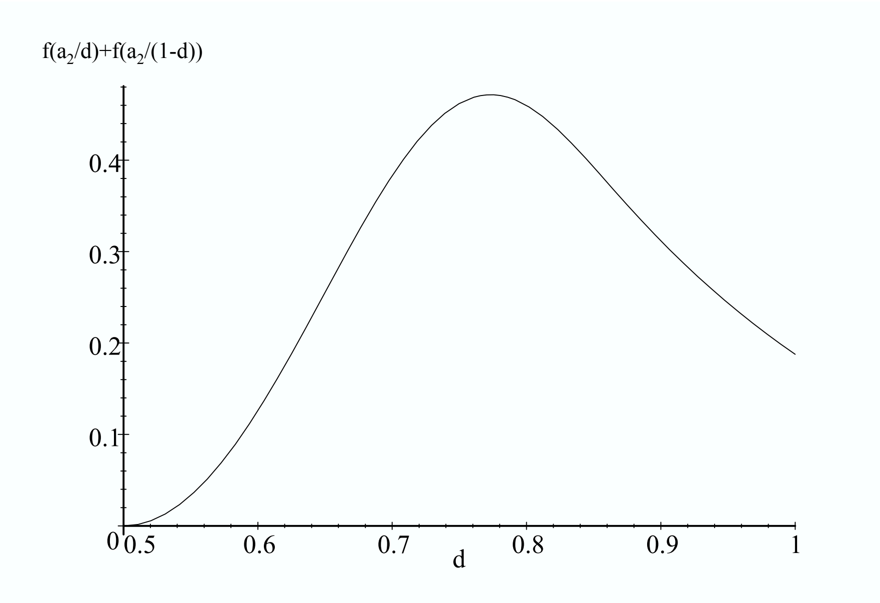

Figure A.2: $f(\frac{a_2}{2d}) + f(\frac{a_2}{2(1-d)})$ for $d \in [0, 1]$.

CASE 2: $b < a_2/2$. In this case we have at $d = 1/2$ that $2f(2b) < 0$, i.e. we are in a local maximum. If there exist any other local maxima at $d > 1/2$, there has to be some $d' \in (1/2, d)$ that constitutes a local minimum. Therefore, if for some d conditions (A.16) and (A.13) hold simultaneously, then there exists $1/2 < d' < d$ such that $f(b/d') + f(b/1-d') > 0$. At d' it has to hold $\lambda_2' = b/(1-d') > a_2$, otherwise $f(\lambda_1') + f(\lambda_2') > 0$ would not be possible. We also know that $\lambda_1' < b/2 < a_2$. Since $d' < d$, we know that $\lambda_1 < \lambda_1'$ and $\lambda_2' < \lambda_2$. Now consider a d' at which $f(\lambda_1') + f(\lambda_2') > 0$. If we increase d to values above d', the derivative of $f(\lambda_1) + f(\lambda_2)$ is

$$\frac{\partial(f(\lambda_1) + f(\lambda_2))}{\partial d} = f'(\lambda_1)\frac{\partial\lambda_1}{\partial d} + f'(\lambda_2)\frac{\partial\lambda_2}{\partial d} \quad \text{(A.17)}$$

$$= \frac{1}{b}[-f'(\lambda_1)\lambda_1^2 + f'(\lambda_2)\lambda_2^2]. \quad \text{(A.18)}$$

If the term in square brackets is positive, then $f(\lambda_1) + f(\lambda_2)$ is increasing as we increase d further. So if we can show that the part in the square brackets is positive for all $(\lambda_1, \lambda_2) \in [0, a_2] \times [a_2, a_3]$, then it is not possible to increase d starting from any d' and achieve a negative value of $f(\lambda_1) + f(\lambda_2)$ (which we would need to arrive at another maximum). Since $max_{[0,a_2]} f'(\lambda)\lambda^2 \leq min_{[a_2,a_3]} f'(\lambda)\lambda^2$, as can be seen in figure A.3, it is not possible to have another local maximum in the interior apart from $d = 1/2$. *QED*

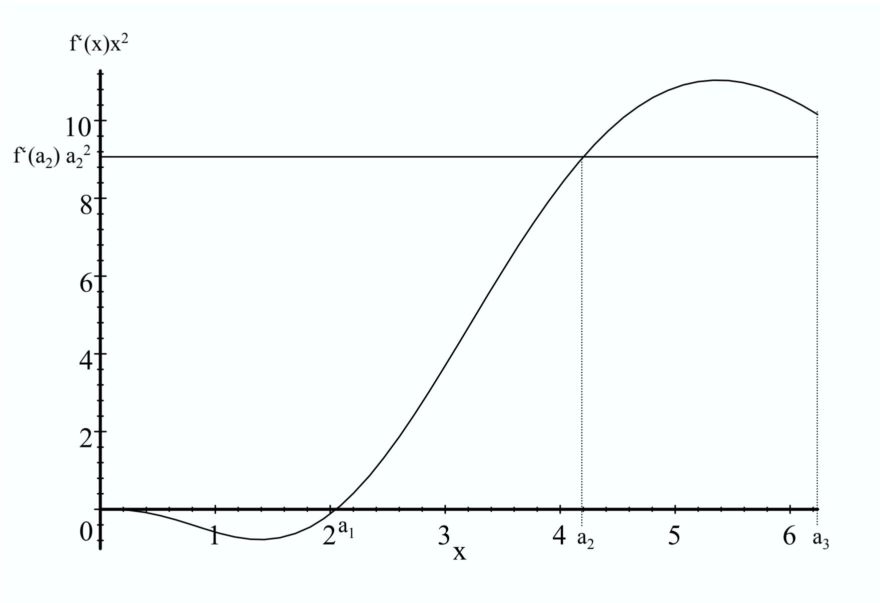

Figure A.3: $f'(x)x^2$ for $x \in [0, a_2]$.

Proof of Proposition 10

The measure of workers who are employed at wage w_i^* is given by $b\ (1-\psi_{i+1})\ p_i \equiv E_i$. Moreover, $E_{i-1} = b\ (1-\psi_i)\ p_{i-1} = b\ (1-\psi_{i+1})\ (1-p_i)\ p_{i-1}$. For the density to be declining, $E_i < E_{i-1}$ has to hold for all i which happens if and only if $p_i < (1-p_i)\ p_{i-1}$. Equal profits imply that $q(w_i^*)\ (1-w_i^*) = q(w_{i-1}^*)\ (1-p_i)\ (1-w_{i-1}^*)$ or $p_i\ \lambda_i\ (1-w_i^*) = p_{i-1}\ \lambda_{i-1}\ (1-p_i)\ (1-w_{i-1}^*)$ yielding the condition $\lambda_i\ (1-w_i^*) > \lambda_{i-1}\ (1-w_{i-1}^*)$. Using the equilibrium conditions $w_i^* = (u_{i-1}-u_{i-2})/p_i + u_{i-1}$ and $w_{i-1}^* = (u_{i-1}-u_{i-2})/p_{i-1} + u_{i-1}$ the inequality becomes $\lambda_i(1-x/p_i) > \lambda_{i-1}(1-x/p_{i-1})$ where $x \equiv (u_{i-1}-u_{i-2})/(1-u_{i-2})$. Therefore, the empirical distribution is decreasing if $g(\lambda) \equiv \lambda\ (1-\lambda\ x/(1-e^{-\lambda}))$ is increasing with respect to the queue length. The first derivative yields $\partial g/\partial\lambda = (1-e^{-\lambda})\ (1-e^{-\lambda}-2\ \lambda\ x)+\lambda^2\ x\ e^{-\lambda}$ which is positive if $\lambda\ x$ is small. Noting that $\lambda_i\ x = (1-e^{-\lambda_i})\ (w_i^*-u_{i-2})/(1-u_{i-2})$ and that the right hand side goes to zero for b large enough the result is established. *QED*

A.2 Appendix to Chapter 2

Properties of the extensive form matching process:

We consider the following matching process. There are T substages. In the first substage, all firms that have at least one applicant choose one and make an offer. We assume firms choose their applicant at random. Workers can accept at most one offer and reject the others. We assume they accept the most attractive offer weakly larger than zero. In every subsequent period, any firm who do not have an offer that is currently accepted can make a new offer. We assume that only those firms that have at least one applicant whom they did not yet make an offer choose one of these applicants at random and make them an offer (the others remain vacant). Workers can accept either one of the new offers or keep accepting a non-rejected offer from an earlier period. We assume they accept the highest offer weakly greater zero.

First observe that the assumed behavior about the firm's decision to make an offer and the worker's decision to accept or reject is individually optimal given the other agents' behavior. Since workers are identical and we have assumed that workers use symmetric application strategies, choosing at random about whom to make an offer next is an optimal strategy for firms. It is also optimal for workers to always accept higher offers over lower offers: It does not affect the offer decisions of firms with even higher wages, and so does not preclude any chance of receiving an even better offer in the future.

To discuss convergence, consider some network that formed from the agents strategies $\{\phi, F, \boldsymbol{\gamma}, \boldsymbol{G}\}$ (for this notation see section 2.1.2 and section 2.4.1). Let N be the maximum number of applications that workers send ($N = 2$ in the basic setup, some finite $N \in \mathbb{N}$ in the extended setup). The focus on workers strategies that are symmetric and anonymous significantly simplifies the analysis. It implies that gross applications are randomly distributed to firms at a given wage. Together with the assumption that the agent space is sufficiently large such that each choice is undertaken by a continuum of agents leads in our specification to a Poisson distribution of gross applications to firms at each wage level. It also allows us to apply the law of large number convention to each wage level, i.e. at each wage the population share of currently matched and unmatched firms and workers develop deterministically. Moreover, note that the process "rolls forward", i.e. the exact end date does not influence the evolution of the system.

We will show that the process converges to a solution in which almost all agents are matched stably. We call a firm stable at the beginning of a given period if one of two conditions holds: Either it is currently matched to a worker whose other applications are with firms that offer a weakly lower wage or that offer a higher wage but are currently matched; or it is unmatched and already offered the job to all its applicants. We call the measure of stable firms at the beginning of subperiod t s_t. Similarly, we call workers stable if they have not applied to any firm with a strictly higher wage that is currently vacant (unemployment is coded as wage zero), and call their measure at the beginning of period t s_t^w. We will show convergence in the following sense: $\boldsymbol{s_t} \to \boldsymbol{v} = \boldsymbol{\phi V}$ and $\boldsymbol{s_t^w} \to \mathbf{1}$. Note that our notion of stability is a local concept involving only the immediate partners in the interaction. Yet the convergence implies that we attain (weak) stability globally if we remove a small (and in the limit measure zero) set of agents from the economy.

To provide some intuition, consider first the case where only a discrete set of wages is offered by workers. At the highest wage let λ_H be the ratio of gross applications to firms. Then the fraction of highest wage firms that are stable at the beginning of the first period is $1 - e^{-\lambda_H}$, as they did not receive any applications. The other highest wage firms make an offer. They either get matched (and then stay matched), run out of applicants or keep proposing the job to additional applicants. If each worker only sends one application to the highest wage, all high wage firms are stable after the first subperiod. Even if not, the fraction of highest wage firms that are not stable in period t is bounded above by $1 - \sum_{\tau=1}^{t}(\lambda_H^\tau e^{-\lambda_H}/\tau!)$ by Poisson matching (as this is the fraction that has applicants left even if no worker ever accepts). This fraction converges to zero. Since for t large nearly all firms at the highest wage are stable, at the second highest wage those firms that are matched are with very high probability stable (because only the highest wage firms could attract the worker away, and most of them are stable). And those second-highest-wage firms that are unmatched propose until they run out of applicants or get matched themselves. If each worker sends at most one application to each of the highest two wages, all firms at the second highest wage are matched after the second subperiod. In any case the fraction of non-stable firms at this wage goes to zero because they exhaust their applicants or get matched with a stable worker. By induction this applies to all offered wages.

Induction does not work with a continuum of offered wages. Nevertheless we can show convergence. It is still straightforward to show that the measure of proposals per period converges to zero. Let $\bar{S}_t$ be the set of currently unmatched firms at the beginning of period t that still have applicants to whom they did not propose yet, i.e. the set of firms that propose in period t. Denote its measure by γ_t. Since there are less applications than proposals, i.e. $\sum_{\tau=1}^{\infty} \gamma_\tau \leq N$, we have $\sum_{\tau=t}^{\infty} \gamma_\tau \to_{t\to\infty} 0$ and $\gamma_t \to 0$. Let a_t denote the measure of new acceptance in period t. Since there are less new acceptances than offers, we also have $\sum_{\tau=t}^{\infty} a_\tau \rightarrow_{t\rightarrow\infty} 0$ and $a_t \rightarrow 0$.

