Learning in the Labor Market

by

Jin Li

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2007

© Jin Li, MMVII. All rights reserved.

The author hereby grants to MIT permission to reproduce and
distribute publicly paper and electronic copies of this thesis document
in whole or in part.

Author ..............................................................

Department of Economics

May 15, 2007

Certified by ..............................................................
Robert Gibbons
Sloan Distinguished Professor of Organizational Economics and
Strategy

Thesis Supervisor

Certified by ..............................................................
Bengt Holmstrom
Paul A. Samuelson Professor of Economics

Thesis Supervisor

Accepted by ..............................................................
Peter Temin, Elisha Gray II Professor of Economics
Chairman, Department Committee on Graduate Students
Abstract

This thesis is a collection of three independent essays that study the implication of learning on labor mobility, labor supply, wage distribution, wage dynamics, and allocations of workers under different assumptions about the nature of employer learning.

The first essay develops a model of job mobility and wage dispersion under the assumption that the current employers have superior information about their workers over outside firms. The superior information of the workers does not lead to market collapse. Instead, there is a unique mixed strategy equilibrium which leads to a positive amount of turnover and a nondegenerate wage distribution. This model implies that a skill-biased technology change that also favors general skill can lead to increase both in job mobility and wage dispersion. This sheds light on the joint evolution of job mobility and wage dispersion in the U.S. in the past 30 years.

The second essay studies the wage distribution and wage dynamics under matching and symmetric Pareto learning. I develop a model that contains pure learning and pure matching as limiting cases. In addition, the model generates effects that arise from the interaction of learning and matching. In particular, the model generates an earning profile typically obtained in a Mincerian regression. Moreover, the model predicts that the wage residuals are more likely to be serially correlated in younger workers in industries with increasingly convex wage schedules. This helps reconcile the conflicting findings that positive correlations are found in small, homogenous samples but not large, heterogeneous samples.

The third essay, jointly with Peter Schnabl, develops a model that examines the optimal solution to the problem of assigning workers into jobs under adverse selections. Workers differ by their disutility of effort. Jobs differ by their productivity and ease of effort-monitoring. Firms would like to assign hard workers to higher level jobs because efforts on these jobs are harder to monitor. To prevent the lazy workers from mimicking the hard workers, we study the use of two instruments firms may use: requiring long hours and distorting job assignments. The model has an essentially unique separating equilibrium. In equilibrium, workers are required to exert inefficiently high levels of effort and firms commit to promote only a fraction of qualified workers.
Thesis Supervisor: Robert Gibbons
Title: Sloan Distinguished Professor of Organizational Economics and Strategy

Thesis Supervisor: Bengt Holmstrom
Title: Paul A. Samuelson Professor of Economics
Acknowledgments

I have been incredibly lucky to have worked under the guidance and supports of Robert Gibbons, Bengt Holmstrom, and Daron Acemoglu. I cannot possibly hope for a better trio as my mentors during my PhD study. Bob is in every dimension a perfect advisor. He is extremely knowledgeable about the subject, always encouraging with his comments, and exceptionally generous with his time. Bengt mentored me from my first day in MIT. His enormous intellectual power and his vision of economics left indelible imprints on the way I think of the subject. Daron sets a perfect example as a teacher and as a scholar. His passion and dedication to the subject made my each interaction with him an inspiring and humbling experience.

I would like to thank David Autor, Josh Angrist, Glen Ellison, Sergei Izmalkov, Panle Jia, Steve Pishke, Eric Van den Steen, Jean Tirole, and Muhamet Yildiz. My thesis has benefited tremendously from their comments, discussion, and insights.

One of the most remarkable aspects of my life at MIT is the amazing group of friends I have made. I am very fortunate to have met Peter Schnabl. Peter has provided me with intellectual stimulus constantly and he also coauthored the third essay in this thesis. I hope that our collaboration can continue in the future. I am very grateful to my officemates of three years, Caterina Reis and Jose Tessada, who have helped foster a perfect working environment and have given me helps on numerous instances. I hope the bonds we have formed during the past several years can remain strong for the rest of our lives. I especially thank Ding Wu, for constant supports. Our friendship has made my PhD study a much more pleasant experience.

I would also like to thank Nathan Barczi, Miriam Bruhn, Sylvain Chassang, Jessica Cohen, Alex Debs, Florian Ederer, Ruimin He, Peter Hinrichs, Rongzhu Ke, Jeanne Lafortune, Jackie Chou Lem, Konrad Menzel, Vasia Panousi, Marek Pycia, Casey Rothschild, Nancy Qian, Heikki Rantakari, Hui Shan, Henry Tang, and Pierre Yared, and Tim Watts. They have made learning and researching about economics extremely fun.

Finally, I would like to thank my family. My parents, Zhihong Li and Meiying
Zhang, taught me the value of education and instill in me the love for knowledge. Without their sacrifice and supports, I cannot imagine the completion of this thesis. My wife, Jienan Han, has been an ideal companion both in my personal and professional endeavor. Her encouragements, insights, patience, and boundless love and faith in me have made this thesis possible. I dedicate this thesis to them.
Contents

1 Job Mobility, Wage Dispersion, and Asymmetric Information 9
  1.1 Introduction ................................................................. 10
  1.2 Model Setup ................................................................. 17
    1.2.1 Worker and Firms ..................................................... 17
    1.2.2 Timing and Information Structure .................................. 19
    1.2.3 Perfect Bayesian Equilibrium ....................................... 21
  1.3 Equilibrium of the Model ................................................ 22
    1.3.1 Existence of a Mixed Strategy PBE ................................ 23
    1.3.2 Uniqueness .............................................................. 31
  1.4 Job Mobility and Wage Dispersion ..................................... 46
    1.4.1 Turnover Probability ................................................. 47
    1.4.2 Wage Distribution ..................................................... 51
  1.5 Inequality and Job Mobility ............................................. 55
    1.5.1 Background Facts and Theories ..................................... 56
    1.5.2 Implications for Inequality and Job-to-Job Mobility ........... 59
  1.6 Extension to Multiple Sectors ......................................... 68
    1.6.1 Multi-Sector Model: Setup and General Results ................. 68
    1.6.2 A Two-Sector Model ................................................... 73
  1.7 Conclusion and Discussion ............................................. 81

2 A Theory of Wage Distribution and Dynamics with Matching and Pareto Learning 89
  2.1 Introduction ................................................................. 90
2.2 Literature Review .................................................. 95
2.3 Model Setup .......................................................... 98
  2.3.1 Workers and Firms ................................................. 98
  2.3.2 Information Structure and Timeline ......................... 100
2.4 Analysis of the Model ............................................... 101
  2.4.1 CEEA and Pareto Learning .............................. 102
  2.4.2 Analysis of the Equilibrium .................................. 105
  2.4.3 Firm-Worker Match and the Wage Formula ................ 109
2.5 Wage Distribution .................................................. 113
2.6 Wage Dynamics ..................................................... 116
  2.6.1 Results Requiring Both Learning and Matching .......... 117
  2.6.2 Results Requiring Learning Only ............................ 124
2.7 Conclusion ............................................................ 129
2.8 Appendix .............................................................. 131

3 Promotion, Turnover, and Adverse Selection 141
  3.1 Introduction .......................................................... 142
  3.2 Model Setup .......................................................... 146
    3.2.1 Workers and Firms ............................................. 146
    3.2.2 Information Structure and Timeline ....................... 148
    3.2.3 Separating Equilibrium .................................... 151
  3.3 Analysis of the Model .............................................. 153
  3.4 Application of the Model ......................................... 169
  3.5 Discussion and Conclusion ....................................... 173
Chapter 1

Job Mobility, Wage Dispersion, and Asymmetric Information

This paper develops a model of job mobility and wage dispersion. Worker ability affects both firm-specific and general productivity. Employers learn more about the abilities of their workers than do outside firms. The superior information of current employers creates a standard lemons problem in the second-hand labor market. Contrary to existing work, the lemons problem does not lead to market collapse in the model presented here. Instead, there exists a unique equilibrium outcome in which the current employer offers a wage equal to the average output of all types below the ability of the worker, and outside firms compete for workers by using mixed strategies. These mixed strategies lead to a non-degenerate wage distribution for all types of workers. This unique equilibrium outcome determines both the allocation of workers with heterogeneous abilities to different firms and also how wages change when workers change jobs.

In addition to providing a new model of turnover, selection, and wage dispersion, the paper sheds light on the joint evolution of wage inequality and job mobility in the United States over the past 30 years. The model implies that, in the presence of technological change that is skill-biased and also favors general skills over firm-specific skills, the wage distribution will become more spread out (corresponding to greater inequality) and both overall and sectoral job mobility will increase. The model also
suggests that the increase in job mobility should be larger for older workers. These patterns are consistent with recent empirical evidence on changes in job mobility in the United States.

1.1 Introduction

The labor market for young workers has two salient features. First, information about the abilities of workers is imperfect and new information arrives frequently. Second, match qualities between workers and firms are uncertain and mismatches can occur. These two features generate two well-documented empirical regularities about this labor market. First, young workers frequently change jobs: a typical U.S. worker holds 7 jobs, or two-thirds of his career total, in the first ten years of his labor market experience; see for example Topel and Ward (1992). Second, the wage distribution of young workers widens over time, both unconditionally and conditionally (on the observables); see for example Farber and Gibbons (1996).

Many models have used imperfect information and uncertain match qualities as the key ingredients to explain job mobility. For example, one important model of job mobility where information about the worker is learned by the market over time is Jovanovic (1979). In this model, the output of a given worker depends on the uncertain match quality between that worker and a given employer, and match quality is learned publicly over time. In equilibrium, a worker switches to a new job only when his expected match quality with the current employer falls below a duration-dependent threshold. One important prediction of the model, that turnover (eventually) decreases with tenure, has received strong empirical support; see for example Farber (2003) for a survey.

In Jovanovic (1979), the public-learning assumption is convenient, and it seems to describe the labor market of high school graduates in the U.S. fairly well; see for example Schoenberg (2005). In many other labor markets, however, it appears natural to assume that the current employer learns more about the worker's ability than
prospective employers do. Models of such markets involve asymmetric information and have received significant empirical support; see for example the evidence in Gibbons and Katz (1991), Acemoglu and Pischke (1998), Schoenberg (2005) and Von Wachter and Bender (forthcoming).

As is well-known, job-mobility models with asymmetric information can create a winner's curse, where an outside firm is able to hire a worker only when its wage offer is higher than the worker's output at the current firm; see for example the models in Greenwald (1986), Gibbons and Katz (1991), and Acemoglu and Pischke (1998, 1999a,b). When the worker's output is common across all firms (i.e., there is no firm-specific human capital or firm-specific matching), this lemons problem may lead to market collapse: there is no turnover in equilibrium. To produce plausible turnover rates, asymmetric-information models typically assume that non-trivial numbers of workers leave for exogenous reasons.

The existing models of job mobility with either symmetric or asymmetric information rarely discuss their implications for wage dispersion, even if the underlying forces in these models (namely imperfect information and matching) should affect dispersion. Perhaps this is because these models are both too complicated and not flexible enough to describe wage distributions. For example, the Jovanovic model relies heavily on the assumptions that match quality and output are normally distributed. These assumptions both restrict the possible equilibrium wage distributions and make the model hard to extend (for example, to multiple sectors).

In addition, models with asymmetric information typically produce a very compressed wage distribution: in Greenwald (1986), for example, all workers have the same wage in equilibrium; in Gibbons and Katz (1991), all movers have the same wage and all stayers have the same wage: and in Acemoglu and Pischke (1998, 1999a,b), all movers have the same wage and the wage distribution of the stayers is determined by the (exogenous) bargaining power of the workers. In contrast to the predictions of these models, much evidence shows that similar workers can be paid very differently in real life.
Explaining why similar workers are paid differently has generated a huge literature, which is masterfully summarized by Mortensen (2001). Two important models are Burdett and Judd (1983) and Burdett and Mortensen (1998). These models show that identical workers can be paid differently if there are search frictions that prevent firms from making offers to all workers. In equilibrium, firms randomize their wage offers because they are indifferent along a schedule that trades off the probability of hiring a worker against the profit made from the worker. This randomization generates wage dispersion for identical workers. In addition to wage dispersion, these two models also generate job mobility. However, the turnover rates in these models depend crucially on the frequency of job offer arrival, which is exogenous.

In short, the literature offers important but separate insights into job mobility and wage dispersion, and yet one might expect these two phenomena to be closely linked because both are related to uncertain worker abilities and imperfect worker-employer matches. In this paper, therefore, we develop a single model that addresses both job mobility and wage dispersion.

Our model applies to asymmetric-information settings where the current employer learns more about a worker’s ability than prospective employers do. In our model, however, the presence of a winner’s curse does not lead to market collapse. Instead, there exists a unique equilibrium outcome in which the current employer offers a wage equal to the average outside output of all types below the worker’s ability, and outside firms compete for the worker by using mixed strategies. These mixed strategies lead to a non-degenerate wage distribution for all types of workers. The unique equilibrium outcome determines both the allocation of workers with heterogeneous abilities to different firms and how wages change when workers change jobs.

Our model is similar to Greenwald’s (1986) classic asymmetric-information model, but with one important difference (described below). There are two periods, a single worker, and finitely many firms. At the beginning of period 1, firms compete for the worker using wage competition a la Bertrand. The worker picks a period 1 employer. At the end of period 1, the incumbent (period 1 employer) learns about the worker’s
ability while outside firms do not. Moreover, the worker may accumulate firm-specific human capital, which makes him more productive at the incumbent in period 2. Fully aware of the informational and production advantage of the incumbent, all firms offer wage contracts simultaneously to the worker at the beginning of period 2.

The key difference between our model and Greenwald's involves the timing of offers at the beginning of period 2. Greenwald assumes (as do many subsequent models) that the incumbent knows all the offers made by the outside firms before making a counteroffer. This offer timing makes the lemons problem in the second-hand labor market so severe that market collapse when there are no exogenous movers or differences in match qualities. In our model, in contrast, we assume that the incumbent and outside firms make their offers simultaneously. This difference in timing alleviates the lemons problem and produces endogenous turnover and a non-degenerate wage dispersion.

Our assumption about the timing of offers is fairly standard and has been used in the literature; see for example Waldman (1984). We believe our assumption is reasonable because employees typically cannot credibly communicate the total value of outside offers to the incumbent: even if the exact monetary values of the outside offers are known to the incumbent, it is unlikely that the worker's non-monetary preferences can be known exactly. In other words, this difference, while stated formally in terms of the model's timing, should actually be interpreted in terms of the information structure. For example, we obtain identical results even if the incumbent makes its offer after the outside firms do, as long as the incumbent does not know the outside offers.

This simultaneity of wage offers makes the structure of the basic model identical to the auction model analyzed by Engelbrecht-Wiggans, Milgrom, and Weber (EMW) (1983). EMW consider a sealed-bid first-price auction with one informed bidder and a finite number of uninformed bidders. In equilibrium, the informed bidder bids the conditional expected value below the object's value, and the uninformed bidders randomize. This basic model can be considered an application of EMW (1983) to
the labor market: the incumbent firm corresponds to the informed bidder and offers its worker the average outside output of all types below the worker's ability, and the outside firms correspond to the uninformed bidders and randomize their wage offers.

In the unique equilibrium outcome of this model, the support of randomized outside offers is the closed interval between the minimal incumbent offer and the maximal incumbent offer. Consequently, some workers will receive outside offers higher than the incumbent's offer. This creates endogenous turnovers, without requiring exogenous movers or differences in match qualities. Moreover, the randomization of outside offers leads to wage dispersion for identical workers, and thus providing an alternative explanations for why similar workers are paid differently, a subject that has received a lot of interests; see for example Burdett and Judd (1983) and Burdett and Mortensen (1998).

The equilibrium in this model should be compared to that in Burdett and Judd (1983) and Burdett and Mortensen (1998): both involve randomization, but they differ in important respects. In the Burdett-Judd-Mortensen models, workers are identical and the number of wage offers they receive follows from an exogenous Poisson process. Firms randomize to trade off the probability of hiring a workers against the profit made from the worker. In equilibrium, all firms make positive expected profit, which converges to zero as the job arrival rate goes to infinity. In this model, in contrast, workers are heterogeneous and receive offers from all firms. Firms randomize to trade off the average quality of the workers hired against the wage paid to the workers, taking into account the productivity and selection effects that arise from worker heterogeneity and asymmetric information. In equilibrium, all outside firms make zero expected profit.

The zero expected profit of outside firms in this model leads to explicit formulas for turnover probabilities. The turnover probabilities are determined by a ratio of the firm-specific output to the difference between the marginal and the average general output of worker. The firm-specific human capital of the worker reflects the incumbent's production advantage; the difference between the marginal and the average
general output of the worker reflects the incumbent’s *information* advantage because the incumbent’s wage offer equals the worker’s average general output. Therefore, we can interpret this ratio as a comparison of the incumbent’s profit from the worker’s firm-specific output to its profit from the worker’s general output. The turnover probability of workers becomes uniformly larger if this ratio becomes smaller for all ability levels. One instance that this can happen is when there is a technological change that is both *log-skill-biased* and *general-skill-biased*.

We also derive explicit formulas for the wage distributions of the movers, of the stayers, and of the two types combined. These formulas enable us to compare the wage distribution of the movers with that of the stayers. It is unclear ex ante which group has a higher average wage because the stayers are of higher average ability, while the movers are luckier in receiving higher outside wage offers. It turns out that the profit ratio plays a key role in the comparison. In particular, if the ratio is increasing (decreasing) in the worker’s ability, the wage distribution of the stayers first order stochastic dominates (dominated by) that of the movers, so the the average wage of the stayers is higher (lower). If the firm-specific output is more sensitive to ability in larger firms, this suggests that the tenure effect is more likely to be observed in larger firms. We also show that when the firm-specific human capital is absent, the better abilities of stayers exactly cancel out with better luck of movers. In this case, the wage distributions of the movers and stayers are identical and a worker of ability \( q \) has a turnover probability of \( 1 - q \).

The explicit formulas for turnover probabilities and the wage distributions enable us to apply the model to study the joint evolution of within-group inequality and job mobility in the U.S. in the past 30 years. Recent increases in wage inequality in the U.S. have attracted attention from both the popular press and academic researchers. The popular press also suggests that job mobility has increased, although evidence from economic research is less clear cut. Nevertheless, it appears that job-to-job mobility has increased significantly; see for example Stewart (2002).

Many hypotheses for the cause of increases in wage inequality involve changes in
technology. In Section 5, we take two such hypotheses seriously. First, we suppose that the technological change has been log-skill-biased, so it favors workers of higher abilities over those of lower abilities. Second, we suppose that the technological change has been general-skill-biased, so it favors general skills over specific skills in production. In the presence of such technological changes, our model predicts that: 1) the wage distribution becomes more spread out in the sense of Bickel and Lehmann (1979), which corresponds to greater inequality; 2) job mobility of all types of workers increases; and 3) the proportionate increase in job mobility is larger for workers with higher levels of firm-specific human capital. These patterns are consistent with recent empirical evidence on changes in job mobility in the United States, which will be reviewed in Section 5.

In Section 6, we generalize the model to multiple sectors, where returns to ability are heterogeneous in different sectors. The multi-sector model delivers three sets of results. First, it provides us with a framework to analyze the sector choice of workers. We show that the worker's sector choice in period 1 is ex ante efficient. Second, the multi-sector model gives a new explanation for inter-industry wage differentials. Different from the existing literature, the source of wage differences here is adverse selection, so that firms in some sectors would be unwilling to lower their wage offers because lower wages result in hiring lower ability workers. Finally, the multi-sector model has closed-form solutions for the probabilities of both between-sector and within-sector turnover, enabling us to have a more detailed look at the impact of technological changes on job mobility and wage dispersion. We show in Section 6.2 that the predictions of the multi-sector model are consistent with the empirical findings of between-sector job mobility by Kambourov and Manovskii (2005).

In the rest of the paper, we proceed as follows. We set up the model in Sector 2. Section 3 solves the mixed equilibrium and proves its uniqueness (in outcomes). We derive the equilibrium turnover probabilities and wage distributions in Section 4. Section 5 explores the model's predictions about job mobility and wage distribution in the face of technological changes. Section 6 extends the model to multiple sectors and
explores how skill-biased technology change affects within-sector and between-sector mobilities. Section 7 concludes and discusses further applications of the model.

1.2 Model Setup

We set up the model formally in this section. Subsection 2.1 describes the model basics, including the types of players and their respective objective functions. Subsection 2.2 specifies the timing and information structure of the model and introduces notations for the strategies of the players. The solution concept of the model is given in Subsection 2.3.

1.2.1 Worker and Firms

There is a single worker who lives for two periods. The worker has ability \( a \), unknown at the beginning of period 1, that is drawn uniformly from \([0, 1]\). The utility of the worker is

\[
U = w_1 + w_2,
\]  

(1.1)

the sum of his wage incomes in the two periods.

We have three comments about the assumptions on the worker. First, the single worker assumption is made for simplicity. When there are multiple workers, we can carry the same analysis as long as the production function has constant return to scale in the number of the workers and that the abilities of the workers are independently distributed. Second, the worker's ability level should not be interpreted as an absolute level but instead the relative rank of the worker in the distribution. For example, a worker of ability 0.3 means that his ability is ranked at the (bottom) 30th percentile. Third, the assumptions that the worker is risk-neutral, has no disutility of effort, and does not discount the future are made completely for simplicity and can be relaxed.

There are \( N \) ex ante identical firms, \( 2 < N < \infty \). The payoff of each firm is the sum of its payoffs in the two periods. A firm’s period 1 payoff is 0 if it does not hire
the worker. If it hires the worker, its period 1 payoff is

\[ \pi_1 = y(a, t) - w_1 \quad (1.2) \]

where \( y(a, t) \), the period 1 output of the worker, depends on the worker’s ability \( a \) and an index \( t \) that reflects the state of technology common to all firms. We assume that

\[ y(a, t) \geq 0, \quad \frac{\partial y(a, t)}{\partial a} > 0, \quad \frac{\partial y(a, t)}{\partial t} > 0 \quad \text{for all } a \text{ and } t, \quad (1.3) \]

so the output is (uniformly) strictly higher if the worker is more able or the technology index is larger. The technology index \( t \) plays no role in the basic model analyzed in Section 3, but is central to the comparative static results presented in Section 4.

A firm’s period 2 payoff is 0 if it does not hire the worker. If it hires the worker, its period 2 payoff is

\[ \pi_2 = y(a, t) + 1_{\{\text{incumbent}\}}s(a, t) - w_2, \quad (1.4) \]

where \( 1_{\{\text{incumbent}\}} \) is an indicator function that takes the value of 1 if the firm is an incumbent (the worker’s period 1 employer) and 0 otherwise. In other words, a firm’s output equals \( y(a, t) + s(a, t) \) if it is an incumbent; its output is \( y(a, t) \) if it is an outside firm (i.e. a firm that does not hire the in period 1). We assume that

\[ s(a, t) \geq 0, \quad \frac{\partial s(a, t)}{\partial a} \geq 0, \quad \frac{\partial s(a, t)}{\partial t} \geq 0 \quad \text{for all } a \text{ and } t, \quad (1.5) \]

so the \( s(a, t) \) is weakly higher if the worker is more able or the technology index is larger. In the analysis below, we interprete \( y(a, t) \) as the general output (from the worker’s general-purpose human capital) and treat \( s(a, t) \) as the firm-specific output (from the worker’s firm-specific human capital)\(^1\).

\(^1\)There are other interpretations as well. For example, if we allow for \( s(a, t) \) to be negative, then it may be interpreted as the match quality.
We also assume that
\[ y(0, t) + s(0, t) < \int_0^1 y(a, t) da = E[y(a, t)], \] (1.6)
which says that the lowest inside output is smaller than the average outside output. This is a standard assumption in the literature to rule out trivial cases; see for example Gibbons and Katz (1991).

1.2.2 Timing and Information Structure

At the beginning of period 1, all firms simultaneously offer contracts to the worker. The contracts are restricted to be nonnegative, non-contingent, single-period wage offers. The assumptions on the set of contracts imply that the period 1 action of each firm \( j \) (\( j \in \{1, \ldots, N\} \)) is a distribution \( G_j \in \Delta R^+ \), where \( \Delta R^+ \) is the set of probability distributions on non-negative real numbers. In other words, each firm \( j \) draws its wage offer randomly from \([0, \infty)\) according to distribution \( G_j \), chosen by itself.

Concerning the restrictions on the available contracts, we make the non-negativity assumption to rule out equilibria in which the wages accepted by the worker are not bounded below. The non-contingent assumption is made to fit with the assumption that outside firms cannot observe the output of the worker (so contracts based on outputs cannot be verified by courts). The single-period assumption is made to reflect the lack of commitment power of the firms (and the worker) and the associated lack of enforceability of long-term contracts. These restrictions on the contracts are standard in the literature; see for example Greenwald (1986) and Gibbons and Katz (1991).

After all wage offers are made, the worker’s action is a function \( D_1 : (R^+)^N \rightarrow \Delta\{1, \ldots, N\} \), where \( (R^+)^N \) denotes the set of \( N \) nonnegative wage offers and \( \Delta\{1, \ldots, N\} \) is the set of probability distributions on \( \{1, \ldots, N\} \). In other words, the worker picks one firm from the \( N \) wage offers it receives. Note that the worker is allowed to randomize in anyway he wants when choosing the offer.
Once the worker picks a firm, period 1 production takes place and the wage is paid. Through production, the incumbent observes the exact ability level of the worker. On the other hand, outside firms observe nothing about the worker’s ability.  

At the beginning of period 2, all firms simultaneously offer contracts to the worker. For each firm \( j \in \{1, \ldots, N\} \), if it hires the worker in period 1, its wage offer, \( w_j : [0, 1] \rightarrow \Delta R^+ \), is a correspondence from the worker’s ability to the set of nonnegative real numbers. In other words, the incumbent is allowed to randomize its wage offer for any ability level. If firm \( j \) does not hire the worker in period 1, it draws its offer from \([0, \infty)\) according to distribution \( F_j \), chosen by itself.

After all offers are made in period 2, the worker’s action is a function \( D_2 : (R^+)^N \rightarrow \Delta\{1, \ldots, N\} \). In other words, the worker picks one offer from the \( N \) period 2 wage offers and works for the firm whose offer is chosen.

Finally, after the worker chooses a period 2 employer, period 2 production takes place, the wage is paid, and the game ends.

Our assumption of the simultaneity of offers in period 2 is the key difference from earlier models (Greenwald (1986), Gibbons and Katz (1991)), which assumes that the incumbent observes all outside offers to the worker and can make counteroffers. This simultaneity assumption is standard in the literature; see for example Waldman (1984). In the current context, it captures that employees typically cannot credibly communicate the total value of outside offers to the incumbent: even if the exact monetary values of the outside offers are known to the incumbent, it is unlikely that the worker’s non-monetary preferences can be known exactly. In other words, this difference, while formally about the model’s timing, should actually be interpreted in terms of difference about information structure. For example, we obtain identical results even if the incumbent makes its offer after the outside firms, as long as the incumbent does not know the outside offers.

\(^{2}\)These information assumptions are extreme and are made for simplicity. We can also adapt our analysis to more a general setting, which only assumes that the information set of the incumbent is finer than that of outside firms. The essential equilibrium structure remains in the more general setting.
We summarize the sequence of events as follows:

1. At the beginning of period 1, firms simultaneously offer contracts to the worker. Each firm \( j \ (j \in \{1, \ldots, N\}) \) draws its offer randomly from \([0, c_0]\) according to a distribution \( G_j \), chosen by itself.

2. The worker makes a decision \( D_1 : (R^+)^N \rightarrow \Delta\{1, \ldots, N\} \) to choose a firm from the \( N \) wage offers. Production takes place and wage is paid. The incumbent observes the ability of the worker while outside firms don’t.

3. At the beginning of period 2, all firms simultaneously offer contracts to the worker. For each firm \( j \in \{1, \ldots, N\} \), if firm \( j \) hires the worker in period 1, it chooses a wage offer \( w_j : [0, 1] \rightarrow \Delta R^+ \) based on the worker’s ability. If firm \( j \) does not hire the worker in period 1, it draws its offer randomly from \([0, 0_b]\) according to a distribution \( F_j \), chosen by itself.

4. The worker makes a decision \( D_2 : (R^+)^N \rightarrow \Delta\{1, \ldots, N\} \) to choose a firm from the wage offers. Production takes place and the wage is paid.

1.2.3 Perfect Bayesian Equilibrium

According to the timing and information structure, the strategy of the worker is a 2-tuple \((D_1, D_2)\), and the strategy of firm \( j \in \{1, \ldots, N\} \) is a 3-tuple \((G_j, F_j, w_j)\). Given the strategies, we solve the Perfect Bayesian Equilibrium (PBE) of the model. The PBE requires that the strategies of the worker and the firms to be sequentially optimal given their beliefs and that their beliefs are determined from the Bayes Rule wherever possible. In particular, we have,

1. The worker’s equilibrium period 2 contract choice \( D_2^* \) are optimal given any period 1 strategy \( D_1 \) of the worker and any strategies of the firms \( \prod_{j=1}^{N} (G_j, F_j, w_j) \) given his belief.
2. For each firm $j$, its period 2 strategy $(F_j^*, w_j^*)$ is optimal given any period 1 strategy $D_1$ of the worker, any period 1 strategy of the firms $\prod_{j=1}^{N} (G_j)$, period 2 equilibrium strategy of the worker $D_2^*$, period 2 strategies of all other firms $(F_{-j}^*, w_{-j}^*)$, and its belief.

3. The worker's contract choices $(D_1^*, D_2^*)$ are optimal given the strategies of the firms: $\prod_{j=1}^{N} (G_j, F_j^*, w_j^*)$ and its belief.

4. For each firm $j$, its strategy $(G_j^*, F_j^*, w_j^*)$ is optimal given the worker's equilibrium strategy $(D_1^*, D_2^*)$, the equilibrium strategies of all other firms: $(G_{-j}^*, F_{-j}^*, w_{-j}^*)$, and its beliefs.

5. For each firm $j$, at the beginning of period 1, its belief about the worker's ability is the prior distribution of the worker's ability. In period 2, if $j$ is the incumbent, it knows the exact ability of the worker. Otherwise, $j$'s belief is the prior distribution of the worker's ability. The worker does not know his ability in period 1. The worker knows his ability in period 2.

1.3 Equilibrium of the Model

We solve the equilibrium of the model in this section. Subsection 3.1 shows that the model does not have a pure strategy PBE. Instead, there is a mixed strategy PBE, which is described in Theorem 1. Subsection 3.2 shows that the equilibrium described in Theorem 1 is essentially unique: every equilibrium of the model leads to the same outcome in job mobility and the wage distribution. The proof of Theorem 2 is technical and is of independent interest. Readers who are more interested in the substantive results of the paper can skip Subsection 3.2 and jump directly to Section 4.
1.3.1 Existence of a Mixed Strategy PBE

For ease of exposition, we simplify some notations. First, the technology index $t$ plays no role in establishing the equilibrium, so we write $y(a)$ and $s(a)$ instead of $y(a,t)$ and $s(a,t)$ in this section. Second, since all firms are ex ante identical, we let firm 1 be the incumbent in period 2 and write its equilibrium wage offer as $w_{In}(a)$.

Before describing the mixed strategy PBE, our first observation, reported in Lemma 1, is that the model does not have a pure strategy PBE. The logic of Lemma 1 is similar to that of Akerlof’s lemon model. In particular, suppose $w$ is the highest outside offer in period 2. If a pure strategy equilibrium exists, the incumbent would keep the worker whenever his inside output $(y(a) + s(a))$ is above $w$. This implies that a worker’s outside output $(y(a))$ must be below $w$ when he leaves the incumbent because the worker’s inside output is greater than or equal to his outside output. In other words, when an outside firm hires the worker with wage $w$, the expected outside output of the worker must be less than $w$. This gives the outside firm a negative payoff, which cannot happen in the equilibrium. Lemma 1 proves this formally.

**Lemma 1** There is no pure strategy PBE.

**Proof.** We prove by contradiction. Suppose instead there is a pure strategy PBE.

Let $w$ be the highest outside offer in period 2 in this equilibrium, so the incumbent can keep the worker if it offers any wage above $w$. Because the incumbent earns zero (in period 2) if it does not keep the worker, the incumbent will keep the worker in equilibrium if his inside output $(y(a) + s(a))$ is greater than $w$. This implies that outside firms never get the worker when $y(a) + s(a) > w$, so the expected profit of any outside firm, conditional on hiring the worker, is at most

$$E[y(a)|y(a) + s(a) \leq w] - w < 0.$$ 

Therefore, if an outside firm hires the worker with positive probability, it must have a negative payoff. Since outside firms can always guarantee themselves nonnegative
payoffs (by offering zero wages), this implies that the outside firms must hire the worker with zero probability, or equivalently, the incumbent must keep the worker with probability 1 in this pure strategy PBE.

This implies that the incumbent must offer a wage greater or equal to $w$ with probability 1. But any wage offer strictly greater than $w$ cannot be optimal for the incumbent (because it can be replaced by a smaller wage, say the average of $w$ and itself, that also keeps the worker but is smaller in amount), the incumbent must offer $w$ with probability 1 in this equilibrium. Now consider an outside firm that deviates by offering $w' = w + \epsilon$ for some $\epsilon > 0$. This deviation hires the worker with probability 1 and gives to the deviating firm an expected profit of

$$E[y(a)] - w - \epsilon = E[y(a)] - y(0) - s(0) - \epsilon.$$ 

The deviation is profitable for small enough $\epsilon$ because $E[y(a)] - y(0) - s(0) > 0$ by the production assumption (1.6). This leads to a contradiction.

The absence of pure strategy equilibrium here stands in contrast with earlier results; see for example Greenwald (1986) and Gibbons and Katz (1991). The difference arises because the incumbent in this model cannot make counteroffers. If there were an pure strategy equilibrium with low outside wage offers, this implies that the incumbent cannot "respond" to the deviation of outside offers, so the equilibrium cannot be sustained. On the other hand, this lemma is similar to earlier results in illustrating the negative consequence of adverse selection on turnover: when firms play pure strategy, no turnover exists without exogenous movers or differences in match qualities.

Although no pure strategy PBE exists in this model, there is a mixed strategy PBE where in period 2 outside firms randomize in their offers. Theorem 1 below proves the existence by formally constructing one equilibrium. The key to this construction depends on 1): the incumbent's period 2 offer so that outside firms are willing to randomize, and 2): the randomization of outside firms in period 2 so that
the incumbent’s offer is optimal. These two features form the essential ingredients of the strategies constructed in Theorem 1 to be an equilibrium. We describe and discuss these two features carefully before stating Theorem 1. Readers who are more interested in the substantial results of the paper can read these discussions and skip the proof of Theorem 1.

To better describe the incumbent’s equilibrium period 2 offer, we first introduce the following definition.