That means that almost all unmatched firms have no applicants left to propose to when t becomes large. We will show that this implies $s_t \rightarrow v$. Assume not, i.e. $s_t \nrightarrow v$. Then there exists $\delta > 0$ and subsequence $\{t_m\}_{m=1}^{\infty}$ such that $\bar{s}_m := v - s_{t_m} > \delta \; \forall m \in \mathbb{N}$, where we denote by $\bar{s}_m$ the measure of unstable firms. Of these firms only the subset $\bar{S}_{t_m}$ is unmatched. Since their measure γ_{t_m} converges to zero, the measure of unstable but matched firms has to be greater than δ for all $m > M$ for some $M \in \mathbb{N}$, and all these firms have positions filled with unstable workers who also applied to firms in $\bar{S}_{t_m}$ and would rather take a job there. Call the set of these workers $\bar{S}_m^1$. We will show a contradiction because all these workers apply to the few firms in $\bar{S}_{t_m}$, therefore firms in $\bar{S}_{t_m}$ get matched quickly and therefore a large fraction of the $\bar{S}_m^1$ workers (who were unstable before) then become permanently stable.

Call the set of applications which make the $\bar{S}_m^1$ workers unstable "unstable applications (in t_m)". The measure of unstable applications in t_m is obviously larger than δ for all $m > M$. Let $\bar{S}_m^2 \subseteq \bar{S}_{t_m}$ denote the subset of firms that hold at least one of the unstable applications (in t_m). These firms hold on average some number, say x_m, of unstable applications. Let $\bar{S}_m^3 \subseteq \bar{S}_m^2$ be the subset of firms that hold at least $x_m/2$ unstable applications. Note that any fraction α of the firms in $\bar{S}_m^3$ receives at least $\alpha\delta/2$ unstable applications. Of these $\bar{S}_m^3$ firms, we consider the half which have the least gross applications and call it $\bar{S}_m^4$. They receive at least $\delta/4$ unstable applications. Of these $\bar{S}_m^4$ firms, we consider the half which have the highest probability of hiring an $\bar{S}_m^1$ worker permanently, conditional on making an offer to him, and call it $\bar{S}_m^5$. Firms in $\bar{S}_m^4$ (and thus in $\bar{S}_m^5$) have a probability of making an offer to an $\bar{S}_m^1$ worker of at least $(\delta/4)/(2N)$, since each fraction α' of these firms have at least $(\delta/4)\alpha'$ unstable applications and at most $2N\alpha'$ other applications (otherwise the firms in $\bar{S}_m^3 \backslash \bar{S}_m^4$ would hold a measure of unstable applications greater than N, but at most a measure of N applications is sent). Let M satisfy $\sum_{\tau=M}^{\infty} a_\tau \leq \delta/8$. A firm in $\bar{S}_m^5$ that makes an offer to a worker in $\bar{S}_m^1$ has a probability larger than 1/2 that its offer is accepted permanently. Therefore at least a fraction $\delta/(16N)$ of $\bar{S}_m^5$ firms gets matched permanently in period t_m. Since each fraction α'' of $\bar{S}_m^5$ firms receives at least $(\delta/8)\alpha''$ unstable applications, at least $(\delta/8)(\delta/(16N))$ unstable applications are no longer unstable because they are now with firms that are permanently matched. Since in each period $t_m,\ m > M$, we permanently "loose" a strictly positive measure of unstable applications, we would need an infinite measure of unstable applications to sustain $\bar{s}_{t_m} > \delta \ \forall m > M$, yielding the desired contradiction. Therefore $s_t \rightarrow v$.

Similarly, $s_t^w \rightarrow 1$, because otherwise we would again have a set of unstable workers similar to $\bar{S}_m^1$ that applies to the few firms in $\bar{S}_{t_m}$, and the same argument applies.

This establishes convergence. In the limit this implies that almost all firms get matched only when the higher wage firms to which their applicants applied are matched, i.e. we can remove firms "top-down" from the market. Since the random application and offer process at each wage implies that all firms at a given wage have equal chance of being the first to propose to a worker that applied to both, the process works as if we select one at random to make the offer first.

Proof of Proposition 11:

For wages strictly below u_1 the result is immediate because the Market Utility cannot be obtained. At wage $w = u_1$ $\mu(w) > 0$ would imply that the Market Utility cannot be reached. Wages strictly above u_1 have $\mu(w) > 0$, as otherwise $p(w)w = w > u_1$ and workers would receive more than the Market Utility when applying there.

We have shown that it is optimal to send low applications to wages below $\bar{w}$, which implies that (2.9) has to hold for all wages in $(u_1, \bar{w}]$ in order to provide the Market Utility. $u_1 \equiv \sup_{w \in \mathcal{V}} p(w)w$ since the optimum has to be obtained at some wage that is actually offered.

For $\gamma_2 > 0$, it is optimal to apply with the high application to wages above $\bar{w}$, and the effective queue length is therefore governed by (2.10). Again the optimum is attained for wages that are actually offered. The effective queue length has to be continuous at $\bar{w}$, as otherwise the job finding probability $p(w)$ for workers would be discontinuous and some wage in the neighborhood of $\bar{w}$ would offer a utility different from the Market Utility. Therefore $\bar{w}$ is determined as the wage where both (2.9) and (2.10) hold.

For $\gamma_2 = 0$, all wages above u_1 that are in the offer set $\mathcal{V}$ have to conform to (2.9) because they receive single applications. So $\bar{w} \geq \sup \mathcal{V}$. But if a higher (not offered) wage would be offered, workers might prefer to send a second application there rather than relocating their first one. Assume the queue length would be governed by (2.9) for all wages in $(u_1, 1]$. If $p(w)w + (1 - p(w))u_1 - c_2 \geq u_1$, workers would like to send a second application. Therefore $\bar{w}$ is the smallest wage for which that inequality holds (but at most 1). At higher wages the inequality would be strict, i.e. workers would get a utility higher than the Market Utility by sending a second application. To fulfill the Market Utility Assumption the additional utility of the second application has to equal its cost, so $u_2 - u_1 = c_2$ has to hold at high wages and the effective queue length is again governed by (2.10). *Q.E.D.*

Proof of Proposition 12:

Proof: Consider a (candidate) equilibrium in which all active firms offer wage $w^* \in (0,1)$. Almost all applications are sent to w^* because of (2.3) and worker optimality 2*b*). $w^* > 0$ then implies $u_1 = p(w^*)w^* > 0$. Moreover $w^* = \bar{w}$. If not, i.e. $\bar{w} > w^*$ or $\bar{w} < w^*$, then a mass of applications would be sent strictly above or below the offered wage, yielding a contradiction. Then profits for wages above w^* are given by (2.15), for wages in $[p(w^*)w^*, w^*]$ by (2.14), and for wages below $p(w^*)w^*$ profits are zero.

The left derivative of the profits with respect to the queue length at $\bar{\mu} = \mu(w^*)$ is obtained by the differentiating (2.14) to get $\pi'_-(\bar{\mu}) = e^{-\bar{\mu}} - u_1$, and the right derivative by differentiating (2.15) which yields $\pi'_+(\bar{\mu}) = e^{-\bar{\mu}}(1 - u_1) - (u_2 - u_1)$. In equilibrium it needs to hold that firms will neither want to increase their wage nor decrease their wage. This leads to $\pi'_+(\bar{\mu}) \leq 0 \leq \pi'_-(\bar{\mu})$. But $\pi'_+(\bar{\mu}) \leq \pi'_-(\bar{\mu})$ implies

$$-e^{-\bar{\mu}}u_1 - (u_2 - u_1) \leq -u_1. \tag{A.19}$$

For a single market wage it holds that $u_2 = u_1 + (1-\bar{p})u_1$ with $\bar{p} = \frac{1-e^{-\bar{\mu}}}{\bar{\mu}}$. We can therefore write $u_2 - u_1 = (1 - \bar{p})u_1$. Then (A.19) reduces to

$$(1 - e^{-\bar{\mu}} - \bar{\mu}e^{-\bar{\mu}})u_1 \leq 0. \tag{A.20}$$

We know that $u_1 > 0$. It is easily shown that the term in brackets is strictly positive for any $\bar{\mu} > 0$, yielding the desired contradiction.

For the extremes, consider $w^* = 1$ first. At $w^* = 1$ firms make zero profits. Since the effective queue length at wages close to 1 is positive by (2.12), wages below one provide profitable deviations. Now consider $w^* = 0$. Equilibrium profits are strictly smaller than one because not all firms get matched. (2.13) implies that at wages $w' > 0$ firms can hire for sure, i.e. the effective queue length at wages above zero is infinity. Therefore, small increases in the wage are profitable. *Q.E.D.*

Proof of Lemma 4, part 1:

Instead of equations (2.22) and (2.23), we now have

$$1 - e^{-\mu_1} - \mu_1 e^{-\mu_1} = \pi^*, \tag{A.21}$$

$$(1 - e^{-\mu_2} - \mu_2 e^{-\mu_2})(1 - e^{-\mu_1}) = \pi^*, \tag{A.22}$$

for some endogenous profit π^*. Consider π^* as a free parameter. For a given π^* (A.21) and (A.22) uniquely determine the measure $\hat{v}_1$ and $\hat{v}_2$ of firms in the low and high group. That is, π^* is supported by a unique measure $\hat{v} = \hat{v}_1 + \hat{v}_2$ of firms. We want to show that there is only a single π^* that is supported by $\hat{v} = v$, which then establishes uniqueness v_1 and v_2 (and thus of F as in part 2). By (A.21) μ_1 strictly increases in π^*. Equal profits at high and low wage firms implies

$$1 - e^{-\mu_2} - \mu_2 e^{-\mu_2} = 1 - \mu_1 e^{-\mu_1}/(1 - e^{-\mu_1}), \tag{A.23}$$

which implies that μ_2 is strictly increasing in π^*, since $\mu_1 e^{-\mu_1}/(1 - e^{-\mu_1})$ is strictly decreasing in μ_1. Since $\mu_2 = \gamma_2/\hat{v}_2$, $\hat{v}_2$ is strictly decreasing in π^*. We have proven the lemma if we can show that also $\hat{v}_1 + \hat{v}_2$ is decreasing in π^*. Since $\mu_1 = (\gamma_1 + \gamma_2 - \gamma_2 p_2)/\hat{v}_1$ we get $\partial\mu_1/\partial\pi^* = -[\mu_1/\hat{v}_1][\partial\hat{v}_1/\partial\pi^*] - (1/\hat{v}_1)(1 - e^{-\mu_2} - \mu_2 e^{-\mu_2})[\partial\hat{v}_2/\partial\pi^*]$. By the prior argument this derivative has to be strictly positive, which together with $\mu_1 > 1 - e^{-\mu_2} - \mu_2 e^{-\mu_2}$ implies $\partial\hat{v}_1/\partial\pi^* + \partial\hat{v}_2/\partial\pi^* < 0$. $\mu_1 > 1 - e^{-\mu_2} - \mu_2 e^{-\mu_2}$ holds because it is by (A.23) equivalent to $\mu_1 > 1 - \mu_1 e^{-\mu_1}/(1 - e^{-\mu_1})$, which is equivalent to $1 > (1 - e^{-\mu_1})/\mu_1$. The latter is true for all $\mu_1 > 0$. Since for $\pi^* \to 0$ we have $\hat{v} \to \infty$ and for $\pi^* \to 1$ we have $\hat{v} \to 0$, there is exactly one π^* supported by a measure $\hat{v} = v$ of firms. *Q.E.D.*

Proof of Proposition 16:

For $\gamma_0 = 1$ or $v = 0$, the result is trivial as matches are always zero. When workers send one application ($\gamma_2 = 0$), one group of firms with equal hiring probability is optimal because of strict concavity of the matching probability $1 - e^{-\lambda}$ (this is a special case of Shimer (2005)). For $\gamma_2 > 0$ we will prove that two groups of firms of which one receives all high applications and the other all low applications will be sufficient to achieve the same number of matches as any other optimal wage setting and application behavior $\{F, \boldsymbol{G}\}$.

Take $\{F, \boldsymbol{G}\}$ as a starting point. Consider some wage $w \in \mathcal{V} \cup \mathcal{W}$ with queue length $\mu(w)$ (other wages do not contribute to the matching). By assumption a continuum of firms offer this wage, and all face the same queue length. We will split the firms at this wage into two subgroups, and reshuffle the application behavior of the workers that send applications to this wage, such that their high application is randomly sent to some firm in the first and their low applications to some firm in the second subgroup. Workers that send both applications to w send one to each group, and accept offers from the first over offers from the second. We leave the application behavior towards other wages unchanged. We will show that for an appropriate choice of the relative size of the subgroups the overall matching is unchanged. Let $\lambda_h(w)$ denote the ratio of workers that only send their high application to wage w to firms offering w under $\{F, \boldsymbol{G}\}$. Let $\lambda_b(w)$ denote the worker/firm ratio for workers that send both applications to w. Let $(1-\bar{\psi})\lambda_l(w)$ be the ratio of worker/firm ratio for workers who send their low or single application to w and do not get a strictly better offer. If $\{F, \boldsymbol{G}\}$ is optimal, then neither of these ratios is infinity, and not all of them are zero (except possibly for some wages that attract a zero measure of agents, which we can neglect without loss of optimality). If only $\lambda_h(w)$ (respectively $(1-\bar{\psi})\lambda_l(w)$) is strictly positive, then we can trivially avoid to change the matching by having a zero fraction of the firms in the second (resp. first) subgroup. Therefore consider the case where at least two of the ratios are strictly positive.

First, we will show that if the two subgroups of firms face some identical effective queue length μ', then $\mu' = \mu(w)$. In this case we clearly have not changed the overall matching in the economy. The prove is by contraposition: Assume $\mu' > \mu(w)$. That means that strictly more firms then before get matched at wage w. On the other hand it becomes strictly harder for workers to get an offer, and since we did not change the application behavior at other wages, strictly less workers get matched at wage w. Since workers and firms are matched in pairs, this yields the desired contradiction. Similarly $\mu' < \mu(w)$ can be ruled out.

Next, we will show that we can indeed equalize the effective queue lengths for both subgroups. Let the fraction of firms in the first subgroup be d. Then the effective queue length for these firms is $\mu_h(d) = \frac{\lambda_h(w)+\lambda_m(w)/2}{d}$, because all applications are effective. For those firms in the second group it is $\mu_l(d) = \frac{(1-\bar{\psi})\lambda_l(w)}{1-d} + \frac{1}{2}\frac{1-e^{-\mu_h(d)}}{\mu_h(d)}\frac{\lambda_b(w)}{1-d}$. For d close to zero $\mu_h(d) > \mu_l(d)$, while for d close to 1 $\mu_h(d) < \mu_l(d)$. By the intermediate value theorem it is possible to equalize both at some $d(w)$.

This shows that for any wage we can conceptually split the firms into some that only receive high and some that only receive low applications without altering the overall matching. Doing this for all wages, we are left with a group of firms comprising all subgroups that only receive high applications (and have $d(w) > 0$), and a group of firms comprising all subgroups that only receive low applications (and have $d(w) < 1$). This resembles two-group matching except for the fact that firms in the same group but from different subgroups may still face different ψ's and λ's.

Consider low (or single) application firms first. Consider two subgroups, one with matching probability $1 - e^{-(1-\psi)\lambda}$, and one with $1 - e^{-(1-\psi')\lambda'}$. From the larger of the subgroups select a subset of firms with equal size to the smaller subgroup. We will show that in an optimal allocation both have the same queue length by shifting firms from one group to the other while leaving the applications that each group receives the same. Let d be the fraction of firms in the first subgroup, and γ and γ' the gross queue length per group. Then the average matchings across both groups is given by

$$d(1 - e^{-(1-\psi)\frac{\gamma}{\nu d}}) + (1-d)(1 - e^{-(1-\psi')\frac{\gamma'}{\nu(1-d)}}). \tag{A.24}$$

Since both subgroups have a strictly positive effective queue length, it cannot be optimal to place all firms in only one subgroup (as otherwise few firms placed in the other would be matched nearly for certain). Therefore, to achieve optimal matching d is characterized by the first order condition

$$\nu[(1 - e^{-\mu}) - (1 - e^{-\mu'}) - \mu e^{-\mu} + \mu' e^{-\mu'}] = 0, \tag{A.25}$$

where $\mu = (1-\psi)\frac{\gamma}{\nu d}$ and $\mu' = (1-\psi')\frac{\gamma'}{\nu(1-d)}$. Since $1 - e^{-\mu} - \mu e^{-\mu}$ is strictly increasing in μ (and similar for μ'), we have $\mu = \mu'$ in the optimal allocation of firms. That means that almost all low or single application firms have the same effective queue length. Reshuffling all effective applications randomly over all firms in the group that receive low or single applications without changing the applications to other firms does therefore not change the overall matching, and we have for this group matching as for the non-preferred group under two-group search.

By this construction, for low and single application firms only the average matching probability at high application firms matters. If we keep the size of low and single application firms constant and leave the gross queue length for them unchanged, but match more workers already at high wage firms, this clearly improves the matching (despite some negative externality on the low or single application firms). By the strict concavity of $1 - e^{-\lambda}$ the average matching probability at high wage firms is maximized if the gross queue length (and thus the effective queue length) is identical for all of them. Therefore the optimal allocation can be achieved by having one group of high wage firms to which workers randomly send their high application, which corresponds to the preferred group in two-group-search.

By the two-group-efficiency of the equilibrium matching, the equilibrium matching is constrained efficient. *Q.E.D.*

Proof of Proposition 17:

Given v and $\boldsymbol{\gamma}$ with $\gamma_2 > 0$, consider two tuples $\{F', \boldsymbol{G}'\}$ and $\{F'', \boldsymbol{G}''\}$ that lead to equal hiring probabilities η' respectively η'' for all firms. Similar to the argument in the previous lemma $\eta' = \eta'' = \bar{\eta}$, since otherwise one tuple would match more workers but fewer firms than the other.

We can again split the firms into two groups, called first and second, all workers send their high application to the second and their low or single application to the first, and accept offers from the second over those from the first. Let d be the fraction of firms in the second group. Again $\mu_2 = \lambda_2 = \gamma_2/(vd)$ and $\mu_1 = [1 - \gamma_2 p_2/(\gamma_1 + \gamma_2)]\lambda_1 = [\gamma_1 + \gamma_2 - \gamma_2(1 - e^{-\lambda_2})/\lambda_2]/(v(1-d))$. Since for $d \approx 0$ clearly $\mu_2 > \mu_1$ and for $d \approx 1$ $\mu_2 < \mu_1$, there exists a $\hat{d}$ such that effective queue length and thus the hiring probability of both groups is equalized. It is easy to show that $\mu_1 - \mu_2$ is strictly increasing in d around $\mu_2 \approx \mu_1$, so that $\hat{d}$ is unique. This two-group process has $\mu_1 = \mu_2$, but the optimal two group process fullfils (2.27), which requires $\mu_1 < \mu_2$, i.e. a strictly smaller preferred group. *Q.E.D.*

Explanation to equation (2.7′):

Consider a particular worker who applies to h firms that offer wage w. For simplicity assume he applies to no other firms. Conditional on the fact that a firm makes an offer to the worker sometime during the recall process, we want to determine the probability σ that this firm hires the worker. The unconditional probability that the worker does not get any offer is $(1 - p(w))^h$, and so the unconditional probability that he gets an offer is $\varsigma = 1 - (1 - p(w))^h$. The unconditional probability of getting an offer from any specific firm is ς/h. This unconditional probability can also be written as the probability of hiring conditional on making an offer multiplied by the unconditional probability of making an offer. So we have $\sigma p(w) = \varsigma/h$, or $\sigma = [1 - (1 - p(w))^h]/[hp(w)]$, which explains the formula used for equation (2.7′). The argument is based on insights from Burdett, Shi and Wright (2001).

Proof of Proposition 20:

We start out by fixing $\boldsymbol{\gamma}$ and denote by $\hat{i}$ the highest integer for which $\gamma_i > 0$; i.e. $\hat{i}$ is the maximum number of applications that workers send. By straightforward extension of the analysis of the workers' best response in section 3 it can be established that the workers' best response to a wage offer distribution is now given by $\hat{i}$ intervals such that every worker sends exactly one application to each interval. More specific, let the utility from sending the first i applications optimally be defined recursively by $u_i \equiv \max_{w\in[0,1]} p(w)w + (1-p(w))u_{i-1}$ for all $i \in \{1, 2, ...\hat{i}\}$, with $u_0 \equiv 0$. Then for any wage offer distribution the effective queue length is characterized by $(u_1, .., u_n, \bar{w}_0, ..., \bar{w}_N)$ such that

$$p(w) \;=\; 1 \qquad \forall w \in [0, \bar{w}_0],\; \textit{and} \tag{A.26}$$

$$p(w)w + (1-p(w))u_{i-1} \;=\; u_i \quad \forall w \in [\bar{w}_{i-1}, \bar{w}_i] \;\; \forall i \in \{1, ..., N\}, \tag{A.27}$$

where $\bar{w}_0 = u_1$ and $\bar{w}_N = 1$. The indifference implies $\bar{w}_i = u_{i-1} + [u_i - u_{i-1}]^2/(2u_i - u_{i-1} - u_{i+1})$ for intermediate $i \in \{1, ..., \hat{i}\}$. Clearly $\bar{w}_{\hat{i}} \geq \sup \mathcal{V}$, because any wages that are actually offered receive applications from workers that at most send $\hat{i}$ applications. At higher (not offered) wages, workers might start sending additional applications. The Market Utility Assumption implies that they cannot receive more than the Market Utility, which implies that $u_i - u_{i-1} = c_i$ for $i > \hat{i}$. The indifference then yields $\bar{w}_i = u_{i-1} + [u_i - u_{i-1}]^2/(u_i - u_{i-1} - c_{i+1})$. If this is in $[0, 1]$ then this gives the appropriate boundary, otherwise $\bar{w}_i = 1$ and workers would strictly refrain from sending this many applications.

Using (A.27), we can rewrite the profit function for a firm who offers a wage $w \in [\bar{w}_{i-1}, \bar{w}_i]$ with $\bar{w}_{i-1} < 1$ as

$$\pi(\mu) = (1 - e^{-\mu})(1 - u_{i-1}) - \mu(u_i - u_{i-1}), \tag{A.28}$$

where $\mu = \mu(w)$. The logic is similar to (2.15). If $\bar{w}_{i-1} = 1$ the profit is trivially zero. Proposition 12, stating that there exists no equilibrium in which only one wage is offered, can now easily be shown with similar techniques whenever $\gamma_i > 0$ for some $i > 1$. By a similar argument it is straightforward that at least i wages have to be offered in equilibrium whenever $\gamma_i > 0$. Given that (A.28) is strictly concave, it is also immediate that all firms within the same interval will offer the same wage, yielding exactly $\hat{i}$ wages when workers send at most $\hat{i}$ applications.

We call the group of firms that ends up offering the i'th highest wage as group i and will index all their variables accordingly. It will be convenient to denote by $\Gamma_i = \sum_{k=i}^{\hat{i}} \gamma_k$ the fraction of workers who apply to at least i firms. Then at wage i the probability of retaining an applicant is $(1-\psi_i) = \sum_{j=i}^{\hat{\imath}} \frac{\gamma_j}{\Gamma_i}[\prod_{k=i+1}^{j}(1-p_k)]$, since a fraction γ_j/Γ_i of applicants sends j applications and does not get a better job with probability $\prod_{k=i+1}^{j}(1-p_k)$. The effective queue length at wage i is given by $\mu_i = (1-\psi_i)\lambda_i$, where $\lambda_i = \Gamma_i/v_i$ is the gross queue length. For $i < \hat{\imath}$ the unique offered wage in $[\bar{w}_{i-1}, \bar{w}_i]$ is obtained by the first-order-conditions of (A.28), which are given by

$$u_i - u_{i-1} = e^{-\mu_i}(1 - u_{i-1}). \tag{A.29}$$

Therefore (A.28) can be rewritten as

$$\pi_i = (1 - e^{-\mu_i} - \mu_i e^{-\mu_i})(1 - u_{i-1}). \tag{A.30}$$

Free entry implies that $\pi_i = K$, which together with (A.29) implies that $\mu_i = \mu_i^*$ and $u_i = u_i^*$ as defined above. By a similar argument as for (2.22) and (2.23) the condition $\pi_i = K$ defines for a given vector $\boldsymbol{\gamma}$ of applications the unique measure v_i of firms in each group, and the wage $w_i = u_{i-1}^* + (u_i^* - u_{i-1}^*)/p_i$ that each group of firms offers (when $i < \hat{\imath}$). Note that the equal profit condition $\pi_i = \pi_{i-1}$ together with (A.29) implies

$$1 - e^{-\mu_i} - \mu_i e^{-\mu_i} = 1 - \frac{\mu_{i-1}e^{-\mu_{i-1}}}{1 - e^{-\mu_{i-1}}}, \tag{A.31}$$

which corresponds to (A.23).

To determine the equilibrium, only $\boldsymbol{\gamma}$ has to still be determined. Recall that i^* denotes the number of applications for which $c_{i^*} < u_{i^*}^* - u_{i^*-1}^*$ and $c_{i^*+1} > u_{i^*+1}^* - u_{i^*}^*$. Consider first the case where $\boldsymbol{\gamma}$ is such that $\hat{\imath} < i^*$. Since profits are determined by first order conditions, there cannot be any wage in $[0, \bar{w}_{\hat{\imath}}]$ that offers higher profits. Therefore a deviating firm has to consider a deviation within $(\bar{w}_{\hat{\imath}}, 1)$. In this region the effective queue length is (at least) given according to $p(w)w + (1-p(w))u_{\hat{\imath}}^* = u_{\hat{\imath}}^* + c_{\hat{\imath}+1}$ (it may even be larger if workers send two or more additional applications to not-offered high wages). Therefore the profit for a deviating firm is (at least) $\pi(\hat{\mu}) = (1 - e^{-\hat{\mu}} - \hat{\mu}e^{-\hat{\mu}})(1 - u_{\hat{\imath}}^*)$, where $\hat{\mu}$ is given by the first order condition $c_{\hat{\imath}+1} = e^{-\hat{\mu}}(1 - u_{\hat{\imath}}^*)$. Since $c_{\hat{\imath}+1} < e^{-\mu_{\hat{\imath}+1}^*}(1 - u_{\hat{\imath}}^*)$, we have $\hat{\mu} > \mu_{\hat{\imath}}$. This implies $\pi(\hat{\mu}) > (1 - e^{-\mu_{\hat{\imath}+1}^*} - \mu_{\hat{\imath}+1}^* e^{-\mu_{\hat{\imath}+1}^*})(1 - u_{\hat{\imath}}^*) = K$, where the equality follows from the definition of $\mu_{\hat{\imath}+1}^*$. The optimal deviating wage can indeed be shown to lie above $\bar{w}_{\hat{\imath}}$ and therefore the deviation is profitable.