**Definition 1:** The average outside output\(^3\) of a worker of ability \(a\) is defined as

\[
B(a) = \frac{\int_a^a y(x)dx}{a} = E[y(x)|x \leq a].
\] (1.7)

The first feature of the equilibrium states that in period 2 the incumbent offers wage that equals the average outside output of the worker. More formally,

\[
w_{In}(a) = B(a), \quad \text{for all } a > 0. \tag{1.8}
\]

It is easy to see that \(B(a)\) is strictly increasing in worker’s ability. Moreover, \(B(a)\) is completely determined by \(y(a)\). For example, if \(y(a) = a\), then \(B(a) = \frac{1}{2}a\). When the incumbent offers \(w_{In}(a) = B(a)\), it implies that the expected payoff of an outside firm by offering \(w < E[y(a)]\) is always 0.

To see this, take the example above of \(y(a) = a\) and \(B(a) = \frac{1}{2}a\). If an outside firm offers a wage of \(w < E[y(a)] = \frac{1}{2}\) and hires the worker, this implies that the incumbent’s offer to the worker must be less than or equal to \(w\). Since the incumbent offers \(B(a) = \frac{1}{2}a\), the ability of the worker must satisfy \(a \leq 2w\). This implies that the hired worker’s ability is uniformly distributed between 0 and \(2w\), so the expected ability of the worker hired by the outside firm equals \(w\). Since the output of a worker equals his ability in this case, the expected output of the worker is also \(w\). Therefore,

---

\(^3\)We omit the modifier "below his ability" in the definition for simplicity. This also draws parallel with the average cost term in price theory.
the expected payoff of this outside firm is $w - w = 0$. Because $w$ is arbitrary, the expected payoff of an outside firm is always zero for any wage offer less than $\frac{1}{2}$.

When the output of a worker does not equal his ability, the same logic still goes through. In general, as long as the incumbent offers $w_{In}(a) = B(a)$, the expected payoff of an outside firm offering wage in $[0, E[y(a)])$ is always zero. This implies that outside firms are willing to randomize their offers in $[0, E[y(a)])$ when the incumbent offers wage equals the average outside output.

Next, we describe the randomization of outside firms in the equilibrium in period 2. Again, we introduce a notation first.

**Definition 2:** The maximum outside offer distribution $F$ is defined as

$$F(w) = \prod_{j=2}^{N} F_j^*(w),$$

where $F_j^*$ is the equilibrium offer distribution of firm $j$. This expression follows because the maximum outside offer is less than $w$ if and only if the offer from each of the outside firm is less than $w$, so the maximum outside offer distribution is the multiplication of the equilibrium offer distribution of all outside firms. Also define the "boundary" of support of $F$ as

$$w = \inf\{w : F(w) > 0, w \geq w_{In}(a) \text{ for some } a\};$$

$$\bar{w} = \sup\{w : F(w) < 1\}.\quad (1.10)$$

Note that the incumbent keeps a worker if and only if the maximum outside offer is less than its offer. In other words, the probability that the incumbent keeps the worker is completely determined by the distribution of the maximum outside offer.

The second feature of the equilibrium states that in period 2 outside firms randomize their offers in $[0, E[y(a)])$ so that the maximum outside offer distribution $F$ satisfies
\[
\frac{1}{y(a) + s(a) - B(a)} = \frac{f(B(a))}{F(B(a))}, \quad \text{for all } a > 0.
\] (1.11)

This feature implies that when the distribution of maximum outside offer satisfies (1.11), the incumbent finds it optimal to offer wage equals the average outside output.

To see this, the incumbent's expected payoff in period 2 by offering \( w \) to a worker of ability \( a \) is given by
\[
F(w)(y(a) + s(a) - w)^4
\] (1.12)
where \( y(a) + s(a) - w \) is its profit for keeping the worker and \( F(w) \) is the probability that it keep the worker because the incumbent keeps the worker if and only the maximum outside offer is less than \( w \). The maximization of the expression above implies that incumbent's optimal wage choice must satisfy
\[
\frac{1}{y(a) + s(a) - w} = \frac{f(w)}{F(w)}.
\] (1.13)

Equation (1.11) replaces \( w \) with \( B(a) \) in the equation above. In other words, \( w = B(a) \) satisfies the first order condition of the incumbent's maximization problem when (1.11) holds. It can be seen the second order condition also holds, so that \( w_{in}(a) = B(a) \) is the optimal wage choice for the incumbent when (1.11) holds.

It is important to note that equation (1.11) describes the distribution of maximum outside offer and does not specify the distribution of offer for each outside firm. In general, there are infinitely many possible individual wage distributions \( F^*_j \) \( (j \in \{2, \ldots, N\}) \) that sustain \( F \), and we cannot pin down the exact equilibrium wage offer distribution of each outside firm. In Theorem 1, we construct an equilibrium in which the offer distribution of each outside firm is identical with \( F^*_j = F_{\text{eq}} \) for all \( j \in \{2, \ldots, N\} \).

In summary, the incumbent’s offer and the randomization of outside firms provides the essential ingredients to the equilibrium construction in Theorem 1. When the

\[\text{4This assumes (for now) that } F \text{ does not have atoms.}\]
incumbent offers $w_{ln}(a) = B(a)$, any wage offer in $(0, E[y(a)])$ is an optimal response for each outside firm. When the maximum offer distribution satisfies the differential equation (1.11), the incumbent finds it optimal to offer $w_{ln}(a) = B(a)$. Therefore, this is an equilibrium strategy profile in period 2. To fully describe a PBE, we also need to specify period 1 strategies and beliefs. These are done in Theorem 1.

**Theorem 1:** The following strategies and beliefs form a PBE:

(i) In period 2, the worker chooses the maximum wage offer. If there are multiple maximum offers, the worker a) stays with the incumbent if its offer is one of the maximum offers; b) randomize otherwise.

(ii) At the beginning of period 2, the incumbent firm offers

$$w_{ln}(a) = B(a) \quad \text{for all } a,$$

and each outside firm $j \ (j \in \{2, \ldots, N\})$ offers a wage drawn independently from the distribution

$$F_j^*(w) = F(w)^{\frac{1}{n-1}},$$

(1.15)

where

$$F(w) = 0 \quad \text{for } w < w_1;$$

$$F(w) = C \exp \left( \int_{w_1}^{w} \frac{1}{y(B^{-1}(x)) + s(B^{-1}(x))} dx \right) \quad \text{for } w \in [w_1, \bar{w}];$$

$$F(w) = 1 \quad \text{for } w > \bar{w},$$

$$[w_1, \bar{w}] = [y(0), E[y(a)]] \quad \text{and } C = \exp \left( - \int_{w_1}^{\bar{w}} \frac{1}{y(B^{-1}(x)) + s(B^{-1}(x)) + B(a)} dx \right).$$

(iii) In period 1, the worker chooses the maximum wage offer. If there are multiple maximum offers, the worker randomizes among them.

(iv) At the beginning of period 1, all firms offer

$$E[y(a)] + \int_0^1 F(B(a))(y(a) + s(a) - B(a)) da.$$

(1.17)

(v) Each firm’s belief about the worker’s ability equals the prior if it has not hired
the worker. The incumbent knows the worker’s ability at the end of period 1. The worker does not know his ability in period 1. The worker knows his ability in period 2.

Proof. We start with the beliefs. Since the worker does not know his own ability in period 1, the beliefs described by (v) is the only beliefs consistent with the (degenerated) Bayes rule.

Next, we examine the strategies in period 2. In period 2, the worker’s utility is maximized by choosing the maximum wage offer. Therefore, the strategy described in (i) is optimal for the worker.

Now given the equilibrium strategy of the worker and the maximum outside offer distribution, the incumbent’s payoff by offering $w$ to a worker of ability $a$ is

$$(y(a) + s(a) - w)F(w).$$

Maximization of the incumbent’s profit gives the following first order condition:

$$\frac{1}{y(a) + s(a) - w_{In}(a)} = \frac{f(w_{In}(a))}{F(w_{In}(a))}. \tag{1.18}$$

The first order condition is a necessary condition for optimality, and it is easy to check that if a solution satisfies it and is also increasing in $a$ (which is the case for $B(a)$), then the second order condition is also satisfied and the solution is optimal. Let $w_{In}(a) = B(a)$, then by the definition of $F$ in (ii) we can check that

$$\frac{1}{y(a) + s(a) - B(a)} = \frac{f(B(a))}{F(B(a))}.$$

Therefore, $w_{In}(a) = B(a)$ satisfies (1.18) and thus maximizes the incumbent’s expected payoff.

Given the incumbent’s wage offer and the worker’s strategy, the expected profit of an outside firm by offering $w$ is

$$P(\text{Not Hiring}) \cdot 0 + P(\text{Hiring})(E[y(a)|B(a) < w] - w). \tag{1.19}$$
This expression is zero for all \( w \leq E[y(a)] \) because \( E[y(a)\mid B(a) < w] = B(B^{-1}(w)) = w \). When \( w > E[y(a)] \), this expression is negative because \( E[y(a)\mid B(a) < w] - w = E[y(a)] - w < 0 \). Therefore, the optimal response of each outside firm is to randomize over \((0, E[y(a)])\). This completes proving the optimality of the strategies in period 2.

Moving back to period 1, it is clear that the worker maximizes his utility by choosing the maximum wage offer because all the firms are ex ante identical. Therefore, the worker’s strategy described in (iii) is optimal.

For the firms, wage competition implies that the equilibrium wage offer will be bid up to the expected output of the worker in period 1 plus the expected profit the firm makes if it is the incumbent in period 2. Therefore, it is an equilibrium that each firm offers in period 1

\[
E[y(a)] + \int_0^1 F(B(a))(y(a) + s(a) - B(a))da.
\]

This finishes the proof. ■

The PBE constructed in Theorem 1 is similar to the equilibrium in Burdett and Judd (1983), in which firms also randomize in their wage offers. In Burdett and Judd (1983), however, all workers are identical. There is also a search friction, so that the number of wage offers received by each worker is a Poisson random variable. Each firm chooses its wage to maximize the product of profit made from the worker and the probability of hiring the worker (by being the highest wage offer received by the worker). In Burdett and Judd’s equilibrium, when a firm raises its wage, the benefit of increasing the probability of hiring the worker exactly cancels out with the cost of paying the worker a higher wage. Therefore, all firms randomize. In addition, all firms receive positive expected profits.

In this model, the outside firms have zero expected profits regardless of the wages they offer, so the probability of hiring the worker becomes irrelevant. Instead, the benefit of offering higher wages comes from hiring workers of higher average ability. It is interesting to note that although firms that pay higher wages hire better workers on
average, they are not necessarily more profitable. This is consistent with the findings of Abowd, Kramarz and Margolis (1999).

Another difference of our PBE from that in Burdett and Judd (1983) is concerned with how the number of firms affect the equilibrium turnover and wage distribution. In Burdett and Judd (1983), the wage distribution converges to the wage from perfect competition when the number of firms goes to infinity. In this model, in contrast, the incumbent offer and the maximum outside offer distribution is independent of the number of firms. Consequently, the turnover probability and the wage distribution are independent of the number of firms. Since turnover is always inefficient in this model due to loss of firm-specific human capital, it suggests that asymmetric information creates a source of inefficiency that cannot be cured by simply increasing the number of firms.

1.3.2 Uniqueness

The PBE in Theorem 1 is just one of the infinitely many PBEs of the game. However, in each PBE of the model, as shown by Theorem 2 below, two properties must be satisfied. First, the incumbent’s wage offer must equal the average outside output. Second, the maximum outside offer distribution is unique and must satisfy the differential equation in (1.11). Since the worker always chooses the firm with the highest wage offer, job mobilities and the wage distributions are completely determined by the incumbent’s offer and the distribution of the maximum outside offers. Therefore, the two properties above imply that the model has a unique equilibrium outcome in job mobility and the wage distribution.

The proof of Theorem 2 depends on two observations. First, the expected payoff of all outside firms in period 2 are zero (Lemma 3). Second, the distribution of the maximum outside offer $F$ are well-behaved, i.e. $F$ does not have an atom except possibly at the bottom (Lemma 5) and $F$ has a connected support (Lemma 6). The

\footnote{In fact, "connectedness" is a misnomer that has been used widely in the literature. The correct mathematical statement is that the closure of the maximum outside offer is connected.}
zero expected profit lemma (Lemma 3) together with the connected support property (Lemma 6) implies that the incumbent's wage offer must equal the average outside output for workers of all ability levels. This equilibrium incumbent offer condition and the no-atom condition (Lemma 5) guarantee that the randomization choice of the worker when there are multiple maximum offers is irrelevant, so the equilibrium maximum outside offer distribution $F$ must satisfy equation (1.11). Equation (1.11), together with the left and right end point of the support of $F$ given in Lemma 9 and 10, completely characterizes the maximum outside offer distribution, so it is unique.

As mentioned above, the structure of our PBE is similar to the equilibrium in Burdett and Judd (1983). In fact, our lemmas and proofs are also similar to those in Burdett and Judd (1983) in many aspects. However, there are differences both in the results and in the proof techniques. We make frequent references to Burdett and Judd (1983) and compare our results and proofs with theirs whenever possible.

To prove Theorem 2, we first show that the incumbent's wage offer is strongly increasing in the worker's ability. Strongly increasing is an order on sets$^6$. We use this order here because we have not shown that the incumbent's wage offer is single-valued yet.

**Lemma 2** If $w_{I_n}(a_2) > w$, we have $w_{I_n}(a_1) \geq w_{I_n}(a_2)$ for all $a_1 > a_2$.

**Proof.** We prove by contradiction. Take two arbitrary ability levels $a_1 > a_2$ such that $w_{I_n}(a_2) > w$. Let $w_1$ and $w_2$ be two equilibrium wage offers of the incumbent when the worker is of ability $a_1$ and $a_2$ respectively; i.e. $w_1 \in w_{I_n}(a_1)$ and $w_2 \in w_{I_n}(a_2)$. Suppose instead we have $w_1 < w_2$.

Define $\tilde{F}(w)$ as the probability that the incumbent keeps the worker if it offers $w$. Clearly we have $\tilde{F}(w_1) \leq \tilde{F}(w_2)$, where recall $\tilde{F}(w)$ is the probability that the incumbent keeps the worker if it offers $w$. Now if $\tilde{F}(w_1) = \tilde{F}(w_2)$, the incumbent's payoff by offering $w_1$ is

$$\tilde{F}(w_1)(y(a_2) + s(a_2) - w_1) > \tilde{F}(w_2)(y(a_2) + s(a_2) - w_2)$$

\footnote{Let $X$ and $Y$ be two sets. $X \geq Y$ if $x \geq y$ for all $x \in X$ and $y \in Y$.}
since $\tilde{F}(w_1) = \tilde{F}(w_2) > 0$. In other words, the incumbent strictly prefers offering $w_1$ to $w_2$ at $a_2$ if $\tilde{F}(w_1) = \tilde{F}(w_2)$, violating the assumption that $w_2 \in w_{In}(a_2)$. Therefore, we must have $\tilde{F}(w_1) < \tilde{F}(w_2)$.

Now let $y_{In}(a) = y(a) + s(a)$. The expected profit of the incumbent by offering $w_2$ when the worker has ability $a_1$ is

$$
(y_{In}(a_1) - w_2)\tilde{F}(w_2) = (y_{In}(a_1) - y_{In}(a_2))\tilde{F}(w_2) + (y_{In}(a_2) - w_2)\tilde{F}(w_2) \geq (y_{In}(a_1) - y_{In}(a_2))\tilde{F}(w_2) + (y_{In}(a_2) - w_1)\tilde{F}(w_1) > (y_{In}(a_1) - y_{In}(a_2))\tilde{F}(w_1) + (y_{In}(a_2) - w_1)\tilde{F}(w_1) = (y_{In}(a_1) - w_1)\tilde{F}(w_1).
$$

The first inequality follows from $w_2 \in w_{In}(a_2)$, and the strict inequality follows because $y_{In}(a_1) - y_{In}(a_2) > 0$ and $\tilde{F}(w_2) > \tilde{F}(w_1)$.

This chain of inequality implies that the incumbent would strictly prefer offering $w_2$ to $w_1$ when the worker has ability $a_1$, thus contradicting the assumption that $w_1 \in w_{In}(a_1)$. Therefore, we must have $w_1 \geq w_2$. □

The purpose of Lemma 2 is to help calculate the expected payoff of an outside firm if it hires the worker because calculating the expected payoff requires the (hired) worker’s conditional ability distribution, which depends on the incumbent’s equilibrium offer. Lemma 2 shows that the incumbent’s offer is monotonic in worker’s ability so that the (hired) worker’s conditional ability distribution must be a left-truncation of the original ability distribution. In particular, Lemma 2 implies that the conditional expected output of a worker hired by an outside firm that offers $w$ is increasing in $w$.

Using the monotonicity property, we show that each outside firm must have zero expected profits. This is the content of Lemma 3.

**Lemma 3** The expected profit of each outside firm is zero.
Proof. It suffices to show that the expected profit of one outside firm is zero. Because if one outside firm earns zero expected profit and another one earns positive profit, the one with zero expected profit can always mimic the one with positive profit to earn positive profit as well. We consider two cases.

Case 1: $F(w) = 0$, where recall that $F$ is the CDF of the maximal outside offers.

Since $F(w) = 0$, the definition of $w$ together with right-continuity of a CDF implies that for each $n > 0$, there exists $w_n > w$ such that a): $w_n$ is offered by an outside firm and b): $F(w_n) < \frac{1}{n}$.

By offering $w_n$, the expected profit of the outside firm is less than

$$F(w_n)(E[y(a) | w_{I_n}(a) \leq w_n]) - w_n$$

because the probability that the firm hires the worker is no greater than $F(w_n)$ and the expected profit from hiring the worker is no greater than $E[y(a) | w_{I_n}(a) \leq w_n] - w_n$ because the incumbent’s offer is monotone in the worker’s ability (Lemma 2).

Equation (1.22) goes to zero as $n$ goes to infinity because $F(w_n) < \frac{1}{n}$ and $E[y(a) | w_{I_n}(a) \leq w_n] - w_n$ is bounded. Because we have only finite many outside firms, this implies that there must be one outside firm whose profit is less than $\frac{1}{n}$ for all $n$ and its profit must thus be zero. This finishes the proof for the case with $F(w) = 0$.

Case 2: $F(w) > 0$.

First, if an outside firm offers a wage less than $w$, this wage offer hires the worker with probability zero (by the definition of $w$) and gives the firm a profit of zero. Since all offers in the support of its equilibrium offers must yield the same profit for the firm, this implies the outside firm must have an expected profit zero (and we are done). Therefore, we may assume all outside firms offer wages greater than or equal to $w$.

When all outside firms offer wages greater than or equal to $w$, our assumption $F(w) > 0$ implies that each outside firm must offer $w$ with positive probability. In other words, $F_j(w) > 0$ for all $j \in \{2, \ldots, N\}$ (recall the firm 1 is the incumbent). When all outside firms offer $w$ (and the incumbent offers no greater than $w$), without
loss of generality, we may assume that the worker chooses firm 2 with probability \( q \leq \frac{1}{2} \).

Now consider the payoff of firm 2 when it offers \( w \). Its expected output is less than or equal to its expected profit conditional on hiring the worker:

\[
q \prod_{j=3}^{N} F_j^*(w) \Pr(w_{in}(a) \leq w)(E[y(a)|w_{in}(a) \leq w] - w).
\] (1.23)

This expression follows because firm 2 can hire the worker only if all other outside firms offer \( w \) (with probability \( \prod_{j=3}^{N} F_j^*(w) \)), the incumbent offers less than or equal to \( w \) (with probability \( \Pr(w_{in}(a) \leq w) \)), and the worker chooses firm 2 in face of this tie (with probability \( q \)). Hence \( \prod_{j=3}^{N} F_j^*(w) \Pr(w_{in}(a) \leq w) \) is the probability that firm 2 hires the worker with wage \( w \). And the firm 2’s expected profit conditional on hiring the worker is at most \( E[y(a)|w_{in}(a) \leq w] - w \).

Suppose firm 2 offers \( w + \varepsilon \) instead. Firm 2’s expected profit with this offer is at least

\[
\prod_{j=3}^{N} F_j^*(w) \Pr(w_{in}(a) \leq w)(E[y(a)|w_{in}(a) \leq w] - w - \varepsilon).
\]

If \( E[y(a)|w_{in}(a) \leq w] - w > 0 \), this is a profitable deviation for \( \varepsilon \) small enough because there is a discrete increase in the probability of hiring the worker. Therefore, we must have \( E[y(a)|w_{in}(a) \leq w] - w = 0 \), so firm 2 must have zero expected profit and we are done.

Lemma 3 stands in contrast with Burdett and Judd (1983) where all firms have a positive expected profit. This difference in profits results from the difference in the technologies of how the worker receives wage offers. In Burdett and Judd (1983), the number of offers a worker receives is a Poisson random variable. This implies that there is a positive probability that the worker receives only one offer. When this happens, the firm that is the only wage offerer becomes the de facto monopsonist. This positive probability of becoming the monopsonist is the source of a positive ex ante profit for firms in Burdett and Judd (1983). In contrast, in this model, no such monopsonist possibility exists because the worker always receives its offer from all

35
firms. This Bertrand-type competition suggests that all outside firms will have zero expected profits in equilibrium.

The zero expected profits result in Lemma 3 plays two important roles in our proof of Theorem 2. First, it is the key property used in Lemma 7 below that determines the equilibrium wage offer of the incumbent. Second, it helps prove Lemma 5, which states that the maximum outside offer distribution does not have an atom except possibly at the bottom.

Before proving Lemma 5, we first prove a technical lemma that states that the incumbent’s offer does not have an atom at any $w > w$. This is the content of Lemma 4.

**Lemma 4** If $w > w$, then $\Pr(w_{in}(a) = w) = 0$.

**Proof.** We prove by contradiction. If instead there exists a wage $w > w$ such that $\Pr(w_{in}(a) = w) > 0$, then Lemma 2 implies that there exists $a_1 < a_2$, such that $w_{in}(a_1) = w_{in}(a_2) = w$.

Now take any $\varepsilon > 0$ and compare the expected profit of an outside firm between offering $w + \varepsilon$ and $w - \varepsilon$. Since by offering $w + \varepsilon$ ($w - \varepsilon$) the incumbent keeps the worker with a probability of at least (most) $F(w)$, the extra expected profit in offering $w + \varepsilon$ is at least

$$F(w)(B(a_2) - w - \varepsilon) - F(w)(B(a_1) - w + \varepsilon)$$

$$= F(w)(B(a_2) - B(a_1) - 2\varepsilon).$$

Since $F(w) > 0$ and $B(a_2) - B(a_1) > 0$, for small enough $\varepsilon$ the term above is positive. In other words, for small enough $\varepsilon$, offering $w - \varepsilon$ is strictly dominated by offering $w + \varepsilon$. Therefore, no outside firm would offer a wage in $(w - \varepsilon, w)$ for small enough $\varepsilon$.

Since no outside firms offer wage in $(w - \varepsilon, w)$, the incumbent that offers $w$ has a profitable deviation by offering $w - \frac{\varepsilon}{2}$ for small enough $\varepsilon$, as long as the maximum
outside offer distribution does not have a mass in \( w \) (so deviating to \( w - \frac{\varepsilon}{2} \) does not affect the incumbent’s probability of retaining the worker).

When the maximum outside offer distribution has a mass at \( w \), let \( q \) be the probability that the worker leaves the incumbent if both the incumbent offer and the maximum outside offer is \( w \). If \( q = 1 \), the incumbent never keeps the worker when the worker randomizes over \( w \). In this case, the incumbent strictly prefers offering \( w - \frac{\varepsilon}{2} \) to \( w \), because it is cheaper and keeps the worker with the same probability.

When \( q < 1 \), the outside firm that offers \( w \) (there is such a firm because the maximum outside offer distribution has a mass at \( w \)) can profitably deviate by offering \( w + \delta \), which increase the probability of hiring the worker and discretely increase the expected ability of the workers hired.

These cases combined show that there is always a profitable deviation when \( \Pr(a|w_{I_n}(a) = w) > 0 \) for \( w > w \). Therefore, we must have \( \Pr(w_{I_n}(a) = w) = 0 \) if \( w > w \).  

**Lemma 5** \( F(w) = F(w_-) \) for all \( w > w \), where \( F(w_-) \) is the left limit of \( F \) at \( w \).

**Proof.** We prove by contradiction. Suppose instead \( F \) has an atom at \( w > w \). To get a contradiction, it suffices to show that the incumbent will not offer a wage in \( (w - \varepsilon, w) \). Because in this case, an outside firm that offers \( w - \frac{\varepsilon}{2} \) has an expected profit (conditional on hiring the worker) of at least

\[
E[B(a)|w_{I_n}(a) \leq w - \varepsilon] - w + \frac{\varepsilon}{2} \leq E[B(a)|w_{I_n}(a) \leq w] - w + \frac{\varepsilon}{2} = \frac{\varepsilon}{2},
\]

where the first equality follows from that the incumbent will not offer a wage in \( (w - \varepsilon, w) \) and that incumbent’s offer does not have an atom (Lemma 3), and the second equality follows from that the expected payoff of all outside firms must be zero (so the conditional profit of an outside firm offering \( w \) must be zero). This leads
to a contradiction because it implies that an outside firm can have positive expected profit by offering $w - \frac{\varepsilon}{2}$, violating Lemma 3.

Now we show that the incumbent will not offer a wage in $(w - \varepsilon, w)$ for small enough $\varepsilon$ and thus complete the proof. Since the incumbent’s expected profit is increasing in the worker’s ability, there exists an $M$ such that if $w_{In}(a) \in (w - \frac{1}{M}, w)$, we must have $y_{In}(a) > w + \frac{1}{M}$. Now take $\varepsilon = \frac{1}{2M}(F(w) - F(w_-))$. Note that $\varepsilon < \frac{1}{M}$, so if the incumbent offers a worker of ability $a$ a wage in $(w - \varepsilon, w)$, we must have $y_{In}(a) > w + \frac{1}{M}$.

Suppose the incumbent offers a worker of ability $a$ a wage in $(w - \varepsilon, w)$, its expected profit is at most $F(w_-)(y_{In}(a) - w + \varepsilon)$. If the incumbent deviates and offers $w + \varepsilon$ instead, its expected profit is at least $F(w)(y_{In}(a) - w - \varepsilon)$. Therefore, the difference in the expected profit after the deviation is at least

$$F(w)(y_{In}(a) - w - \varepsilon) - F(w_-)(y_{In}(a) - w + \varepsilon)$$

$$= (F(w) - F(w_-))(y_{In}(a) - w) - (F(w) + F(w_-))\varepsilon$$

$$> (F(w) - F(w_-))\frac{1}{M} - (F(w) + F(w_-))\varepsilon$$

$$> 0,$$

where the first inequality follows because $y_{In}(a) > w + \frac{1}{M}$ and the second inequality follows because $\varepsilon = \frac{1}{2M}(F(w) - F(w_-)) < \frac{F(w) - F(w_-)}{M(F(w) + F(w_-))}$. This shows that the incumbent would not offer a wage in $(w - \varepsilon, w)$ and finishes the proof.

The statement of Lemma 5 is identical to a corresponding result in Burdett and Judd (1983), which also states that the distribution of the maximum offer received by the worker has no atoms except at the bottom. However, there are differences between ours and theirs both in the environments and in the technique of proofs. In Burdett and Judd (1983), all firms are infinitesimal. The wage offer of a single firm has no effect on the equilibrium distribution of maximum wage offer received by the worker. If there is an atom in the wage distribution, a firm that offers a wage at the
atom can “jump up” its wage by epsilon, which discretely increases its probability of
hiring the worker. This “jump up” strictly increases the firm’s expected profit because
the firm’s conditional expected profit of hiring the worker is positive in Burdett and
Judd (1983).

This “jump up” argument, however, does not work in our environment. Partly
this is because there are only finite firms in our environment, so a change in the
wage offer of any firm may affect the equilibrium maximum outside offer distribution.
More importantly, the conditional expected profit of hiring the worker in our model
may not be positive, in contrast with that in Burdett and Judd (1983). In fact,
the conditional expected profit of outside firms is always zero when the incumbent
offers \( w_{ln}(a) = B(a) \), as is seen in (1.8). Therefore, a “jump up” that increases the
probability of hiring the worker may not increase the firm’s expected profit in this
model.

Our proof of the no atom result is based on a combination of “jump up” and “jump
down” arguments and relies heavily on the zero expected profit result (Lemma 3).
The sketch of the proof can be presented as follows. Suppose instead there is an atom
at \( w > w \). Because the incumbent’s offer is unrelated to outside offer distribution and
that the incumbent’s profit of keeping the worker is positive, we can use the same
“jump up” argument to show that the incumbent would not want to offer a wage
in \((w - \varepsilon, w)\) for some small \( \varepsilon \) if there is an atom at \( w \). Together with Lemma 4
which shows that the incumbent’s offer has no atom except possibly at the bottom,
it follows that the probability that the incumbent offers a wage in \((w - \varepsilon, w]\) is 0.

The key implication of above is that any wage in \((w - \varepsilon, w]\) leads to the same
expected ability of the hired worker for outside firms. Now by assumption there is at
least one outside firm that offers \( w \) in equilibrium. Since \( w > w \), this outside firm hires
the worker with positive probability, so Lemma 3 implies that the conditional expected
profit of hiring the worker is zero. When this outside firm deviates by “jumping down”
to a wage in \((w - \varepsilon, w)\), it lowers its wage offer but keeps the expected ability of the
worker hired. So the expected profit of the outside firm becomes positive, and this
contradicts Lemma 3. Hence, there cannot be an atom in the maximum outside offer
distribution.

Another nice property of the maximum equilibrium outside offer distribution is
that (the closure) of its support is connected in $[w, \overline{w}]$. This is the content of Lemma
6.

**Lemma 6** For $w_1 < w_2 \in [w, \overline{w}]$, we have $F(w_1) < F(w_2)$.

**Proof.** We prove by contradiction. Suppose instead we have $F(w_1) = F(w_2)$ for
some $w_1 < w_2 \in [w, \overline{w}]$. Without loss of generality, we may assume that $w_2$ is the
largest wage such that no outside firms makes offer between $(w_1, w_2)$ (with positive
probability), i.e. $w_2 = \sup\{w : F(w) = F(w_1)\}$. This implies that, for any $\varepsilon > 0$, we
can find a wage $w \in [w_2, w_2 + \varepsilon)$ that is offered in equilibrium by an outside firm.

Now take an outside firm that offers wage $w \in [w_2, w_2 + \varepsilon)$, its expected profit is
at most

$$F(w_2 + \varepsilon) \Pr(w_{In}(a) \leq w_2 + \varepsilon)(E[B(a)|w_{In}(a) \leq w_2 + \varepsilon] - w_2),$$

(1.27)

where $F(w_2 + \varepsilon) \Pr(w_{In}(a) \leq w_2 + \varepsilon)$ is an upper bound of the probability of hiring
the worker and $E[B(a)|w_{In}(a) \leq w_2 + \varepsilon]$ is an upper bound of the expected output
of the worker hired.

Suppose the firm deviates and offers $\frac{w_1 + w_2}{2}$ instead. Its expected profit is at least

$$F(w_1) \Pr(w_{In}(a) \leq w_1)(E[B(a)|w_{In}(a) \leq w_1] - \frac{w_1 + w_2}{2}),$$

(1.28)

where $F(w_1) \Pr(w_{In}(a) \leq w_1)$ is a lower bound of hiring the worker and $E[B(a)|w_{In}(a) \leq
w_1]$ is a lower bound of the expected output of the worker hired.

Since $F(w_1) = F(w_2)$, it is clear that the incumbent would not offer a wage in
$(w_1, w_2)$ in equilibrium. Furthermore, Lemma 4 implies that the incumbent offers $w_2$
with probability zero. Therefore, we have

\[
\Pr(w_{I\alpha}(a) \leq w_1) = \Pr(w_{I\alpha}(a) \leq w_2); \quad (1.29)
\]

\[
E(B(a)|w_{I\alpha}(a) \leq w_1) = E(B(a)|w_{I\alpha}(a) \leq w_2). \quad (1.30)
\]

Substituting these expressions in (1.27) and (1.28), we see that the expected profit from offering \(w_1 + w_2 \) is

\[
F(w_2) \Pr(w_{I\alpha}(a) \leq w_2)(E[B(a)|w_{I\alpha}(a) \leq w_2] - \frac{w_1 + w_2}{2})
\]

\[
> F(w_2 + \varepsilon) \Pr(w_{I\alpha}(a) \leq w_2 + \varepsilon)(E[B(a)|w_{I\alpha}(a) \leq w_2 + \varepsilon] - w_2)
\]

for small enough \(\varepsilon\). This implies that the outside firm has a profitable deviation, and we have a contradiction. ■

The proof of Lemma 6 is similar to Burdett and Judd (1983): if there is a gap in the support of the distribution, then the outside firm that offers a wage at the top of the gap can deviate by slightly lowering its wage offer. Lemma 4 and Lemma 5 guarantee the probability of hiring the worker is not changed by a slightly lowering of the wage offer, so such deviation is profitable.

Lemma 6, together with the zero expected profit lemma (Lemma 3), implies that the incumbent’s wage offer must equal the average output for workers for all ability levels \(a\) such that \(B(a) \in (w, \bar{w})\). This is stated formally in Lemma 7.

**Lemma 7** If \(B(a) \in (w, \bar{w})\), then \(w_{I\alpha}(a) = B(a)\).

**Proof.** Suppose an outside firm offers a wage \(w \in (w, \bar{w})\). It hires the worker with positive probability. This implies that the firm’s conditional expected profit of hiring the worker is zero (Lemma 3). Since the incumbent’s wage offer is strictly increasing
by Lemma 2 and Lemma 4, the conditional expected profit of the firm is

\[ E[y(a)|w_{I_n}(a) \leq w] - w \]

\[ = E[a|a \leq w_{I_n}^{-1}(w)] - w \]

\[ = B(w_{I_n}^{-1}(w)) - w \]

\[ = 0, \]  \hspace{1cm} (1.32) 

where the last equality follows from Lemma 3.

By Lemma 6, the support of maximum outside offer is dense in \((w, w]\). Therefore, \(B(w_{I_n}^{-1}(w)) = w\) for all \(w \in (w, w]\). Since \(B\) is strictly increasing, this implies that \(w_{I_n}(a) = B(a)\) if \(B(a) \in (w, w]\). \(\blacksquare\)

Lemma 7 determines the incumbent’s offer without assuming that the incumbent’s wage offer is an optimal response. The profit maximization of the incumbent trades off its wage cost with the probability of keeping the worker, which is determined by the distribution of the maximum outside offers. Since Lemma 7 already gives us the solution to this maximization problem \((w_{I_n}(a) = B(a))\), we can use it to back out the equilibrium maximum outside offer distribution \(F\). This is the content of Lemma 8.

**Lemma 8** The equilibrium maximum outside offer distribution satisfies

\[ F(w) = C \exp \left( \int_{w}^{w} \frac{1}{y(B^{-1}(x)) + s(B^{-1}(x)) - x} dx \right) \text{ for } w \in [w, w], \]  \hspace{1cm} (1.33) 

where \(C = \exp \left( - \int_{w}^{w} \frac{1}{y(B^{-1}(x)) + s(B^{-1}(x)) - x} dx \right) \).

**Proof.** Since \(F\) does not have an atom above \(w\) (Lemma 5), the incumbent’s expected profit by offering \(w\) to a worker of ability \(a\) is

\[ (y(a) + s(a) - w)F(w). \]  \hspace{1cm} (1.34) 

Since \(F\) is strictly increasing, it is differentiable almost everywhere so the incumbent’s
profit maximization condition leads to the following first order condition:

\[
\frac{1}{y(a) + s(a) - w_{In}(a)} = \frac{f(w_{In}(a))}{F(w_{In}(a))}.
\]  

(1.35)

Since we have \(w_{In}(a) = B(a)\) for all \(B(a) \in [\underline{w}, \overline{w}]\), the expression above can be written as

\[
\frac{1}{y(B^{-1}(w)) + s(B^{-1}(w)) - w} = \frac{f(w)}{F(w)}.
\]

Integrating this equation, we obtain (1.33) by using that \(F(\overline{w}) = 1\) and \(F\) is right-continuous at \(\underline{w}\).

Lemma 8 pins down the distribution of the maximum outside offer in its support. To completely determine the maximum outside distribution, we need to specify the end points of \(F\). This is done in Lemma 9 and 10. Lemma 9 says that the highest possible outside offer equals the ex ante expected outside output of the worker \((E[y(a)])\).

It is clear that outside firms that offer \(w > E[y(a)]\) would have negative payoffs. The reason why \(\overline{w}\) cannot be lower than \(E[y(a)]\) is identical to that in Lemma 1: if \(\underline{w} < E[y(a)]\), then the highest incumbent offer must also be less than \(E[y(a)]\) since any offer greater than \(\overline{w}\) keeps the worker with probability 1. This leads to a contradiction because now an outside firm can offer a slightly higher wage than the highest incumbent offer to make sure that it hires the worker with probability 1 and thus makes a positive profit (This contradicts the zero expected profit result in Lemma 3).

**Lemma 9** \(\overline{w} = E[y(a)]\).

**Proof.** Outside firms that offer \(w > E[y(a)]\) clearly have negative profits, so we must have \(\overline{w} \leq E[y(a)]\).

Suppose instead \(\overline{w} < E[y(a)]\). Let \(a^* = y^{-1}(\overline{w})\). Since the incumbent’s offer is strongly increasing in worker’s ability (Lemma 2), all workers with ability greater
than $a^*$ must be offered at least $\bar{w}$. Moreover, identical reasoning as in Lemma 1 shows that the incumbent will choose $w_{I_n}(a) = \bar{w}$ for all $a > a^*$.

Now if an outside firm offers $w = \bar{w} + \varepsilon$, it hires the worker with probability 1 and its expected profit is

$$E[y(a)] - \bar{w} - \varepsilon > 0. \quad (1.36)$$

for small enough $\varepsilon$. This is a contradiction to the zero expected profit condition (Lemma 3) of outside firms. ■

Finally, Lemma 10 determines the left-end point.

**Lemma 10** $w = y(0)$.

**Proof.** We prove by contradiction. If $w < y(0)$, then an outside firm can deviate by offering $w + \varepsilon < y(0)$ for some small $\varepsilon$. This outside firm hires the worker with positive probability (by the definition of $w$) and earns positive conditional profit because the wage is lower than the lowest possible output. This violates the zero expected profit of outside firms (Lemma 3).

If $w > y(0)$, let $a_1 = B^{-1}(w) > 0$, so $w_{I_n}(a_1) = w$. Also define $a_2$ as the unique ability level such that: $y(a_2) + s(a_2) = w$. Because $B(a) < y(a)$ for all $a > 0$, it follows that $a_1 > a_2$.

Now we must have $w_{I_n}(a) \geq w$ for $a > a_2$ because otherwise the incumbent keeps such worker with probability zero. On the other hand, we also have $w_{I_n}(a_1) = w$. Therefore, by the monotonicity lemma (Lemma 2), we must have $w_{I_n}(a) = w$ for all $a \in (a_2, a_1]$.

Furthermore, we must have $F(w) > 0$ because otherwise when it offers $w$ its profit is zero. This implies that when an outside firm offers $w$ (which happens with positive probability because $F(w) > 0$), it cannot hire any worker with ability greater than
Therefore, its conditional expected profit is

\[
E[y(a)|a \leq a_2] - w < E[y(a)|a \leq a_1] - w = B(a_1) - w = 0.
\]

Furthermore, the outside firm hires the worker with positive probability because \( w > y(0) \). This leads to a contradiction because all outside firms must have expected profit of zero by Lemma 3. 

Lemma 10 states that \( w \) equals the outside output of the lowest-ability worker. This result is a special case of the more general case where \( s(a) \) may be negative. In the more general case, the left end point is determined by the right-most intersection of the average outside output curve \( (B(a)) \) and the marginal inside output curve \( (y(a) + s(a)) \). Since \( y(a) + s(a) > B(a) \) for all \( a \) in this model, so the two curves do not intersect except possibly at \( a = 0 \). Therefore, \( w \) equals the outside output of lowest-ability worker \( (a = 0) \).

One implication of Lemma 10 is that \( F \) will typically have an atom at \( w \). Moreover, it can be seen that when the atom occurs, the worker must choose the incumbent with probability 1 if there is a tie at \( w \), except possibly when \( s(a) = 0 \). Otherwise, when the incumbent offers \( w \), it keeps with worker with a probability strictly less than \( F(w) \), so the incumbent has a profitable deviation to \( w + \varepsilon \) for small enough \( \varepsilon \). For any wage \( w > w \), the probability that a tie occurs at \( w \) is zero in equilibrium, so any tie-breaking rule of the worker at \( w \) is possible in equilibrium.

Lemma 8-10 completely determine the distribution of the maximum outside offer by specifying the differential equation that governs it (Lemma 8) and its two end points (Lemma 9 and 10). Therefore, the distribution of the maximum outside offer is unique. We state the result formally in Theorem 2.
Theorem 2: In each PBE, all outside firms have zero expected profits in period 2. The incumbent’s offer must satisfy \( w_{1n}(a) = B(a) \) for all \( a \in (0, 1] \), and the distribution of the maximum outside offer must satisfy

\[
F(w) = \begin{cases} 
0 & \text{for } w < w_1(x) \\
C \exp \left( \int_{w_1(x)}^{w} \frac{1}{y(B^{-1}(x)) + s(B^{-1}(x)) - x} dx \right) & \text{for } w \in [w_1(x), w] \\
1 & \text{for } w > w,
\end{cases}
\]

where \([w, \bar{w}] = [y(0), E[y(a)]]\) and \( C = \exp( -\int_{w_1(x)}^{\bar{w}} \frac{1}{y(B^{-1}(x)) + s(B^{-1}(x)) - x} dx) \).

1.4 Job Mobility and Wage Dispersion

We study the implication of the model on job mobility and wage dispersion in this section. In Subsection 4.1, we derive explicit formulas for turnover probability of workers of all ability levels (Proposition 1). We show that the turnover probability is completely determined by a profit ratio that compares the incumbent’s production advantage over the worker’s firm-specific human capital with its informational advantage over the worker’s general human capital. In Subsection 4.2, we obtain explicit formulas for the wage distributions of the stayers, of the movers, and the two types combined. We compare the wage distribution of the stayers with that of the movers and give sufficient conditions on when the average wage of stayers is higher or lower. When there is no firm-specific output, we show that the wage distribution of the stayers is identical to that of the movers, so the average wage of the movers equal that of the stayers.

Before proceeding to the next subsection, we make two remarks here. First, we use the word "workers" in our discussion as if the model had multiple workers. We use the plural form because even if the model has only one worker, the same analysis in Section 3 can be applied to multiple workers when the production has constant return to scale. We make the constant return to scale assumption in this section, so
we can interpret the turnover probability of the worker as the proportion of workers who move to outside firms and $F(w)$ as the proportion of workers whose maximum outside wage offer is less than or equal to $w$. Second, we re-include in this section the technology index $t$ into the expressions and write $y(a, t)$, $s(a, t)$, $B(a, t)$ to study the role of technology on turnover probability and wage distribution.

### 1.4.1 Turnover Probability

We study the turnover probability of the worker in this section. Because the formulas are easier for the probability that a worker stays with the firm, we state the results in staying probability instead.

**Definition 3:** Let $P(a, t)$ be the equilibrium probability that a worker of ability $a$ stays with the incumbent in period 2 when the technology level is $t$.

In the unique equilibrium outcome, the incumbent offers wage $w_{ln}(a, t) = B(a, t)$ to a worker of ability $a$ at the beginning of period 2. This implies that the worker stays with the incumbent if the maximum outside offer he receives is less than $B(a, t)$. This happens with probability $F(B(a, t))$ by the definition of $F$. Therefore, we have

$$P(a, t) = F(B(a, t)).$$

This is the content of Proposition 1.

**Proposition 1:** For all $a$ and $t$, $P(a, t) = \exp(-\int_a^1 \frac{B'(x,t)}{s(x,t)+xB'(x,t)} dx)$.

**Proof.** For a worker with ability $a > 0$, his offer from the incumbent is $B(a, t)$ (Lemma 7). This worker stays with the incumbent if and only if the maximum outside offer he receives is less than $B(a, t)$, which occurs with probability $F(B(a, t))$. In other words, we have

$$P(a, t) = F(B(a, t)).$$

Let $p(a, t) = \frac{\partial P(a, t)}{\partial a}$. Then (1.28) implies that $p(a, t) = f(B(a, t))B'(a, t)$. 

47
By Lemma 8,

\[
p(a, t) = \frac{f(B(a, t)) B'(a, t)}{F(B(a, t))} \frac{B'(a, t)}{y(a, t) + s(a, t) - B(a, t)}
\]

where the last equality uses \( B'(a, t) = \frac{y(a,t) - B(a,t)}{a} \).

Integrating equation (1.39), we obtain that

\[
\ln P(1, t) - \ln P(a, t) = \int_a^1 \frac{B'(x, t)}{s(x, t) + xB'(x, t)} dx \quad \text{for all } a > 0. \tag{1.40}
\]

Since \( P(1, t) = F(B(1, t)) = 1 \), the equation above gives

\[
P(a, t) = \exp\left(- \int_a^1 \frac{B'(x, t)}{s(x, t) + xB'(x, t)} dx \right) \quad \text{for all } a > 0. \tag{1.41}
\]

Finally, for a worker with ability \( a = 0 \), the comment following Lemma 10 implies that \( P(0, t) = P(0+, t) = F(B(0+, t)) \), so the formula remains correct (but the integral maybe improper and should be interpreted as a limit).

The formula of turnover probability in this model differs qualitatively with results in the literature; see for example Greenwald (1986) and Gibbons and Katz (1991). In earlier asymmetric information models, the turnover probability of a worker typically takes a zero-one form: a worker leaves the firm with probability 1 if his ability is below certain threshold and otherwise leaves with the firm with probability 0. Here, the turnover probability decreases continuously with the worker’s ability and only the highest ability worker has a zero turnover probability.

Another feature of the formula is that the turnover probability is completely de-
Let us define \( B'(a,t) \)

\[
B'(a,t) = \frac{s(a,t)}{y(a,t) - B(a,t)}.
\]

then \( r(a,t) \) can be interpreted as a ratio of profits. The numerator equals the worker’s firm-specific output, which is the incumbent’s profit from its production advantage. The denominator is the difference between the marginal and average outside output of the worker. Since the incumbent’s wage offer equals average outside output, the denominator reflects the incumbent’s profit source from its informational advantage on the worker’s general output. This profit ratio of production and information advantage is the key to the comparative statics results.

For example, if the profit ratio in one environment is uniformly larger than in the other, then the turnover probability is uniformly smaller in the environment with the larger ratio. There are two basic cases for the profit ratio to be uniformly larger in one environment than the other. In the first case, the productional advantages are the same in both environments, but the informational advantage is uniformly larger in one environment. Then the turnover probability is larger where the informational advantage is larger. This is stated formally in Corollary 1.

**Corollary 1:** If \( s(a,t_1) = s(a,t_2) \) and \( y(a,t_1) - B(a,t_1) > y(a,t_2) - B(a,t_2) \) for all \( a \), then \( P(a,t_1) < P(a,t_2) \) for all \( a > 0 \).

**Proof.** Note that \( y(a,t) - B(a,t) = aB'(a,t) \). Therefore, \( y(a,t_1) - B(a,t_1) > y(a,t_2) - B(a,t_2) \) for all \( a \) implies that \( B'(a,t_1) > B'(a,t_2) \) for all \( a \). Consequently, when \( s(a,t_1) = s(a,t_2) \), we have \( \frac{B'(a,t_1)}{s(a) + aB'(a,t_1)} > \frac{B'(a,t_2)}{s(a) + aB'(a,t_2)} \) for all \( a \). Proposition 1 immediately implies that \( P(a,t_1) < P(a,t_2) \) for all \( a > 0 \). \( \blacksquare \)

Corollary 1 is closely related to skill-biased technology changes that favor high ability worker over low ability worker. Suppose \( \frac{\partial y(a,t_1)}{\partial a} > 0 \) and the technological level is changed from \( t_2 \) to \( t_1 > t_2 \). It can be checked that this implies \( y(a,t_1) - B(a,t_1) > y(a,t_2) - B(a,t_2) \) for all \( a \) (this is proved in Theorem 4 below). In other words, a technological change that favors high-ability workers over the low-ability workers increases
the difference between marginal and average general output. Therefore, Corollary 1 implies that a skill-biased technology change increases the turnover probability if it does not change the firm-specific output.

The second basic case for the profit ratio to be uniformly larger is when the general outputs are the same in both environments but the specific output is uniformly larger in one. In this case, the environment with the higher firm-specific output has uniformly lower turnover probabilities. This is the content of Corollary 2.

**Corollary 2:** Suppose \( y(a, t_1) = y(a, t_2) \) and \( s(a, t_1) > s(a, t_2) \) for all \( a \), then \( P(a, t_1) > P(a, t_2) \) for all \( a \).

**Proof.** The conditions above imply that \( \frac{B'(a, t_1)}{s(a, t_1) + aB'(a, t_1)} < \frac{B'(a, t_2)}{s(a, t_2) + aB'(a, t_2)} \) for all \( a \). The corollary then follows immediately from Proposition 1. ■

Corollary 2 implies that turnover probability decreases with firm-specific human capital, which is a standard result in labor economics; see for example Hashimoto (1981). But the logic here is different from the existing literature, which is typically cast in a partial equilibrium framework. Hashimoto (1981), for example, obtains the result by assuming that wage offers from outside firms are exogenous and do not respond to the level of firm-specific human capital of workers. In such case, when the firm-specific human capital of a worker increases, the incumbent will raise its offer because its marginal benefit from keeping the worker becomes higher. In contrast, in this model, the incumbent’s offer is independent of the firm-specific human capital level. This result follows from a general equilibrium reasoning, i.e. outside wage offers respond to the level of firm-specific human capital. In this model, the outside wage offers adjust downwards in such a way that the incumbent’s offer remains unchanged.

Although both the partial equilibrium and the general equilibrium models yield the same prediction on job mobility, they can have different implications on wage distributions of the movers and stayers. For example, in this (general equilibrium) model, it is possible for the average wage of the stayers to be lower than that of the movers while typical partial-equilibrium models have the opposite prediction.
Understanding the mover-stayer wage differential is useful because it sheds light on the role of job changes on wage growth of workers, an interesting topic especially given the big cross-country differences in the mover-stayer wage gap; see for example Topel and Ward (1992) and Acemoglu and Pischke (1998). We study the mover-stayer wage gap in the next subsection. It turns out that the profit ratio between firm-specific and general outputs also plays an important role in the wage comparison.

1.4.2 Wage Distribution

In period 2, the worker either leaves the incumbent and becomes a mover, or continues to work for the firm and becomes a stayer. We characterize the wage distributions of the movers, of the stayers, and of the two types combined in this section. We also compare the wage distribution of the movers with that of the stayers and provide sufficient conditions on when one dominates the other in the sense of First Order Stochastic Dominance (FOSD). In the special case where the firm-specific output is missing, we show that the two distributions are identical.

First, we introduce the notations for wage distributions of different types.

Definition 4: Let $G(w)$ be the wage distributions of the stayers and movers combined; let $G_S(w)$ be the wage distribution of the stayers; let $G_M(w)$ be the distribution of the movers.

Proposition 2 derives the wage distributions of the three types. To avoid complicated expressions, the formulas are written in terms of the average outside output ($B(a, t)$) and the staying probability ($P(a, t)$). All of the formulas follow from straightforward reasoning.
Proposition 2:

\[ G(B(a, t)) = aP(a, t); \]  
(1.43)

\[ G_s(B(a, t)) = \frac{\int_0^a P(x, t) dx}{\int_0^1 P(x, t) dx}; \]  
(1.44)

\[ G_M(B(a, t)) = \frac{\int_0^a (1 - P(x, t)) dx}{\int_0^1 (1 - P(x, t)) dx}. \]  
(1.45)

Proof. We first calculate \( G(B(a, t)) \). If a worker’s wage is less than \( B(y(a, t)) \), it must be the case that both the incumbent offer and the maximum outside offer he has received are less than \( B(a, t) \). The probability that the incumbent offer is less than \( B(a, t) \) is \( a \). The probability that the maximum outside offer is less than \( B(a, t) \) is \( F(B(a, t)) = P(a, t) \). Since the outside offers are independent of the incumbent’s offer, this gives that \( G(B(a, t)) = aP(a, t) \).

Next, we calculate the wage distribution of the stayers. Since a worker of ability \( a \) (who is offered \( B(a, t) \) by the incumbent) stays with the incumbent with probability \( P(a, t) \), the total number of stayers who receive less than \( B(a, t) \) is \( \int_0^a P(x, t) dx \). The wage distribution of the stayers in (1.44) follows immediately.

Finally, \( G \) is a weighted average of \( G_s \) and \( G_M \), so that we have

\[ G_s(B(a, t)) \int_0^1 P(x, t) dx + G_m(B(a, t)) \int_0^1 (1 - P(x, t)) dx = G(B(a, t)). \]

This gives the wage distribution of the movers. \( \blacksquare \)

The expressions of wage distributions in Proposition 2 enables us to compare the wage distribution of the movers with that of the stayers. This comparison reflects two conflicting forces. On the one hand, since the maximum outside offer distribution is independent of worker types, a worker is of higher ability if he stays. In other words, stayers on average have higher abilities. On the other hand, for two workers of the same ability (so they receive the same offer from the incumbent), if one worker stays while the other one moves, then the movers must have received a better outside offer. In other words, movers on average have better lucks in receiving wage offers.
Therefore, the comparison of the mover-stayer wage distribution can be thought of as a comparison between abilities and lucks as sources of wage growth.

It turns out that the key variable for the comparison is the profit ratio \( r(a, t) \) in (1.42). Corollary 3 shows that when this ratio is increasing in ability, the wage distribution of the stayers FOSD that of the movers, so the stayers have a higher average wage. And if this ratio is decreasing, the wage distribution of the movers FOSD that of the stayers, so the movers have a higher average wage. In particular, when the firm-specific output is linear in ability, the profit ratio \( r(a, t) \) equals \( \frac{1}{B'(a, t)} \), the inverse of the slope of the average outside output curve. In this case, the stayers have a higher (lower) average wage than the movers if the average outside output curve \( B(a, t) \) is concave (convex).

**Corollary 3:** The wage distribution of the stayers FOSD that of the movers if \( r(a, t) = \frac{s(a, t)}{y(a, t) - B(a, t)} \) is increasing in \( a \); the wage distribution of the movers FOSD that of the stayers if \( r(a, t) = \frac{1}{B'(a, t)} \) is decreasing in \( a \). When \( s(a, t) = sa \), the stayers have higher (lower) average wage if the average output curve \( (B(a, t)) \) is concave (convex).

**Proof.** By Proposition 2, the average wage of the stayers is

\[
\int_0^1 B(a, t) \left( \frac{P(a, t)}{\int_0^1 P(x, t)dx} \right) da. \tag{1.46}
\]

The average wage of the movers is

\[
\int_0^1 B(a, t) \left( \frac{ap(a, t)}{\int_0^1 (1 - P(x, t))dx} \right) da. \tag{1.47}
\]

The likelihood ratio of movers over stayers is

\[
\left( \frac{\int_0^1 P(x, t)dx}{\int_0^1 (1 - P(x, t))dx} \right) \left( \frac{ap(a, t)}{P(a, t)} \right) \tag{1.48}
\]

\[
= \left( \frac{\int_0^1 P(x, t)dx}{\int_0^1 (1 - P(x, t))dx} \right) \frac{1}{s(a, t) \frac{1}{y(a, t) - B(a, t)} + 1} \tag{1.49}
\]
The term in the big parenthesis is a constant. Therefore, if \( \frac{s(a,t)}{B'(a,t)} \) is increasing, this implies that the likelihood ratio is decreasing in \( a \), so the order of monotone likelihood ratio implies the order of FOSD.

When \( s(a,t) = sa \), the likelihood ratio in (1.48) is increasing if \( B'(a,t) \) is increasing in \( a \). The convexity result follows immediately. ■

A positive mover-stayer wage gap corresponds to a positive tenure effect. Corollary 3 suggests that the tenure effect is likely to be positive when the profit ratio is increasing in ability, which may happen if the numerator of the profit ratio, \( s(a,t) \), is increasing very fast in ability. If we believe that more able worker can better leverage their abilities in larger firms because there are more positions and opportunities there, then the firm-specific output is likely to increase faster in larger firms. In this case, Corollary 3 suggests that we are more likely to observe a positive tenure effect in larger firms. If we also believes that the level of the firm-specific human capital is positively correlated with its slope, then we expect to see positive tenure effects in industries where the turnover rates are low.

Finally, the better luck of the movers completely cancels out with the better ability of the stayers when the specific output is absent. Corollary 4 shows that when there is no firm-specific output (\( s(a,t) \equiv 0 \)), the wage distributions of the stayers and the movers are identical and the worker moves to an outside firm with ex ante probability \( \frac{1}{2} \).

**Corollary 4:** If \( s(a,t) \equiv 0 \), then \( P(a,t) = a \) for all \( a \), and \( G_s(w) = G_m(w) \) for all \( w \). Therefore,

\[
\int_0^1 P(a,t)da = \frac{1}{2}. \tag{1.50}
\]

**Proof.** By Proposition 1, \( P(a,t) = \exp(- \int_a^1 \frac{B'(x,t)}{s(x,t) + xB'(x,t)} dx) \). When \( s(a,t) \equiv 0 \),

\[
\frac{B'(x,t)}{s(x,t) + xB'(x,t)} \equiv \frac{1}{x}.
\]
This implies that \( P(a, t) = a \) for all \( a \). Now by Proposition 2, the wage distribution of the stayers and movers combined is

\[
G(B(a, t)) = aP(a, t) = a^2 = \frac{\int_0^a P(x, t)dx}{\int_0^1 P(x, t)dx} = G_s(B(a, t)).
\]

This implies that the movers and stayers have the same wage distribution. Equation (1.50) follows immediately.

The mathematical structure of determining turnover probability and mover-stayer gap is equivalent to that of comparing two lotteries. More precisely, we can think of the incumbent’s wage offers (according to the worker’s ability) as the prizes in the first lottery and maximum outside wage offers as the prizes in the second lottery. Then the turnover probability equals the probability that a draw in the first lottery is smaller than that in the second one. The average wage of the stayers is the average winning draw in the first lottery; the average wage of the movers is the average winning draw in the second lottery. Of course, the prize distribution of the two lotteries is determined by the equilibrium strategies that depend on the underlying production function. Corollary 4 implies that when there is no firm-specific output \( s(a, t) \equiv 0 \), the two lotteries have the identical prize distribution, i.e. the distribution of the incumbent’s offer (according to the worker’s ability) is identical to that of the maximum outside offers. From here, it is easy to see why the wage distributions of the stayers and the movers are identical and why the worker moves to an outside firm with ex ante probability \( \frac{1}{2} \).

### 1.5 Inequality and Job Mobility

This section applies the predictions of the model on turnovers and wage distributions in Section 4 to shed light on the joint evolution of wage inequality and job mobility in the U.S. in the past 30 years. In Subsection 5.1, we review the facts and theories on wage inequality and technological changes. In addition, we discuss why a technological
change that favors general skills may lead to a decrease in job mobility in symmetric learning models. In Subsection 5.2, we show that if technological changes are skill-biased and also favor general skills, this model predicts that the wage distribution will become more spread out (corresponding to greater inequality) and job mobility will rise (Theorem 3 and Theorem 4). Moreover, the proportional increase in job mobility will be larger for older workers (Theorem 5). These predictions are consistent with the empirical findings of Stewart (2002).

1.5.1 Background Facts and Theories

Wage inequality in the U.S. has increased substantially in the past 30 years (Bound and Johnson (1992), Katz and Murphy (1992), Murphy and Welch (1993)). At least half of the increase in wage inequality results from the rise in residual inequality, the dispersion of wages in observationally equivalent groups (Juhn, Murphy, and Pierce (1993)). Moreover, much of the change in within-group inequality appears to concentrate on the top end of the distribution. For example, Lemieux (2006) shows that the within-group inequality has increased for the college-educated and but has changed little for other groups since the 1990s. Autor, Katz, and Kearney (2005) also report that the within-group inequality has continued to rise after the 1990s for the 90/50 wage ratio while it has declined for the 50/10 ratio.

One hypothesis for the cause of this increase in within-group inequality is skill-biased technology change (SBTC). SBTC states that, if we order workers by their abilities, the advancement of technology favors high-ability workers and raises the relative demand for them (Berman, Bound, and Griliches (1994), Autor, Katz, and Krueger (1998)). The computer and internet revolution, together with associated technological and organizational changes, lend direct support to this hypothesis.

It has been recognized, however, that ability is multidimensional (Gardner (1983)). Therefore, ordering workers in a one dimensional ability is an oversimplification that may fail to capture the complexity and the full impacts of technological changes.
(Acemoglu (2002)). For example, different industries and occupations require different mixes of ability types, so a technological change can have different impacts on different sectors (where a sector here can either be an industry or an occupation).

One approach to alleviate this oversimplification is to decompose ability into a general component and a specific component. The general ability contributes to all types of production activities, while the specific ability is useful to production only in a particular firm, industry, or occupation. Casual empiricism suggests that the technological innovation has become more frequent over the years, so production favors workers who are quick to learn and are more flexible, i.e. those with higher general ability. In the same vein, sociologists have argued that aptitude, the general capacity for learning, has become increasingly valuable; see for example Sennett (2003).

In this section, we study the labor market implications of the general-skill-biased technology change (GSBTC) hypothesis, which states that the increase in technology has made general ability more important in production. One version of this hypothesis is put forward by Eric Gould (2002), who examines a two-sector model of comparative advantage a la Roy (1951). Gould proves that wage inequality increases if abilities used in the different sectors become more correlated. Gould also carries out empirical analysis using CPS data by decomposing the whole economy into three sectors (professional, service, and blue-collars). He shows that the ratio of the variance of log-wages within each sector to the variance of log-wage of the whole economy has decreased and the implied abilities in these sectors have become more correlated over the years. These findings point to a diminished role for comparative advantage and an increased role for general ability.

The increased correlation of abilities across sectors naturally affects the job mobility of workers. When abilities required in different sectors become more similar, one might think that mobility between jobs should become larger because skills are now more substitutable. This view focuses on the labor demand side of job changes. If the abilities required by different firms are more substitutable, then a firm with a positive labor demand shock can more easily find workers from other firms. Taking
In this view, Kambourov and Manovskii (2004) argue that the increases in variability of productivity shocks to occupations can explain the increases in between-occupation mobility in the U.S.

It has been argued, however, that an important source of job mobility comes from labor supply reasons that involve firm-worker matching; see for example Jovanovic and Moffitt (1990). By embedding Gould’s idea into a simplified turnover model of Jovanovic (1979), we find that GSBTC decreases, rather than increases turnovers. This is because the increased correlation of abilities between sectors implies that the productivity of the worker is more likely to be similar across sectors, thus the return of switching into a new sector becomes smaller. For example, suppose there are two jobs: farming and hunting. If abilities used in farming and hunting are independent, then a bad farmer is as likely to be a good hunter as a good farmer. Now if abilities in these two sectors become correlated, then a bad farmer is more likely to be a bad hunter. This reduces the incentive for the bad farmer to change jobs so over all job mobility becomes smaller.

It is worthwhile to note that in Jovanovic (1979) and subsequent symmetric information models on turnover, the absolute level of general ability does not affect job mobility. In these models, a worker’s turnover decision is determined by differences in his output at different firms. Because general ability can be taken as a constant common in the output of the worker at all firms, changes in levels of general ability do not affect the relative productivity of the worker in different sectors and thus do not affect the worker’s turnover decision. In contrast, the absolute level of general ability matters in job mobility in this model. As is discussed in Subsection 4.1, the turnover probability in this model depends on the ratio of the firm-specific output to the difference between marginal and average general output, which relates to the absolute level of general ability. In Subsection 5.2, we show that GSBTC decreases this ratio and hence increases turnover and discuss the relevant empirical evidence on turnover.
1.5.2 Implications for Inequality and Job-to-Job Mobility

In this subsection, we show formally how the predictions of our model relate with the empirical findings on wage inequality and job mobility. Theorem 3 below shows that if a technological change is log-skill-biased and general-skill-biased, then job mobility increases and the wage distribution becomes more spread out in the sense of Bickel and Lehmann (1979). Theorem 4 shows that if the technological change does not affect firm-specific output, then the same results in Theorem 3 can be attained by weakening the log-skill-biased assumption to the skill-biased assumption. Theorem 5 shows that if general output is sufficiently important in production, and if a technological change is skill-biased, then the increase in job mobility is larger for workers with higher levels of firm-specific human capital. Before stating these theorems, we first define the different types of technological changes formally.

**Definition 5:** A technological change is skill-biased if the increase in technology raises the output of the higher ability workers more than that of the lower ability workers:

\[
\frac{\partial^2 y(a,t)}{\partial a \partial t} > 0 \quad \text{for all } a \text{ and } t. \tag{1.51}
\]

**Definition 6:** A technological change is log-skill-biased if the increase in technology raises the output of the higher ability workers proportionately more than that of the lower ability workers:

\[
\frac{\partial^2 \log y(a,t)}{\partial a \partial t} > 0 \quad \text{for all } a \text{ and } t. \tag{1.52}
\]

One can show that a log-skill-biased technological change is also skill-biased, so log-skill-biased is a stronger notion. Both of the definitions above compare how technological changes affect the general output at different ability levels. The next definition focuses on how technological changes affect the general component and firm-specific component of the output differently.
Definition 7: A technological change is *general-skill-biased* if an increase in technology raises the general component of output proportionately more than the firm-specific part:

\[
\frac{\partial \log y(a, t)}{\partial t} > \frac{\partial \log s(a, t)}{\partial t} \quad \text{for all } a \text{ and } t.
\] (1.53)

The definitions above introduce different types of technological changes. To state the theorems of how technological changes affect the wage distributions, we next introduce the definition by Bickel and Lehmann (1979) of "spread out", which compares the dispersion of distributions.

Definition 8: A distribution $H_1$ to be more *spread out* than distribution $H_2$ if

\[
H_1^{-1}(q) - H_1^{-1}(q') \geq H_2^{-1}(q) - H_2^{-1}(q') \quad \text{for all } 0 < q' < q < 1.
\] (1.54)

In other words, the distance between the values corresponding to the two quantiles is larger for any two quantiles in the more "spread out" distribution. Alternatively, suppose there is a class of distributions $H(x, t)$, let $x(q, t)$ be the value of the $q$th quantile under index $t$. The distributions are more spread out as $t$ increases if

\[
\frac{\partial^2 x(q, t)}{\partial q \partial t} > 0 \quad \text{for all } q \text{ and } t.
\] (1.55)

A unique feature of "spread out" is that its order is preserved under translation. In other words, if $H_1(x)$ is more spread out than $H_2(x)$, then $H_1(x + t)$ is also more spread out than $H_2(x)$ for all $t$. This stands in contrast with Second Order Stochastic Dominance (SOSD), a popular measure of dispersion, which only compares distributions with the same mean. When two symmetric distributions have the same mean, then "spread out" implies SOSD. For our purpose, technological changes typically affect both the mean and the dispersion of the wage distribution, so we use spread out as the measure of wage inequality.

The major result in this section is that when a technological change is both log-skill-biased and general-skill-biased, turnover probability increases for workers of all
ability levels and the wage distribution becomes more spread out. The intuition of the increased turnover is that the technological change makes the firm specific output relatively less important than the general output, so turnover increases. In particular, recall in Proposition 1 that the turnover probability decreases with \( r(a, t) = \frac{s(y(a,t))}{y(a,t) - B(a,t)} \).

A technological change that is both log-skill-biased and general-skill-biased decreases this ratio and thus increases turnover. On the other hand, the intuition of the increased wage dispersion is that the technological change makes the output distribution more spread out. It is natural to guess that shape of the wage distribution changes in the same direction as the output distribution, and we confirm this conjecture in Theorem 3 using explicit calculation made possible by Proposition 2.

**Theorem 3:** Let the period 2 outside output be \( y(a, t) > 0 \) and the inside output be \( y(a, t) + s(a, t) \), where \( s(a, t) > 0 \). If a technological change is log-skill-biased \( (\frac{\partial^2 \log y(a,t)}{\partial a \partial t}) > 0 \) and general-skill-biased \( (\frac{\partial \log y(a,t)}{\partial t} > \frac{\partial \log s(a,t)}{\partial t}) \), then the turnover probability increases and the wage distribution (of the movers and stayers combined) becomes more spread out as \( t \) increases.