Clearly also $\hat{\imath} > i^*$ cannot be an equilibrium, because for workers who send $\hat{\imath}$ applications the marginal costs of the last application do not cover its marginal benefit.

For the case where $\hat{\imath} = i^*$ all workers want to send exactly i^* applications. We will show that $\bar{w}_{i^*} = 1$, which implies that firms do not have a profitable deviation, which establishes the existence and uniqueness result. If $\bar{w}_{i^*} = 1$, it means that the queue length in $[\bar{w}_{i^*-1}, 1]$ is determined by $p(w)w + (1-p(w))u^*_{i^*-1} = u^*_{i^*}$. Note that is implies $p(1) = [u^*_{i^*} - u^*_{i^*-1}]/[1 - u^*_{i^*-1}] = e^{-\mu^*_{i^*}}$, where the second equality follows from the definition of $\mu^*_{i^*}$. Determining the effective queue length this way is in accordance with the Market Utility Assumption if and only if it is not profitable to send an additional application. If an additional application is sent, by a logic similar to lemma 3 it is optimal to send it to the highest wage. The marginal benefit would be $p(1) + (1-p(1))u^*_{i^*} - u^*_{i^*}$, or $e^{-\mu^*_{i^*}}[1 - u^*_{i^*}]$. Since $c_{i^*+1} < e^{-\mu^*_{i^*+1}}[1 - u^*_{i^*}]$ and $\mu^*_{i^*+1} > \mu^*_{i^*}$ it is not profitable for workers to send another application to a deviant firm. Therefore $\bar{w}_{i^*} = 1$ is indeed the correct specification.
By similar arguments it is easy to see that for $c_{i^*} = u^*_i - u^*_{i-1}$ equilibria exist if and only if $\boldsymbol{\gamma}$ has $\gamma_{i^*} + \gamma_{i^*-1} = 1$, $\gamma_{i^*} \in [0,1]$; i.e. workers randomize over i^* and $i^* - 1$ applications.

To show constrained efficiency, we will first consider search efficiency for given $\boldsymbol{\gamma}$ and v. Let $\hat{\imath}$ still denote the maximum number of applications that workers send. Consider $\hat{\imath}$ groups of firms, with $d_i v$ firms in each groups, which are ordered by their attractiveness for workers. That is, a worker who applies to i firms applies once to each of the lowest i groups and accepts an offer from a higher group over an offer from a lower group. We call an allocation of firms across groups that leads to the maximum number of matches $\hat{\imath}$-group-efficient. Compare two adjacent groups i and $i-1$ with total measure $\nu = v_i + v_{i-1}$. We show that the only efficient way of dividing this measure up between the two groups is the equilibrium division. The maximal total number of matches within these groups is given by

$$\max_{d \in [0,1]} M(d) = \nu d(1 - e^{-\mu_i}) + \nu(1-d)(1 - e^{-\mu_{i-1}}). \tag{A.32}$$

It can be shown that a boundary solution cannot be optimal, as it means that one application is waisted. Noting that $(1 - \psi_{i-1}) = \gamma_{i-1}/\Gamma_{i-1} + (1-\psi_i)(1-p_i)\Gamma_i/\Gamma_{i-1}$ we can write $\mu_i = (1-\psi_i)\lambda_i$ and $\mu_{i-1} = [\gamma_{i-1}/\Gamma_{i-1} + (1-\psi_i)(1-p_i)\Gamma_i/\Gamma_{i-1}]\lambda_{i-1}$. The first derivative is then

$$\begin{aligned}\frac{\partial M(d)}{\partial d}\frac{1}{\nu} &= 1 - e^{-\mu_i} - (1 - e^{-\mu_{i-1}}) + e^{-\mu_i} d(1-\psi_i)\frac{\partial \lambda_i}{\partial d} \\ &+ e^{-\mu_{i-1}}(1-d)\left[(1-\psi_{i-1})\frac{\partial \lambda_{i-1}}{\partial d} - \frac{\Gamma_i}{\Gamma_{i-1}}(1-\psi_i)\frac{\partial p_i}{\partial d}\lambda_{i-1}\right]\end{aligned}$$

We can use similar substitutions as for (2.27), with the adjustment that now $\partial \mu_i/\partial d =$

$-\mu_i/d = -\nu\mu_i^2/[(1-\psi_i)\Gamma_i]$, to show that the last term in the first line equals $-\mu_i e^{-\mu_i}$, and the second line reduces to $e^{-\mu_{i-1}}[\mu_{i-1} - (1 - e^{-\mu_i} - \mu_i e^{-\mu_i}]$. Therefore we again have

$$\frac{\partial M(d)}{\nu \partial d} = (1 - e^{-\mu_i} - \mu_i e^{-\mu_i})(1 - e^{-\mu_{i-1}}) - (1 - e^{-\mu_{i-1}} - \mu_{i-1} e^{-\mu_{i-1}}) = 0. \qquad \text{(A.33)}$$

The first order condition implies equality with zero. For given ν this uniquely characterizes the optimal interior d, since similar substitutions as above yield $\partial^2 M/\partial d^2 = -\nu[\mu_2^2 e^{-\mu_2}(1 - e^{-\mu_1})/d + e^{-\mu_1}(1 - e^{-\mu_2} - \mu_2 e^{-\mu_2} - \mu_1)^2/(1-d)] < 0$. It is straightforward to show that for a given measure v of firms there exists an $\hat{i}$-group efficient allocation across all $\hat{i}$ groups. A similar construction as in the proof of proposition 16 shows that $\hat{i}$ groups are sufficient to achieve the constrained optimal search outcome. A similar construction as in proposition 17 shows that the outcome of a random process (i.e. one wage) could also be achieved with $\hat{i}$ groups, but the division of firms across groups would not be optimal. More generally, such an argument establishes that the optimal allocation cannot be achieved with less than $\hat{i}$ wages given the number of applications summarized in $\boldsymbol{\gamma}$.

It is very tedious to analyze whether (A.33) - which coincides with profit equality as in (A.31) - determines the allocation of firms to the $\hat{i}$ groups uniquely for any v. Therefore we will not consider the efficiency of search in the case without free entry. We will establish that the overall entry of firms and the measure of firms in each group under equilibrium conditions 1a), 1b), 2a) and 3) yields optimal entry and optimal search simultaneously, taking $\boldsymbol{\gamma}$ as given. The important insight from the previous analysis of constrained optimal search is that (A.33) has to hold in the optimal search outcome for all $i \in \{2, ..., \hat{i}\}$, and that we can apply the envelop theorem. Let $\boldsymbol{d}(v) = (d_1(v), d_2(v), .., d_{\hat{i}}(v))$ be the fraction of firms in each of the $\hat{i}$ groups under constrained optimal search given v and $\boldsymbol{\gamma}$. Again let $M^*(\boldsymbol{\gamma}, v, \boldsymbol{d}(v))$ denote the constrained efficient number of matches given v and $\boldsymbol{\gamma}$. Similar to (2.28) the objective function is given by $\max_{v \geq 0} M^*(\boldsymbol{\gamma}, v, \boldsymbol{d}(v)) - vK$. When $\hat{i} > 0$, then $K < 1$ ensures that the optimal solution is in the interior of $[0, V]$. We will show that the first order condition uniquely determines the solution and corresponds to the free entry condition.

By the envelope theorem the impact of a change of the fraction $d_i(v)$ of firms in each group on the measure of matches can be neglected, i.e. $\frac{\partial M^*}{\partial d_i}\frac{\partial d_i}{\partial v} = 0$ at the $\hat{i}$-efficient d_i. We get as first order condition

$$\partial M^*(\boldsymbol{\gamma}, v, \boldsymbol{d})/\partial v = K, \qquad \text{(A.34)}$$

where $\boldsymbol{d} = \boldsymbol{d}(v)$. Writing $M^*(\boldsymbol{\gamma}, v, \boldsymbol{d}) = [1 - \sum_{i=1}^{\hat{i}}[\gamma_i \prod_{j=1}^{i}(1-p_j)]$ we have

$$\begin{aligned}\frac{\partial M^*(\boldsymbol{\gamma}, v)}{\partial v} &= \sum_{i=1}^{\hat{i}}\Big[\gamma_i \sum_{j=1}^{i}[\frac{\partial p_j}{\partial v}\prod_{\substack{k\leq i\\k\neq j}}(1-p_k)]\Big] \\ &= \sum_{i=1}^{\hat{i}}\Big[\frac{\partial p_i}{\partial v}\Gamma_i(1-\psi_i)\prod_{k<i}(1-p_k)\Big], \qquad (A.35)\end{aligned}$$

where the second line is obtained by rearranging the terms for each $\partial p_i/\partial v$. To simplify notation, define the partial sum

$$\xi_{i'} = \sum_{i \geq i'}^{N}[\frac{\partial p_i}{\partial v}\Gamma_i(1-\psi_i)\prod_{k<i}(1-p_k)]. \qquad (A.36)$$

Since $p_{\hat{i}} = (1-e^{-\mu_{\hat{i}}})/\mu_{\hat{i}}$ we have $\partial p_{\hat{i}}/\partial v = -(1/\mu_{\hat{i}}^2)(1-e^{-\mu_{\hat{i}}} - \mu_{\hat{i}}e^{-\mu_{\hat{i}}})(\partial\mu_{\hat{i}}/\partial v)$. Since $\mu_{\hat{i}} = \gamma_{\hat{i}}/(d_{\hat{i}}v)$, we have $\partial\mu_{\hat{i}}/\partial v = -\gamma_{\hat{i}}/(d_{\hat{i}}v^2) = -d_{\hat{i}}\mu_{\hat{i}}/\gamma_{\hat{i}}$. So we get $\partial p_{\hat{i}}/\partial v = -d_{\hat{i}}(1 - e^{-\mu_{\hat{i}}} - \mu_{\hat{i}}e^{-\mu_{\hat{i}}})/\gamma_{\hat{i}}$. Noting that $\Gamma_{\hat{i}}(1-\psi_{\hat{i}}) = \gamma_{\hat{i}}$, we have established that

$$\xi_{\hat{i}} = d_{\hat{i}}(1-e^{-\mu_{\hat{i}}} - \mu_{\hat{i}}e^{-\mu_{\hat{i}}})\prod_{k<\hat{i}}(1-p_k). \qquad (A.37)$$

By induction we can establish the following lemma, which we will prove subsequently because it would distract from the argument at this point.

Lemma A1 *For all i it holds that*

$$\xi_i = (\sum_{k=i}^{N} d_k)(1-e^{-\mu_i} - \mu_i e^{-\mu_i})\prod_{j<i}(1-p_k). \qquad (A.38)$$

This implies that $\xi_1 = 1 - e^{\mu_1} - \mu_1 e^{-\mu_1}$. The first order condition $\xi_1 = K$ uniquely defines μ_1, and corresponds to the free entry condition of the lowest wage firms. By (A.33) (or respectively by (A.31)) it also determines μ_i uniquely for all $i \in 2, ..., \hat{i}$, which in turn determines v_i uniquely for all $i \in 1, ..., \hat{i}$. Thus, the measure of firms in each group under equilibrium conditions 1a), 1b), 2a) and 3) coincides with the measure of firms in each group implied by the first order conditions for optimal entry (incorporating optimal subsequent search). Since there is only one allocation fulfilling the first order conditions, and boundary solutions are not optimal, this again characterizes the global maximum. Thus equilibrium entry and search is constrained optimal given $\boldsymbol{\gamma}$.