**Proof.** We first show that the turnover probability increases with \( t \). First note that \( aB'(a, t) = ay(a, t) - \int_0^a y(x, t)dx \). Therefore,

\[
\frac{\partial \log (aB'(a, t))}{\partial t} = \frac{\partial \log y(a, t)}{\partial t} \left( \frac{a(\partial y(a, t)/\partial t) - \int_0^a (\partial y(x, t)/\partial t)dx}{ay(a, t) - \int_0^a y(x, t)dx} \right)
\]

\[
= \left( \frac{y(a, t) \int_0^a y(x, t)dx}{ay(a, t) - \int_0^a y(x, t)dx} \right) \frac{\partial y(a, t)/\partial t}{y(a, t)} - \frac{\int_0^a (\partial y(x, t)/\partial t)dx}{\int_0^a y(x, t)dx} < 0,
\]

where the last inequality follows from the log-skilled biased assumption, which implies that \( \frac{\partial y(a, t)/\partial t}{y(a, t)} \) is increasing in \( a \), so \( \frac{\partial y(a, t)/\partial t}{y(a, t)} \) is greater than an average of smaller ratios. The inequality above together with the general-skill biased assumption
\[
\frac{\partial \log y(a,t)}{\partial t} > \frac{\partial \log s(a,t)}{\partial t}
\] implies that
\[
\frac{\partial \log aB'(a,t)}{\partial t} > \frac{\partial \log s(a,t)}{\partial t}.
\] (1.57)

With the inequality above, we see that
\[
\frac{\partial}{\partial t} \left( \frac{aB'(a,t)}{s(a,t) + aB'(a,t)} \right)
= \frac{1}{(s(a,t) + aB'(a,t))^2} \left( \frac{\partial aB'(a,t)}{\partial t} s(a,t) - \frac{\partial s(a,t)}{\partial t} aB'(a,t) \right)
= \frac{aB'(a,t)s(a,t)}{(s(a,t) + aB'(a,t))^2} \left( \frac{\partial \log aB'(a,t)}{\partial t} - \frac{\partial \log s(a,t)}{\partial t} \right)
> 0.
\] (1.58)

This immediately implies that turnover probability increases with \( t \) (by Proposition 1).

Next, we show that the wage distributions become more spread out as \( t \) increases. Let \( w(q, t) \) be the value of the \( q \)th quantile in the wage distribution with technology index \( t \). We need to show that \( \frac{\partial^2 w(q, t)}{\partial q \partial t} > 0 \). Take \( w = B(a, t) \), then Proposition 2 implies that the proportion of workers receiving wage less than \( w \) is \( q(a, t) = a \exp(- \int_a^1 \frac{r(z, t)}{z} \, dz) \), where \( r(a, t) = \frac{aB'(a,t)}{s(a,t) + aB'(a,t)} \). In other words, we have
\[
w(q(a, t), t) = B(a, t),
\] (1.59)
where \( q(a, t) = a \exp(- \int_a^1 \frac{r(z, t)}{z} \, dz) \). This implies that
\[
\frac{\partial w}{\partial q} = \frac{\partial B}{\partial a}
= \frac{aB'(a,t)}{\partial q/\partial a}
= \frac{aB'(a,t)}{a^2 \exp(- \int_a^1 \frac{r(z, t)}{z} \, dz)(1 + r(a, t))},
\] (1.60)
Take derivative with respect to \( t \) to the above expression, we have

\[
\frac{\partial^2 w}{\partial q \partial t} = \frac{\exp(-\int_a^1 \frac{r(z,t)}{z} dz)\{\frac{\partial aB'(a,t)}{\partial t}(1 + r(a,t)) - aB'(a,t)\left[\frac{\partial r(a,t)}{\partial t} - \int_a^1 \frac{\partial r(z,t)}{\partial t} dz\right]\} - \frac{1}{a^2} \exp(-2\int_a^1 \frac{r(z,t)}{z} dz)(1 + r(a,t))^2}{a^2 \exp(-2\int_a^1 \frac{r(z,t)}{z} dz)(1 + r(a,t))^2}.
\]

The denominator of is clearly positive. In the numerator, \( \exp(-\int_a^1 \frac{r(z,t)}{z} dz) > 0 \), and the term in the bracket is

\[
\frac{\partial aB'(a,t)}{\partial t}(1 + r(a,t)) - aB'(a,t)\left[\frac{\partial r(a,t)}{\partial t} - \int_a^1 \frac{\partial r(z,t)}{\partial t} dz\right] = \frac{\partial aB'(a,t)}{\partial t} + aB'(a,t)\int_a^1 \frac{\partial r(z,t)}{\partial t} dz + (\frac{\partial aB'(a,t)}{\partial t} r(a,t) - aB'(a,t)\frac{\partial r(a,t)}{\partial t}) > 0,
\]

where the inequality holds because \( \frac{\partial aB'(a,t)}{\partial t} > \frac{\partial \log s(a,t)}{\partial t} \geq 0 \), \( aB'(a,t) \int_a^1 \frac{\partial r(z,t)}{\partial t} dz > 0 \) (\( \frac{\partial r(z,t)}{\partial t} > 0 \) follows from the mobility part of the proof), and

\[
\frac{(\partial aB'(a,t) r(a,t) - aB'(a,t)\frac{\partial r(a,t)}{\partial t})}{\partial t} = \left(\frac{(\partial aB'(a,t) r(a,t))}{\partial t} r(a,t)\right)^2 = (\frac{\partial aB'(a,t)}{\partial t} + \frac{\partial s(a,t)}{\partial t})^r(a,t)^2 > 0
\]

This shows that the numerator is also positive, so we have \( \frac{\partial^2 w}{\partial q \partial t} > 0 \). Thus the wage distribution becomes more spread out when \( t \) increases. 

The media has suggested that job mobility in the U.S. has increased recently, although evidence from economic research has been less definitive. A special symposium on job stabilities and securities is reported in the Journal of Labor Economics (1999). Evidence from different authors using different data sources suggests that the “short-term” job stability, defined as shares of workers with less than 18 months of tenure (Jaeger and Stevens (1999)), four year retention rate (Neumark, Polsky, and Hansen (1999)), or one year retention rate (Gottschalk and Moffit (1999)), has not
changed much since 1983. On the other hand, the “long-term” job stability, defined as shares of workers with less than ten years of tenure (Jaeger and Stevens (1999)) or as eight year retention rate (Neumark, Polsky, and Hansen (1999)), has decreased somewhat since the 1990s.

Perhaps the more relevant empirical measure of turnover probability for this model is job-to-job mobility, where workers directly work for new firms when they change jobs. Stewart (2002) examines job changes in this direction and classifies a job change into changes that occur from a): employment to employment (EE), b): employment to unemployment (EU), and c): employment to not-in-labor-force-participation (EN). He shows that the EU and EN transition rates in the U.S. have decreased over the years. But most strikingly, the EE transition rate in the U.S. has increased dramatically. According to Stewart (2002), between 1975-2000 the EE transition rate in the U.S. has increased 45% for men and 58% for women.

Theorem 3 analyzes the general case where technological changes affect both general and firm-specific outputs. In some cases, the firm-specific output may be less sensitive to or even independent of the technological change. When this happens $(s(a, t) = s(a)$ for all $a$), Theorem 4 shows that technological changes lead to both increased turnovers and a more spread wage distribution as long as they are skill-biased. Since any skill-biased technology change is naturally general-skill-biased when $s(a, t) = s(a)$, the contribution of Theorem 4 is to weaken the log-skill-biased requirement in Theorem 3.

**Theorem 4:** Let the period 2 output be $y(a, t) > 0$ for the outside firms and $y(a, t) + s(a), (s(a) > 0)$ for the incumbent. If the technological change is skill-biased $(\frac{\partial y(a, t)}{\partial s(a)} > 0)$, then the turnover probability increases and the wage distribution (of the movers and stayers combined) becomes more spread out as $t$ increases.

**Proof.** We first show that the turnover probability increases with $t$. First note
That \( aB'(a, t) = ay(a, t) - \int_0^a y(x, t)dx \). Therefore,

\[
\frac{\partial(aB'(a, t))}{\partial t} = \frac{\partial(ay(a, t) - \int_0^a y(x, t)dx)}{\partial t} = \int_0^a \left( \frac{\partial y(a, t)}{\partial t} - \frac{\partial y(x, t)}{\partial t} \right)dx = \int_0^a \int_x^a \frac{\partial^2 y(z, t)}{\partial z \partial t}dzdx > 0. \tag{1.64}
\]

Since \( s(a) > 0 \), (1.64) implies

\[
\frac{\partial}{\partial t} \left( \frac{aB'(a, t)}{s(a) + aB'(a, t)} \right) > 0.
\]

Therefore, the turnover probability increases with \( t \) by Proposition 1.

To show that increase in \( t \) makes the distribution more spread out, we proceed as in Theorem 3. Again we have

\[
w(q(a, t), t) = B(a, t), \tag{1.66}
\]

where \( q(a, t) = a \exp(-\int_a^1 \frac{r(z, t)}{z}dz) \). The same calculation as in Theorem 3 gives that

\[
\frac{\partial^2 w}{\partial q \partial t} = \exp(-\int_a^1 \frac{r(z, t)}{z}dz) \left\{ \frac{\partial aB'(a, t)}{\partial t}(1 + r(a, t)) - aB'(a, t) \left[ \frac{\partial r(a, t)}{\partial t} - \int_a^1 \frac{\partial r(z, t)/\partial t}{z}dz \right] \right\} \frac{a^2 \exp(-2 \int_a^1 \frac{r(z, t)}{z}dz)(1 + r(a, t))^2}{a^2 \exp(-2 \int_a^1 \frac{r(z, t)}{z}dz)(1 + r(a, t))^2} \tag{1.67}
\]

Again, the denominator is positive. In the numerator, \( \exp(-\int_a^1 \frac{r(z, t)}{z}dz) > 0 \), and the term in the bracket is

\[
\frac{\partial aB'(a, t)}{\partial t}(1 + r(a, t)) - aB'(a, t)\left[ \frac{\partial r(a, t)}{\partial t} - \int_a^1 \frac{\partial r(z, t)/\partial t}{z}dz \right] \tag{1.68}
\]

\[
= \frac{\partial aB'(a, t)}{\partial t} + aB'(a, t) \int_a^1 \frac{\partial r(z, t)/\partial t}{z}dz + (\frac{\partial aB'(a, t)}{\partial t}r(a, t) - aB'(a, t)\frac{\partial r(a, t)}{\partial t}) > 0,
\]

65
where the inequality holds because \( \frac{\partial aB'(a,t)}{\partial t} > 0 \), \( aB'(a,t) \int_a^1 \frac{\partial r(z,t)}{\partial t} dz > 0 \), and

\[
\left( \frac{\partial aB'(a,t)}{\partial t} r(a,t) - aB'(a,t) \frac{\partial r(a,t)}{\partial t} \right)
= \left( \frac{\partial (aB'(a,t))}{\partial t} / r(a,t) \right) r(a,t)^2
= \left( \frac{\partial aB'(a,t)}{\partial t} + \frac{\partial s(a)}{\partial t} \right) r(a,t)^2
> 0.
\]

This shows that the numerator is also positive, so we have \( \frac{\partial^2 w}{\partial t \partial k} > 0 \). Thus, the wage distribution becomes more spread out when \( t \) increases.

The next theorem shows that if general ability is sufficiently important in production, when there is a SBTC, the increase in turnover is larger for workers with higher levels of firm-specific human capital (which positively correlates with worker age). The logic of Theorem 5 can be seen from the benchmark case in Corollary 4 where there is no firm-specific human capital. In this case, the aggregate turnover probability is always \( 1 \). Consequently, for workers with little firm-specific human capital, their average turnover probability is always close to \( 1 \), and thus any technological change cannot have a large effect on the aggregate turnover probability of such workers. When workers have more firm-specific human capital, their aggregate turnover probability is lower, and technological changes can have a larger impact on it.

**Theorem 5**: *Let the output be \( y(a,t) > 0 \) for the outside firms and \( y(a,t) + ks(a) \) for the incumbent, where \( ks(a) > 0 \). If the technological change is skill-biased \( \left( \frac{\partial^2 y(a,t)}{\partial a \partial t} > 0 \right) \) and \( y(a,t) - B(a,t) > ks(a) \) for all \( a \), then the proportionate increase in turnover increases with \( k \), i.e.,

\[
\frac{\partial^2 \log(1 - P(a,t,k))}{\partial t \partial k} > 0 \text{ for all } a.
\]

66
Proof. The proof is based entirely on computation. It is easy to check that
\[
\frac{\partial^2 \log(1 - P(a, t, k))}{\partial t \partial k} = \frac{-\frac{\partial P(a, t, k)}{\partial k}(1 - P(a, t, k)) - \frac{\partial P(a, t, k)}{\partial k}}{(1 - P(a, t, k))^2}.
\]

By Corollary 2 and Theorem 4, we have
\[
\frac{\partial P(a, t, k)}{\partial k} > 0 \quad \text{and} \quad \frac{\partial P(a, t, k)}{\partial t} < 0,
\]
so \(\frac{\partial P(a, t, k)}{\partial t} < 0\). Therefore, it suffices to show that \(\frac{\partial^2 P(a, t, k)}{\partial t \partial k} < 0\) to prove the theorem.

Now since \(P(a, t) = \exp(-\int_a^{1} \frac{B(x, t)}{s(x) + xB'(x, t)} dx)\), it can be checked that
\[
\frac{\partial^2 P(a, t, k)}{\partial t \partial k} = P(a, t, k)^2 \frac{\partial P(a, t, k)}{\partial k} \frac{\partial P(a, t, k)}{\partial t} + P(a, t, k) \frac{\partial^2 \log(P(a, t, k))}{\partial t \partial k},
\]
where the \(\frac{\partial P(a, t, k)}{\partial k} \frac{\partial P(a, t, k)}{\partial t} < 0\), so it suffices to show that \(\frac{\partial^2 \log P(a, t, k)}{\partial t \partial k} < 0\) to prove the theorem. To do so, we check that
\[
\frac{\partial^2 \log P(a, t, k)}{\partial t \partial k} = -\int_a^{1} \frac{ks(x) \frac{\partial B(x, t)}{\partial x} (xB'(x, t) - ks(x))}{(ks(x) + xB'(x, t))^3} dx < 0,
\]
where the inequality follows because \(\frac{\partial B(x, t)}{\partial x} > 0\) (by Theorem 4) and \(xB'(x, t) = y(x, t) - B(x, t) > ks(x)\) (by assumption). This finishes the proof.

The prediction of Theorem 5 is also consistent with the empirical findings of Stewart (2002). He finds that the proportionate increase in employment to employment (EE) transition rate has been strongly increasing with experience even if the absolute EE transition rate decreases with experience. From 1975 to 2000, the proportionate increase in EE rate for men is 23% for the age group 25-35, 79% for the age group of 35-45, and 144% for the age group of 45-55.
1.6 Extension to Multiple Sectors

We extend the model to multiple sectors in this section. Most of the theorems in the one-sector have their natural corresponding results in the multi-sector model so many proofs are omitted. In Subsection 6.1, we set up the multi-sector model and show the existence and uniqueness of the equilibrium. In addition, we show that a): the worker's sectoral choice in period 1 is efficient and b): there is an inter-industry wage differentials in this model. In Subsection 6.2, we examine a two-sector model in detail. First, we continue the study on sectoral choice and show that when two sectors have the same expected output, in equilibrium the worker always chooses in period 1 the sector with a steeper slope to ability. Second, we derive closed-form solutions for both between-sector and within-sector turnover probabilities. This enables us to have a more detailed look at the impact of technological changes on job mobility and the wage distribution.

1.6.1 Multi-Sector Model: Setup and General Results

Setup

The multi-sector model is identical to the one-sector model except in the composition of the firms. As before, the model has two periods. There is a single worker with unknown ability a, which is drawn uniformly from $[0, 1]$. Now there are $i = 1, 2, ..., I$ sectors. Each sector i has $N_i$ firms, with $2 < N_i < \infty$.

Production takes place when the worker is matched with a firm. The output of the worker in period 1 is $y_i(a, t)$ if he is in sector i. In period 2, the worker's output in sector i and firm j is

\[
Y_{ij} = y_i(a, t) + s_i(a, t) + s_{ij}(a, t) \quad \text{if with the incumbent;}
\]

\[
= y_i(a, t) + s_i(a, t) \quad \text{if in an outside firm, same sector;}
\]

\[
= y_i(a, t) \quad \text{if in a different sector,}
\]

where we assume that $y_i(a, t) \geq 0, s_i(a, t) \geq 0$, and $s_{ij}(a, t) \geq 0$ for all $a, t, i, and
The term $s_t(a, t)$ reflects the view that some of the human capital the worker accumulates is sector-specific (Neal (1995), Parent (2000), Kambourov and Manovskii (2005)), where a sector here can be interpreted either as an industry or as an occupation. To simplify notation, we define $y_{jn}(a, t) = y_t(a, t) + s_t(a, t) + s_{ij}(a, t)$ as the worker's output with the incumbent in period 2.

The timing and information structure are identical to the one-sector model. As before, the ability of the worker is not known in period 1. Moreover, only the incumbent observes the ability of the worker at the beginning of period 2 and outside firms observe nothing. It is possible to enrich the information structure to allow for the possibility that firms in the same sectors have better information about the workers than firms in different sectors.

The strategies of the worker and the firms are also defined just as in the one-sector model. The strategy of the worker is $(D_1, D_2)$, where $D_i$ is the worker's choice of firms in period $i = 1, 2$. The strategy of firm $j$ in sector $i$ is a triple $(G_{ij}, F_{ij}, w_{ij})$, where $G_{ij}$ is the distribution of firm $ij$'s period 1 wage offer, $F_{ij}$ is the distribution of firm $ij$'s period 2 wage offer if it isn't the incumbent, and $w_{ij}$ is its period 2 wage offer if it is the incumbent. We solve the PBE of the game.

Existence and Uniqueness

The existence and uniqueness results (Theorems 1 and 2) in the one-sector model have a natural correspondence in the multi-sector model. Similar to the one-sector model, we define the average outside output in sector $i$ in period 2 as

$$B_i(a, t) = \begin{cases} \int_0^a y_t(x, t)dx & \text{if the incumbent is not in sector } i; \\ a & \text{if the incumbent is in sector } i. \end{cases} \quad (1.70)$$

---

7In contrast to the one-sector model, the outcome of the model would be different if we assume instead that the worker knows his ability. The two major differences are, first, the sectoral choice depends not only on the wage offered but also on his own ability. Second, the sector choice of the worker in period 1 signals the ability of the worker and affects the wage offer of firms in period 2.
Note that $B_i(a, t)$ depends on the worker's period 1 sector: it is larger if the incumbent is in sector $i$. Now define the average outside output, the upper-envelope of the average output of all sectors, as

$$B(a, t) = \max_{i \in \{1, \ldots, I\}} \{B_i(a, t)\}. \quad (1.71)$$

In the multi-sector model, we can construct a mixed strategy equilibrium in the same way as in the one-sector model. As before, the constructed equilibrium has two key features. First, the incumbent offers a worker of ability $a$ in period 2

$$w_{In}(a, t) = B(a, t). \quad (1.72)$$

Using the same reasoning as in Theorem 1, we can show that the incumbent's offer implies that all outside firms must have non-negative payoffs. In particular, outside firms in sector $i$ have exactly zero expected profit if their wage offers are in the range where the average outside output of sector $i$ is the maximum of all sectors, i.e. $w \in \{B_i(a, t) : B_i(a, t) = B(a, t)\}$. And outside firms in sector $i$ would have a negative payoff if it offers a wage $w$ such that $w \notin \{B_i(a, t) : B_i(a, t) = B(a, t)\}$. This implies that given the incumbent's offer, all outside firms in sector $i$ are willing to randomize in the wage range $\{B_i(a, t) : B_i(a, t) = B(a, t)\}$.

Second, the randomization of outside firms induces a maximum outside offer distribution that satisfies

$$\frac{1}{y_{In}(a, t) + B(a, t)} = \frac{f(B(a, t))}{F(B(a, t))} \quad \text{for all } a > 0, \quad (1.73)$$

where $y_{In}(a, t)$ is the output of the worker with the incumbent and $F(w) = \prod_{i=1}^{I} \prod_{j=1, j \neq i'}^{N_i} F_{ij}(w)$ is the distribution of the maximum outside offers. This differential equation is very similar to (1.11) in the one-sector model, and identical reasoning shows that under the distribution in (1.73) the incumbent would find it optimal to offer $w_{In}(a, t) = B(a, t)$.

Theorem 6 formally states the existence of a PBE. Since the proof of Theorem 6 is almost identical to that of Theorem 1, we omit it here.
Theorem 6: There exists a PBE such that in period 2 the incumbent offers \( w_{In}(a) = B(a,t) \). The maximum outside offer distribution satisfies

\[
\frac{f(w)}{F(w)} = \frac{1}{y_{In}(B^{-1}(w), t) - w}, \text{ for all } w \in [\underline{w}, \overline{w}],
\]

(1.74)

where \( \overline{w} = B(1,t) \) and \( \underline{w} = \max\{B(0,t), \max\{B(a,t) : B(a,t) \geq y_{In}(a,t)\}\} \).

As in the one-sector model, the multi-sector model has infinitely many PBE. And just as in the one-sector model, all the PBEs must have the same two features: 1): the incumbent offers wage equals average outside output, and 2): the maximum outside offer distribution satisfies (1.73). This is the content of Theorem 7, which is a natural correspondence to Theorem 2. The proof of Theorem 7 is almost identical to that of Theorem 2. Both theorems can be proved by using the same sequence of lemmas (Lemma 3-10). Furthermore, all lemmas can be proved in the same way except for the zero expected profit lemma (Lemma 3). But the proof of the zero expected profit lemma requires some modifications.

Lemma 11 The expected profit of each outside firm is zero.

Proof. We distinguish two cases.

Case 1: \( F(\underline{w}) = 0 \).

Since \( F(\underline{w}) = 0 \) and \( F \) is right-continuous, for any \( n > 0 \), there exists an outside firm offering \( w_n > \underline{w} \) such that \( F(w) < \frac{1}{n} \).

By offering \( w_n \), the outside firm has an expected profit less than

\[
F(w_n)(E[B(a)|w_{In}(a) \leq w_n] - w_n) \leq \frac{1}{n}(E[B(1)]).
\]

(1.75)

Let \( n \) goes to infinity, the right hand side of the inequality goes to zero. Since we have only finite sectors, this implies that at least one sector has profit zero.

Now if there is an outside firm that has positive profit in equilibrium, its wage offers must be larger than \( \underline{w} + \varepsilon \) for some \( \varepsilon > 0 \). Otherwise, there will be offers from the firm that lead to profits arbitrarily close to zero. However, since all offers from
the firm is greater than \( w + \varepsilon \), this violates the definition of \( w \), which is the infimum of wage offers that hires the worker with positive probability.

Case 2: \( F(w) > 0 \).

Identical arguments as in Lemma 3 show that the firm that offers \( w \) must have zero profit. Now if there is an outside firm that has positive profit, any of its wage offer must be larger than \( w \). This implies that the maximum outside offer must be always higher than \( w \), which contradicts the assumption that \( F(w) > 0 \).

We now state the uniqueness theorem formally. Since the proof of Theorem 7 is identical to Theorem 2, we omit it here.

**Theorem 7:** In each PBE, the incumbent must offer \( w_{1n}(a) = B(a, t) \) for all \( a \in (B^{-1}(w), B^{-1}(\bar{w})) \) in period 2. The maximum outside offer distribution must satisfy

\[
\frac{f(w)}{F(w)} = \frac{1}{y_{1n}(B^{-1}(w), t) - w} \quad \text{for all } w \in [w, \bar{w}],
\]

where \( w = \max\{B(0, t), \max\{B(a, t) : B(a, t) \geq y_{1n}(a, t)\}\} \), and \( \bar{w} = B(1, t) \).

Theorem 7 implies that the multi-sector model has a unique equilibrium outcome in terms of turnover probability and the wage distribution. We discuss these in more detail in a two-sector model in Subsection 6.2. For the rest of the section, we discuss two other implication of the theorem. First, Theorem 7 implies that the period 1 sectoral choice of the worker is ex ante efficient. This is because the zero expected profit result in period 2 (Lemma 11) together with the Bertrand-type competition in period 1 imply that each firm must have zero expected profit in equilibrium. Therefore, all surpluses go to the worker, who will thus choose in period 1 a sector that maximizes the total ex ante expected surplus. We state this result formally in the following corollary. Since the proof follows exactly from the discussion above, we omit it here.

**Corollary 5:** The worker's sector choice in period 1 is ex ante efficient.

The second implication of the uniqueness theorem is that the randomization of outside firms gives an explanation for the inter-industry or inter-occupation wage
differentials. This is because when an outside firm offers a wage \( w = B_i(a, t) \), its expected payoff conditional on hiring the worker is \( B_i(a, t) - B(a, t) \leq 0 \). Therefore, outside firms in sector \( i \) would only offer wages in \( w \in \{ B_i(a, t) : B_i(a, t) = B(a, t) \} \). Since firms in different sectors randomize in different ranges, this leads to inter-industry or inter-occupation wage differentials. Corollary 6 states this formally.

**Corollary 6:** Suppose there exists \( w = t_1 < t_2 < \ldots < t_N < t_{N+1} = \bar{w} \) such that \( B_i(a, t) = B(a, t) \) if and only if \( B_i(a, t) \in [t_i, t_{i+1}] \). In this case, firms in sector \( i \) only offer wage in \( [t_i, t_{i+1}] \) or below \( w \). The average wage and worker ability in sector \( i \) increase with \( i \).

The condition of Corollary 6 can be satisfied in a number of scenarios. For example, it can occur if the slopes of the outputs (with respect to ability) in different sectors \( (y_i(a, t)) \) are ordered uniformly. Different from the existing literature, the source of wage differences in this model is adverse selection, which forces firms in different sectors to offer wages in different ranges. Our explanation yields predictions consistent with the findings that sectors more sensitive to ability hire more able workers and pay higher wages; see for example Gibbons Katz, Lemieux, and Parent (GKLP) (2005). Different from GKLP (2005), firms in this model are unwilling to lower their wage offers even if they could hire a worker with the lower wage.

### 1.6.2 A Two-Sector Model

In this subsection, we examine the sector choice, job mobility, and wage distribution in a two-sector model. First, we show that when the output in one sector is uniformly more sensitive to ability than the other sector, the worker chooses in equilibrium the more "sensitive" sector if the average output in that sector is at least as high as the other one. In this case, the within-sector movers have a higher average wage than both the stayers and between-sector movers if sector-specific and firm-specific human capital is absent. Second, we give explicit formulas for staying and moving probabilities. This allows us to examine how technological changes affect both job mobility and
the wage distribution. In particular, a skill-biased technology change increases both
the within-sector and total job mobility. Furthermore, its impact on between-sector
mobility is larger for workers with higher levels of sector-specific human capital. For
workers with sufficient sector-specific human capital, their aggregate between-sector
mobility also increases. These predictions are consistent with the findings of Kam-
bourov and Manovskii (2005).

When there are two sectors with equal average outputs, it appears that the worker
should always choose the sector with a steeper slope (of output in ability) because
there is an option value of having high ability in the steeper sector. In this model,
however, asymmetric information restricts the mobility of the worker. In particular,
by choosing a steeper sector in period 1, the worker bears the cost of being more
likely to stay in the steeper sector if his ability happens to be low. It is unclear ex-
ante whether the option value exceeds this cost. Nevertheless, Proposition 3 below
suggests that when sector-specific and firm-specific human capital accumulation is
less important, the worker will prefer the steeper sector in period 1. Furthermore,
Proposition 3 also implies that the average wage of the within-sector movers are
higher than the average wage of both the stayers and between-sector movers. Since
this result is independent of the technology index, we omit $t$ in its statement.

**Proposition 3:** If $y'_1(a) < y'_2(a)$, $s_i(a) = s_{ij}(a) = 0$ for all $a \in [0,1]$, $i = 1, 2$ and
$E[y_1(a)] \leq E[y_2(a)]$, then each worker chooses sector 2 in period 1. There are three
types of workers in period 2: stayers (who stay with the incumbent), within-sector
movers, and between-sector movers. In period 2, the within-sector movers have the
highest average wage among the three groups.

**Proof.** Without loss of generality, we may assume that $B_1(0) = y_1(0) < y_2(0) <
B_2(0)$. Because otherwise, the output is higher in sector 2 for all ability levels and
the Proposition is trivial. Note that $E[y_1(a)] \leq E[y_2(a)]$ implies that $B_1(1) \leq B_2(1)$.
If $y'_1(a) < y'_2(a)$ for all $a$, it can be checked that $B'_1(a) < B'_2(a)$ for all $a$. Because
$B_1(0) > B_2(0)$, let us define $a^\ast$ as the unique ability level where

$$B_1(a^\ast) = B_2(a^\ast).$$  \hfill (1.77)

By choosing (a firm in) sector 1 in period 1, the worker's expected total outputs in the two periods are

$$E[y_1(a)] + \int_0^{a^\ast} P_1(a^\ast)y_1(a) + (1 - P_1(a^\ast))y_2(a) \, da + \int_{a^\ast}^1 P_1(a)y_1(a) + (1 - P_1(a))y_2(a) \, da,$$

where $P_1(a)$ is the probability that a worker with ability stays with the incumbent, the subscript refers to the fact that the worker is in a sector-1 firm in period 1. By choosing sector 2 in period 1, the worker's expected total outputs in the two periods are

$$E[y_2(a)] + \int_0^{a^\ast} (P_2(a^\ast) - P_2(a))y_1(a) + (1 - P_2(a^\ast) + P_2(a))y_1(a) \, da + \int_{a^\ast}^1 y_2(a) \, da.$$  \hfill (1.79)

Comparing these two terms above, we find that $E[y_2(a)] > E[y_1(a)]$;

$$\int_{a^\ast}^1 y_2(a) \, da > \int_{a^\ast}^1 P_1(a)y_1(a) + (1 - P_1(a))y_2(a) \, da,$$

because $y_2(a) > y_1(a)$ for $a > a^\ast$, and finally the difference of the two middle terms is

$$\int_0^{a^\ast} (P_2(a) + P_1(a^\ast) - P_2(a^\ast))(y_2(a) - y_1(a)) \, da$$

$$= \int_0^{a^\ast} (B_1(a) - B_2(a))aP_2(a) \, da$$

$$> 0,$$

where the equality follows by integration by parts and using that $B_1(a^\ast) = B_2(a^\ast)$. This implies that the total surplus is higher if the worker chooses sector 2 in period 1. Since the worker captures all of the surplus, he will choose sector 2 is period 1.

To compare the average wages, first note that the average wage of within-industry movers is clearly higher than the between-industry movers. The average wage of the
stayers and of the within-industry movers are

\[ W_s = \frac{\int_0^1 B(a)P(a)da}{\int_0^1 P(a)da}; \quad (1.82) \]

\[ W_{MI} = \frac{\int_a^1 B(a)ap(a)da}{\int_a^1 P(a)ap(a)da} = \frac{\int_a^1 B(a)P(a)da}{\int_a^1 P(a)da}. \quad (1.83) \]

Since \( B(a) \) is strictly increasing, it is easy to see that \( W_{MI} > W_s \).

It is well-known that wage losses of laid-off workers who switch industries are larger than those who remain in the same industry (Neal 1995). Proposition 3 points to a similar prediction for job-to-job movers. It also suggests that the average wage of within-sector movers is higher than that of the stayers if firm-specific and sector-specific human capital is less important. This is a new prediction that can be taken to data, which I hope to do in future work.

Next, we analyze how technological changes affect job mobilities and wage distributions. In general, technological changes affect the outputs in different sectors differently. Here, we consider a special case in which the relative difference in outputs between sectors is independent of the technological changes. In particular, suppose the output in sector 2 is uniformly more sensitive to ability than that in sector 1 and the worker is in sector 2 in period 1. The output of the worker in period 2 can then be written as

\[ Y = y(a,t) \quad \text{if in sector 1;} \]

\[ = y(a,t) + d(a) + I(a) \quad \text{if in sector 2;} \quad (1.84) \]

\[ = y(a,t) + d(a) + I(a) + s(a) \quad \text{if with the incumbent,} \]

where \( d(a) \) is the difference in general output between sector 2 and sector 1, \( I(a) \) reflects the sector-specific capital in sector 2, and \( s(a) \) reflects the firm-specific output. We assume that \( I'(a) \geq 0 \) and \( s'(a) \geq 0 \) for all \( a \) so that the sector and firm

\[ ^6 \text{We choose the worker to be in sector 2 because Proposition 3 states that the workers choose into a sector with steeper slope in ability when the sector also has higher average expected output.} \]
specific human capitals are nondecreasing in the worker's ability. Note that there is no technology index in $d, I, \text{ and } s$ because the relative difference in outputs between sectors is independent of the technological changes. We also have $d'(a) > 0$ for all $a$ because output in sector 2 is uniformly more sensitive to ability.

This specification of output functions induce the average outside outputs of the two sectors as follows:

$$B_1(a, t) = \frac{\int_0^a y(x, t) dx}{a}$$  \hspace{1cm} (1.85)

$$B_2(a, t) = \frac{\int_0^a y(x, t) + d(x) + I(x) dx}{a}$$  \hspace{1cm} (1.86)

Since $d'(a) > 0$ and $l'(a) \geq 0$, the slope of average outside output in sector 2 is increasing faster than that in sector 1. Therefore, there exists a unique ability level $a^*$ (satisfying $\int_{a^*}^\infty d(x) + I(x) dx = 0$) such that $B_1(a, t) < B_2(a, t)$ for $a < a^*$ and $B_1(a, t) > B_2(a, t)$ for $a > a^*$. It follows that the the average outside output $B(a, t)$, which is the upper-envelop of the two, satisfies

$$B(a, t) = B_1(a, t) \quad \text{for } a < a^*$$

$$= B_2(a, t) \quad \text{for } a \geq a^*$$  \hspace{1cm} (1.87)

The characterization of the average outside output enables us to write explicit formulas for the probability of staying, of within-sector moving, and of between-sector moving in period 2, which is reported in Proposition 4. These formulas are obtained in the same way as in the one-sector case. The computation of turnover probabilities is greatly simplified by the fact that outside firms in different sectors do not overlap in their ranges of wage offers. For example, firms in sector 1 only offer wages $w \in \{B(a, t) : a \leq a^*\}$ and firms in sector 2 offer wages $w \in \{B(a, t) : a \geq a^*\}$. Note that since the incumbent offers $w_{in}(a, t) = B(a, t)$, so workers with ability greater than $a^*$ never moves to sector 1 in this two-sector model. Before stating Proposition 4, we introduce some notations first.
Definition 8: Let $P_s(a, t)$ be the probability that the worker stays with the incumbent in period 2. Let $P_{MW}(a, t)$ be the probability that the worker moves to an outside firm within the same industry. Let $P_{MB}(a, t)$ be the probability that the worker moves to an outside firm in a different industry.