Finally, when we endogenize $\boldsymbol{\gamma}$, again note that the number of applications of other workers in equilibrium is not important for the marginal benefits of each individual worker, which are always $u_i^* - u_{i-1}^*$. Therefore again the decision on the number of applications is constrained efficient, establishing constrained efficiency overall. *Q.E.D.*

Proof of Lemma A1:

We are left to show that the following holds for all $i \in \{1, ..., \hat{\imath} - 1\}$:

$$\xi_{i+1} = (\sum_{k=i+1}^{\hat{\imath}} d_k)(1 - e^{-\mu_{i+1}} - \mu_{i+1} e^{-\mu_{i+1}}) \prod_{j<i+1} (1 - p_k). \tag{A.39}$$

It clearly holds for $i = \hat{\imath} - 1$ by $(A.37)$. Now assume it holds for some i. We will consider ξ_i. We know that

$$\xi_i = \xi_{i+1} + \Gamma_i (1 - \psi_i) \frac{\partial p_i}{\partial v} \prod_{k<i} (1 - p_k). \tag{A.40}$$

The second summand can be written as

$$\frac{\partial p_i}{\partial v} \prod_{k<i} (1 - p_k) \quad = \quad -\frac{1 - e^{-\mu_i} - \mu_i e^{-\mu_i}}{\mu_i^2} [\frac{\partial \mu_i}{\partial v} \prod_{k<i} (1 - p_k)] \tag{A.41}$$

Since $\mu_i = \lambda_i (1 - \psi_i) = \lambda_i (\sum_{j=i}^{\hat{\imath}} \frac{\gamma_j}{\Gamma_i} (\prod_{k=i+1}^{j} (1 - p_k)))$ we can write the term in square brackets in (A.41) as

$$\begin{aligned} \frac{\partial \mu_i}{\partial v} \prod_{k<i} (1 - p_k) &= -\frac{\Gamma_i (1 - \psi_i)}{d_i v^2} \left[\prod_{k<i} (1 - p_k) \right] + \xi_{i+1} \frac{\lambda_i}{\Gamma_i (1 - p_i)} \\ &= -\frac{d_i \mu_i^2}{\Gamma_i (1 - \psi_i)} \left[\prod_{k<i} (1 - p_k) \right] + \xi_{i+1} \frac{\lambda_i}{\Gamma_i (1 - p_i)}. \end{aligned}$$

Observing that $\frac{1}{\mu_i}(1 - e^{-\mu_i} - \mu_i e^{-\mu_i}) = p_i - e^{-\mu_i}$, we can substitute the prior equation into (A.41) and multiply by $\Gamma_i(1 - \psi_i)$ to get

$$\Gamma_i (1 - \psi_i) \frac{\partial p_i}{\partial v} \prod_{k<i} (1 - p_k) = \frac{p_i - e^{-\mu_i}}{1 - p_i} \xi_{i+1} + d_i (1 - e^{-\mu_i} - \mu_i e^{-\mu_i}) \prod_{j<i} (1 - p_j).$$

We can substitute this into (A.40), and use (A.39) and the property of $\hat{\imath}$-group-efficient search in (A.31) to obtain

$$\xi_i = (\sum_{k=i}^{N} d_k)(1 - e^{-\mu_i} - \mu_i e^{-\mu_i}) \prod_{j<i} (1 - p_k). \tag{A.42}$$

Q.E.D.

Proof of Proposition 21:

First we show that for $i^* \to \infty$ the (weakly) shorter side of the market gets matched with probability approaching 1. Since equilibrium search is always more efficient than a process of random applications and acceptances, we will show this for the latter. As $i^* \to \infty$ it cannot happen that workers and firms both are matched with probabilities bounded away from one. If that were the case, than some fraction $\alpha > 0$ of firms would always remain unmatched. But then the chance that a worker applies to such a firm with any given application is α, so that the probability that he applies to such a firm with at least one of his applications converges to 1, yielding a contradiction. With unequal sizes it is obviously the shorter side whose probability of being matched converges to one; with equal sizes the probability of being matched is the same and agents from both sides get matched with probability converging to one.

For the next arguments, recall that the marginal utility gain (excluding the marginal application cost) of the i^*'th application, given by $u^*_{i^*} - u^*_{i^*-1}$, converges to zero as $i^* \to \infty$. We will use this to establish the limit for the average wage if firms are either on the long or on the short side of the market.

Case 1: We will show that $w(i^*) \to 0$ if firms are strictly on the short side of the market. Assume there exists a subsequence of i^*'s such that $v(i^*) < 1 - \epsilon$ for all i^* and some $\epsilon > 0$. That implies $\varrho(i^*) < \alpha$ for some $\alpha < 1$. If $w(i^*) \nrightarrow 0$, then there exists a subsequence such that $w(i^*) \to \omega > 0$ and $\pi(i^*) \to 1 - \omega$ (since $\eta(i^*) \to 1$). Now consider a deviant firm that always offers wage $w' = \omega/2$. As workers send more applications, the hiring probability for the deviant has to converge to 1. This is due to the fact that for workers the marginal utility of sending the last application converges to zero, which implies that the probability of getting the job at the deviant firm has to become negligible as otherwise each worker would like to send his last application there to insure against the $1 - \alpha$ probability of not being hired. With the hiring probability approaching 1 the profit of the deviant converges to $1 - \omega/2$, i.e. the deviation is profitable. Thus it has hold that $w(i^*) \to 0$.

Case 2: We will show that $w(i^*) \to 1$ if firms are strictly on the long side of the market. Assume there exists a subsequence of i^*'s such that $v(i^*) > 1+\epsilon$ for all i^* and some $\epsilon > 0$. In this case $\eta(i^*) < \alpha$ for some $\alpha < 1$ and all i^*. If $w(i^*) \nrightarrow 1$, then there exists a subsequence such that $w(i^*) \to \omega < 1$ and $\pi(i^*) \to \pi < \alpha(1 - \omega)$. Consider a firm that always offers wage $w' \in (\omega, 1)$ such that $1 - w' > \alpha(1 - \omega)$. Again the hiring probability of the deviant converges to 1, because if there were a non-negligible chance of getting the job at w' worker's

would rather send there last application to this higher than average wage. But then the deviant's profit converges to $1-w'$ and the deviation is profitable. So $w(i^*) \to 1$.
This immediately implies that $v(i^*) \to 1$. Otherwise a subsequence of i^*'s according either to case 1 or to case 2 has to exist, but in case 1 profits are above entry costs and in case 2 they are below entry costs, violating the free entry condition. Finally, since $v(i^*) \to 1$ and firms get matched with probability close to one, $\pi = K$ implies that the average paid wage $w(i^*)$ has to converge to $1-K$. This directly implies that $u^*_{i^*} \to 1-K$.
To show that the individual search effort converges to zero, i.e. that also $U^*(i^*) = u^*_{i^*} - c^{i^*}(i^*) \to 1-K$, rewrite the workers' utility as $U^*(i^*) = \sum_{i=1}^{i^*}[u^*_i - u^*_{i-1} - c^{i^*}_i] = \sum_{i=1}^{I}[u^*_i - u^*_{i-1} - c^{i^*}_i] + \sum_{i=I+1}^{i^*}[u^*_i - u^*_{i-1} - c^{i^*}_i]$ for some $I \leq i^*$, where $c^{i^*}_i = c^{i^*}(i) - c^{i^*}(i-1)$ again denotes marginal costs. For a given i the difference $u^*_i - u^*_{i-1}$ is simply a number independent of i^* (and the associated cost function). It converges to zero for large i, which entails that $u^*_{i^*} - u^*_{i^*-1} \to_{i^*\to\infty} 0$. Moreover $c^{i^*}_i \leq c^{i^*}_{i^*} \leq u^*_{i^*} - u^*_{i^*-1}$ for all $i \leq i^*$, which only restates that that we consider changing cost functions with $c^{i^*}_i \to_{i^*\to\infty} 0$. Therefore the partial sum $\sum_{i=1}^{I}[u^*_i - u^*_{i-1} - c^{i^*}_i] \to_{i^*\to\infty} \sum_{i=1}^{I}[u^*_i - u^*_{i-1}]$ for any fixed $I \in \mathbb{N}$. On the other hand we have $0 \leq \sum_{i=I+1}^{i^*}[u^*_i - u^*_{i-1} - c^{i^*}_i] \leq \sum_{i=I+1}^{\infty}[u^*_i - u^*_{i-1}]$, but $\sum_{i=I+1}^{\infty}[u^*_i - u^*_{i-1}] \to_{I\to\infty} 0$ since $\sum_{i=1}^{\infty}[u^*_i - u^*_{i-1}] \leq 1$. Therefore $\lim_{i^*\to\infty} U(i^*) = \lim_{I\to\infty}\lim_{i^*\to\infty}[\sum_{i=1}^{I}[u^*_i - u^*_{i-1} - c^{i^*}_i] + \sum_{i=I+1}^{i^*}[u^*_i - u^*_{i-1} - c^{i^*}_i]] = \sum_{i=1}^{\infty}[u^*_i - u^*_{i-1}] = \lim_{i^*\to\infty} u^*_{i^*} = 1-K$. *Q.E.D.*

A.3 Appendix to Chapter 3

It will be convenient to prove lemma 7 prior to lemma 6.

Proof of Lemma 7: We consider first the case where the consumer is promised the same service $s \in \{0, \bar{s}\}$ from every firm in every period. Having characterized the consumer search behavior for this case, it is straightforward to extend it to the case where low-quality firms promise less. We will work with average discounted payoffs. The functional equation for sampling with recall, given that the best quality the consumer has yet encountered is q, and given the current shock ρ, can be written as

$$\begin{aligned} V^C(q,\rho) \;=\; \max\{&(1-\delta)(q+s+\rho) + \delta E_{\rho'} V^C(q,\rho'), \\ &(1-\delta)(E_{q|\gamma}(q)+\rho) + \delta E_{\tilde{q}|\gamma} E_{\rho'} \max\{V^C(q,\rho'), V^C(\tilde{q},\rho')\}, \\ &(1-\delta)u_0 + \delta E_{\rho'} V^C(q,\rho')\}, \end{aligned} \tag{A.43}$$

where the first line describes the utility from returning to a known firm with quality q, the second line describes random sampling and the last line consumption of the numeraire. E_x denotes the expectation operator with regard to variable x. $x = q|\gamma$ refers to variable q when the probability of a high quality is γ.[23] We drop the decision-irrelevant constant $u(y-P)$. The right hand side of (A.43) defines an operator $T : Z \to Z$. If $[\underline{\rho}, \bar{\rho}]$ is bounded, $Z = \{\nu : \{q_l, q_h\} \times [\underline{\rho}, \bar{\rho}] \to \Re | \nu \text{ is continuous and bounded}\}$ and it is easily checked that T fulfills Blackwell's (1965) sufficient conditions for a contraction. Therefore, a solution to the problem exists and is unique. For $\bar{\rho} = \infty$ note that for all $\rho > \bar{\rho}' = u_0 - q_l + \frac{\delta}{1-\delta}(q_h + \bar{s} - q_l)$ the consumer will consume the indivisible good, because even if he loses high quality and high service forever (and only gets q_l now instead of u_0) the taste shock today outweighs the forgone benefits. Therefore we do not alter his decision problem if we restrict $[\underline{\rho}, \bar{\rho}]$ to $[\underline{\rho}, \bar{\rho}']$ and assume a distribution $F'(\rho) = F(\rho)$ for all $\rho < \bar{\rho}'$ and $F'(\rho) = 1$ for all $\rho \geq \bar{\rho}'$. Similarly we can bound the support from below by $\underline{\rho}' = u_0 - q_h - \bar{s} - \frac{1}{1-\delta}(q_h + \bar{s} - q_l)$ without altering the consumer's choice. On the restricted problem the contraction property establishes that (A.43) has a unique solution, and so the unrestricted problem has a unique solution.

From (A.43) note that $V^C(q, \rho)$ is weakly increasing in q. Therefore for $q = q_h$ the first line in the max-operator is larger than the second. Thus, whenever a consumer with state variable q_h enters the market, he will return to the firm with quality q_h rather than sample a new one. He enters the market if the taste shock is high enough, i.e., higher than $\hat{\rho}_h \in (\underline{\rho}, \bar{\rho})$ that makes the player indifferent between not consuming (line 3 in equation (A.43)) or going into the market (line 1), so that

$$(1-\delta)(q_h + s + \hat{\rho}_h) + \delta E_{\rho'} V^C(q_h, \rho') = (1-\delta)u_0 + \delta E_{\rho'} V^C(q_h, \rho')$$

or $\hat{\rho}_h = u_0 - q_h - s$.

Then in any given period the ex ante probability that this player will enter the market is $[1 - F(u_0 - q_h - s)]$, while the ex ante probability of not consuming is $F(u_0 - q_h - s)$. Knowing this, the expected average discounted payoff is

$$\begin{aligned} E_{\rho'} V^C(q_h, \rho') &= \int_{\underline{\rho}}^{\bar{\rho}} \max\{q_h + s + \rho, u_0\} dF(\rho) \qquad \text{(A.44)} \\ &= F(u_0 - q_h - s)u_0 + [1 - F(u_0 - q_h - s)](q_h + s) + \int_{u_0 - q_h - s}^{\bar{\rho}} \rho dF(\rho) \\ &= u_0 + \int_{u_0 - q_h - s}^{\bar{\rho}} [1 - F(\rho)] d\rho. \end{aligned}$$

[23] We used the shortcut γ for $q|\gamma$ in the main body of the text.

The second equality is simply the probability of not consuming times the opportunity cost, plus the probability of going into the market times the value from quality and service of doing so, plus the expected value of the taste shock when going into the market. The last line follows by integration by parts.

Now consider $q = q_l$. Assume that searching for a higher quality firm is preferable to staying at the low quality firm and obtaining service in the next period. (We will show in the subsequent proof of lemma 6 that this is indeed optimal). The threshold $\hat{\rho}_l$ for the taste shock is now given by the equality of line 2 and 3 in (**??**), so that

$$(1-\delta)(E_{q|\gamma}(q)+\hat{\rho}_l)+\delta(1-\gamma)E_{\rho'}V^C(q_l,\rho')+\delta\gamma E_{\rho'}V^C(q_h,\rho')=(1-\delta)u_0+\delta E_{\rho'}V^C(q_l,\rho'),$$

or

$$\frac{\delta\gamma}{1-\delta}\left[E_{\rho'}V^C(q_h,\rho')-E_{\rho'}V^C(q_l,\rho')\right]=u_0-E_{q|\gamma}(q)-\hat{\rho}_l. \tag{A.45}$$

Taking $\hat{\rho}_l$ as given, we can express the expected value as

$$\begin{aligned} E_{\rho'}V^C(q_l,\rho') &= F(\hat{\rho}_l)[(1-\delta)u_0+\delta E_{\rho'}V^C(q_l,\rho')] \\ &\quad +[1-F(\hat{\rho}_l)](1-\delta)\left[E_{q|\gamma}(q)+E_{\rho'}(\rho'|\rho'\geq\hat{\rho}_l)\right] \\ &\quad +[1-F(\hat{\rho}_l)]\delta\left[\gamma E_{\rho'}V^C(q_h,\rho')+(1-\gamma)E_{\rho'}V^C(q_l,\rho')\right]. \end{aligned}$$

The first line weights the opportunity cost of consumption by the probability $F(\hat{\rho}_l)$ of not consuming. The term $[1-F(\hat{\rho}_l)]$ in the second and third line reflects the probability of entering the market. The utility from doing so is comprised of two components. Line 2 reflects the instantaneous expected value from entering the market due to quality and taste shock, while line 3 represents the expected continuation value after encountering a firm with high or low quality respectively. After rearranging terms we have

$$\begin{aligned} E_{\rho'}V^C(q_l,\rho') &= F(\hat{\rho}_l)u_0+[1-F(\hat{\rho}_l)]E_{q|\gamma}(q)+\int_{\hat{\rho}_l}^{\bar{\rho}}\rho dF(\rho) \\ &\quad +[1-F(\hat{\rho}_l)]\frac{\delta\gamma}{1-\delta}\left[E_{\rho'}V^C(q_h,\rho')-E_{\rho'}V^C(q_l,\rho')\right]. \end{aligned}$$

Inserting (A.45) and rearranging gives

$$E_{\rho'}V^C(q_l,\rho')=u_0-[1-F(\hat{\rho}_l)]\hat{\rho}_l+\int_{\hat{\rho}_l}^{\bar{\rho}}\rho dF(\rho)=u_0+\int_{\hat{\rho}_l}^{\bar{\rho}}[1-F(\rho)]d\rho. \tag{A.46}$$

Substituting (A.46) and (A.44) into (A.45), we obtain an implicit function characterizing the threshold shock value $\hat{\rho}_l\in(u_0-q_h-s,u_0-E_{q|\gamma}(q))$:

$$\hat{\rho}_l-u_0+E_{q|\gamma}(q)+\frac{\delta\gamma}{1-\delta}\int_{u_0-q_h-s}^{\hat{\rho}_l}[1-F(\rho)]d\rho=0. \tag{A.47}$$

By the intermediate value theorem there is a solution to this equation, and the solution is unique as the left hand side is strictly increasing in $\hat{\rho}_l$.

Finally, note that when both firms offer service $s = \bar{s}$, the customer will not return to a low quality firm (see lemma 6). Since service is not provided in the first period, the customer will never experience service from any low quality firm, even if it promises to provide service should the customer return. Therefore the results also hold for the case where only high quality firms promise service $\bar{s}$, while low quality firms may not. *Q.E.D.*

Proof of Lemma 6: Consider a consumer of type θ with opportunity cost u_0^θ who has experienced only low quality firms. If low quality firms do not offer service, the consumer would search for a high quality firm, as nothing is lost by doing so.

If both types of firms offer service, the cost of searching consists of the forgone service (recall that first period service at a new firm is zero). Assume searching for a high quality firm is not optimal, given that the best firm encountered so far is low quality and all firms offer service. In other words a consumer always returns to the first firm he encounters. Similar to a derivation as in equation (A.44), the expected value at $q = q_l$ is then $E_{\rho'} V^C(q_l, \rho') = u_0^\theta + \int_{u_0^\theta - q_l - \bar{s}}^{\bar{\rho}} [1 - F(\rho)] d\rho$. The condition under which returning to the low quality firm rather than searching is optimal is then

$$\begin{aligned} & (1-\delta)(q_l + \bar{s} + \rho) + \delta E_{\rho'} V^C(q_l, \rho') \\ \geq \; & (1-\delta)(E_{q|\gamma}(q) + \rho) + \delta E_{\tilde{q}|\gamma} E_{\rho'} \max\{V^C(q_l, \rho'), V^C(\tilde{q}, \rho')\}, \end{aligned}$$

or

$$(1-\delta)\left(q_l - E_{q|\gamma}(q) + \bar{s}\right) \geq \delta\gamma\left[E_{\rho'} V^C(q_h, \rho') - E_{\rho'} V^C(q_l, \rho')\right].$$

Substitution and division by γ yields

$$(1-\delta)\left(q_l - q_h + \frac{\bar{s}}{\gamma}\right) \geq \delta \int_{u_0^\theta - q_h - \bar{s}}^{u_0^\theta - q_l - \bar{s}} [1 - F(\rho)] d\rho. \tag{A.48}$$

Since $\int_{u_0^\theta - q_l - \bar{s}}^{u_0^\theta - q_h - \bar{s}} [1 - F(\rho)] d\rho > 0$ and independent of δ, and $\gamma > 0$, there exists δ^θ such that for $\delta > \delta^\theta$ condition (A.48) cannot hold, where δ^θ is defined as the survival probability that solves (A.48) with equality. For $\bar{s} < \gamma(q_h - q_l)$, $\delta^\theta \leq 0$. If $\delta > \underline{\delta} \equiv \max\{0, \delta^w, \delta^p\}$, all consumers will search for high quality firms. This establishes lemma **??**. Note that for $\gamma \geq \lambda$ a bound $\underline{\delta}$ can be established independently of the exact value of γ by finding the fixed point of the equality in (A.48) when γ is replaced by λ. *Q.E.D.*

Proof of Lemma 9 : Consider the mapping $\tau : [\lambda, 1] \times [u_0^w - q_h - \bar{s}, u_0^p - q_l] \rightarrow [\lambda, 1] \times [u_0^w - q_h - \bar{s}, u_0^p - q_l]$ such that $\tau(\gamma, \hat{\rho}) = \begin{pmatrix} \tau_1(\gamma, \hat{\rho}_l) \\ \tau_2(\gamma, \hat{\rho}_l) \end{pmatrix}$. Similar to equation (3.10) let $\tau_1(\gamma, \hat{\rho}_l)$ be defined as

$$\tau_1(\gamma, \hat{\rho}_l^\theta) = 1 - \frac{\left[1 - \delta + \delta B^\theta\right][1 - \lambda]}{1 - \delta + \delta(1 - \lambda + \gamma)B^\theta + \left[\delta\lambda + \frac{\delta^2}{1-\delta}B^\theta\gamma\right]C^\theta}, \tag{A.49}$$

with $B^\theta \equiv 1 - F(\hat{\rho}_l^\theta)$ and $C^\theta \equiv 1 - F(u_0^\theta - q_h - s)$. When the wealthy follow other wealthy consumers, $\theta = w$. However, the analysis holds similarly for the poor following other poor, i.e., $\theta = p$. For $\gamma \in [\lambda, 1]$ the multiplier of $(1-\lambda)$ is strictly smaller than 1 and so $\tau_1(\gamma, \hat{\rho}_l^\theta) > \lambda$. Clearly $\tau_1(\gamma, \hat{\rho}_l^\theta) < 1$. Similar to equation (3.1), let $\tau_2(\gamma, \hat{\rho}_l^\theta) = \tau_2(\gamma)$ be implicitly defined by

$$\tau_2(\gamma) = u_0^\theta - E_{q|\gamma}(q) - \frac{\delta\gamma}{1-\delta}\int_{u_0^\theta - q_h - s_h(\theta)}^{\tau_2(\gamma)} [1 - F(\rho)]d\rho. \tag{A.50}$$

The function τ is continuous. For τ_1 this is easy to see. For τ_2, note that in (A.50) γ as a function of τ_2 is continuous and strictly monotone. Therefore $\tau_2(\gamma)$ is also continuous. Domain and codomain of τ are identical, and they are compact subsets of $\Re^2$. By Brouwer's fixed point theorem there exists a fixed point of τ. *Q.E.D.*

Proof of Proposition 22: Consider $\theta \in \{p, w\}$. For $\gamma \in (\lambda, 1)$ we have $\gamma^\theta > \lambda$ (see discussion in proof of lemma 9, where $\tau_1(\gamma, \hat{\rho}_l^\theta)$ corresponds to γ^θ). To compare γ^w and γ^p consider the general form of (3.10) with w replaced by θ, where $\theta \in \{p, w\}$. Some algebra reveals that $(\partial\gamma^\theta/\partial B^\theta) > 0$ iff $(\gamma - \lambda)\delta(1-\delta) + \delta^2(\gamma - \lambda)C^\theta > 0$, which holds since $\gamma \in (\lambda, 1)$. Clearly $(\partial\gamma^\theta/\partial C^\theta) > 0$. Therefore $\gamma^w > \gamma^p$ if $C^p < C^w$ and $B^p < B^w$, which is by lemma (8) the case for $s_h(w) \geq s_h(p)$ or $\bar{s} < u_0^p - u_0^w$. By the same lemma $s_h(p) - s_h(w) = \bar{s} > \xi_\gamma$ implies $C^p > C^w$ and $B^p > B^w$, which in turn implies $\gamma^w < \gamma^p$. In the intermediate case of $s_h(p) - s_h(w) = \bar{s} \in (u_0^p - u_0^w, \xi_\gamma)$ we have $C^p > C^w$ but $B^p < B^w$. If $s_h(p) - s_h(w) \approx u_0^p - u_0^w$, then $C^p \approx C^w$ but $B^p < B^w$ and therefore $\gamma^w > \gamma^p$. If $s_h(p) - s_h(w) \approx \xi_\gamma$, then $C^p > C^w$ but $B^p \approx B^w$ and therefore $\gamma^w < \gamma^p$. If $s_h(p) - s_h(w) = \bar{s} \in (u_0^p - u_0^w, \xi_\gamma)$ an increase in $\bar{s}$ increases C^p and B^p but leaves C^w and B^w unchanged, and there exists unique $\hat{\xi}_\gamma \in (u_0^p - u_0^w, \xi_\gamma)$ for which $s_h(p) - s_h(w) = \hat{\xi}_\gamma$ implies $\gamma^w < \gamma^p$. *Q.E.D.*

Proof of Proposition 23: To illustrate how Π is calculated, consider the following candidate stationary equilibrium: All consumers follow the wealthy, the wealthy are promised service by high quality and not by low quality firms, and no firm promises service to the poor. In this case, the benefit of a wealthy consumer to a high quality firm, denoted Π^{wh}, comprises the wealthy consumer's own contribution $P - c$, plus the life-time contributions of his followers.

The expected number N_w of wealthy followers in the next period is given by the number of consumers who are searching in that period divided by the number of all wealthy who are consuming, i.e., $N_w = \frac{\varphi_l^w}{\varphi_l^w + \varphi_h^w} = 1 - \gamma$. In subsequent periods they consume with probability $1 - F(u_0^w - q_h - \bar{s})$ conditional on surviving. They generate benefit Π^{wh} every time they visit. These followers do not get service on their first visit to the firm, and finally, there are $N_p = \frac{\varphi_l^p}{\varphi_l^w + \varphi_h^w} \frac{1-\alpha}{\alpha}$ poor consumers who follow in the next period. In every subsequent period they consume with probability $1 - F(u_0^p - q_h)$ if they survive. They generate benefit P each time they consume. Thus, the contribution of a wealthy consumer is given by

$$\begin{aligned}\Pi^{wh} = P - c \qquad &+ \beta N_w \Pi^{wh}\left[1 + \frac{\delta\beta}{1-\delta\beta}(1 - F(u_0^w - q_h - s))\right] + \beta N_w c \qquad \text{(A.51)}\\ &+ \beta N_p P \left[1 + \frac{\delta\beta}{1-\delta\beta}(1 - F(u_0^p - q_h))\right].\end{aligned}$$

The proof of the proposition is divided into three lemmata. The following lemma establishes that a leader's benefit to a firm can be arbitrarily high if he is followed by sufficiently many customers of the other type.