Proposition 4: Suppose the worker is in sector 2 in period 1. If a worker has ability $a > a^*$,

$$P_s(a, t) = \exp(-\int_a^1 \frac{B'(x, t)}{y_{in}(x, t) + xB'(x, t)} dx),$$

$$P_{MW}(a, t) = 1 - \exp(-\int_a^1 \frac{B'(x, t)}{y_{in}(x, t) + xB'(x, t)} dx),$$

and

$$P_{MB}(a, t) = 0$$

If a worker has ability $a < a^*$,

$$P_s(a, t) = \exp(-\int_a^1 \frac{B'(x, t)}{d(x) + y_{out}(x, t) + zB'(x, t)} dx),$$

$$P_{MW}(a, t) = 1 - P(a^*, t),$$

and

$$P_{MB}(a, t) = P(a^*, t)(1 - \exp(-\int_a^{a^*} \frac{B'(x, t)}{d(x) + y_{out}(x, t) + zB'(x, t)} dx))$$

Proof. Just as the one-sector case, Theorem 6 implies that the probability the worker stays with the incumbent is

$$\exp(-\int_a^1 \frac{B'(x, t)}{y_{in}(x, t) - y_{out}(x, t) + xB'(x, t)} dx),$$

(1.90)

where $y_{out}(a, t) = y(a, t)$ for $a < a^*$ and $y_{out}(a, t) = y(a, t) + d(a) + I(a)$ for $a > a^*$. This immediately leads to the statements in the proposition.

Given the turnover probabilities, we can analyze how technological changes affect job mobilities and the distribution, taking into account that some job changes happen within the sector and others happen between the sectors. We show that the intuition of the one-sector model continues to hold in the multi-sector model. For a skill-biased technological change that does not change the relative outputs of the two sectors, it increases both within-sector and aggregate turnover probabilities. It also makes the wage distribution more spread out.
Theorem 8: Suppose $\frac{\partial^2 B(a,t)}{\partial a \partial t} > 0$ and $d(0) + I(0) + s(0) \geq 0$. Then an increase in $t$ increases the turnover probability of all ability levels and the aggregate within-sector turnover. It also makes the wage distribution of the economy more spread out.

Proof. Identical argument as in Theorem 4 gives that

$$\frac{\partial^2 B(a,t)}{\partial a \partial t} > 0.$$ (1.91)

This implies that $\frac{\partial(B(a(t))}{d(a)+s(a))}/dt > 0$ for all $a > a^*$ and $\frac{\partial (B_1(a(t))}{d(x)+s(x)} > 0$ for all $a < a^*$. Therefore, by Proposition 4, $P_s(a,t)$ is decreasing in $t$ for all $a$ and thus the probability of turnover increases for all ability levels.

Now the aggregate within-sector turnover probability is

$$P_{MW} = \int_0^{a^*} (1 - P_s(a^*,t))da + \int_{a^*}^{1} (1 - P_s(a,t))da.$$ (1.92)

Since $P_s(a,t)$ decreases with $t$, it is clear that the aggregate within-sector turnover probability increases.

The proof on the wage distribution of the economy becomes more spread out is identical as that in Theorem 4 and is omitted. □

The predictions of Theorem 8 is consistent with the empirical findings. For example, Kambourov and Manovskii (2005) reports that the occupational mobility in the U.S. has increased from 10%-15% (per year) at the one digit level, 12%-17% at the two digit level, and 16%-20% at the three digit level between 1968 and 1997. The corresponding increase in industry mobility is 7%-12% in one digit level, 8%-13% in two digit level, and 10%-13% in three digit level. It is worth noting that these numbers include any types of job change, and a more relevant measure is job-to-job mobility.

Theorem 8 does not address how technological changes affect between-sector job mobility. In general, we cannot make definite statements about changes in between-sector inequality without imposing extra assumptions. Nevertheless, Theorem 9 shows that change in between-sector turnover probability increases with the level of sector-
specific human capital. Moreover, the aggregate between-sector turnover probability increases when the level of sector-specific human capital is high enough. These results are stated formally in Theorem 9.

**Theorem 9:** Consider a family of sector-specific outputs \( I(a, k) = I_0(a) + kh(a) \). Suppose \( h(a) = 0 \) for \( a < a^* \) and \( h'(a) > 0 \) for \( a > a^* \), where \( a^* \) satisfies \( \int_{a^*}^a I_0(a) + s(a) \, da = 0 \). If \( \frac{\partial^2 P(a, t)}{\partial a \partial t} > 0 \) and \( d(0) + I(0) + s(0) \geq 0 \), then \( \frac{\partial^2 \log \text{APMB}(k, t)}{\partial a \partial k} > 0 \), where \( \text{APMB}(k, t) \) is the aggregate between-sector turnover probability. Furthermore, \( \frac{\partial \log \text{APMB}(k, t)}{\partial t} > 0 \) for large enough \( k \).

**Proof.** The proof is purely computational. The aggregate between-sector turnover probability is

\[
\text{APMB}(k, t) = \int_a^a P_S(a^*, t, k) - P_S(a, t, k) \, da
\]

The term \( \int_a^a 1 - \exp\left( - \int_a^{a^*} \frac{B'(x, t)}{f(x) + s(x) + I_0(x) + xB'(x, t)} \, dx \right) \, da \) is independent of \( k \). Therefore, it suffices to show that \( \frac{\partial^2 \log P(a^*, k, t)}{\partial a \partial k} > 0 \).

Now

\[
P(a^*, t, k) = \exp\left( - \int_{a^*}^1 \frac{B'(x, t)}{s(x) + xB'(x, t, k)} \, dx \right),
\]

so it can be checked that

\[
\frac{\partial^2 \log(P(a^*, t, k))}{\partial t \partial k} = s(a^*) \left[ \frac{2((s(a^*) + a^*B'(a^*, t, k))a^*(h(a^*) - \int_0^{a^*} h(x) \, dx) \frac{\partial B(a^*, t, k)}{\partial a})}{(s(a^*) + a^*B'(a^*, t, k))^4} > 0, \right.
\]

because \( \frac{\partial^2 B(a^*, t, k)}{\partial a \partial t} > 0 \) (by Theorem 4) and \( h(a^*) - \int_0^{a^*} \frac{h(x) \, dx}{a^*} > 0 \) (which follows form \( h'(a) > 0 \)).

This finishes the proof that \( \frac{\partial \log \text{APMB}(k, t)}{\partial t} > 0 \).

Finally, as \( k \) goes to infinity, \( \frac{\partial \log(P(a^*, t, k))}{\partial t} \) goes to zero, so the aggregate between-sector turnover increases.

The predictions of Theorem 9 is consistent with empirical findings. Kambourov
and Manovskii (2005) report significant increases in between-sector (both industry and occupation) mobility at both 1-digit and 2-digit industry and occupation levels for workers with at least 12 years of education. Moreover, the increase in between-sector mobility is increasing with the age of the workers. Interesting enough, they also report that there have been no significant changes in between-sector mobility for workers with less than 12 years of education. This would be consistent with our interpretation if one believes that abilities of more educated workers are harder to observe and hence these workers are subject to a more severe adverse selection problem.

1.7 Conclusion and Discussion

This paper develops a model of job mobility and wage distribution. Worker ability affects both firm-specific and general productivity. Young workers decide whether to quit or remain with their current employer after information about their abilities is revealed. Employers learn more about the abilities of the young workers than do outside firms. The superior information of current employers creates a standard lemons problem in the second-hand labor market. Contrary to existing work, the current paper establishes that this lemons problem does not lead to market collapse. Instead, there exists a unique equilibrium outcome in which the current employer offers a wage equal to the average output of all types below the ability of the worker and outside firms compete for labor services of the worker by using mixed strategies. These mixed strategies lead to a non-degenerate wage distribution for all types of workers. This unique equilibrium outcome determines both the allocation of workers with heterogeneous abilities to different firms and also how wages change when workers change jobs (due to both selection and productivity effects).

In addition to providing a new model of turnover, selection, and wage distribution, the paper sheds light on the joint evolution of wage inequality and job mobility in the United States over the past 30 years. The model implies that in the presence of technological change that is skill-biased and also favors general skills over firm-specific
skills, the wage distribution will become more spread out (corresponding to greater inequality) and both overall and sector job mobility will increase. The model also suggests that mobility should increase more for older workers. These patterns are consistent with recent empirical evidence on changes in job mobility in the United States.

We also extend the model to a multi-sector setting that delivers three sets of results. First, we show that the worker's sector choice in period 1 is ex ante efficient. In particular, the worker is more likely to choose a sector whose output is more sensitive to ability in period 1. Second, we give a new explanation for inter-industry wage differentials based on adverse selection. Third, we discuss issues of wage and job changes depending on whether the changes are within-sector or between-sector.

The model points to a few new results that can be tested empirically. For example, it predicts that the wage change is larger for within-industry movers than between-industry movers. Moreover, our model also predicts that the average wage of within-sector movers is higher than that of the stayers if firm-specific and sector-specific human capital is less important. This suggests that the tenure effect may be increasing with the experience level of workers.

Finally, the model is a general framework that may help understand and analyze various labor market outcomes. For example, our model can be applied to study training decisions of firms. Previous studies have not looked at how training decisions are affected by the output distribution of workers, while our model is very suited to examine such issues. In particular, it can be shown that training in general human capital may affect turnover and firms may overinvest in general human capital.
Bibliography


Chapter 2

A Theory of Wage Distribution and Dynamics with Matching and Pareto Learning

This paper develops a model of the wage distribution and wage dynamics based on matching and Pareto learning. In particular, we analyze learning about ability and matching into a hierarchy of sectors with different returns to ability. In a competitive labor market, information about a worker’s ability is revealed publicly through Pareto learning. Workers with higher expected abilities are employed in sectors with higher returns to abilities. With Pareto learning, a worker’s expected output is completely determined by his maximum output in the past and his length of experience in the labor market. This feature enables us to explicitly compute the wage distribution of each cohort, which is shown to follow a generalized Pareto distribution with increasing variance in experience.

Our model contains both pure learning and pure matching models as limiting cases, so the model reproduces the predictions of learning models and the predictions of matching models. The model also generates results from the interaction of learning and matching. First, we prove that the expected wage increase is proportional to the current wage level plus a constant, which is similar to the standard wage regression
specification. Numerical results using the formula suggest that better matching due to improved information contributes significantly to wage growth, especially in the early careers of the workers. Second, we prove that changes in wage residuals are serially correlated. Our model predicts that positive serial correlation in wage changes is more likely to be found among younger workers or in industries with increasingly convex wage schedules. These results help reconcile the conflicting findings that serial correlations are found in small, relatively homogeneous samples and not in large, heterogeneous samples.

2.1 Introduction

This paper develops a model of the wage distribution and wage dynamics based on matching and Pareto learning. The wage distribution and wage dynamics are subject to a common set of forces. Information about the abilities of workers is imperfect and new information arrives over time. In addition, match qualities between workers and firms are uncertain and mismatches can occur. In response to the new information about ability and the potential for mismatch, workers frequently change jobs, especially in the earlier stage of their careers. And, at the aggregate level, while each worker experiences his wage changes through job-hopping, their collective effects result in the widening of the wage distribution of the entire cohort (workers entering the labor market at the same time) over time.

Many formal models have been built to capture one or more of the facts concerning wage distribution and wage dynamics. On the one hand, models of wage dynamics are often cast in a learning framework, where a worker's wage changes because new information is revealed about the worker's productivity; see for example, Holmstrom (1982), Harris and Holmstrom (1982) and Farber and Gibbons (1996). On the other hand, the wage distribution is most easily explained by models with matching, which takes place through either sorting or selection between firms and workers; see Sattinger (1993) and Neal and Rosen (2003) for excellent reviews. Several models combine both learning and matching and are able to explain more patterns of the
wage distribution and wage dynamics; see for example Jovanovic (1979), MacDonald (1982), and Gibbons, Katz, Lemieux, and Parent (2005), and see Section 2 for a detailed literature review.

This paper adds to the existing literature on the wage distribution and wage dynamics by developing a model of matching and Pareto learning. There is a continuum of workers and a continuum of firms. The abilities of workers are unknown and are drawn independently from a Pareto distribution. Firms differ by their technologies, which are drawn from a generalized Pareto distribution. Production takes place when a worker is matched with a firm. The output from production is complementary in the worker's ability and the firm's technology. The output of the worker is observed publicly, so firms learn about the worker's ability through Pareto learning. We solve for the market equilibrium where a) workers with the same expected ability receive the same wage, b) each firm maximizes its profit, and c) the demand for each type of worker equals its supply.

The most novel feature of the model is its use of Pareto learning. In Pareto learning, the ability of a worker is drawn from a Pareto distribution. The exact value of the worker's ability is unknown, but it can be learned from observing a sequence of signals that are drawn uniformly between zero and the worker's ability. An attractive property of Pareto learning is that the conditional distribution of the worker's ability is again a Pareto distribution, which depends on only the maximum value of the signals and the number of signals observed. Moreover, when there is a continuum of workers, Pareto learning implies that the distribution of the expected ability of the workers is a generalized Pareto distribution (i.e., Pareto with a mass at the bottom).

We choose to use Pareto learning on several grounds. First, Pareto learning implies that the change in the worker's expected ability is "downward rigid." More precisely, when a new signal about the worker's ability arrives, the largest possible decrease in his expected ability is bounded by a constant divided by the number of signals observed so far. In contrast, there is no upper bound on the largest possible increase in the worker's expected ability. If ability is a good proxy for the wage, then Pareto
learning suggests that the wage change is also downward rigid, which is a well-known labor market regularity. Second, if the ability of each worker is drawn from a “regular distribution,” then extreme value theory implies that the upper tail of the extremal ability distribution is approximately Pareto; see Gabaix and Landier (2006) for exact statements. Finally, Pareto learning implies that the upper tail of the derived wage distribution is approximately Pareto, which fits the real wage distribution fairly well.

Pareto learning enables us to obtain explicit formulas for the wage distributions of cohorts of all experience levels. Using these formulas, we are able to explain several stylized facts and empirical regularities in the wage distribution and wage dynamics. Our most important result on the wage distribution is that the equilibrium wage distribution of each cohort and of the whole economy is approximately Pareto: the number of workers with income higher than $x$ equals $C(x - b)^{-\alpha}$, where $\alpha, b$, and $C$ are constants that depend on the parameters of production, ability distribution, and technology distribution. $b$ and $C$ also depends on the cohort age; but $\alpha$ is independent of it. These predicted wage distributions fit the real distribution fairly well (see for example Lydall (1958)), and the explicit formulas of the distributions enable us to relate our results to existing theories and to carry out comparative statics.

We compare the derived wage distributions to the ability distributions of workers and find that the wage distribution lies to the right of the ability distribution. In other words, the wage is convex in ability. Of course, ability is unobservable, but our convexity result does provide an explanation for why similar workers, especially for those at the top of their professions, can earn dramatically different wages; see also Rosen (1981) for another explanation the same phenomenon.

Our final result on the wage distribution is concerned with how the recent increase in the wage inequality in the U.S. may be explained by changes in parameters that describe the production technology, the ability distribution, and the technology distribution. First, we show that the wage distribution becomes more dispersed if the production function becomes more convex in ability, which may result from a skill-biased technology change. Second, we show that the wage distribution becomes more
dispersed if the technological distribution across firms has a fatter tail, which may result from a higher capital/labor ratio. Thirdly, we show that the wage distribution becomes more dispersed if the ability distribution of workers has a fatter tail, which may result from a higher average schooling level of workers.

Concerning wage dynamics, this paper suggests explanations for the following empirical regularities. First, a cohort's median wage increases with experience. Second, for any cohort, the wage increase from experience is larger at higher quantiles. Third, the wage dispersion of a cohort of workers, measured either in the wage variance or quantile ratios, increases with cohort age. Fourth, if two workers of different ages have the same wage, the younger one is more likely to obtain a (sufficiently) higher wage in any future given time. Fifth, the return to experience is larger if the production is more convex in ability, which may result from a skill-biased technology change. The first four predictions match well-known empirical regularities in the labor market. The fifth prediction is consistent with the findings about the recent labor market experience of young cohorts in the U.S.; see for example Card and Lemieux (2001).

Our model also delivers two sets of results on wage dynamics that cannot be obtained by pure learning models. The first set of results centers on our findings that the expected wage of each worker increases with age. This result by itself is not novel and has been obtained by Jovanovic (1979), MacDonald (1982), and Gibbons, Katz, Lemieux, and Parent (2005). Our paper differs from the previous ones by providing an explicit formula for the expected wage increase, which leads to quantitative estimates of the wage increase from better matching due to improved information. In particular, if a worker's wage is above an age-dependent threshold, then the expected wage increase is a linear function of his current wage level. This formula is similar to the standard wage regression specification; see for example Mincer (1974). In other words, some of the regression coefficient of experience in a Mincerian regression can be interpreted as wage growth from better matching due to improved information.

Estimates of return to experience typically range from 5 to 12%; see for example Abraham and Farber (1987), Altonji and Shakotko (1987), and Topel (1991). In our
model, the implied wage growth rate ranges from 1% to 20% in the first five years depending on different parameter choices. This suggests that a significant portion of the wage growth may be explained by better matching of workers due to improved information. Our numerical results also suggest that a) the wage increase is larger when the production function is more convex in ability, b) wage increase decreases with experience, and c) the wage increase is larger when the technological distribution has a fatter tail. These results suggest that the return to experience is larger both in jobs with higher sensitivities to abilities and in industries with more diverse technologies. These predictions can be tested directly, which I hope to do in the future.

The second set of results that require both learning and matching centers on the findings that wage residuals are positively correlated. In pure learning models, the wage is a martingale, so the expected wage of next period always equals the current wage. Therefore, changes in wages are serially uncorrelated. In contrast, when there is matching, the wage is a submartingale, so the expected wage of next period is higher than the current wage. Moreover, if the expected wage increase is larger at higher wage levels, this leads to positive correlation in wage changes. Under Pareto learning, we show that the expected wage increase is larger at higher wage levels.

When the distribution of ability is not necessary Pareto, we provide a sufficient condition for wage changes to be positively serially correlated. This condition requires (a) the variance of wage change to be increasing in wage level and (b) the wage is increasingly convex in ability. These sufficient conditions suggest that positive serial correlation is more likely to be found in younger workers or in industries with increasingly convex wage schedules, shedding light on why certain studies have found changes in wage residuals are serially correlated (Baker, Gibbs, Holmstrom (1994b) (hereafter BGH), Hause (1980), and Lillard and Weiss (1979)), and others have not (Abowd and Card (1989), Topel (1991), and Topel and Ward (1992)).

The rest of the paper is organized as follows. Section 2 reviews the literature. Section 3 develops the model and solves for the market equilibrium. In Section 4,
we derive the explicit wage functions and the wage distributions. Patterns of wage
dynamics are analyzed in Section 5. Section 6 concludes.

2.2 Literature Review

As noted above, models of wage dynamics are often cast in a learning framework. For
example, Farber and Gibbons (1996) explores the pure effect of learning on wage dy-
namics. In their model, there is a perfectly competitive economy of firms and workers.
Workers differ in their abilities, which equal their expected marginal productivities.
Workers' abilities are initially unknown but can be inferred through their histories of
outputs. No long-term contract is available in this economy, so the wage of a worker
equals his expected marginal productivity, which in turn equals the worker's expected
ability. With this setup, standard results in learning imply that the expected ability
(and thus the wage) of a worker is a martingale. Furthermore, learning implies that
the wage distribution of a cohort becomes more dispersed as it gains experience.

On the other end, models of the wage distribution are often cast in a matching
framework. Matching in these models is attained through either sorting or selection.
In a typical sorting model, workers differ in ability and firms differ in technology. The
ability and the technology are both one-dimensional and are usually complementary to
each other. The complementarity implies that workers of higher abilities are matched
with firms of higher technology in equilibrium. Consequently, the wage distribution
is skewed to the right of ability distribution. There is a large literature of matching
models; both Sattinger (1993) and Neal and Rosen (2003) offer excellent reviews.

An important variant to the basic sorting model, often called one-sided match-
ing, groups individuals in a two-stage procedure: typically, in the first stage indi-
viduals decide whether to become a worker or a manager, and in the second stage
managers are matched with workers; see for example, Lucas (1978), Rosen (1982),
Garicano(2000), Garicano and Rossi-Hansberg(forthcoming), and Antras, Garicano,
and Rossi-Hansberg (2006). Kremer (1993) does not have the two-step procedure,
but the production function in his model requires workers to work together and their abilities are complementary to each other. In equilibrium, workers of similar abilities are grouped together and similar results on wage distribution are obtained.

In contrast to sorting models, the ability of a worker is typically multidimensional in selection models. There are multiple sectors in the economy, and each sector values different dimensions of ability differently. Individuals select into different sectors according to their comparative advantages in abilities. Because workers will choose the sectors they are relatively best at, the resulting wage distribution is more unequal than if the workers were randomly assigned into sectors; see for example, Roy (1951) and Borjas (1987), and see Neal and Rosen (2003) and Sattinger (1993) for reviews.

Several integrative models combine both selection and learning and are able to explain more empirical regularities. In Gibbons, Katz, Lemieux, and Parent (2005), hereafter GKLP, there is a perfectly competitive economy with heterogeneous workers and jobs. Workers differ in their abilities, and firms have different kinds of jobs. The production is stochastic, and the expected output of each job is an affine function of worker ability. The jobs differ in their slopes (in ability) and intercepts, and a job with higher slope than another will have a smaller intercept, so no job is completely dominated by another. The ability of a worker is unknown to everyone but can be estimated from his history of outputs. Workers choose jobs according to their expected abilities. For example, a worker might initially select a job with middle level of slope and intercept, and then move to a job with a higher or lower slope depending on his outputs. Higher level of output signals higher ability, and a worker with higher expected ability will prefer a job with larger slope.

GKLP predicts various patterns of the wage distribution and dynamics. First, the wage distribution is skewed to the right of ability distribution because workers of higher abilities will choose jobs of higher slopes. Second, the average wage of the cohort increases over time through better matching. Third, the wage becomes more dispersed over time, measured by wage ratios of different quantiles, because workers of different abilities can be better separated through longer histories of outputs.
A similar model by MacDonald (1982) studies learning and selection in a general equilibrium context. Workers in his model are either of type A or B and there is a continuum of jobs between 0 and 1. Type A workers are best at job 0 and type Bs are best at job one. The types of the workers are unknown but can be learned over time. At the end of each period, each worker receives a signal of either a or b, assuming that type A workers have a higher probability of receiving a. Each worker lives N periods, so workers can be partitioned into $2N+1$ information classes, differentiated by the probability of being an A type. Consequently, jobs are also partitioned into $2N+1$ intervals, with intervals closer to 0 being assigned in equilibrium to the information class more likely to be of type A. Results from MacDonald (1982) are similar to that from GKLP because both models aggregate learning and matching. However, the general equilibrium feature of the MacDonald model further strengthens some of the results. In particular, younger workers have low wages because they in general stay in information sets with more people, who are substitutes to the worker and bid down the wage.

Integrative models with learning and matching have also been used to explain the turnover rates in labor market; see for example Jovanovic (1979) and Woodcock (2003). In those models, the productivity of a firm-worker pair depends both on the worker’s ability and the match quality between the two. A worker switches jobs if the expected match quality is lower than a threshold that possibly depends on the worker’s experience and tenure at the firm. The average wage of a cohort increases because there are fewer and fewer “bad” matches for the cohort as it ages.

This paper also presents an integrative model that matches similar patterns of the wage distribution and wage dynamics as is done in MacDonald and GKLP. In addition, we obtain explicit formulas of the wage distributions and wage dynamics, enabling us to a) compare our predicted wage distribution with the wage distribution in real world and b) quantify the importance of better matching due to improved information in contributing to the wage growth of workers.
2.3 Model Setup

We set up the model in this section. Subsection 3.1 describes the model basics, including the objective functions of workers and firms. Subsection 3.2 specifies the timing and information structure of the model.

2.3.1 Workers and Firms

The economy has infinite number of periods. Time is discrete. At each time \( t \), a measure \( 1 - \rho \) of new workers enter the economy and each of the existing workers exits with probability \( 1 - \rho \). New workers are in cohort 0 and has a cohort size of \( 1 - \rho \). In general, workers who have been in the economy for \( n \) periods are in cohort \( n \) and has a cohort size of \( (1 - \rho)\rho^{n-1} \). Time extends all the way to negative infinity, so that total measure of workers in the economy is 1.

Workers differ in their innate abilities, which we denote as \( a \). The innate ability of each worker is drawn independently from a Pareto distribution with parameter \( (1, \alpha) \), i.e.

\[
\begin{align*}
\Pr(x \geq a) &= a^{-\alpha} \quad \text{for } a > 1; \\
\Pr(x \geq a) &= 1 \quad \text{for } a \leq 1.
\end{align*}
\]

(2.1)

We assume \( \alpha > 2 \), so the distribution has a finite mean and variance.

The utility function of each worker is

\[
U(w_0, w_1, ...) = \sum_{n=0}^{\infty} \delta^n (1_{\{\text{hired, } n\}} u(w_n) + (1 - 1_{\{\text{hired, } n\}}) u(1)),
\]

(2.2)

where \( \delta \in (0, 1) \) is the discount factor, \( 1_{\{\text{hired, } n\}} \) is an indicator function that equals 1 if the worker is hired in period \( n \) and zero otherwise, \( w_n \) is the worker’s wage income when he is in cohort \( n \), and \( u \) is a strictly increasing and strictly concave function. In other words, if the worker is hired by a firm that offers \( w_n \), his utility in that period is \( u(w_n) \). Otherwise, the worker receives his outside option that offers a wage of 1.\(^1\)

\(^1\)We normalize the outside option of the workers to be 1 for simplicity. As will be clear after the
There is unit mass of infinitesimal firms. Each firm lives for only one period and hires at most one worker. The firms differ by their technology $s$, which measures the productivity of the firms. The distribution of $s$ across firms follows a generalized Pareto distribution with parameter $(1, \alpha, \sigma)$, i.e.

$$\Pr(x \geq s) = \sigma^\gamma s^{-\gamma} \text{ for } s > 1;$$

$$\Pr(x \geq s) = 1 \text{ for } s \leq 1.$$  \hspace{1cm} (2.3)

In other words, this distribution is a linear combination of degenerate distribution at 1 (of mass $\sigma^\gamma$) with a Pareto distribution with parameter $(1, \gamma)$.\(^2\) We assume $\gamma > 2$ for the distribution to have a finite mean and variance.

Production takes place when one firm is matched with one worker. The output $y$ is linear in the firm’s technology and depends stochastically on the worker’s ability. If a firm with technology $s$ hires a worker of innate ability $a$, the output equals

$$Y(s, a, z) = \frac{sz^\beta}{1 + \beta},$$  \hspace{1cm} (2.4)

where $z$ is a random variable drawn uniformly from $[0, a]$ and $\beta > 0$ is a production parameter that measures the sensitivity of output to ability levels. The $1 + \beta$ in the denominator is a normalizing term.

Each firm maximizes its expected payoff. The expected payoff of a firm of technology $s$ is

$$1_{\{\text{hire}\}}(E[Y(s, a, z)] - w)$$  \hspace{1cm} (2.5)

where $1_{\{\text{hire}\}}$ is an indicator function that equals 1 if the firm hires a worker and zero otherwise, $E[Y(s, a, z)]$ stands for expected output when the hired worker has innate ability $a$, and $w$ is the wage paid to the worker. In other words, a firm receives 0 if

\(^2\)The assumption of a point mass at the bottom helps simplify the expressions for the matching formula and wage formula significantly. This assumption is not crucial to our analysis.
it does not hires a worker. When it hires a worker, the firm’s expected payoff equals its expected output minus the wage payment.

The specification of the production function leads to a simple expression for the expected output: when a worker of ability $a$ matches with a firm of technology $s$, then

$$E[Y(s, a, z)] = \int_0^a \frac{sz^\beta}{(1 + \beta)a} dz = sa^\beta. \quad (2.6)$$

The expected output increases with both the technology level $s$, the ability level $a$, and the production parameter $\beta$. Note that the production is complementary in innate ability and technology level, so it is more efficient to match high ability workers with high technology firms. In addition, note that $\beta$ is the elasticity of expected output to ability: outputs are more sensitive in innate ability when $\beta$ is larger.

To simplify our analysis below, it is convenient to define the effective ability of a worker as

$$e = a^\beta. \quad (2.7)$$

Therefore, the expected output of a firm-worker pair equals the direct multiplication of the technology index and the effective ability, i.e.

$$E[Y(s, e, z)] = se. \quad (2.8)$$

It is easy to see that the effective ability is distributed as $\text{Pareto}(1, \frac{\alpha}{\beta})$ because the innate ability is distributed as $\text{Pareto}(1, \alpha)$.

### 2.3.2 Information Structure and Timeline

Firms do not observe the ability of any worker but they know the distribution of the effective ability of the workers. In additions, all firms observe the past histories of each existing worker, including his past outputs levels and the technological levels of the firms he worked for.
At the beginning of each period, firms simultaneously offer contracts to workers. Each contract consists of a single fixed wage. There is no limit on the number of contracts a single firm can offer, and we allow firms to offer different contracts to different workers.

The single fixed wage offer reflects three separate aspects of the model. First, wages are not contingent on output: this assumption is without loss of generality because the workers have no disutility of effort and have a strictly concave utility function. Second, there is long-term contract: this reflects the fact that firms are short-lived. Finally, firms do not offer a lottery in the contracts: this is because the workers have a strictly concave utility function.

Once the wage contracts are offered, workers pick their offers according to a pre-specified order. Each worker can only pick at most one contract. After the worker accepts an offer from a firm, the firm retracts the rest of the contracts offered.

Production takes place after a worker has accepted the contract from a firm. If a firm’s wage offers are accepted by no workers, the firm receives its outside option, which equals zero. Similarly, if a worker does not pick any contracts, he receives his outside option, which equals 1.

### 2.4 Analysis of the Model

We solve the subgame perfect equilibrium of the model in this section. Subsection 4.1 introduces the concept of conditional expected effective ability (CEEA) and characterizes its distributions. Subsection 4.2 uses the concept of CEEA and shows that the SPE of the model can be identified with a market equilibrium involving demands and supplies of CEEAs. We solve this market equilibrium and show that it is unique and efficient. Subsection 4.3 derives the wage formula (as a function of CEEA) under the market equilibrium and analyzes the properties of wage formulas.

---

3 Any arbitrary order works here.
4 This is because each firm can hire at most worker.
2.4.1 CEEA and Pareto Learning

This subsection introduces the concept of CEEA and shows that it depends only on the maximum observed output of the worker. As will be seen in the next subsection, the wage of a worker is completely determined by his CEEA. We also give explicit formula for the distribution of CEEA of each cohort and of the entire economy. These formulas will be used in Section 5 to calculate the wage distribution.

Each worker’s history of past outputs and past employers is observable to all firms. The histories of workers enable firms to update the conditional distribution of effective ability of each worker using Bayes Rule. In particular, firms can compute the CEEA of any worker. We use $\eta_n$ to denote the CEEA of a worker in cohort $n$, so

$$\eta_n = E[a^\beta | (y_0, s_0), \ldots, (y_{n-1}, s_{n-1})], \quad (2.9)$$

where recall that $a$ is the innate ability of the worker and $a^\beta$ is his effective ability. Also recall that the production specifies

$$y_m = y(s_m, a, z_m) = \frac{s_m z_m^\beta}{1 + \beta} \quad \text{for } 0 \leq m < n,$$

where $z_m$ is drawn uniformly from $[0, a]$. Therefore, when $s_m$ is observable, the only information available about the worker’s ability is given by $z_m$. In other words, we can write the CEEA of a worker in cohort $n$ as

$$\eta_n = E[a^\beta | z_0, \ldots, z_{n-1}]. \quad (2.10)$$

In general, the CEEA of a worker depends on his entire histories of $z$. With the assumptions of Pareto prior and uniform draws, CEEA depends only on the maximum of the $z_n$, matters for the conditional distribution of effective ability. Since we will use the notation of the maximum $z$ repeatedly, we define it formally.

**Definition 1:** Let $m(z_1, \ldots, z_n) = \max\{z_1, \ldots, z_m, 1\}$ be the worker’s maximum
observed ability.

The reason that the CEEA depends only on \( m \) is that the conditional distribution of \( a \) is completely determined by \( m \) : when the ability \( a \) has a Pareto distribution and \( z \) is a uniform drawn from \([0, a]\), a well-known result from statistics (Pareto Learning) says that \( m \) is a sufficient statistic for the conditional distribution of the innate ability of the worker. In other words, two workers’ conditional distribution of their innate abilities are identical if and only if they have the same maximum observed ability \( m \). Moreover, the innate ability of a cohort \( n \) worker of with maximum observed ability \( m \) has a conditional distribution of \( \text{Pareto}(m, \alpha + n) \):

\[
\Pr(a \geq a_n|m) = \left( \frac{a_n}{m} \right)^{\alpha + n} \quad \text{for } a_n \geq m \\
= 1 \quad \text{for } a_n < m
\]

The result above then implies that

\[
\eta_n = E[a^\beta|z_0, \ldots z_{n-1}]
\]

\[
= E[a^\beta|m(z_0, \ldots z_{n-1})]
\]

\[
= \int_{m}^{\infty} \frac{(\alpha + n)m^{\alpha + n}}{x^{\alpha + n + 1}}x^\beta dx
\]

\[
= \frac{\alpha + n}{\alpha - \beta + n} m^\beta,
\]

where in the third equility, \( \frac{(\alpha + n)m^{\alpha + n}}{x^{\alpha + n + 1}} \) is the conditional probability that the innate ability equals \( x \). This allows us to define \( \eta_n(m) \) as the CEEA of a worker of cohort \( n \) and maximum observed ability \( m \) alone.

Equation (2.12) provides a link between CEEA and the maximal observed ability, so the distribution of the maximal observed ability can help calculate the induced distribution of CEEA. Lemma 1 below shows that the CEEA distribution of each cohort is a generalized Pareto distribution: there is a point mass at the bottom, and
the probability that CEEA is greater than a level $x$ is proportional to $x$ to the power of $\frac{\beta}{\alpha}$, where $\alpha$ is the parameter for the innate ability distribution and $\beta$ is the elasticity of expected output to ability.

**Lemma 1:** For workers in cohort $n$, their CEEA distribution is a generalized Pareto distribution with parameters $\left(\frac{\alpha+n}{\alpha+\beta+n}, \frac{\alpha}{\alpha+n}\right)$, i.e.

$$
\Pr(\eta \geq \eta_n) = \frac{n(\frac{\alpha+\beta+n}{\alpha+n} \eta_n)^{-\frac{\alpha}{\beta}}} {\alpha+n} \quad \text{for } \eta_n \geq \frac{\alpha+n}{\alpha+\beta+n};
$$

$$
\Pr(\eta \geq \eta_n) = 1 \quad \text{for } \eta_n < \frac{\alpha+n}{\alpha+\beta+n}.
$$

**Proof.** The probability that the maximum observed ability of a worker in cohort $n$ is less than $m > 1$ is

$$
\Pr(x \leq m) = \int_1^m \frac{\alpha}{x^{\alpha+1}} dx + \int_m^\infty \left(\frac{m}{x}\right)^n \frac{\alpha}{x^{\alpha+1}} dx = 1 - \frac{n}{\alpha+n} m^{-\alpha}.
$$

(2.14)

And the probability that $m = 1$ is

$$
\Pr(x = m) = \frac{\alpha}{\alpha+n} \quad \text{for } m = 1.
$$

(2.15)

Since the maximum observed ability $m$ is a sufficient statistic for the effective ability, by (2.12) we see that for a worker in cohort $n$,

$$
\eta_n > x \quad \Leftrightarrow \quad m > \left(\frac{\alpha - \beta + n}{\alpha+n} x\right)^{\frac{1}{\beta}} \quad \text{for } x > \frac{\alpha + n}{\alpha - \beta + n}.
$$

(2.16)

Now we can use equation (2.14) and (2.16) to compute the CDF of $\eta_n$ and obtain the desired result. 