Lemma 10 *Fix $M > 0$. Assume type θ customers are being followed by consumers of the other type $\bar{\theta} \neq \theta$. Assume the type $\bar{\theta}$ consumers do not receive service. Then for any $\delta \in (0,1)$ there exists $\bar{\alpha} > 0$, such that for all $\alpha^\theta \in (0, \bar{\alpha})$ the benefit Π of a type θ customer to a firm is greater than M (independent of the service strategies toward type θ consumers).*

Proof: Since the type $\bar{\theta}$ followers do not receive service by assumption, they will search for high quality firms. The value of next-period type $\bar{\theta}$ followers to any firm due to a visit by a leader is at least

$$(1-\delta)\alpha^{\bar{\theta}}[1 - F(u_0^{\bar{\theta}} - q_l)]^2(1-\lambda)\delta\frac{1}{\alpha^\theta}P\beta, \qquad \text{(A.52)}$$

where $\alpha^w = \alpha$ and $\alpha^p = 1 - \alpha$. In every period there will be $(1-\delta)\alpha^{\bar{\theta}}$ newborn followers of type $\bar{\theta}$ who go into the market with probability greater than $[1 - F(u_0^{\bar{\theta}} - \underline{q})] > 0$, do not find a sufficiently good firm with probability $(1-\lambda)$, survive another period with probability δ, and consume again with probability of at least $[1 - F(u_0^{\bar{\theta}} - q_l)]$. This time they follow a leader who was in the market the previous period, of whom there are at most α^θ. They pay price P, and since they follow a period later than the visit of the leader, their value is discounted by β. The expression goes to infinity as α^θ going to zero.

The firm might incur service costs for the leader, but these are easily offset by his immediate type $\bar{\theta}$ followers. The leader might also have followers of his own type, which themselves bring a benefit larger than M in the period after and will therefore increase this consumers benefit even more. *Q.E.D.*

Recall that a high quality firm can induce a customer to return by promising service. The following lemma shows that a high quality firm will provide service when the customer's profit contribution is sufficiently large. Let $\Pi^{\theta h}$ denote the benefit of one-time consumption of a type θ consumer for a high quality firm.

Lemma 11 *There exists $M > 0$ such that in any stationary equilibrium with $\Pi^{\theta h} > M$, a high quality firm will promise service $\bar{s}$ in any period to type θ consumers.*

Proof: Let s be the equilibrium strategy of a high quality firm to a type θ customer that generates the benefit of $\Pi^{\theta h}$ for the firm. Let $\check{s} \neq s$ be a one-shot deviation in the service promise.

If promising $\check{s}$ instead of $s = \bar{s}$ results in the customer searching for another firm and never returning, then for $\Pi^{\theta h} > M = c$ offering the service is optimal, since by offering service the firm retains the business of this consumer and gains $\Pi^{\theta h} - c$ when he returns. As discussed above, our restrictive equilibrium concept necessitates the discussion of the case where the consumer would return even if service were not promised for one period.[24] For this case the proof is divided into two parts. The first establishes that increasing (decreasing) the service promise increases (decreases) the probability with which the consumer returns by a finite amount. The second provides the lower bound for the profitability of the consumer such that the threat of a potential time delay warrants service promises.

For the first part we discuss the consumer's reaction to a deviation. In equilibrium the customer returns whenever $\rho \geq \hat{\rho}_h^\theta = u_0^\theta - q_h - s$, otherwise he does not consume. Assume that the customer also chooses the firm rather than random sampling at $\check{s} \in \{0, \bar{s}\} \backslash \{s\}$. Let $V^C(q_h, \rho) = V^C(\rho)$ and V^F be the flow payoff of this strategy for the customer and the firm respectively. Consider the customer's response to a one-shot deviation by the firm. The value function $\check{V}^C(\rho)$ of the customer for the period directly after the deviation is

$$\check{V}^C(\rho) = \max\{(1-\delta)(q_h + \check{s} + \rho) + \delta E V^C(\rho), (1-\delta)u_0^\theta + \delta E \check{V}^C(\rho)\}. \tag{A.53}$$

[24]Since the equilibrium strategy of a firm is a function $s(\theta)$ independent of the history, after a one-shot deviation a consumer still expects to get the equilibrium service promise whenever he returns. This belief makes it hard to sustain an equilibrium at $s = \bar{s}$ and easy to sustain an equilibrium at $s = 0$, because the consumer expects the change for only a single period and reacts little (compared e.g. to the case where he expects the change to continue forever).

Let $\check{\rho}$ be the value for which the first term in the max operator is equal to the second term, i.e.,

$$(1-\delta)(q_h+\check{s}+\check{\rho})+\delta EV^C(\rho)=(1-\delta)u_0^\theta+\delta E\check{V}^C(\rho). \tag{A.54}$$

This implies that the customer will return to the firm when $\rho \geq \check{\rho}$, and will not consume otherwise. Then

$$E\check{V}^C(\rho)=\int_{\check{\rho}}^{\bar{\rho}}\left[(1-\delta)(q_h+\check{s}+\rho)+\delta EV^C(\rho)\right]f(\rho)d\rho+\int_{\underline{\rho}}^{\check{\rho}}\left[(1-\delta)u_0^\theta+\delta E\check{V}^C(\rho)\right]f(\rho)d\rho.$$

Therefore

$$\begin{aligned}(1-\delta F(\check{\rho}))E\check{V}^C(\rho) &= (1-F(\check{\rho}))\left[(1-\delta)(q_h+\check{s})+\delta EV^C(\rho)\right] \\ &\quad+\int_{\check{\rho}}^{\bar{\rho}}(1-\delta)\rho f(\rho)d\rho+F(\check{\rho})(1-\delta)u_0^\theta.\end{aligned} \tag{A.55}$$

Substituting (A.55) into the equation (A.54), integration by parts and rearranging yields:

$$(1-\delta)(\check{\rho}+q_h+\check{s})-\delta\int_{\check{\rho}}^{\bar{\rho}}(1-F(\rho))d\rho-u_0^\theta+\delta EV^C(\rho)=0.$$

The value of $EV^C(\rho)$ is given by lemma (7). Substitution leads to

$$(1-\delta)(\check{\rho}+q_h+\check{s}-u_0^\theta)+\delta\int_{u_0^\theta-q_h-s}^{\check{\rho}}(1-F(\rho))d\rho=0. \tag{A.56}$$

Equation (A.56) has a unique solution. It also reveals that for $s=0$ and $\check{s}=\bar{s}$ we have $\check{\rho}<u_0^\theta-q_h$, which implies that the frequency of consumption is increased by the deviation. Let ζ_s be the probability of returning each period under the equilibrium strategy s, and let $\check{\zeta}_s$ be the probability of returning next period after a one-shot deviation in the service promise. Then $\check{\zeta}_0-\zeta_0=(1-F(\check{\rho}))-(1-F(u_0^\theta-q_h))>0$. On the other hand for $s=\bar{s}$ and $\check{s}=0$ equation (A.56) reveals that $\check{\rho}<u_0^\theta-q_h-\bar{s}$, which implies that the deviation decreases the frequency of consumption. That is, $\check{\zeta}_{\bar{s}}-\zeta_{\bar{s}}\equiv(1-F(\check{\rho}))-(1-F(u_0^\theta-q_h-\bar{s}))<0$. Hence, a change in service provision changes the frequency of consumption by a finite amount, i.e. $\triangle\zeta\equiv\min\{\check{\zeta}_0-\zeta_0,|\check{\zeta}_{\bar{s}}-\zeta_{\bar{s}}|\}>0$.

For the second part we discuss the firm's incentive to deviate. We show that for Π large enough $s=0$ cannot be an equilibrium strategy since a one-shot deviation would be profitable. We also show that $s=\bar{s}$ is an equilibrium strategy.

Consider first the case where the candidate equilibrium strategy is $s=0$, the one-shot deviation is $\check{s}=\bar{s}$. In this case $(\check{\zeta}_0-\zeta_0)>0$. Note that the effective discount factor for the firm in this case is $\delta_F=\delta\beta$ because the firm discounts with β and the survival probability of the customer is δ. Normalizing profits by $(1-\delta_F)$, the equilibrium value to the firm is $V^F = \zeta_0 \Pi^{\theta h}$. The value to the firm from period $t+1$ onward after a one-shot deviation in period t is

$$\check{V}^F = \check{\zeta}_0((1-\delta_F)(\Pi^{\theta h} - c) + \delta_F V^F) + (1-\check{\zeta}_0)\delta_F \check{V}^F.$$

A one-shot deviation is *profitable* if $\check{V}^F > V^F$, or equivalently $\Pi^{\theta h} > \frac{\check{\zeta}_0}{\check{\zeta}_0 - \zeta_0}c$. This is fulfilled if $M \geq \frac{1}{\triangle\zeta}c$.

Consider now the case of $s = \bar{s}$ and $\check{s} = 0$. In this case $(\check{\zeta}_{\bar{s}} - \zeta_{\bar{s}}) < 0$. The equilibrium (flow) value to the firm is $V^F = \zeta_{\bar{s}} \Pi^{\theta h}$. The flow value to the firm from period $t+1$ onward after a one-shot deviation in period t is

$$\check{V}^F = \check{\zeta}_{\bar{s}}((1-\delta_F)(\Pi^{\theta h} + c) + \delta_F V^F) + (1-\check{\zeta}_{\bar{s}})\delta_F \check{V}^F.$$

A one-shot deviation is *not profitable* if $\check{V}^F \leq V^F$, or equivalently $\Pi^{\theta h} \geq \frac{\check{\zeta}_{\bar{s}}}{\zeta_{\bar{s}} - \check{\zeta}_{\bar{s}}}c$. This is fulfilled if $M \geq \frac{1}{\triangle\zeta}c$. Therefore for $\Pi^{\theta h} > M \geq \frac{1}{\triangle\zeta}c$ the only equilibrium strategies for high quality firms is $s = \bar{s}$. *Q.E.D.*

Finally we show that high quality firms will always outbid low quality firms:

Lemma 12 *Let $\delta \in (\underline{\delta}, 1)$. In any stationary equilibrium, either $s_h(\theta) \geq s_l(\theta)$ for $\theta \in \{p, w\}$, or $s_h(\theta) < s_l(\theta)$ but type θ consumers nevertheless do not return to low quality firms.*

Proof: Assume $s_h(\theta) < s_l(\theta)$ and type θ customers stop searching when they have found a low quality firm. There are two possibilities: Either they stop searching at the first firm they encounter, in which case $\gamma^\theta = \lambda$. Or they do not return to high quality firms but keep searching for a low quality firm. In this case $\gamma^\theta < \lambda$.

Call type θ consumers group Y, and type $\bar{\theta}$ consumers group Z. Group Y consumers must have some consumers that follow them or service would not be profitable. It cannot be that every consumer who is searching follows group Y, because group Z would then not receive service as it has no followers, and would therefore look for high quality firms. If $\gamma^\theta = \lambda$, by Proposition 22 $\gamma^{\bar{\theta}} > \lambda$ and group Z consumers should follow members of their own group. If $\gamma^\theta < \lambda$, group Z consumers are better off sampling on their own. Therefore either group Y members are only followed by group Y members, or they are only followed by group Z members.

In the second case, both groups would have to continue searching for low qualities (plus service) after finding a high quality firm. Assume group Y did not; then they would stay at the first firm they patronize. But then group Z has no followers, thus receives no service and will look for high quality firms. But then it is not optimal for group Z to follow group Y. Assume group Z did not leave high quality firms; then they either stay at the first firm they encounter and do not follow group Y, or they only look for high quality firms, in which case following group Y is suboptimal. Therefore it must be that group Z searches for low quality firms, which implies that group Z also receives service from low quality firms. To receive service, it must be that they are followed by group Y. So both groups receive service form low quality firms and not from high quality firms. In this case, if the poor are leaving high quality firms to search for low quality plus service, then the wealthy strictly prefer to leave high quality firms to receive low quality plus service (as their higher frequency of consumption is similar to a lower discount factor). Yet by an argument similar to proposition 22, the signal of the wealthy is more informative about finding low quality firms when both types search for them and get identical service. Therefore the wealthy would not follow the poor, and this case cannot constitute an equilibrium.

We are therefore left with the case in which each group Y member is, in equilibrium, followed by some expected number N_θ of other members of its own group (and none of the other group). The candidate equilibrium profit contribution $\Pi^{\theta l}$ that a low quality firm receives from a group Y customer returning one more time is

$$\Pi^{\theta l} = P - c + \beta N_\theta \Pi^{\theta l} \left[1 + \frac{\delta\beta}{1-\delta\beta}(1 - F(u_0^\theta - q_l - \bar{s}))\right] + \beta N_\theta c. \tag{A.57}$$

The derivation is similar to that of equation (A.51). In the stationary setting

$$\beta N_\theta \left[1 + \frac{\delta\beta}{1-\delta\beta}(1 - F(u_0^\theta - q_l - \bar{s}))\right] < 1.$$

Solving for $\Pi^{\theta l}$ yields:

$$\Pi^{\theta l} = \frac{P - c + \beta N_\theta c}{1 - \beta N_\theta [1 + \frac{\delta\beta}{1-\delta\beta}[1 - F(u_0^\theta - q_l - \bar{s})]]}. \tag{A.58}$$

For this to be an equilibrium, $\Pi^{\theta l} \geq 0$. Consider first the case where $\Pi^{\theta l} > 0$, i.e. $P - c + \beta N_\theta c > 0$. In this case high quality firms have an incentive to deviate and also offer service to group Y consumers, which upsets the equilibrium. To see this, note that for a high quality firm the candidate equilibrium profit contribution from a group Y consumer is zero after the first period of consumption, because he does not consume there again. Deviating and offering service to the customer and all his followers generates the profit contribution

$$\Pi' = P - c + \beta N_\theta \Pi' \left[1 + \frac{\delta\beta}{1-\delta\beta}(1 - F')\right] + \beta N_\theta c,$$

where $1 - F'$ is the probability with which a customer that is offered service is returning. Since $\bar{\rho} > u_0^\theta - q_h - \bar{s}$, the frequency $1 - F' > 0$. Since $P - c + \delta N_\theta c > 0$, it follows that $\Pi' > 0$.[25] But then high quality firms would offer service.