The bottom mass of the CEEA distribution corresponds to workers whose maximum observed performance $m$ equals 1. In other words, there has been so far “no evidence” proving that these workers have innate ability greater than one. Moreover, the CEEA level of the bottom mass decreases with the cohort’s age and approaches
one as \( n \) goes to infinity. This reflects that the CEEA of a worker decreases if he fails to improve his maximum observed ability. In particular, if a worker never produces above 1, his CEEA level eventually approaches 1, which is the lowest possible ability level in its range of prior. Finally, the bottom mass of the CEEA shrinks to zero as the cohort age goes to infinity. This is because the abilities of workers will be fully revealed as the cohort age goes to infinity and there is 0 measure of workers with an innate ability of 1.

The formula in Lemma 1 shows that the CEEA distribution of each cohort has the same coefficient \( (\frac{\alpha}{\beta}) \) of power in the generalized Pareto distributions. In other words, the number of workers with CEEA level above a threshold decreases in the same rate for all cohorts as the threshold level increase. The implies that CEEA distribution aggregated across all cohorts also has a generalized Pareto distribution with the same coefficient of power. Define \( G(\eta) \) as the distribution of CEEA in the economy. Lemma 2 gives the exact formula of \( G \).

**Lemma 2:** Let \( \lambda(a, \beta, \rho) = \left( \sum_{\rho=0}^{\infty} (1 - \rho) \rho^n \frac{n}{a+n} \frac{(\alpha-\beta+n)}{a+n} \right)^{\frac{\beta}{\alpha}}. \) For \( \eta > \frac{\alpha}{a-\beta} \), we have

\[
1 - G(\eta) = \Pr(x > \eta) = \left( \frac{\eta}{\lambda(a, \beta, \rho)} \right)^{-\frac{\beta}{\alpha}}.
\]  

**Proof.** This follows directly from Lemma 1 by aggregation.

Lemma 2 only gives the CEEA distribution of \( \eta > \frac{\alpha}{a-\beta} \). In principle, we can also write the formula for \( 1 < \eta < \frac{\alpha}{a-\beta} \). However, each cohort has a bottom point mass (see Lemma 1) in this range, so there will be countable many point masses between 1 and \( \frac{\alpha}{a-\beta} \). Therefore, formula of \( G \) in this range is complicated and it does not affect our analysis, so we omit it here.

### 2.4.2 Analysis of the Equilibrium

We solve the equilibrium of the model in this subsection. The solution of the model has two steps. First, Theorem 1 shows that the SPE of the model can be identifies
as a market equilibrium involving demand and supply of the CEEA of the workers. Second, Theorem 2 solves this market equilibrium. In particular, it gives formulas for the equilibrium matching function between CEEA and technologies levels and wage as a function of CEEA. These formulas are general and do not depend on the particular distribution of technologies and CEEAs. The rest of the paper applies these general formulas to derive more explicit expressions when the CEEAs and technologies are Pareto distributed.

Recall that no long-term contracts can be offered and the expected output of a firm-worker pair equals the firm’s technology times the worker’s CEEA, so each worker’s value to a firm depends solely on his CEEA. Therefore, it is reasonable to guess that workers with the same CEEA will receive the same wage level in equilibrium. Furthermore, since the number of workers equals the number of firms and each worker has positive CEEA, it is plausible that in equilibrium every worker will be employed. Indeed, the proposition below shows that both are true and the analysis of wages can be simplified by looking at “the market wage” for each CEEA level.

**Theorem 1**: Every Subgame Perfect Equilibrium of the model satisfies the following:

(i): There exists a monotone function of $W(\eta)$ which maps CEEA to wages.

(ii): Each firm maximizes its profit: for a firm of technology $s$, it chooses a worker of productivity potential of to maximize $s\eta - W(\eta)$.

(iii): Labor supply equals labor demand: if in equilibrium firm of technology $s$ chooses worker of CEEA $\eta(s)$, then

$$G(\eta) = \int_0^{\infty} 1_{\{\eta(s) \leq \eta\}} dF(s),$$

where $G(\eta)$ is the CDF of the CEEA, $1_{\{\eta(s) \leq \eta\}}$ is an indicator function that equals 1 if $\eta(s) \leq \eta$ and 0 otherwise, and $F(s)$ is the CDF of the technology distribution.

**Proof.** See Appendix.

Condition (i) is intuitive: higher CEEA gives larger expected output, so wage
increases with CEEA. Condition (ii) is the standard expected profit maximization of the firms. Note that is a correspondence. In other words, it is possible for firms with the same technology to choose workers of different CEEA. Condition (iii) states that the demand for workers equals supply for every level of CEEA. Since many CEEA levels have measure zero, we use Lebesgue integral to represent that the labor demand for workers of CEEA below a certain level should equal the supplies.

To solve for the market equilibrium, we first observe that $W(\eta)$ must be strictly increasing because the expected output of a worker is strictly increasing in $q$ for any firm he is matched with. Second, we can rewrite the condition (ii) in the market equilibrium as the following:

$$s\eta(s) - W(\eta(s)) \geq s\eta(s') - W(\eta(s')) \quad \text{for all } s, s'$$

(2.18)

This formulation suggests that the problem of finding the equilibrium can be thought of as a mechanism design problem. In particular, each firm’s type is its technology index $s$, and the mechanism requires each firm to announce its type. If a firm of type $s$ announces type $s''$, it will be assigned a worker of CEEA $\eta(s')$ and pay a transfer of $W(\eta(s'))$, which gives the firm a payoff of $s\eta(s') - W(\eta(s'))$.

The solve for this mechanism design problem, we need to find an assignment rule $\eta(s)$ and a transfer rule $W(\eta(s))$, so that in equilibrium each firm finds it optimal to truthfully announce its own type. However, because the distribution of technology $F$ has an atom in the bottom, we need to solve for the equilibrium separately for the atom before applying the standard mechanism design technique to the rest of the technology distribution. This is done in Theorem 2. Since Theorem 2 does not depend on the particular distribution of $F$ and $G$, we state the theorem in terms of $F$ and $G$ even if we can state the results with more explicit expressions. We calculate some of the more explicit expressions for wages in the next section.

**Definition 2**: Let $s(\eta)$ be the technology of the firm that is matched to a worker of CEEA $\eta$ in equilibrium.
Definition 3: Let $\eta(1_+)$ be the unique CEEA such that $G(\eta(1_+)) = F(1)$.

Theorem 2: There exists a unique market equilibrium such that

$$s(\eta) = F^{-1}(G(\eta));$$
$$W(\eta) = \eta \quad \text{for } \eta \leq \eta(1_+);$$
$$W(\eta(s)) = s \eta(s) - \int_1^s \eta(x) dx \quad \text{for } \eta(s) > \eta(1_+).$$

The market equilibrium is efficient in the sense that it maximizes the total expected outputs of the economy.

Proof. We first show that firms of higher technology must be matched with workers of higher CEEA in equilibrium. If not, then there exist two firms, firm 1 and 2, with technology $s_1 > s_2$. Firm 1 hires worker 1 and firm 2 hires worker 2, where the CEEA of the workers satisfy $\eta_1 < \eta_2$.

Now for any wage schedule $W(\eta)$, the sum of equilibrium payoff of firm 1 and 2 is

$$s_1 \eta_1 - W(\eta_1) + s_2 \eta_2 - W(\eta_2)$$
$$< s_1 \eta_2 - W(\eta_2) + s_2 \eta_1 - W(\eta_1).$$

This implies that either firm 1 or firm 2 can find a profitable deviation, which leads to a contradiction. In other words, the equilibrium allocation must be positive assortative, i.e.

$$\eta(s) = G^{-1}(F(\eta))$$

. Since the expected output is complementary in expected output and technology, the resulting allocation must be efficient.

Denote the highest CEEA of workers associated with firm of technology $s = 1$ as $\eta(1_+)$. Positive assortative matching implies that workers with CEEA $\eta \leq \eta(1_+)$ are all matched with firms of $s = 1$. This immediately implies that

$$W(\eta) = \eta \quad \text{for } \eta \leq \eta(1_+).$$
Now consider the matching and wages of firms of technology greater than 1 and workers of CEEA greater than \( \eta(1+) \). These remaining distributions of firm technology and worker CEEA have no atoms, so standard mechanism design technique applies. Let's index a firm's type by their technology levels. Let \( \eta(s) \) and \( W(\eta) \) be the allocation and transfer rule: if a firm of type \( s \) announces that it is of type \( s' \), its payoff is \( s\eta(s') - W(\eta(s')) \). Let \( \Pi(s) = s\eta(s) - W(\eta(s)) \) be the equilibrium payoff of a firm of type \( s \). Then a standard condition for the firms to announce their types truthfully is that

\[
\Pi(s) = \int_1^s \eta(x) \, dx.
\]

Since \( \Pi(s) = s\eta(s) - W(\eta(s)) \), the above implies that

\[
W(\eta(s)) = s\eta(s) - \int_1^s \eta(x) \, dx \quad \text{for} \quad \eta(s) > \eta(1+)
\]

This expression and \( W(\eta) = \eta \) for \( \eta \leq \eta(1+) \) completely pins down wage schedule of CEEA. Together with the positive assortative matching, this wage schedule shows that if an equilibrium exists, it must be unique. It is also easy to check that the above prescribed allocation and transfer rule is a market equilibrium.

2.4.3 Firm-Worker Match and the Wage Formula

The general expressions for equilibrium firm-worker match and wage formulas are already contained in Theorem 2. The assumptions that the ability and technology distributions are Pareto distributed makes the formulas particularly simple. This subsection gives explicit expressions for these formula, which will be used repeated below to derive the wage distributions and results of wage dynamics.

First, we study the equilibrium matching function. Theorem 2 states that in equilibrium a worker of CEEA \( \eta \) is matched with a firm of \( s(\eta) = F^{-1}(G(\eta)) \). Here we compute instead the \( \eta(s) \), the equilibrium CEEA of the worker that matches a firm with technology \( s \). Recall Lemma 2 shows that for CEEA \( \eta > \frac{a}{a-\beta} \), the equilibrium
CEEA distribution $G$ satisfies

$$1 - G(\eta) = \left( \frac{\eta}{\lambda(a, \beta, \rho)} \right)^{-\frac{\gamma}{\alpha - \beta}},$$

and $G$ has infinitely many atoms for $\eta < \frac{\alpha}{\alpha - \beta}$. The matching formula can be greatly simplified if we assume that

$$\sigma^{-\gamma} < \left( \frac{\lambda(a, \beta, \rho)(\alpha - \beta)}{\alpha} \right)^{-\frac{\gamma}{\alpha - \beta}}. \quad (2.26)$$

This assumption assures that every worker with CEEA below $\frac{\alpha}{\alpha - \beta}$ will be matched with a firm of technology $s = 1$.

When this assumption is satisfied, all the point masses in the bottom distribution of CEEA will be matched with firm of technology. For $s > 1$, if a firm of technology $s$ is matched with a worker of CEEA $\eta$, then positive assortative matching in Theorem 2 implies that the number of workers above should equal the number of firms above $s$, i.e.:

$$\sigma^{-\gamma} s^{-\gamma} = \left( \frac{\eta}{\lambda} \right)^{-\frac{\gamma}{\alpha - \beta}} \text{ for } s > 1;$$

$$\Leftrightarrow \eta(s) = \lambda \left( \frac{s}{\sigma} \right)^{\theta} \text{ for } s > 1, \quad (2.27)$$

where $\theta = \frac{\alpha \beta}{\alpha - \beta}$.

This matching formula has several highly intuitive properties. As the power parameter of the ability distribution $\alpha$ increases ($\theta$ decreases), the tail of innate ability distribution becomes thinner, so firms of a given technology level will be matched with workers of lower CEEA. Conversely, as the power parameter of the technology distribution $\gamma$ increases ($\theta$ increases) implies that the tails of the technology distribution is thinner, so a worker of given CEEA level will be matched with a firm of lower technology, or equivalently, firms with given technology will be matched with worker of higher CEEA.

With the explicit matching formula, we can now apply Theorem 2 to calculate
wage as a function of technology. Since this formula is important by itself and is also central to the analysis below, we list it as a separate proposition. For \( s > 1 \), let \( w(s) \) be the equilibrium wage paid by a firm of technology \( s \). Proposition 1 gives the exact formula for \( w(s) \). One intuitive way to derive the formula of \( w(s) \) is as follows.

Observe that the equilibrium output of technology \( s \) is \( \eta(s) \). Now consider a small increase \( \Delta s \) in the firm's technology. This leads the firm to be matched with better workers in equilibrium, and the associated increase in output is

\[
(s + \Delta s)\eta(s + \Delta s) - s\eta(s) \approx \eta(s)\Delta s + s\eta'(s)\Delta s.
\]  

(2.28)

To a first order approximation, the increase in output consists of two terms that correspond to increase in firm's profit and increase in wages. The \( \eta(s)\Delta s \) term captures the benefit of increase in technology and thus determines the profit level for firms with different technology. The \( s\eta'(s)\Delta s \) term reflects the increase in output resulting from employing a worker with higher CEEA. Since \( s\eta'(s)\Delta s \) is the difference between the increase in expected outputs and increase in profits, it is the increase in wages. From the matching function (2.27), we have \( w'(s) = \lambda \theta(\frac{s}{\gamma})^\theta \). Integrating this term and using the zero-profit condition of firms with \( s = 1 \) as a boundary condition, we obtain the following.

**Proposition 1** For \( s > 1 \), the wage as a function of technology can be written as

\[
w(s) = \frac{\lambda s^{-\theta}}{1 + \theta(\theta s^{\theta+1} + 1)},
\]  

(2.29)

and wage is convex in \( s \).

**Proof.** See Appendix. ■

From the formula, we see that the equilibrium wage increases and is convex with technology. If we think of better technology corresponds to higher ranks within a firm, this fits the observations that wage is often convex in job ranks inside a
firm, see for example Treble et al (2001). It is also clear that the workers' wages are lower than their marginal outputs, and the gap between the expected outputs and workers' wages increases with firms' technologies and the corresponding worker abilities. This is because the difference between expected outputs and wages simply equals the firm's profit, and a firm's profit in equilibrium increase with its technology. This increasing gap is often referred to as wage compression. The wage compression in this model results from workers of higher abilities being matched with firms of better technologies, which have a higher rent in equilibrium. This is quite different from the typical source of wage compression, which occurs because some kind of informational asymmetry; see for example Acemoglu and Pischke (1999).

Since we have an explicit matching formula between CEEA and technologies, we can also write wage as a function of CEEA. This is done in Proposition 2 below. One way to obtain the formula is as follows. Note that the marginal increase in wages at a given CEEA level should equal, in equilibrium, the marginal increase in the expected output of the firm matched with this CEEA level. Since the expected output equals the product of the firm's technology and the worker's CEEA, the marginal increase in wages at a CEEA level equals the technology of a firm that's matched with this CEEA level in equilibrium. In other words, we have

$$ W'(\eta) = \frac{\partial (\eta s)}{\partial \eta} = s(\eta) = \sigma \left( \frac{\eta}{\lambda} \right)^{\frac{1}{2}}. $$

(2.30)

Integrating the above expression and using the appropriate boundary condition, we obtain wage as a function worker CEEA.

The interpretation that marginal increase in wages at a CEEA level equals the technology of a firm that's matched with this CEEA level in equilibrium shows that convexity of wage in CEEA is independent of the underlying distribution of firms' technologies and workers' abilities. As long as positive assortative matching occurs in equilibrium, wage is weakly convex in ability because
Proposition 2 \textit{The wage as a function of CEEA can be written as}

\begin{align}
W(\eta) &= \eta \quad \text{for } \eta < \eta(1_+); \\
W(\eta) &= \frac{\lambda\sigma^{-\theta}}{1+\theta} (\theta \sigma^{\theta+1} (\eta \frac{\theta+1}{\theta} + 1) \text{ for } \eta \geq \eta(1_+),
\end{align}

where \( \eta(1_+) = \lambda\sigma^{-\theta} \). \textit{Wage is weakly convex in CEEA.}

\textbf{Proof.} See Appendix. \blacksquare

Although ability is not measurable, casual observations on the wage, especially in entertainment and sports sectors, suggest that small difference in ability at the top of distribution can lead to huge differences in wages. This suggests that wage maybe convex in ability, a topic that has received much attention from the economic literature; see for example Rosen (1981) and Kremer (1993). In this model, wage convexity results from better matching, which is similar to that of Kremer (1993), where convexity of wages results from better matching between workers.

2.5 Wage Distribution

This section derives explicit formula for the wage distributions of each cohort and of the economy. We show that the wage distribution of each cohort is very similar to a Pareto distribution, which shifts to the right as the cohort ages. In addition, we study how the wage distribution are affected by changes in technology and underlying ability distribution.

First, calculation of the wage distributions follows directly from the CEEA distribution in Lemma 1 and 2 and the wage formula in Proposition 2. Since Proposition 2 shows that wage of a worker equals a constant plus his CEEA to a power , the shape
of the wage distribution is similar to that of the CEEA distribution. The proposition below gives the exact formula.

**Proposition 3** The wage distribution of cohort $n$ satisfies the following. Let $\eta(1_+) = \lambda \sigma^{-\theta}$. For $W_n > W(\eta(1_+))$,

$$\Pr(W \geq W_n) = s(n)(\frac{\theta \lambda \sigma}{1 + \theta})^{\frac{\alpha}{\alpha + \beta \gamma}} \lambda^{\frac{-\theta}{\beta \gamma}} (W_n - \frac{\lambda \sigma^{-\theta}}{1 + \theta})^{-\frac{\alpha \gamma}{\alpha + \beta \gamma}}, \quad (2.33)$$

where $s(n) = \frac{n}{\alpha + n} (\frac{\alpha - \beta + n}{\alpha + n})^{\frac{-\theta}{\beta \gamma}}$ increases in $n$. The distribution of wages for all workers satisfies

$$\Pr(W \geq W_n) = (\frac{\theta \lambda \sigma}{1 + \theta})^{\frac{\alpha}{\alpha + \beta \gamma}} (W_n - \frac{\lambda \sigma^{-\theta}}{1 + \theta})^{-\frac{\alpha \gamma}{\alpha + \beta \gamma}}. \quad (2.34)$$

**Proof.** The wage distribution follows from direct computation. To check $s(n)$ increases with $n$, we note that

$$\frac{d \log(s(n))}{dn} = \frac{1}{n} - \frac{1}{\alpha + n} - \frac{\alpha}{\beta}(\frac{1}{\alpha - \beta + n} - \frac{1}{\alpha + n})$$

$$= \frac{\alpha (\alpha - \beta)}{n (\alpha + n) (\alpha - \beta + n)}$$

$$> 0.$$
that \( s(n) \) that increases with \( n \). Therefore, the wage distribution shifts to the right as the cohort ages.

For large enough wages, the distribution of the wages is nearly completely determined by the parameter \( \frac{\alpha \gamma}{\alpha + \beta \gamma} \), which controls how fast the tail of wage distribution shrinks. The wage distribution will be skewed more to the right if \( \frac{\alpha \gamma}{\alpha + \beta \gamma} \) decreases. This happens if \( \alpha \) decreases, or \( \gamma \) decreases, or \( \beta \) increases. This leads to the following corollary:

**Corollary 1:** The wage distribution is more skewed to the right if a) the innate ability distribution is more skewed to the right; or b) the technology distribution is more skewed to the right; or c) the production function is less concave.

Corollary 1 provides several intuitive predictions that are relevant to income inequality. First, income inequality increases as the distribution of innate ability becomes more skewed to the right. If more schooling shifts the ability distribution to the right, this suggests that schooling can increase inequality. Second, a less concave production function corresponds to technologies that magnify the productivity difference between workers of different abilities. One recent example that makes the production function less concave is the introduction of computers. Computers improve the productivity of workers of all abilities, but they raise the productivities of high-skilled workers more than that of low-skilled workers.

Third, we offer two interpretations of the decrease of \( \gamma \). If we think of all firms in this economy as one-job firms, as is in the exact setup our model, then a decrease in \( \gamma \) should be interpreted as more firms are equipped with better technologies. Alternatively, we can think of all firms in the economy are identical, and each firm has a distribution of jobs with different productivities. In this case, a decrease in \( \gamma \) reflects the change in the organizational form of the firms. A smaller \( \gamma \) indicates that more jobs in the firm are having higher productivities, which might result from a greater level of hierarchy or increased scale of production.
Our model also allows comparison between the ability distribution and wage distribution. In our model, we have two kinds of abilities. For innate abilities, the number of workers with innate ability above \( a \) is inversely proportional to the \( a \) power of \( a \). For effective abilities, the number of workers with effective ability above \( e \) is inversely proportional to the \( \alpha / \beta \) power of \( e \).

Since \( \frac{\alpha \gamma}{\alpha + \beta \gamma} < \frac{\alpha}{\beta} \), the wage distribution is always skewed to the right of effective ability distributions because of the positive assortative matching between CEEA and technology. The wage distribution, however, is not always skewed to the right of innate ability distribution because the concavity of production shifts the distribution of effective ability to the left. The wage distribution is skewed to the right of innate ability distribution if and only if \( \frac{\alpha \gamma}{\alpha + \beta \gamma} < \alpha \). In other words, the wage distribution is skewed to the right of innate ability distribution if either the production function is not too concave, or there are sufficiently many high technology firms in the upper tail. We thus obtain the following corollary:

**Corollary 2:** The wage distribution is always skewed to the right of the effective ability distribution. The wage distribution is skewed to the right of the innate ability distribution if \( \frac{\gamma}{\alpha + \beta \gamma} < 1 \).

### 2.6 Wage Dynamics

Our model yields many results on wage dynamics. In the first subsection, we present two results that require both learning and matching. The first result shows that the expected wage increases with time. Furthermore, for workers with a wage above an age-dependent threshold, their expected wage increase is proportional to their current wage level plus a constant. We compute the magnitude of wage increase for various parameter values and find that better matching through additional information can contribute to a 2 to 20 percent wage increase in the first year. The second result shows that changes in wage residuals are serially correlated. We also provide a sufficient condition for changes in wage residuals to be serially correlated when not all of the
assumptions in our model are satisfied. The sufficient condition suggests that positive serial correlation is more likely to be found in younger workers or in industries with increasingly convex wage schedules.

In the second subsection, we list results on wage dynamics that require learning only. The Pareto distribution assumptions make it easy to state the results in terms of quantiles. First, we show that median wage of a cohort increases with age. We also prove that wage level at most quantiles also increases and wage increase is larger at upper quantiles. Second, we show that the wage distribution of a cohort becomes more dispersed over age. This is done both by proving that the wage variance of a cohort increases with age and by showing the wage ratios at different quantiles increase with age. Third, we prove a specific form of personal effects in wage dynamics: for workers who take less time to reach a wage level, they are more likely to reach a higher wage level (in any given time).

2.6.1 Results Requiring Both Learning and Matching

Our first result focus on the sign of change in the average CEEA and average wage of workers. It is well known that the conditional expectation is a martingale; see for example DeGroot (1973). Therefore, the average CEEA does not change. The average wage of a cohort, however, increases with the cohort’s age. This is related to a version of Blackwell’s theorem, which states that the conditional expectation of a maximization problem is a sub-martingale. In the labor market setting, extra information acquired through the history of outputs allows workers being better assigned to firms and thus raise the average productivity and the average wage of workers.

Denote $\eta_n$ as the CEEA of a worker in cohort $n$. Denote $I_n$ as the information set about worker after $n$ periods.
Proposition 4 For any cohort $n$,

$$E[\eta_{n+1}|I_n] = \eta_n$$
$$E[W(\eta_{n+1}|I_n)] > W(\eta_n)$$

Proof. Recall that

$$\eta_n = E[a^\theta|I_n].$$

Consequently,

$$E[\eta_{n+1}|I_n] = E[E[a^\theta|I_{n+1}]|I_n] = E[a^\theta|I_n] = \eta_n,$$

where the 2nd equality follows from law of iterated expectations. Therefore, the CEEA of a worker is a martingale.

Now by Proposition 2, wage is convex in CEEA. Therefore, by Jensen’s Inequality for conditional probability, we have

$$W(\eta_n) = W(E[\eta_{n+1}|I_n]) < E[W(\eta_{n+1})|I_n],$$

where the inequality is strict because $W$ is strictly convex over a range with positive probability.

Proposition 4 shows that the wage of a worker can increase without human capital accumulation. Several earlier papers have made a similar point. For example, Jovanovic (1979) has a model where there are specific matching qualities between firms and workers. The average wage increases because workers with expected low matching qualities leave and move to firms with better expected matches. MacDonald (1983) considers a selection model with two types of workers and two types of jobs. Each type of worker is better at a particular type of jobs, but the type of any worker is unknown at the beginning. The average wages of the older workers are higher than the younger ones because they are more likely to be assigned jobs they are better at. There is also a general equilibrium effect that enlarges the wage increase: younger workers have low wages because they in general stay in information sets with more people, who are substitutes to the worker and bid down the wage. Finally, GKLP
(2006) has a model in which workers experience wage increase because they can better select into jobs to their comparative advantages over time.

Different from the previous models, the parametric assumption in this model provides explicit formula for wage increase. In particular, the wage increase here is almost proportional to the current wage level for most workers. More precisely, recall that \( \eta(1+) \) is the highest CEEA level of workers who are matched with a firm of technology \( s = 1 \). The proposition below states that if a worker's current wage is high enough that his next period wage will surely be above \( W(\eta(1+)) \), then the expected wage increase equals a constant times the worker's current wage minus \( W(\eta(1+)) \).

**Proposition 5** Suppose a worker in cohort \( n \) has CEEA \( \eta \) such that \( \Pr(\eta_{n+1} \geq \eta(1+) | \eta_n = \eta) = 1 \), then

\[
E[W(\eta_{n+1})|\eta_n = \eta] - W(\eta_n) = D(n, \alpha, \beta, \gamma)(W(\eta) - W(\eta(1+))),
\]

where \( D(n, \alpha, \beta, \gamma) \) is a constant independent of \( \eta \).

**Proof.** Let the maximum observed ability of a worker after \( n \) periods is \( m \), so his conditional innate ability \( a_n(m) \) is distributed as \( \text{Pareto}(m, \alpha + n) \). First, we can calculate the conditional distribution of the worker's maximum observed ability after period \( n + 1 \) as

\[
\Pr(m_{n+1} \leq x | m_n = m) = \int_m^x (\alpha + n)m^{\alpha+n}a^{-(\alpha+n+1)}da + \int_x^\infty (\alpha + n)m^{\alpha+n}\left(\frac{x}{a}\right)a^{-(\alpha+n+1)}da
\]

\[
= 1 - \frac{1}{\alpha + n + 1}\left(\frac{m}{x}\right)^{\alpha+n},
\]

and

\[
\Pr(m_{n+1} = m | m_n = m) = \frac{\alpha + n}{\alpha + n + 1}.
\]

Second, we express the wage as a function of the maximum observed ability. Since
\( \eta_n(m) = \frac{\alpha + n}{\alpha - \beta + n} m^\beta \) by (2.16), Proposition 2 implies that

\[
W(\eta_n(m)) - W(\eta(1+)) = C_n m^{\frac{\beta(\theta + 1)}{\theta}},
\]

(2.44)

where

\[
C_n = \frac{\theta + 1}{\theta} (\frac{\alpha + n}{\alpha - \beta + n})^{\frac{\beta + 1}{\theta}}.
\]

Combining these two results, we obtain that

\[
E[W(\eta_{n+1}(m)) - W(\eta(1+)) | m_n = m] = E[C_{n+1} m^{\frac{\beta(\theta + 1)}{\theta}} | m_n = m]
\]

\[
= C_{n+1} m^{\frac{\beta(\theta + 1)}{\theta}} + \int_{x}^{\infty} C_{n+1} m^{\frac{\beta(\theta + 1)}{\theta}} m^{\alpha + n - (\alpha + n + 1)} dx
\]

\[
= C_{n+1} m^{\frac{\beta(\theta + 1)}{\theta}} [1 + \frac{\theta}{\theta(\alpha + n) - \beta(\theta + 1)}] m^{\frac{\beta(\theta + 1)}{\theta}}.
\]

(2.45)

This implies that

\[
E[W(\eta_{n+1}(m)) - W(\eta_n(m)) | m_n = m] = E[W(\eta_{n+1}(m)) - W(\eta(1+)) | m_n = m] - (W(\eta_{n+1}(m)) - W(\eta(1+)))
\]

\[
= \{C_{n+1} m^{\frac{\beta(\theta + 1)}{\theta}} - C_n\} m^{\frac{\beta(\theta + 1)}{\theta}}
\]

\[
= D(n, \alpha, \beta, \gamma)(W(\eta) - W(\eta(1+))),
\]

(2.47)

where

\[
D(n, \alpha, \beta, \gamma) = \frac{C_{n+1}}{C_n} \{\frac{\alpha + n}{\alpha + n + 1} [1 + \frac{\theta}{\theta(\alpha + n) - \beta(\theta + 1)}] - 1\}. \quad \blacksquare
\]

To get a sense of the importance of expected wage increase from better matching due to improved information, we calculate \( D(n, \alpha, \beta, \gamma) \) for \( \alpha = 2, 3, \beta=0.3, 0.7, 1.1, \gamma=2, 3, \) and \( n \) from 1 to 8. Table I below calculates for these parameters values.

Under the column Period \( n \) is the rate of wage change between Period \( n \) and Period \( n-1. \) If \( W(\eta(1+)) \) is small relative to \( W(\eta) \), these number can roughly be thought of percentage increase in wages.

[INSERT TABLE I]

Table I display three patterns about wage changes. First, increases in the wage
decrease over time. This is because the marginal value of information decreases over time in two ways: first, earlier signals have a larger impact on the estimated CEEA of workers and lead to larger changes to wages than later one. Second, marginal return from better matching is smaller if the original matching is better. If we think of each period as one year, then depending on the parameter values, expected wage increase is between 2.25% to 22.32% from year 1 to year 2, between 1.11% to 6.20% from year 2 to 3, and 0.63% to 2.66% from year 3 to 4.

Second, wage increase is larger for more convex production functions (larger $\beta$). Wage increase from a production function with $\beta = 1.1$ can be more than three times as that from $\beta = 0.3$. Firms with more convex production function benefit more from higher abilities, which lead to a more convex wage function in ability. The more convex wage function results in larger increases in expected wage growth.

Third, wage increase is larger when the distribution of firm technology is more diverse (smaller $\gamma$). Productivity gains from matching are small when the technologies of firms are similar. In the extreme case where all firms have the same technology, matching does not matter at all and the expected output increase and the wage increase is zero for all periods. When the technology is diverse, gains from better matching will be important and lead to large increase in wages.

It is somewhat surprising that changes in the distributions of ability do not have a uniform effect on wage increase. As the ability distribution becomes more diverse (smaller $\alpha$), it raises the expected wage increase in earlier periods but lowers it in later ones. One possibility is that since ability distribution is more diverse, the speed of information revelation becomes faster in earlier periods. This may crowd out the information value in later periods and suppress future wage growth.

The second result in this model that requires both matching and learning is that wage changes are serial correlated. In pure learning models, wage is a martingale, so the expected wage of next period always equals the current wage. Therefore, changes in wages are serially uncorrelated. In contrast, when there is matching, wage is a
submartingale, so the expected wage of next period is higher than the current wage. Moreover, if the expected wage increase is larger at higher wage levels, this leads to positive correlation in wage changes. Under the Pareto learning, we can verify that the expected wage increase is larger at higher wage levels. This is stated formally in Proposition 6.

**Proposition 6** \( E[(W(\eta_{n+1}) - W(\eta_n))(W(\eta_n) - W(\eta_{n-1}))] > E[W(\eta_{n+1}) - W(\eta_n)]E[W(\eta_n) - W(\eta_{n-1})] \).

**Proof.** By law of iterated expectation,

\[
E[(W(\eta_{n+1}) - W(\eta_n))(W(\eta_n) - W(\eta_{n-1}))] = E[E[E[(W(\eta_{n+1}) - W(\eta_n))(W(\eta_n) - W(\eta_{n-1}))]|\eta_n]|\eta_{n-1}] (2.48)
\]

For a fixed \( \eta_{n-1} \), \( W(\eta_n) - W(\eta_{n-1}) \) clearly increases with \( \eta_n \). Moreover, (2.46) in Proposition 5 implies that \( E[W(\eta_{n+1})|\eta_n] - W(\eta_n) \) is increasing in \( \eta_n \). Therefore,

\[
E[E[E[W(\eta_{n+1})|\eta_n] - W(\eta_n)](W(\eta_n) - W(\eta_{n-1}))|\eta_{n-1}] > E[E[W(\eta_{n+1})|\eta_n] - W(\eta_n)]E[W(\eta_n) - W(\eta_{n-1})]|\eta_{n-1}] (2.51)
\]

In Proposition 6, we showed that \( E[W(\eta_{n+1})|\eta_n] - W(\eta_n) \) is increasing in \( \eta_n \) by using the explicit calculation in (2.46). I am not aware of any general conditions ensuring \( E[W(\eta_{n+1})|\eta_n] - W(\eta_n) \) increases in \( \eta_n \). However, by a second order approx-
Consequently, if $\text{Var}(\eta_{n+1} - \eta_n)$ is small and is non-decreasing with $\eta_n$, a sufficient condition for $E[W(\eta_{n+1})|\eta_n] - W(\eta_n)$ to increase with $\eta_n$ is $W'''(\eta_{n+1}) > 0$.


Gibbons (1997) noticed that serial correlation is found in small, relatively homogeneous samples and is not found in large, heterogeneous samples. To resolve the differences in findings, Gibbons suggested that one possible explanation is “that only certain small groups of workers (such as the managerial and professional workers in Baker-Gibbs-Holmstrom and Lillard-Weiss) exhibit such a personal effect.”

Our discussion above suggests that serial correlation in changes in wage residuals is most likely to be discovered in two groups. The first group includes young workers where the effect of information on wage growth is larger (see Table I). The second group includes workers in industries where the wage profile can be particularly convex: like a financial, IT, entertainment, and sports industry. Many workers in this group are likely to be managerial or professional workers, as suggested by Gibbons (1997).
Of course, our results suggest that the largest serial correlation will be found in young workers in industries with very convex wage profiles.

2.6.2 Results Requiring Learning Only

Learning models deliver a wide variety of results consistent with the wage dynamics observed in the labor market. However, in most models, it is assumed that wage equals the expected output. When wage is a nonlinear function of expected abilities, few general results are known about wage dynamics. In our model, our parametric assumptions enable us to compute the explicit wage distribution for each cohort. These explicit formulas are used to describe the wage dynamics in our model.