Consider now the case $\Pi^{\theta l} = 0$, i.e. $P - c + \beta N_\theta c = 0$. Therefore, low quality firms are indifferent between promising service or not. In this case high quality firms are also indifferent between offering service or not. By the tie-breaking rule we employed, both types of firms offer service.[26] However, consumers then do not search for low quality firms; consequently high quality firms offer service, and group Y customers would not search for low qualities. *Q.E.D.*

---

[25] Since this is a deviation from a steady state, $\beta N_r \Pi' \left[1 + \frac{\delta\beta}{1-\delta\beta}(1 - F')\right]$ might be larger than 1, in which case the discounted profit from offering service is unbounded.

[26] This is the only place where we use this tie-breaking rule. The result holds also when we employ the assumption that firms do not offer service when indifferent. The point is that both types of firms resolve indifference the same way. Moreover, simple restrictions such as a high survival rate δ, a high cost-price wedge $c - P$ or a modest service influence $\bar{s}$ would also guarantee the result, as they rule out indifference.

Bibliography

[1] Abowd, John M., Francis Kramarz, and David N. Margolis (1999): "High Wage Workers and High Wage Firms." *Econometrica*, 67, 251-333.

[2] Acemoğlu, Daron, and Robert Shimer (1999): "Efficient Unemployment Insurance." *Journal of Political Economy*, 107, 893-928.

[3] Acemoğlu, Daron, and Robert Shimer. "Wage and Technology Disperson," *Review of Economic Studies*, 67 (2000), 585-607.

[4] Albrecht, James, and Bo Axell (1984): "An Equilibrium Model of Search Unemployment." *Journal of Political Economy*, 92, 824-40.

[5] Albrecht, James, Pieter A. Gautier, Serene Tan, and Susan Vroman (2005): "Matching with Multiple Applications Revisited," *Economic Letters*, 84, 311-314.

[6] Albrecht James, Pieter A. Gautier, and Susan Vroman (2005): "Equilibrium Directed Search with Multiple Applications," forthcoming in *Review of Economic Studies.*

[7] Banerjee, Abhijit (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797-817.

[8] Banerjee, Abhijit and Drew Fudenberg (2004): "Word-of-Mouth Learning," *Games and Economic Behavior*, 46, 282-303.

[9] Bearden, William O. and Michael J. Etzel (1982) "Reference Group Influence on Product and Brand Purchase Decisions," *Journal of Consumer Research*, 9, 183-194.

[10] Bearden, William O., R. G. Netemayer and J. E. Teel (1982): "Measurement of Consumer Susceptibility to Interpersonal Influence," *Journal of Consumer Research*, 15, 473-481.

[11] van den Berg, Gerard J., and Geert Ridder (1998): "An Empirical Equilibrium Search Model of the Labor Market." *Econometrica*, 66, 1183-1221.

[12] Bikhchandani, Sushil, David Hirshleifer and Ivo Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992-1026.

[13] Blackwell, David (1965): "Discounted Dynamic Programming," *Annals of Mathematical Statistics*, 75, 352-365.

[14] Bose, S., Orosel, G. O., Ottaviani, and L. Vesterlund (2005) "Monopoly Pricing in the Binary Herding Model," mimeo.

[15] Bose, S., Orosel, G. O., Ottaviani, and L. Vesterlund (2005): "Dynamic Monopoly Pricing and Herding," mimeo.

[16] Bulow, Jeremy and Jonathan Levin (forthcoming): "Matching and Price Competition," *American Economic Review*.

[17] Burdett, Kenneth and Kenneth L. Judd (1983): "Equilibrium Price Distributions," *Econometrica*, 51, 955-970.

[18] Burdett, Kenneth and Dale T. Mortensen (1998): "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39, 257-273.

[19] Burdett, Kenneth, Shouyong Shi and Randall Wright (2001): "Pricing and Matching with Frictions," *Journal of Political Economy*, 109, 1060-1085.

[20] Burnkrant, Robert E. and Alain Cousineau (1975): "Informational and Normative Social Influence in Buyer Behavior," *Journal of Consumer Research*, 2, 206-215.

[21] Butters, Gerad R. (1977): "Equilibrium Distributions of Sales and Advertising Prices," *Review of Economic Studies*, 44, 465-491.

[22] Cason, Timothy and Charles Noussair (2003): "A Market with Frictions in the Matching Proces: An Experimental Study," mimeo.

[23] Chade, Hector, and Lones Smith (2004): "Simultaneous Search," mimeo.

[24] Champley, Christophe P. (2004): *Rational Herds - Economic Models of Social Learning*, Cambridge: Cambridge University Press.

[25] Cohen, Joel B. and Ellen Golden (1972): "Informational Social Influence and Product Evaluation," *Journal of Applied Psychology*, 56, 54-59.

[26] Cole, Harold L., George J. Mailath and Andrew Postlewaite (1992): "Social Norms, Savings Behavior, and Growth," *Journal of Political Economy*, 100, 1092-1125.

[27] Cole, Harold L., George J. Mailath and Andrew Postlewaite (1995): "Incorporating Status Concern in Economic Models," *Quarterly Review of the Minneapolis Federal Reserve Bank*, Fall 1995.

[28] DeGroot, Morris H. (1970): *Optimal Statistical Decisions*, New York: McGraw-Hill.

[29] Delacroix, Alain and Shouyong Shi (2004): "Directed Search On the Job and the Wage Ladder," *International Economic Review*, 45, 579-612.

[30] Diamond, Peter A. (1971): "A Model of Price Adjustments," *Journal of Economic Theory*, 156-168.

[31] Diamond, Peter A. (1982): "Wage Determination and Efficiency in Search Equilibrium," *Review of Economic Studies*, 49, 217-227.

[32] Ellison, Glenn and Drew Fudenberg (1995): "Word-of-Mouth Communication and Social Learning," *Quarterly Journal of Economics*, 110, 93-125.

[33] Fang, Hanming (2001): "Social Culture and Economic Performance," *American Economic Review*, 91, 924-937.

[34] Friedman, Milton (1968): "The Role of Monetary Policy," *American Economic Review*, 58, 1-17.

[35] Gale, David and Lloyd Shapley (1962): "College admissions and the stability of marriage," *American Mathematical Monthly*, 69, 9-15.

[36] Gale, David (1987): "Limit Theorems for Markets with Sequential Bargaining," *Journal of Economic Theory*, 40, 20-54.

[37] Galenianos, Manolis and Philipp Kircher (2005): "Directed Search with Multiple Job Applications," mimeo.

[38] Gautier, Pieter A., and José L. Moraga-González (2004): "Strategic Wage Setting and Coordination Frictions with Multiple Applications," mimeo.

[39] Holzer, Harry J., Lawrence F. Katz, and Alan B. Krueger (1991): "Job Queues and Wages." *Quarterly Journal of Economics* 106, 739-768.

[40] Hosios, Arthur J. (1990): "Competitive Search Equilibrium," *Review of Economic Studies*, 57, 279-298.

[41] Immorlica, Nicole and Mohammad Mahdian (2005): "Marriage, Honesty, and Stability," *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms.*

[42] Julien, Benoit, John Kennes, and Ian King (2000): "Bidding for Labor." *Review of Economic Dynamics*, 3, 619-629.

[43] Lauermann, Stephan (2006): "Efficiency of Decentralized Trading," mimeo.

[44] Mailath, George J. and Andrew Postlewaite (2004): "Social Assets," forthcoming in *International Economic Review.*

[45] McCall, John J. (1965): "The Economics of Information and Optimal Stopping Rules," *Journal of Business*, 38, 300-317.

[46] Moen, Espen R. (1997): "Competitive Search Equilibrium," *Journal of Political Economy*, 105, 385-411.

[47] Montgomery, James D. (1991): "Equilibrium Wage Dispersion and Interindustry Wage Differentials," *Quaterly Journal of Economics*, 106, 163-179.

[48] Mortensen, Dale T. (1982a): "The Matching Process as a Noncooperative Bargaining Game," in *The Economics of Information and Uncertainty*, ed. John J. McCall, Chicago: University of Chicago Press.

[49] Mortensen, Dale T. (1982b): "Property Rights and Efficiency in Mating, Racing, and Related Games," *Amercian Economic Review*, 72, 968-979.

[50] Mortensen, Dale T. and Randall Wright (2002): "Competitive Pricing and Efficiency in Search Equilibrium," *International Economic Review*, 43, 1-20.

[51] Mortensen, Dale T. (2003): *Wage Dispersion: Why Are Similar Workers Paid Differently?* Cambridge: Zeuthen Lecture Book Series.

[52] Nelson, Phillip (1970): "Information and Consumer Behavior," *Journal of Political Economy*, 78, 311-329.

[53] Ochs, Jack (1990): "The Coordination Problem in decentralized Markets: An Experiment," *Quarterly Journal of Economics*, 105, 545-559.

[54] Peters, Michael (1991): "Ex Ante Price Offers in Matching Games: Non-Steady States," *Econometrica*, 59, 1425-1454.

[55] Peters, Michael (1997): "On the Equivalence of Walrasian and Non-Walrasian Equilibria in Contract Markets: The Case of Complete Contracts," *Review of Economic Studies*, 64, 241-264.

[56] Peters, Michael (2000): "Limits of Exact Equilibria for Capacity Constrained Sellers with Costly Search," *Journal of Economic Theory*, 95, 139-168.

[57] Phelps, Edmund S. (1967): "Phillips Curves, Expectations of Inflation and Optimal Unemployment," *Economica*, 34, 254-281.

[58] Phillips, Alban W. (1958): "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom," *Economica*, 25, 283-299.

[59] Pincus, Steven and L.K. Waters (1977): "Informational Social Influence and Product Quality Judgements," *Journal of Applied Psychology*, 62, 615-619.

[60] Pissarides, Christopher A. (1984): "Search Intensity, Job Advertising, and Efficiency," *Journal of Labor Economics*, 2, 128-143.

[61] Pissarides, Christopher A. (1985): "Short-run Equilibrium Dynamics of Unemployment, Vacancies and Real Wages," *American Economic Review*, 75, 676-690.

[62] Pissarides, Christopher A. (2000): *Equilibrium Unemployment Theory*, 2nd ed., Cambridge, MA: MIT Press.

[63] Postel-Vinay, Fabien, and Jean-Marc Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica*, 70, 2295-2350.

[64] Reinganum, Jennifer F. (1979): "A Simple Model of Equilibrium Price Dispersion," *Journal of Political Economy*, 87, 851-858.

[65] Rogers, Everett M. (1995) *Diffusion of Innovation.* 4th ed., Free Press, New York.

[66] Rogerson, Richard, Robert Shimer and Randall Wright (2005): "Search Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature*, 43, 959-988.

[67] Roth, Alvin E. (1984): "The evolution of the labor market for medical interns and residents: A case study in game theory," *Journal of Political Economy*, 92, 991-1016.

[68] Roth, Alvin E. and Elliot Peranson (1999): "The redesign of the matching market for american physicians: Some engineering aspects of economic design," *American Economic Review*, 89, 748-780.

[69] Rothschild, Michael (1974): "Searching for the Lowest Price When the Distribution of Prices is Unknown," Journal of Political Economy, 82, 689-711.

[70] Samuelson, Larry (2004): "Information-Based Relative Consumption Effects," *Econometrica*, 72, 93-118.

[71] Shi, Shouyong (2002): "A Directed Search Model of Inequality with Heterogeneous Skills and Skill-Biased Technology," *Review of Economic Studies*, 69, 467-491.

[72] Shimer, Robert (1996): "Essays in Search Theory," PhD dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.

[73] Shimer, Robert (2005): "The Assignment of Workers to jobs in an Economy with Coordination Frictions," *Journal of Political Economy*, 113/5, 996-1025.

[74] Smallwood, Dennis and John Conlisk (1979): "Product Quality in Markets Where Consumers are Imperfectly Informed," *Quarterly Journal of Economics*, 93, 1-23.

[75] Stigler, George J. (1961): "The Economics of Information," *Journal of Political Economy*, 69, 213-225.

[76] Stigler, George J. (1962): "Information in the Labor Market," *Journal of Political Economy*, 70, 94-105.

[77] Thomson, William (1999): "The Young Person's Guide to Writing Economic Theory," *Journal of Economic Literature*, 37, 157-183.

[78] Weitzman, Martin L. (1979): "Optimal Search for the Best Alternative," *Econometrica*, 47, 641-54.

[79] Wolinsky, Asher (1990): “Information Revelation in a Market with Pairwise Meetings,” *Econometrica*, 58, 1-25.