In the rest of the section, we first show that the median wage of a cohort increases with the cohort age. Moreover, the wage increases are larger at larger quantiles of the distribution. Next, we show that the wage distribution is more dispersed over time, both in terms of the variance of wage and different quantile ratios. Finally, we state a result on individual wage growth: a worker with a high wage growth in the past is likely to experience high wage growth in the future.

Our first result states that the median wage of a cohort increases with the cohort age. It is well known that the median wage of a cohort increases with age empirically. Theoretically, however, this result is not true in any learning model. In our model, the distribution of wage of each cohort has a mass at the bottom and is almost Pareto distributed for the rest. As the cohort ages and receive more information about the workers, the wage mass at the bottom is reduced and spreads to the rest of the distribution. As a result, the upper-tail of the wage distribution becomes fatter and the median wage of the cohort increases.

**Proposition 7** For a cohort with age \( n > \alpha \), the median wage of the cohort increases with cohort age.
Proof. See Appendix.

When the upper-tail of the wage distribution becomes fatter, identical reasoning as above indicates that wages increases at all quantiles above the bottom mass. Moreover, we can even show that wage increase at upper quantiles is larger than that of lower quantile. This follows because when the bottom mass of the CEEA spreads to the rest of the distribution, the thinning tails of the CEEA distribution implies that change in CEEA level is larger at the upper level. Since wage is convex in CEEA, change in wage level is also larger at the upper level.

Furthermore, for sufficiently old cohorts, the wage increase is larger as the production function becomes less concave. The effect of change in concavity of production comes from two sources. First, a less concave function moves the distribution of effective ability, and thus the CEEA corresponding to any quantile to the right. Since wage increases faster at higher level of CEEA, this leads to larger wage increases at any given quantile. Second, the output relies more on worker’s effective ability as the production function becomes less concave. As a result, a good signal of ability gives a larger increase in wages in economies where production functions depend more on abilities. The next Proposition summarized the discussion above.

Definition 4: Let \( q_k \) be the \( k \)th CEEA quantile of cohort \( n \), i.e.

\[
\Pr(x < \eta^k_n) = k. \tag{2.58}
\]

Proposition 8 For \( k > G\left(\frac{\alpha}{\alpha-\beta}\right) \), we have

\[
\frac{\partial(W(\eta^k_{n+1}) - W(\eta^k_{n}))}{\partial k} > 0. \tag{2.59}
\]

Moreover, for sufficiently large \( n \),

\[
\frac{\partial^2 W(\eta^k_n)}{\partial \eta \partial \beta} > 0. \tag{2.60}
\]

Proof. See Appendix.
Previous propositions describe the wage changes of a cohort at different quantiles. As a cohort ages, its wage distribution not only shifts to the right, but also becomes more dispersed. Typical measures of wage dispersion include the variance of the wage distribution and the wage ratios at different quantiles. We show below that the wage distribution becomes more dispersed under both measures. It is a standard result in learning models that wage distribution of a cohort becomes more dispersed over time; see for example Farber and Gibbons (1996). In particular, when workers first entered the labor market, they have the same expected ability and thus have the same wage. As a result, their wage variance is zero in the first period. As more information about the workers is revealed over time, the distribution of CEEA becomes more diverse in the sense of a mean-preserving spread, which leads to a larger variance:

\[
\text{Var}(\eta_{n+1}) = E[\text{Var}(\eta_{n+1}|\eta_n)] + \text{Var}[E(\eta_{n+1}|\eta_n)] = E[\text{Var}(\eta_{n+1}|\eta_n)] + \text{Var}(\eta_n).
\]

(2.61)  (2.62)

When wage is linear in the worker's expected ability, the formula above automatically implies that the wage variance increases over time. In our model, wage isn't linear in CEEA. One sufficient condition for the wage variance to increase is that the expected wage change is increasing in worker's CEEA. To see this, write

\[
W(\eta_{n+1}) = W(\eta_n) + d(\eta_n) + \epsilon,
\]

where \(d(\eta_n) = E[W(\eta_{n+1})|\eta_n] - W(\eta_n)\) and \(\epsilon = W(\eta_{n+1}) - E[W(\eta_{n+1})|\eta_n]\). Then the sufficient condition requires that \(d(\eta_n)\) increases with \(\eta_n\).

\[
\text{Var}(W(\eta_{n+1})) = E[\text{Var}(W(\eta_{n+1})|\eta_n)] + \text{Var}[E(W(\eta_{n+1})|\eta_n)] > \text{Var}(W(\eta_n))
\]

(2.63)  (2.64)  (2.65)

where the inequality follows because both \(W(\eta_n)\) and \(d(\eta_n)\) is increasing in \(\eta_n\). For Pareto learning models, we know that \(d(\eta_n)\) is increasing in \(\eta_n\) from (2.46) in Proposition 5, so the variance of the wage distribution increases with cohort age. This is
stated in the following proposition.

**Proposition 9** *For all* \( n \),

\[
\begin{align*}
\text{Var}(\eta_{n+1}) & > \text{Var}(\eta_n) \quad (2.66) \\
\text{Var}(W(\eta_{n+1})) & > \text{Var}(W(\eta_n)). \quad (2.67)
\end{align*}
\]

**Proof.** Follows directly from the discussion above. ■

Another popular measure of increased wage dispersion is the wage ratio at different quantiles. The formula for wage distribution of cohorts (2.34) enables us to derive explicit expression of wage ratio of different quantiles. The proposition below proves by direct computation that for sufficiently high wage quantiles, the wage ratio increases with time. This is consistent with, but stronger than the result in Proposition 8 that wage increase is larger at larger quantiles.

**Proposition 10** Let \( 0 < q_l < q_h < 1 \). For all \( n \), if \( q_h > \frac{a}{a+n} \), then

\[
\frac{W(\eta_{n+1}^{q_h})}{W(\eta_n^{q_h})} > \frac{W(\eta_{n+1}^{q_l})}{W(\eta_n^{q_l})}. \quad (2.68)
\]

**Proof.** See Appendix. ■

The discussion above summarizes the general properties of the wage dynamics of an entire cohort. Our framework can also be used to describe patterns of individual wage dynamics. We finish this section by stating a person effect in wage growth. The proposition below shows that for two workers with the same current wage, the younger one is more likely to obtain a (sufficiently) higher wage in any given time in future. This result can be tested directly using a panel data set.

**Proposition 11** Suppose a worker in cohort \( n \) has CEEA level \( \eta_h \). For all \( d > 0 \) and \( \eta_h > \exp(\frac{a^2}{(a-\beta)\alpha})\eta_l \), the probability that the worker achieve a CEEA level greater than \( \eta_h \) after \( d \) periods decreases with \( n \).
Proof. If a worker in cohort $n$ has CEEA $\eta$, his maximum observed ability $m_n$ must satisfy

$$\frac{\alpha + n - \beta}{\beta + n} m_n^\beta = \eta. \quad (2.69)$$

For the worker to have a CEEA greater than $\eta_h$ in $d$ periods, his maximum observed ability then must be greater than $m_{n+d}$, where

$$\frac{\alpha + n + d}{\beta + n + d} m_{n+d}^\beta = \eta_h. \quad (2.70)$$

Since the conditional distribution of the worker’s innate ability is $Pareto(m_n, \alpha + n)$, the equation above implies that the probability that the worker’s CEEA level is greater than $\eta_h$ after $d$ periods equals

$$\int_{m_{n+d}}^\infty [1 - \left(\frac{m_{n+d}}{x}\right)^\beta] (\alpha + n) m_n^{\alpha+n} x^{\alpha-n-1} dx \quad (2.71)$$

$$= \frac{d}{\alpha + n + d} \left[ \frac{\alpha - \beta + n}{\alpha + n} \frac{\alpha + n + d}{\alpha - \beta + n + d} \right]^{\alpha+n} \left(\frac{\eta_h}{\eta}\right)^{\alpha+n}. \quad (2.72)$$

Take logs of the equation above and then take derivatives with respect to $n$, we get

$$\frac{d(\alpha + n + d) - (\alpha - \beta + n)^2}{(\alpha - \beta + n)(\alpha + n + d)(\alpha - \beta + n + d)}$$

$$+ \frac{1}{\beta} \left( \frac{\beta d}{\alpha - \beta + n}(\alpha + n + d) \right) + \frac{1}{\beta} \log(\frac{\eta_h}{\eta})$$

$$< \frac{\beta}{(\alpha - \beta)\alpha} + \frac{1}{\beta} \log(\frac{\eta_h}{\eta}). \quad (2.73)$$

Therefore, if $\eta_h > \exp(\frac{\beta^2}{(\alpha-\beta)\alpha})\eta$, then (2.73) is negative, which implies that the probability for a worker to reach $\eta_h$ depends negatively on the time the worker takes to arrive at $\eta_h$.

When a worker experiences fast wage growth, there are two natural explanations for it. First, the worker is of high ability and therefore is more likely to produce good...
performances. This possibility suggests that the worker is likely to continue fast wage growth. Second, the worker has been quite lucky in the past and had several good draws, which implies that the worker is likely to experience below average wage growth in the future. Proposition 11 resolves the conflicting predictions by these intuitions by showing that for two workers with the same wage, as long as future wage threshold is large enough, it is more likely for the worker with faster wage growth to reach the threshold in any given amount of time. To a sense of the upper-bound for the threshold: for $\alpha = 3, \beta = 0.7$, we only need $\eta_h/\eta_l > 1.1$.

Another way to understand Proposition 11 is that for workers with the same CEEA, younger workers have larger variance in expected abilities than older ones. This suggests that the proportion of younger workers who have an expected ability above a sufficiently higher than current CEEA level is larger than that in older workers. Therefore, the probability of a younger worker obtaining a (sufficiently) high wage in future is larger than that of an older worker.

### 2.7 Conclusion

In this paper, we develop a model that combines learning and matching to study wage distribution and dynamics. The predictions from our model match with many of the empirical regularities. Our model is quite tractable and offers closed-form solutions to the wage function and wage distribution. The simplicity of the model helps us perform various comparative statics on the wage distribution and dynamics. We hope to use this model as a basis for building more accurate models.

Our results are mainly cast as descriptions of the wage distribution and dynamics of the entire economy because in the main body of the analysis each firm can only hire one worker. However, we showed in Section VI that the results also apply to an economy where firms can hire multiple workers. This indicates that our results can be interpreted as predictions of wage dynamics within an organization as well. We
find that some of our predictions conform well to the evidence on wage and promotion
dynamics inside firms.

In our model, the learning is symmetric: neither the worker himself nor his em-
ployer has superior information over the rest of the parties in the economy about the
worker's productivity. This assumption is clearly far from realistic, and it will be
interesting to extend the model into asymmetric learning. In this way, we anticipate
that we will be able better capture the empirical evidence on labor turnover and wage
dynamics in the internal labor markets. One possible result from this setup is that
for workers with the same experience level, the one with lower tenure level is likely
to have higher mean and higher variance in his future wage growth.

Another important element left out of the model is human capital accumulation.
The early work of Becker (1964) and Mincer (1974) demonstrated the importance of
education, on the job training, and other source of human capital accumulation in
the earning profiles of workers. We left out human capital to single out the effects
of learning and matching on the wage dynamics of a cohort. Our specification is
very similar to a Mincerian wage regression, and our simulation suggests that better
information constitutes an important source of wage increase.

In summary, we have provided a framework that incorporates matching and sym-
metric learning in explaining the various regularities in wage distribution and dynam-
ics. With appropriate interpretations, our results apply both to the whole economy
and to that inside the firm. We hope to extend our model by including asymmetric
learning, human capital accumulation, incentive provision, and other factors to draw
a more accurate picture of wage distribution and dynamics.
2.8 Appendix

Theorem 1: Every Subgame Perfect Equilibrium of the model satisfies the following:

(i): There exists a monotone function of $W(\eta)$ which maps CEEA to wages.

(ii): Each firm maximizes its profit: for a firm of technology $s$, it chooses a worker of productivity potential of to maximize $s\eta - W(\eta)$.

(iii): Labor supply equals labor demand: if in equilibrium firm of technology $s$ chooses worker of CEEA $\eta(s)$, then

$$G(\eta) = \int_0^\infty 1_{\{\eta(s) \leq \eta\}} dF(s),$$

where $G(\eta)$ is the CDF of the CEEA, $1_{\{\eta(s) \leq \eta\}}$ is an indicator function that equals 1 if $\eta(s) \leq \eta$ and 0 otherwise, and $F(s)$ is the CDF of the technology distribution.

Proof: This proof follows from a sequence of lemmas.

Lemma A1: Workers will accept the highest offered wage. If there are several highest offered wages, he will randomize.

Proof. Since every job provides the same learning opportunity, the value of a wage contract to the worker is determined entirely by its wage. ■

Lemma A2: Workers with the same CEEA will be offered the same fixed wage almost surely in any equilibrium.

Proof. We prove by contradiction. Suppose instead there exists worker A and B with the same CEEA, but the equilibrium wage of worker A is higher than that of B: $w_A > w_B$. Now the firm that hires A can deviate by offering $-\infty$ to A, and $\frac{w_A + w_B}{2}$. This deviation allows the firm to hire B. Since the expected output of A and B are the same and the firm now pays less, this is a profitable deviation. ■

Lemma A3: Let $W(\eta)$ be the wage of workers of CEEA receives in equilibrium. Then $W(\eta)$ is monotone and thus piecewise continuous.
Proof. We prove by contradiction. Suppose instead there exists workers with $\eta_1 < \eta_2$ such that $W(\eta_1) > W(\eta_2)$. Now that the firm that hires a worker of $\eta_1$ can deviate by offering $\frac{1}{2} (W(\eta_1) + W(\eta_2))$ to a worker with CEEA $\eta_2$ and $-\infty$ to all other workers. This deviation allows the firm to hire a more productive worker (on average) and pays a lower wage. It is clearly a profitable deviation. Piecewise continuity follows immediately from monotonicity. ■

Lemma A4: Every worker is employed in equilibrium.

Proof. First, we see that if a worker of CEEA is unemployed, then every worker with CEEA smaller than must be unemployed. Second, let $\underline{\eta}$ be the infimum of CEEA of the employed workers. If $\underline{\eta} < 1$, then positive amount of workers will be unemployed and positive amount of firms will have vacancies. Let $j$ be a firm with a vacancy, then by offering $1 + (\underline{\eta} - 1)/3$ to an unemployed worker of CEEA $1 + (\underline{\eta} - 1)/2$ and negative infinity to all others, the firm will make positive profits. So every worker must be employed. ■

Proposition 1: For $s > 1$, the wage as a function of technology can be written as

$$w(s) = \frac{\lambda \sigma^{-\theta}}{1 + \theta} (\theta s^{\theta+1} + 1), \quad (2.74)$$

and wage is convex in $s$.

Proof. By Theorem 2, workers with CEEA $\eta > \eta(1_+)$ are matched with firms of $s > 1$. Moreover, since

$$\Pi(s) = \int_1^s \eta(x)dx, \quad (2.75)$$

so we have

$$w(s) = \eta(s)s - \int_1^s \eta(x)dx \quad (2.76)$$

$$= \int_1^s \eta'(x)dx + w(1_+), \quad (2.77)$$

where $w(1_+) = \lambda \sigma^{-\theta}$ is the right-limit of the wage paid by firms with technology.
\( s = 1. \) Now using \( \eta(s) = \lambda(s)^{-\theta}, \) direct computation gives the wage equation. The convexity follows because \( \theta > 0. \) ■

**Proposition 2:** The wage as a function of CEEA can be written as

\[
W(\eta) = \eta \quad \text{for } \eta < \eta(1_+); \\
W(\eta) = \frac{\lambda^s}{1 + \theta} (\theta \sigma^{\theta+1}(\frac{n}{\chi})^{\theta+1} + 1) \quad \text{for } \eta \geq \eta(1_+),
\]

where \( \eta(1_+) = \lambda \sigma^{-\theta}. \) Wage is weakly convex in CEEA.

**Proof.** For workers with \( 1 < \eta < \eta(1_+) \), that \( W(\eta) = \eta \) is given directly by Theorem 2. For \( \eta > \eta(1_+) \), we use the matching function to obtain that \( s(\eta) = \sigma(\frac{n}{\chi})^{\frac{\theta}{2}}. \) Then the wage formula can be obtained by substituting \( s(\eta) = \sigma(\frac{n}{\chi})^{\frac{\theta}{2}} \) into (2.74). It is clear that the wage is convex in CEEA. ■

**Proposition 7:** For a cohort with age \( n > \alpha \), the median wage of the cohort increases with cohort age.

**Proof.** From Lemma 1, we see that

\[
\Pr(x \geq \eta_n) = \frac{n}{\alpha + n} \left( \frac{\alpha - \beta + n}{\alpha + n} \eta_n \right)^{-\frac{\theta}{2}} \quad \text{for } \eta_n \geq \frac{\alpha + n}{\alpha - \beta + n}; \\
\Pr(x \geq \eta_n) = 1 \quad \text{for } \eta_n < \frac{\alpha + n}{\alpha - \beta + n}.
\]

Since \( n > \alpha \), less than half of the worker have CEEA at the bottom mass. Therefore, the median CEEA level \( \eta_n^{md} \) satisfies

\[
\frac{1}{2} = \Pr(x \geq \eta^{md}) = s(n)(\eta_n^{md})^{-\frac{\theta}{2}},
\]

where

\[
s(n) = \frac{n}{\alpha + n} \left( \frac{\alpha - \beta + n}{\alpha + n} \right)^{-\frac{\theta}{2}}.
\]

Then it is easy to check that \( s(n) \) increases with \( n. \) Therefore, the median CEEA level \( \eta^{md} = (2s(n))^{\frac{\theta}{2}} \) is increasing in \( n. \) Since the wage is monotone with CEEA, so the median wage is increasing with \( n. \) ■
Proposition 8: For \( k > G\left(\frac{\alpha}{\alpha - \beta}\right) \), we have

\[
\frac{\partial(W(\eta_{n+1}^k) - W(\eta_n^k))}{\partial k} > 0. \tag{2.82}
\]

Moreover, for sufficiently large \( n \),

\[
\frac{\partial^2 W(\eta_n^k)}{\partial n \partial \beta} > 0. \tag{2.83}
\]

Proof. Let \( \eta_n^k \) be the \( k \)th quantile CEEA of cohort \( n \). Then by Lemma 1, we see that

\[
\eta_n^k = \left(\frac{s(n)}{1 - k}\right)^{\frac{\alpha}{\alpha}}, \tag{2.84}
\]

where \( s(n) = \frac{n}{\alpha + n} \left(\frac{\alpha - \beta + n}{\alpha + n}\right)^{-\frac{\beta}{\alpha}} \). It is easy to see that

\[
\frac{\partial \eta_n^k}{\partial k} > 0, \quad \frac{\partial \eta_n^k}{\partial n} > 0, \quad \text{and} \quad \frac{\partial^2 \eta_n^k}{\partial n \partial k} > 0. \tag{2.85}
\]

Since the wage is convex in CEEA, we have

\[
\frac{\partial^2 W(\eta_n^k)}{\partial n \partial k} = W''(\eta_n^k) \frac{\partial \eta_n^k}{\partial k} \frac{\partial \eta_n^k}{\partial n} + W'(\eta_n^k) \frac{\partial^2 \eta_n^k}{\partial n \partial k} > 0. \tag{2.86}
\]

This implies that

\[
\frac{\partial(W(\eta_{n+1}^k) - W(\eta_n^k))}{\partial k} > 0. \tag{2.87}
\]

To examine the effect of concavity on the speed of wage increase, we first check that

\[
\frac{\partial \eta_n^k}{\partial \beta} > 0. \tag{2.88}
\]

Moreover,

\[
\frac{\partial^2 \log(\eta_n^k)}{\partial n \partial \beta} = \frac{1}{\alpha} \log(1 - \frac{\alpha}{\alpha + n}) + \frac{1}{\alpha - \beta + n}. \tag{2.89}
\]

which is positive for sufficiently large \( n \). Since \( \frac{\partial \eta_n^k}{\partial \beta} > 0, \quad \frac{\partial \eta_n^k}{\partial k} > 0, \quad \text{and} \quad \frac{\partial^2 \log(\eta_n^k)}{\partial n \partial \beta} > 0, \) we have \( \frac{\partial^2 \eta_n^k}{\partial n \partial \beta} > 0 \).
Finally, we have

$$\frac{\partial^2 W(n^n_k)}{\partial m \partial \beta} = W''(n^n_k) \frac{\partial n^n_k}{\partial m} \frac{\partial n^n_k}{\partial \beta} + W'(n^n_k) \frac{\partial^2 n^n_k}{\partial m \partial \beta} > 0.$$  \hspace{1cm} (2.90)

\[\square\]

Proposition 10: Let $0 < q_i < q_h < 1$. For all $n$, if $q_h > \frac{\alpha}{\alpha + n}$, then

$$\frac{W(n^n_{q_h})}{W(n^n_{q_h+1})} > \frac{W(n^n_{q_h})}{W(n^n_{q_i})}.$$  \hspace{1cm} (2.91)

Proof. Again, define

$$s(n) = \frac{n}{\alpha + n} \left( \frac{\alpha - \beta + n}{\alpha + n} \right)^{-\frac{\beta}{\alpha}}.$$  \hspace{1cm} (2.92)

Since $q_h > \frac{\alpha}{\alpha + n}$, we have $1 - q_h = s(n)(r^n_{q_h})^{-\frac{\beta}{\alpha}}$. Therefore,

$$n^n_{q_h} = \left( \frac{s(n)}{1 - q_h} \right)^{\frac{\beta}{\alpha}}.$$  \hspace{1cm} (2.93)

To calculate $n^n_{q_i}$, there are two possible cases. First, if $q_i > \frac{\alpha}{\alpha + n}$, then we have

$$n^n_{q_i} = \left( \frac{s(n)}{1 - q_i} \right)^{\frac{\beta}{\alpha}}.$$  \hspace{1cm} (2.94)

Therefore, we have

$$\frac{W(n^n_{q_h})}{W(n^n_{q_i})} = \frac{\theta \sigma^{\beta+1} \lambda \frac{\beta - 1}{\theta} \frac{s(n)}{1 - q_h} \frac{\beta + 1}{\theta} \frac{s(n)}{1 - q_h}}{\theta \sigma^{\beta+1} \lambda \frac{\beta - 1}{\theta} \frac{s(n)}{1 - q_i} \frac{\beta + 1}{\theta} \frac{s(n)}{1 - q_i}} + 1.$$  \hspace{1cm} (2.95)

This expression increases with $s(n)$, which increases $n$, so we have

$$\frac{W(n^n_{q_h})}{W(n^n_{q_i+1})} > \frac{W(n^n_{q_h})}{W(n^n_{q_i})}.$$  \hspace{1cm} (2.96)

135
Second, if \( q < \frac{\alpha}{\alpha + n} \), then we have

\[
W(\eta_m^q) = \frac{\alpha + n}{\alpha - \beta + n} > (\frac{s(n)}{1 - q})^\frac{\alpha}{\alpha - n}. \tag{2.97}
\]

Now if \( q < \frac{\alpha}{\alpha + n + 1} \), then

\[
W(\eta_{n+1}^{q_i}) = \frac{\alpha + n + 1}{\alpha - \beta + n + 1} < W(\eta_m^q). \tag{2.98}
\]

Since \( W(\eta_{n+1}^q) > W(\eta_m^q) \), we get

\[
\frac{W(\eta_{n+1}^q)}{W(\eta_{n+1}^{q_i})} > \frac{W(\eta_m^q)}{W(\eta_m^q)}. \tag{2.99}
\]

Now if \( q > \frac{\alpha}{\alpha + n + 1} \), then we have

\[
\frac{W(\eta_m^q)}{W(\eta_{n+1}^{q_i})} < \frac{\theta \sigma^{\theta+1} \lambda^{\frac{\theta+1}{\theta}} (\frac{s(n)}{1 - q})^{\frac{\theta+1}{\theta} - 1} + 1}{\theta \sigma^{\theta+1} \lambda^{\frac{\theta+1}{\theta}} (\frac{s(n)}{1 - q})^{\frac{\theta+1}{\theta} - 1} + 1} \tag{2.100}
\]

\[
< \frac{\theta \sigma^{\theta+1} \lambda^{\frac{\theta+1}{\theta}} (\frac{s(n+1)}{1 - q})^{\frac{\theta+1}{\theta} - 1} + 1}{\theta \sigma^{\theta+1} \lambda^{\frac{\theta+1}{\theta}} (\frac{s(n+1)}{1 - q})^{\frac{\theta+1}{\theta} - 1} + 1} \tag{2.101}
\]

\[
= \frac{W(\eta_m^q)}{W(\eta_{n+1}^{q_i})}. \tag{2.102}
\]

This finishes the proof. ■

Lemma 2 gives an explicit formula of the CEEA distribution \( G \), which enables us to write a formula that describes the equilibrium matching between CEEA and firm tec.
Bibliography


Chapter 3

Promotion, Turnover, and Adverse Selection

This paper develops a model that examines the optimal assignment of workers into jobs under adverse selection. Workers differ by their disutility of effort; jobs differ by their productivity and ease of effort-monitoring. Firms would like to assign hard workers to "managerial" jobs because efforts in these jobs are harder to monitor. To prevent the lazy workers from mimicking the hard workers, we study the use of two instruments at firms' disposal: requiring long hours and distorting job assignments. The model has an essentially unique separating equilibrium. In equilibrium, workers are required to exert inefficiently high levels of effort in earlier stages of their career and firms commit to promote only a fraction of qualified workers.

We also consider two basic comparative statics: the effects of easier mobility across firms, and an increase in the relative productivity of managerial jobs. When mobility costs decline, work effort increases unambiguously. Task assignment also becomes more efficient when the decrease in mobility costs is sufficiently large. Similarly, an increase in the relative productivity of managerial jobs unambiguously raises the effort levels required of promotion track workers. Because higher productivity in these jobs can worsen the adverse selection problem, however, task assignment need not become more efficient on balance. These predictions are consistent with stylized facts about
the hours and turnover patterns over time within the U.S, as well as the comparison between U.S. and Europe.

3.1 Introduction

When jobs vary in the degree to which productive effort is observable, firms face a hard problem. Some people can work harder than others, and efficiency often demands that hard workers (high types) be assigned to jobs that are difficult to monitor: for example, firms would like to put hard workers into management positions. The problem arises when the hard workers’ efficient career path is more attractive to everyone, so that firms are faced with an adverse selection problem. This paper examines the optimal solution to this problem by analyzing the properties of separating contracts in market equilibrium (Rothschild & Stiglitz, 1976).

In particular, we study the use of two instruments at the firm’s disposal. First, firms can demand that "promotion track" workers put in long hours (or high effort more generally) at the beginning of their careers. While the marginal hour is inefficient even for hard workers, it is especially onerous for lazier workers and so helps to screen them out. Akerlof (1976) pioneered the idea that workers may be induced to enter an inefficient rat race in order to signal their productivity, under the assumption that the disutility of poor work conditions is negatively correlated with ability. Holmstrom (1982) showed that the same result can obtain when workers themselves do not know their productivity, since signals of high work effort and high ability are conflated for firms. Holmstrom’s model differs from rat race models in that there is no heterogeneity in the disutility of effort and no asymmetric information; this paper adopts the more classical view that work effort functions as a screen.

There is evidence that educated and professional workers view long work hours as important for their careers. A review by the United Kingdom’s Institute for Employment Studies (Kodz et al., 2003) summarizes much of the evidence on the causes
and consequences of long hours, with a focus on the UK and Europe. While at least as many workers seem to express a desire for longer work hours (with a commensurate increase in pay) as for shorter work hours, this pattern reflects large numbers of part-time workers who feel they are underemployed. As the number of hours worked increases, the proportion of survey respondents who would like to work fewer hours for less pay goes up sharply; a substantial majority of those working more than 48 hours per week would like to work less. In a narrower context, Lander, Rebitzer & Taylor (1996) find that two thirds of associates at large American law firms would prefer to take a wage increase in the form of reduced hours rather than increased earnings. They argue that the associates do not reduce their hours because long hours are viewed as a requirement for making partner. This perception appears to be correct: current partners are less likely to recommend hypothetical candidates for promotion when the candidate restricts hours for family reasons.

The second instrument we study is the use of noise in the promotion process. In our model, it is efficient for the hard workers to be promoted to managerial positions, but these promotions are valued more highly by lazier workers. While our model takes a reduced form approach to explaining this preference, our justification is that workers with a high disutility of effort have more to gain from job assignments where effort is difficult to monitor. In any case, the adverse selection problem will be at its most severe when this ranking of preferences holds. On the flip side, such preferences allow the firm to resolve the adverse selection problem by committing itself to leave some hard workers unpromoted. The efficiency cost of inefficient future task assignment is counterbalanced by the benefit that work hours can be distorted less than they would otherwise need to be.

To the best of our knowledge, this is a novel explanation for inefficient assignment of workers to jobs. Waldman (1984) views such inefficiencies as arising from incumbent firms’ informational advantage over outside firms. Because outside firms are able to observe workers’ jobs but only the incumbent can directly observe ability, there is an incentive to distort the assignment to jobs in order to lower outside wage
offers. Malcolmson & Fairburn (2001) argue that firms may prefer to use promotions rather than monetary rewards as incentives; this results in greater accountability for the managers who distribute rewards, because promoting favorites will lower the future productivity of the manager’s division. However, the use of promotions as an incentive device means that some employees will be inefficiently moved into a higher job after getting lucky. Finally, Lazear (2002) studies the appearance of inefficient task assignment expressed in the Peter Principle\(^1\) and explains it as a manifestation of mean reversion.

We consider two basic comparative statics within the model. We first consider the effects of easier mobility across firms. Much of the literature views turnover as arising from new information about a worker’s relative productivity across firms or sectors (e.g., Jovanovic (1979), Gibbons et al. (2005), Li (2007a)). The resulting job changes are therefore welfare enhancing because of better allocation of workers to jobs. In our model, by contrast, turnover is always inefficient in the sense that workers’ productivity is always lower in outside firms for a given job\(^2\). Turnover occurs nonetheless because mobility costs are unknown to the incumbent firm, which tries to extract ex post rents from the worker (Hashimoto (1981), Li (2007b)). Turnover also occurs because the incumbent firm uses inefficient job assignment as a solution to the adverse selection problem: hard workers who fail to get promoted will leave for a managerial job in another firm unless their mobility cost is particularly high. When mobility costs decline, a larger proportion of workers who take the promotion track will therefore end up in managerial jobs, undermining the purpose of distorting task assignment in the first place. In effect, the “threat” of non-promotion is no longer as effective against lazier workers who try to mimic the hard workers. This affects the firm’s tradeoff between work effort distortions and job assignment distortions. As a result, when mobility costs decline, work effort increases unambiguously. Task assignment becomes more efficient when the decrease in mobility costs is sufficiently

\(^1\)“Every employee tends to rise to the level of incompetence” (Peter & Hull. 1970).
\(^2\)Though see Bernhardt & Scoones (1998) for a model where the potential for mobility across employers generates inefficient investment in human capital.
stark, though interestingly the same result may fail to hold locally.

We next consider the effects of an increase in the relative productivity of managerial jobs. We interpret this increase as a form of skill-biased technical change (SBTC) within the context of our model; SBTC is widely believed to have been an important phenomenon in recent decades. More specifically, tasks where effort is difficult to observe (our broad conceptualization of the model's managerial jobs) are likely to overlap considerably with non-routine cognitive tasks, and there is evidence that SBTC takes the form of increased demand for the latter (Autor, Levy & Murnane, 2003). In our model, increases in the productivity of managerial jobs make distortions in task assignment less attractive relative to distortions in work effort. This unambiguously raises the effort levels required of promotion track workers. Because higher productivity in managerial jobs can worsen the adverse selection problem, however, task assignment need not become more efficient on balance.

We are motivated to examine these comparative statics by two sets of stylized facts. First, work hours tend to be higher in the United States than in continental European countries, but there has been substantial disagreement over the source of the difference. One possible explanation is that excessive hours are used more often in the United States as a solution to adverse selection problems. Our model supports such an interpretation, since the cost of job mobility seems to be higher and the return to high-level jobs lower in continental Europe. Secondly, work hours at the top of the wage distribution have been increasing over the past several decades within the United States. Over a very similar time frame, the returns to high level jobs have increased, and the popular conception is certainly that employer stability has declined.

The remainder of the paper is organized as follows. Section II sets up the model and defines a separating Perfect Bayesian Equilibrium in our context. Section III

---

3 Plausibly, task assignment is more efficient in the United States, but the model does not predict this relationship unambiguously.

4 The academic evidence on this point is not clear cut. Most papers find that the job mobility has increased at most moderately. However, Stewart (2002) shows that between 1975-2000 the employment to employment transition rate in the U.S. has increased 45% for men and 58% for women.
is the core of the paper and solves the model. Section IV presents applications to the two comparative statics discussed above, while Section V discusses interpretation issues and concludes.

3.2 Model Setup

We set up the model formally in this section. Subsection 2.1 describes the basic environment, including the objective functions of the firms and the workers. Subsection 2.2 gives the information structure and timeline. The solution concept of the model is introduced in Subsection 2.3.

3.2.1 Workers and Firms

There is a continuum of workers who live for two periods. The workers differ in their costs of effort. With probability \( \mu \), a worker is of high type, and his cost of effort is \( c(e) \), where \( c(0) = 0, c'(e) > 0, \) and \( c''(e) > 0 \). With probability \( 1 - \mu \), a worker is of low type, and his cost of effort is \( \theta c(e) \), where \( \theta > 1 \). The utility of each worker is additive across periods. In each period, a worker’s utility equals the wage he receives minus his cost of effort, \( e \). In summary, the utility function of the two types can be written as follows:

\[
U_h(w_1, w_2, e_1, e_2) = \sum_{t=1}^{2} w_t - c(e_t); \\
U_l(w_1, w_2, e_1, e_2) = \sum_{t=1}^{2} w_t - \theta c(e_t).
\]

There are \( N > 2 \) ex ante identical firms. Each firm has two types of job. Job 1 is a "standard" job, in the sense that a) its output can be contracted perfectly, and b) its output does not depend on the worker’s type conditional on effort. Specifically,
we assume that the output of a worker in job 1 equals his effort level, \( e \). Moreover, job 1 is the only type that can be offered in period 1\(^5\).

In period 2, each firm also has the option of offering job 2 besides offering job 1. Job 2 is a "managerial" job in the sense that a) its output cannot be contracted at all, and b) its output depends only on the type of the worker. In particular, the worker's output on Job 2 is \( Y_h \) if he is a high type, and his output is \( Y_l < Y_h \) if he is a low type. Because the outputs cannot be contracted upon, we may assume that job 2 requires \( e = 0 \).

This specification of job 2 captures the reduced form of a more complex model, where effort in job 2 is harder for the principal to observe than effort in job 1. The optimal job 2 contract design will trade off incentives for production against other concerns (e.g., multitasking or, with different utility functions, risk aversion). Output and efficiency are higher if job 2 is filled by a high type, because the low types' higher disutility of effort makes the incentive problem more severe. At the same time, however, low types benefit more from promotion to job 2 than high types do, provided the types are offered identical contract terms. This is because low types gain from distorting effort downward to a larger degree. This fundamental dilemma—that the types the firm would least like to promote are the types who stand to gain from promotion the most—is reflected in our reduced form setup.

When a worker works for a different firm in period 2, there is an efficiency loss \( k \) associated with the switching of the firms, where \( k \) is a random value drawn uniformly from \([0, K]\). Possible interpretation of \( k \) includes firm-specific human capital, substitutability of jobs across firms or moving costs across firms; we discuss its interpretation more in Section V. We assume that

\[
Y_h - K > \max_e \left( e - c(e) \right); \tag{3.2}
\]

\[
Y_l < \max_e \left( e - \theta c(e) \right).
\]

\(^5\)This assumption is made for simplicity. One can show that if the proportion of high type workers is small enough, firms will only want to offer Job 1 in period 1.
Therefore, it is efficient in period 2 to assign the high type to job 2 and the low type to job 1 if the types were known.

The output of a firm is additive across its outputs on the two jobs. The output on each job has constant return to scale in the number of workers. The payoff of a firm in each period equals its total outputs minus the total wages it pays.

### 3.2.2 Information Structure and Timeline

Before discussing the timeline, we first specify the information structure on worker types and the efficiency loss $k$. We assume that workers know about their own types, but firms do not. Moreover, we assume that the incumbent knows the distribution of the efficiency loss $k$, but not the exact realization of $k$ at the beginning of period 2. On the other hand, outside firms know the exact realization of $k$.

At the beginning of $t = 1$, all firms offer contracts to all workers simultaneously. Each contract takes the form of a triple $(w, e, p)$, where $w$ is the wage paid to the worker in period 1, $e$ is the effort required on job 1 in period 1, and $p$ is the probability that the worker will be "promoted" to job 2 in period 2, which means that he is offered a contract on job 2 by his period 1 employer. We assume that each firm may commit to the promotion probability $p$.

The firms' ability to commit to $p$ but not to contract terms in period 2 merits comment. There are at least two justifications for this assumption: First, we can conceptualize this model as the description of a subgame between one cohort of workers and an infinitely-lived firm. If past promotion decisions are observable to future cohorts of workers, then firms may not want to deviate from the "announced" $p$ for fear of reputational consequences. Period 2 contract terms may not be observable to the same degree as job assignments. Secondly, firms may sink investments into their production structure at the beginning of the game, and each production structure may require a fixed proportion of managerial workers in period 2. In the context of
an overlapping cohort model, this explanation also requires some rigidity in the firm’s ability to adjust the number of new hires into job 1.\footnote{Such rigidity could come from search frictions in the labor market or from imperfect substitution between old and young workers performing job 1.}

For simplicity, we assume that for each firm $i \in \{1, \ldots, N\}$, it is restricted to offer only one of the two types of contracts: $\{(w_{1h}^i, e_{1h}^i, p_h^i), (w_{1l}^i, e_{1l}^i, 0)\}$.\footnote{In principle, each firm $i$ can offer a menu of contracts $\{(w_{1m}^i, e_{1m}^i, p_m^i)\}_{m=1}^{M_i}$. But since there are only two types, it is standard to show that it is without loss of generality to restrict the set to admissible contracts to consist of two types of contracts only. Moreover, it can be shown that one can restrict the promotion probability of one type of contract to be zero. This greatly simplifies the notations.} We call the first type “high type” contracts because it offers the opportunity for promotion and thus can be thought of as intended for the high types. We call the second type ”low type” contracts because it does not allow for promotion and is thus intended for low types.

Once all contracts are offered, each worker chooses a contract and works for the associated firm. In particular, each worker’s strategy is a function $D_1 : R^{6N} \rightarrow \{1, \ldots, N\}$, where $\tau \in \{h, l\}$ stands for the worker’s type, $R^{6N}$ is the $6N$–dimensional space of real numbers, and $\{1, \ldots, N\}$ is the set of $N$ firms. We have $6N$ here because there are $N$ firms and the contracts offered by each firm form a 6-tuple $\{(w_{1h}^i, e_{1h}^i, p_h^i), (w_{1l}^i, e_{1l}^i, 0)\}$. At the end of period 1, outputs are realized and workers are paid. The contract choice of the workers are observed by all firms.

At the beginning of $t = 2$, each incumbent (the period 1 employer of the worker), say firm $i$, promotes proportion $p_h^i$ of workers who accept its ”high-type” contract in period 1: these promoted workers are offered to work in job 2 with a wage of $w_{2p}^i$. For the remaining $1 - p_h^i$ workers who accept the ”high-type” contract, firm $i$ offers a contract of $(w_{2h}^i, e_{2h}^i)$. For workers who accepted the ”low type” contract, firm $i$ offers a contract of $(w_{2l}^i, e_{2l}^i)$. The incumbent’s promotion decisions and wage offers are observed by all outside firms. After observing the incumbent’s choices, outside firms offer contracts to both the promoted and non-promoted workers. We restrict each firm to offer ”job 2 contracts” to outside workers who took a high-type contract in period 1 and to offer ”job
to outside workers who took a low-type contract in period 1. This restriction is without loss of generality and helps to simplify the notation. In particular, firm $i$'s strategy can be written as $((w_{out,p}^i, 0), (w_{out,n}^i, 0), (w_{out,l}^i, e_{out,l}^i))$, where $w_{out,p}^i$ is the wage firm $i$ offers to outside workers who are promoted (by their incumbent firm), the 0 behind it reflects the fact that the effort in Job 2 is normalized to 0, $w_{out,n}^i$ is firm $i$'s offer to outside workers who are not promoted but had taken the "high-type" contract, and $w_{out,l}^i$ and $e_{out,l}^i$ are firm $i$'s wage and the associated effort level offer to outside workers who took "low-type" contracts.

We do not allow the incumbent to make a counter offer to a worker who is successfully hired away by an outside firm. The incumbent would typically want to do so ex post, since workers always have higher productivity if they remain with their original employer. However, if the incumbent cannot observe outside firms' wage offers directly, it will want to commit against counter offers ex ante: otherwise all workers would claim to have received an offer from a $k = 0$ rival. More generally, the exact source of turnover is not central to the main conclusions of our model, and so there is little gain to modeling offers and counter offers more elaborately. Turnover here functions primarily as a "second chance" for non-promoted high types to get into job 2.

Note also that we constrain low types who mimic high types in period 1 and fail to be promoted to receive the same contract offers as non-promoted high types. Once outside offers have been made and a low type has rejected them (because $k$ is very high), he certainly has an incentive to reveal his type to the incumbent firm, and this revelation is credible. The incumbent firm could then offer a new job 1 contract that specifies a lower (and more efficient) level of effort appropriate to the low type's disutility. Ex ante, of course, firms would like to commit themselves against this kind of renegotiation, since renegotiation makes it more attractive for a low type to choose the high type contract in period 1. We assume that firms are able to make such commitments, designing institutional barriers that prevent workers from switching
from one "career track" to another. This modeling choice is not central to our argument.

After receiving all the offers, workers choose a contract and the associated employer. In particular, the worker's contract can be written as \((D_{2pr}, D_{2nt}, D_{2lt})\), where \(\tau \in \{h, l\}\) stands for the two different types, \(D_{2pr} : R^N \rightarrow \{1, \ldots, N\}\) is the choice of the worker of type \(\tau\) who were promoted by the incumbent firm, \(D_{2nt} : R^{N+1} \rightarrow \{1, \ldots, N\}\) is the choice of the worker of type \(t\) who took high-type contract in period 1 but not promoted, and \(D_{2lt} : R^{2N} \rightarrow \{1, \ldots, N\}\) is the choice of the worker who took low-type contract in period 1.

After the workers accept contracts, outputs are realized, wages are paid, and the game ends.

### 3.2.3 Separating Equilibrium

According to the timing and information structure, the strategy of a worker of type \(\tau\) is a 2-tuple \((D_{1\tau}, D_{2\tau})\), where \(D_{2\tau} = (D_{2pr}, D_{2nt}, D_{2lt})\) is the decision of a worker of type \(\tau\) depending on whether he's promoted, not promoted (but took the high-type contract), or took the low-type contract in period 1. The strategy of firm \(i \in \{1, \ldots N\}\) is a 3-tuple \((C_i, C_{in}, C_{out})\), where \(C_i = ((w_{ih}, e_{ih}, p_{ih}), (w_{il}, e_{il}, 0))\) stands for firm \(i\)'s contract offer in period 1, \(C_{in} = ((w_{2p}, 0), (w_{2n}, e_{2n}), (w_{2l}, e_{2l}))\) is firm \(i\)'s period 2 offer to its own workers, depending on whether the worker is promoted, not promoted (but took the high-type contract), or took the low-type contract in period 1, and finally \(C_{out} = ((w_{out,p}, 0), (w_{out,n}, 0), (w_{out,l}, e_{out,l}))\) is firm \(i\)'s contract offer to outside workers.

Given the strategies, we solve the Perfect Bayesian Equilibrium (PBE) of the model. We focus on the separating equilibrium where the high type workers take
the high type contracts and low type workers take the low type contracts. The PBE requires that the strategies of the worker and the firms to be sequentially optimal given their beliefs and that their beliefs be determined from the Bayes’ Rule wherever possible. In particular, we require a separating PBE to satisfy

1. The worker’s equilibrium period 2 contract choice \( D_{2t}^* \) is optimal given any period 1 strategy \( D_{1t} \) of the worker and any strategies of the firms \( \prod_{i=1}^{N} (C_i^1, C_i^1, C_{out}^i) \), given his belief.

2. For each firm \( i \), its period 2 strategy \( C_{out}^{i*} \) is optimal given any period 1 strategy \( D_{1t} \) of the worker, any period 1 and beginning of period 2 strategy of the firms \( \prod_{i=1}^{N} (C_i^1, C_i^1) \), the period 2 equilibrium strategy of the worker \( D_{2t}^* \), the period 2 equilibrium strategies of all other firms \( C_{out}^{-i*} \), and firm \( i \)’s belief.

3. For each firm \( i \), its period 2 strategy \( (C_i^{i*}, C_i^{i*}) \) is optimal given any period 1 strategy \( D_{1t} \) of the worker, any period 1 strategy of the firms \( \prod_{i=1}^{N} (C_i^1) \), the period 2 equilibrium strategy of the worker \( D_{2t}^* \), the period 2 strategies of all other firms \( (C_{out}^{-i*}, C_{out}^{-i*}) \), and its belief.

4. The worker’s contract choices \( (D_{1t}^*, D_{2t}^*) \) are optimal given the strategies of the firms: \( \prod_{i=1}^{N} (C_i^1, C_i^1, C_{out}^i) \) and its belief. Moreover, the worker of type \( h \) selects contracts with \( p > 0 \) and the worker of type \( l \) selects contracts with \( p = 0 \).

5. For each firm \( i \), its strategy \( (C_i^{i*}, C_i^{i*}, C_{out}^{i*}) \) is optimal given its beliefs, the worker’s equilibrium strategy \( (D_{1t}^*, D_{2t}^*) \), and the equilibrium strategies of all other firms \( (C_{out}^{-i*}, C_{out}^{-i*}) \).

6. The players update their beliefs according to Bayes’ rule whenever possible. This means that the worker knows his type. Each firm’s belief about the worker’s ability equals the prior in period 1. At the beginning of period 2, each firm believes that a worker who took a high-type contract \( (p > 0) \) is of high type with probability 1, and that a worker who took a low-type contract \( (p = 0) \) is of low type with probability 1.
3.3 Analysis of the Model

In this section, we find the separating PBE by construction. Since a PBE requires all players to be sequentially rational given their beliefs, we solve the model by backward induction, moving from the first to the sixth condition listed at the end of Section III. The game involves a large number of potential nodes, although many of them are never reached along the equilibrium path. In particular, note that workers can be in one of six situations at the beginning of period 2: the worker might have taken the low type contract in period 1; the worker might have taken the high type contract and been promoted; the worker might have taken the high type contract and failed to be promoted; and for each of these possibilities, the worker could be either a high type or a low type. Our focus on separating PBE simplifies the analysis, since we know that in equilibrium all firms believe a worker is a high type if and only if he accepted a high type contract in period 1.

First, we observe that a worker will always choose a contract that gives him the highest utility. Workers can be classified into three groups at the beginning of period 2. If the worker has chosen a low-type contract in period 1, then he will receive in period 2 contracts on job 1 only. In this case, a low-type worker chooses a contract that maximizes $w - Oc(e)$, and a high-type worker chooses a contract that maximizes $w - c(e)$. If a worker has chosen a high type contract and is promoted, he will receive in period 2 contracts on job 2 only. Since there is no disutility of effort in job 2, the worker's optimal response is to choose a contract that offers the highest wage. If a worker has chosen a high type contract and is not promoted, he will receive a job 1 offer from the incumbent and job 2 offers from rivals. Suppose that the offer from the incumbent is $(w_{2m}, e_{2n})$ and the maximum wage offer across rival firms is $w_{\text{out},n}^{\text{max}}$; the worker elects to stay with the incumbent firm when $w_{2m} - c(e_{2n}) > w_{\text{out},n}^{\text{max}}$ if he is a high type and $w_{2m} - \theta c(e_{2n}) > w_{\text{out},n}^{\text{max}}$ if he is a low type. In case there are ties, workers randomize. This is stated formally in the following lemma.

**Lemma 12** Suppose firm 1 is the incumbent in period 2. The equilibrium decision
rule of the worker can be written as follows:

\[ D_{2pr}^-\left(\{(w^i, e^i)\}_{i=1}^N\right) \in \{j : w^i \geq w^j \text{ for all } i = 1, \ldots, N\} \]  
\[ D_{2dh}^-\left(\{(w^i, e^i)\}_{i=1}^N\right) \in \{j : w^i - c(e^n) \geq w^i - c(e^i) \text{ for all } i = 1, \ldots, N\} \]  
\[ D_{2th}^-\left(\{(w^i, e^i)\}_{i=1}^N\right) \in \{j : w^i - \theta c(e^n) \geq w^i - \theta c(e^i) \text{ for all } i = 1, \ldots, N\} \]  
\[ D_{2ll}^-\left(\{(w^i, e^i)\}_{i=1}^N\right) \in \{j : w^i - 1_{\{n=1\}} \theta c(e^n) \geq w^i - 1_{\{n=1\}} \theta c(e^i) \text{ for all } i = 1, \ldots, N\} \]  
\[ D_{2ll}^+\left(\{(w^i, e^i)\}_{i=1}^N\right) \in \{j : w^i - 1_{\{n=1\}} c(e^n) \geq w^i - 1_{\{n=1\}} c(e^i) \text{ for all } i = 1, \ldots, N\} \]

Second, we consider the offers of the firms to outside workers. Since all firms are ex ante identical, the identity of the firms are irrelevant (other than when it is an incumbent or an outside firm), so we suppress the superscript \( i \) in our discussion. There are three types of workers at this stage: the promoted, the non-promoted who took a high-type contract in period 1, and those who took a low-type contract.

In the separating equilibrium, the firm believes that workers who took a low-type contract are of low type with probability 1. In this case, suppose the worker is offered a contract of \((w_2, e_2)\) by the incumbent. Therefore, to hire the worker away from the incumbent, an outside firm needs to provide a utility of at least \( w_2 - \theta c(e_2) \). If there were only one outside firm, its maximal profit from hiring away the worker is equals

\[
\max_{e,w} \quad -\theta c(e) - w - k \\
\text{s.t.} \quad w - \theta c(e) \geq w_2 - \theta c(e_2)
\]  

(3.4)

It is easy to see that the solution to this program is to require the worker to put in effort of \( e^*_2 \), where \( \theta c'(e^*_2) = 1 \), and to give the worker \( w = w_2 - \theta(c(e_2) - c(e^*_2)) \) whenever he is hired. It follows that the firm would receive a negative profit (conditional on hiring this worker) whenever

\[ k > e^*_2 - w_2 - \theta(c(e^*_2) - c(e_2)). \]
When this condition holds, one optimal response for the outside firm is to offer a contract that satisfies

\[ w_{\text{out},l}^* = -\infty; \]  
\[ e_{\text{out},l}^* = e_1^*. \]  

When \( k < e^* - w_{21} - \theta(c(e_1^*) - c(e_{21})) \), a single outside firm could obtain positive profits. Because there is more than one outside firm \((N > 2)\), in equilibrium the outside firms "bid up" the wage to reduce their profit to zero. In this case, we must have

\[ w_{\text{out},l}^* = e_1^* - k. \]

In summary, when a worker who took a low-type contract in period 1 is offered \((w_{21}, e_{21})\) by the incumbent, the equilibrium outside offers satisfy

\[ w_{\text{out},l}(w_{21}, e_{21}) = \begin{cases} e_1^* - k & \text{if } k \leq e_1^* - w_{21} - \theta(c(e_1^*) - c(e_{21})) \\ -\infty & \text{otherwise} \end{cases} \]  
\[ e_{\text{out},l}^* = e_1^*. \]

Now consider workers who were promoted. In the separating equilibrium, the firm believes that the promoted workers are of high type with probability 1. Suppose the promoted worker is offered a wage of \(w_{2p}\) by the incumbent. Since the outside firm can hire away the worker only by offering \(w_{\text{out},p} \geq w_{2p}\), it could make nonnegative profit by hiring the worker if and only if

\[ Y_h - k - w_{2p} \geq 0. \]
Similar reasoning as above gives that one optimal strategy of the outside firms is

\[ w_{out,p}^*(w_{2p}) = \begin{cases} Y_h - k & \text{if } k \leq Y_h - w_{2p}; \\ -\infty & \text{otherwise.} \end{cases} \tag{3.8} \]

Finally, consider workers who took the high-type contract in period 1 but were not promoted. In the separating equilibrium, the firm believes that the promoted workers are of high type with probability 1. Suppose this nonpromoted worker is offered a contract of \((w_{2n}, e_{2n})\) by the incumbent. Therefore, an outside firm believes that it can hire away the worker only by offering \(w_{out,n} \geq w_{2n} - c(e_{2n})\). This means that an outside firm believes that it can make non-negative profit if only if

\[ Y_n - k - (w_{2n} - c(e_{2n})) \geq 0. \]

Again, the zero profit condition of outside firms implies that one optimal strategy of the outside firms (given their beliefs) satisfies

\[ w_{out,n}^*(w_{2n}, e_{2n}) = \begin{cases} Y_h - k & \text{if } k \leq Y_n - (w_{2n} - c(e_{2n})); \\ -\infty & \text{otherwise.} \end{cases} \tag{3.9} \]

These three cases complete the characterization of the equilibrium outside offer strategy in period 2. Lemma 2 below summarizes the discussions above.

**Lemma 13** In period 2, one equilibrium outside offer strategy is as follows: for each
firm $i$, 

$$
\begin{align*}
  w_{\text{out},l}^*(w_{2l}, e_{2l}) &= e^*_i - k & \text{if } k \leq e^*_i - w_{2l} - \theta(c(e^*_i) - c(e_{2l})), \\
  &= -\infty & \text{otherwise,}
\end{align*}
$$

$$
\begin{align*}
  e^*_{\text{out},l} &= e^*_i; \\
  w_{\text{out},p}^*(w_{2p}) &= Y_h - k & \text{if } k \leq Y_h - w_{2p}, \\
  &= -\infty & \text{otherwise};
\end{align*}
$$

$$
\begin{align*}
  w_{\text{out},n}^*(w_{2n}, e_{2n}) &= Y_h - k & \text{if } k \leq Y_n - (w_{2n} - c(e_{2n})), \\
  &= -\infty & \text{otherwise.}
\end{align*}
$$

Moreover, any equilibrium strategy must induce the same $\max\{w_{\text{out},l}^*, \max\{w_{\text{out},p}^*, \max\{w_{\text{out},n}^*(w_{2n}, e_{2n})\}\}$ when $k$ is smaller than the respective threshold.

Third, we characterize the equilibrium contract offer of the incumbent firm at the beginning of period 2. There are again three types of workers to analyze: the promoted, the non-promoted who took a high-type contract in period 1, and those who took a low-type contract.

Suppose the incumbent offers $(w_{2l}, e_{2l})$ to a worker who who took a low-type contract in period 1. Given the optimal response of outside firms in equation (3.7), the incumbent keeps the worker if $k > e^*_i - w_{2l} - \theta(c(e^*_i) - c(e_{2l}))$. It follows that in equilibrium the incumbent chooses $(w_{2l}, e_{2l})$ to maximize

$$
\begin{align*}
  \Pr(k > e^*_i - w_{2l} - \theta(c(e^*_i) - c(e_{2l}))) & (e_{2l} - w_{2l}) \\
  &= \left[1 - \frac{e^*_i - w_{2l} - \theta(c(e^*_i) - c(e_{2l}))}{K}\right] (e_{2l} - w_{2l}).
\end{align*}
$$

The solution to this maximization problem gives that

$$
\begin{align*}
  e^*_{2l} &= e^*_i; \\
  w^*_{2l} &= e^*_i - \frac{K}{2}.
\end{align*}
$$
This implies that the equilibrium probability that a worker who took the low-type contract in period 1 stays with the incumbent is

\[ P_l^* = \frac{1}{2}. \]  

(3.16)

Moreover, the threshold of \( k \) above which the incumbent keeps the worker is

\[ k_i^* = \frac{K}{2}. \]  

(3.17)

Furthermore, we can calculate the expected profit of the incumbent in period 2 on this type of worker as

\[ \pi_i^* = \frac{K}{4}. \]  

(3.18)

Next, we consider the promoted workers. Suppose the incumbent offers \( w_{2p} \). Given the optimal response of outside firms in equation (3.8), the incumbent keeps the worker if \( k > Y_h - w_{2p} \). Since the incumbent believes that promoted workers are of high type with probability 1, its maximization problem becomes

\[
\Pr(k > Y_h - w_{2p})(Y_h - w_{2p}) = (1 - \frac{Y_h - w_{2p}}{K})(Y_h - w_{2p}).
\]

(3.19)

The solution to this maximization problem gives that

\[ w_{2p}^* = Y_h - \frac{K}{2}. \]  

(3.20)

This implies that the equilibrium probability that a promoted worker stays with the incumbent is

\[ P_p^* = \frac{1}{2}. \]  

(3.21)

Just as in the case above, the threshold of \( k \) above which the incumbent keeps the promoted worker is

\[ k_p^* = \frac{K}{2}. \]  

(3.22)
and the expected profit of the incumbent in period 2 on a promoted worker is

$$\pi_p^* = \frac{K}{4}. \quad (3.23)$$

Finally, we consider the nonpromoted workers. Suppose the incumbent offers \((w_{2n}, e_{2n})\) to a nonpromoted worker. Given the optimal response of outside firms in equation (3.9), the incumbent keeps the worker if

$$k > Y_n - (w_{2n} - c(e_{2n})) \quad (3.24)$$

It follows that in equilibrium the incumbent chooses \((w_{2n}, e_{2n})\) to maximize

$$\Pr(k > Y_n - (w_{2n} - c(e_{2n})))(e_{2n} - w_{2n}) \quad (3.25)$$

The solution to this maximization problem gives that

$$e_{2n}^* = e_{h}^*, \quad w_{2n}^* = \frac{Y_h + e_{h}^* + c(e_{h}^*) - K}{2}. \quad (3.26)$$

This implies that the equilibrium probability that a nonpromoted worker stays with the incumbent is

$$P_n^* = \frac{1}{2} - \frac{\delta}{2K} \leq \frac{1}{2}, \quad (3.27)$$

where \(\delta = Y_h - (e_{h}^* - c(e_{h}^*)) > 0\) is the productivity difference of the high-type worker between job 2 and job 1 within the same firm. It follows that the threshold of \(k\) above which the incumbent keeps the nonpromoted worker is

$$k_n^* = \frac{K + \delta}{2}, \quad (3.28)$$

and the firm’s expected profit from a nonpromoted worker (who is believed to be a high type) is

$$\pi_n^* = \frac{K}{4} (1 - \frac{\delta}{K})^2 \quad (3.29)$$

These three cases complete the characterization of the equilibrium incumbent offer.
strategy in period 2. Lemma 3 below summarizes the discussions above.

**Lemma 14** At the beginning of period 2, the incumbent’s strategy can be summarized as follows:

\[
\begin{align*}
(w^*_{21}, e^*_{21}) &= (e^*_i - \frac{K}{2}, e^*_i), \\
(w^*_{2p}, 0) &= (Y_h - \frac{K}{2}, 0), \\
(w^*_{2n}, e^*_{2n}) &= (\frac{Y_h + e^*_h + c(e^*_h) - K}{2}, e^*_h),
\end{align*}
\]

and the equilibrium staying probability, cutoff value of \(k\), and the expected profit of the incumbents for the three cases are

\[
\begin{align*}
P^*_l &= \frac{1}{2}, & k^*_l &= \frac{K}{2}, & \pi^*_l &= \frac{K}{4}; \\
P^*_p &= \frac{1}{2}, & k^*_p &= \frac{K}{2}, & \pi^*_p &= \frac{K}{4}; \\
P^*_n &= \frac{1}{2} - \frac{\delta}{2K}, & k^*_n &= \frac{K + \delta}{2}, & \pi^*_n &= \frac{K}{4} (1 - \frac{\delta}{K})^2.
\end{align*}
\]

Lemmas 1-3 characterize the equilibrium strategies of workers and firms in period 2. In each separating PBE, the strategies of the firms and workers are essentially unique\(^\text{10}\). Now we use these results to analyze the equilibrium strategies in period 1. When workers choose contracts in period 1, their decision problems can be classified into two steps: they first decide which types of contracts to accept. Then they choose the best contract within the type. In a separating equilibrium, we require the high-type workers to choose high-type contracts and low types to choose the low-type workers.

Therefore, in the analysis of the separating equilibrium, we need to calculate the expected period 2 payoff of different types of workers by choosing different types

\(^{10}\text{The only nonuniqueness comes from the randomization of the worker when there are ties and from the rival firms’ offers when } k \text{ is so large that it is weakly dominated from them to hire the worker.}\)}
of contracts. First, note that if a worker accepts a low-type contract in period 1, Lemma 3 shows that the incumbent and outside firms will both offer a contract that requires $e_{2l}^* = e_l^*$. Moreover, the incumbent offers $w_{2l}^* = e_l^* - \frac{K}{2}$, and outside firms offer $e_l^* - c(e_l^*) - k$ if $k \leq \frac{K}{2}$ by Lemma 2. Since $k$ is uniformly distributed, we can calculate the expected period 2 utility of both a high type and a low type worker, each of whom accepted a low-type contract in period 1, as

$$ U_{2l}^i = e_l^* - c(e_l^*) - \frac{3K}{8}; $$

$$ U_{2l}^i = e_l^* - \theta c(e_l^*) - \frac{3K}{8}. $$

If a worker accepts a high-type contract in period 1 and is promoted, Lemma 3 shows that the incumbent offers $w_{2p}^* = Y_h - \frac{K}{2}$, and outside firms offer $Y_h - k$ if $k \leq \frac{K}{2}$ by Lemma 2. Given that $k$ is uniformly distributed, we can calculate the expected utility in period 2 of a promoted worker for both a high type and a low type as

$$ U_{2p}^h = Y_h - \frac{3K}{8}; $$

$$ U_{2p}^l = Y_h - \frac{3K}{8}. $$

If a worker accepts a high-type contract in period 1 and is not promoted, Lemma 3 shows that the incumbent offers a contract that requires effort level $e_{2n}^* = e_h^*$ and a wage of $w_{2n}^* = \frac{Y_h + e_h^* + c(e_h^*) - K}{2}$. By Lemma 2, outside firms in equilibrium offer the nonpromoted worker a contract on Job 2 that gives a wage of $w_{out,n}^* = Y_h - k$ whenever $k < k_n^* = \frac{K + \delta}{2}$. It follows that the expected utility of a high type when not promoted in period 2 is

$$ U_{2n}^h = P_n^*(Y_h + e_h^* - c(e_h^*) - K) + (1 - P_n^*)(Y_h - k_n^*), $$

where recall that $P_n^* = \frac{1}{2} - \frac{\delta}{2K}$ is the equilibrium probability that a nonpromoted worker stays with the incumbent in period 2. Note that for a low-type worker who
took the high-type contract but is not promoted, his only difference in utility with the high type in period 2 occurs when the low-type stays with the incumbent. Therefore, we have

\[ U_{2n}^l = U_{2p}^h - P_n^*(\theta - 1)c(e_h^*). \] (3.35)

We should point out that the low types are no more likely to leave the incumbent along the equilibrium path. This results from the equilibrium strategy we choose for outside firms: all outside firms would offer a wage equal to negative infinity when the efficiency loss \( k \) is larger than \( k_n^* \). In other words, both types of workers stay with the incumbent whenever \( k > k_n^* \) and leave otherwise. However, if outside firms deviate by offering wages that barely fail to attract the high types, this deviation will attract the low types. Of course, such a deviation is weakly dominated.

Suppose the high-type contract in period 1 offers a promotion probability of \( p \) in period 2. Then we define the respective period 2 utility of the high and low types who take this contract as

\[
U_{2h}^h(p) = pU_{2p}^h + (1-p)U_{2n}^h; \\
U_{2h}^l(p) = pU_{2p}^l + (1-p)U_{2n}^l.
\] (3.36)

This allows us to write the equilibrium decision rule of the worker at the beginning of period 1 as follows.

**Lemma 15** Let \( D = \{h, l\} \) be the two types of contracts that are intended for the high and low types respectively. The equilibrium period 1 decision of the worker satisfies

\[
D_{1h}^i((w_d^j, e_d^j, p_d^j))_{i=1}^N \in \arg\max\{w_d^j - c(e_d^j) + U_{2h}^h(p_d^j), \ j = 1, ..., N, \ d \in D\}; \\
D_{1l}^i((w_d^j, e_d^j, p_d^j))_{i=1}^N \in \arg\max\{w_d^j - \theta c(e_d^j) + U_{2h}^l(p_d^j), \ j = 1, ..., N, \ d \in D\}.
\]
The last step in our characterization of the separating equilibrium is to specify the equilibrium contract offers at the beginning of period 1. In this model, the utility of the workers are quasilinear in their wage payments. Therefore, it is easy to see that firms must have zero expected payoffs in period 1 in equilibrium since they can compete in their wage offers. Therefore, the contracts that are accepted in period 1 are those that maximize the expected payoffs of each type.

We first consider contracts that cater to the low types. The equilibrium low-type contract maximizes the expected surplus of the low-type workers subject to the incentive constraint that the high types do not want to mimic the low types. As in typical adverse selection models, it can be verified (after the derivation of the optimal contract for the high types) that the incentive constraint of the high-types are slack. Therefore, we look for \((e_{1l}^*, w_{1l}^*)\) that

\[
\max_{e_{1l}, w_{1l}} w_{1l} - \theta c(e_{1l}) + U_{2l}^* \\
\text{such that } e_{1l} - w_{1l} + \pi_{2l}^* \geq 0.
\]

where \(U_{2l}^* = e_i^* - \theta c(e_i^*) - \frac{3K}{6} \) (from equation (3.18)) is the expected utility of a low type worker in period 2 after taking this (low-type) contract, and \(\pi_{2l}^* = \frac{K}{4}\) is the incumbent’s expected profit in period 2 from a low-type contract taker.

The solution of the program above satisfies

\[
e_{1l}^* = e_i^*; \quad w_{1l}^* = e_i^* + \frac{K}{4}.
\]

In other words, the effort level required for the contract is efficient. Moreover, firms lose money in period 1 and recoup their losses in period 2. The equilibrium low-type contract also implies that the expected utility of a low-type worker in the separating equilibrium satisfies

\[
U_i^l = 2e_i^* - 2\theta c(e_i^*) - \frac{K}{8}.
\]
The $\frac{K}{8}$ term reflects the expected loss from moving to an outside firm.

Next, we consider contracts that cater to the high types in period 1. The equilibrium high-type contract maximizes the expected surplus of the high-type workers subject to the incentive constraint that the low types do not want to mimic the high types and that all firms make nonnegative expected profits. Therefore, we look for $(w^*_h, e^*_h, p^*_h)$ to

$$\max_{e,p,w} w - c(e) + pu^*_h(p)$$

such that

$$0 \leq p \leq 1,$$  \hspace{1cm} (3.41)

$$w - \theta c(e) + U^l_2(p) \leq U^l,$$ \hspace{1cm} (3.42)

and

$$e - w + p\pi^*_p + (1 - p)\pi^*_n \geq 0,$$ \hspace{1cm} (3.43)

where recall $U^h_2$ and $U^l_2$ (from equation (3.36)) are the period 2 utility of workers for the high and low types if they take the high-type contracts in period 1, $U^l$ is the total expected utility of a low type worker from taking a period 1 low type contract (from equation (3.39)), and finally $\pi^*_p$ and $\pi^*_n$ are the firm’s period 2 payoff from a promoted and a nonpromoted worker (from equation (3.23)).

To solve this program, we use the observation that firms must have zero expected profit to substitute out for $w$. In this way, we can rewrite objective function (after some substitutions) as

$$\max_{e,p} e - c(e) + Y^h - p\frac{k^*_p}{2} - (1 - p)[(1 - P^*_n)\frac{k^*_n}{2} + P^*_n\delta]$$ \hspace{1cm} (3.44)

where $\delta = Y^h - (e^*_h - c(e^*_h)), k^*_p = \frac{1}{2}, k^*_n = \frac{K + \delta}{2}$, and $k^*_n/K = 1 - P^*_n$. Note that $e - c(e)$ is the period 1 surplus, $Y^h$ is the highest possible period 2 surplus, which happens when the high type worker stays with the incumbent and is assigned Job 2. The $p\frac{k^*_p}{2}$ is the expected loss of a promoted worker from moving to another firm; the $(1 - p)[(1 - P^*_n)\frac{k^*_n}{2}]$ is the loss from moving for a non-promoted worker; and $(1 - p)P^*_n\delta$
is the expected loss of a high type from working on Job 1 with the incumbent.

Now define

\[
A = \frac{1}{2} K(1 - P_n^*)^2 + P_n^* \delta - \frac{K}{8},
\]

\[
= \frac{3\delta}{8} (1 + 2P_n^*),
\]

\[
B = P_n^*(\theta - 1)c(e^*_h),
\]

and

\[
U_0^i = U^i + \frac{K}{8}.
\]

After some algebra, we can write the Lagrangian with the constrained maximization problem as

\[
L = e - c(e) + Y^h - \frac{K}{8} - (1 - p)A + \lambda(U_0^i - Y^h - (e - \theta c(e)) + (1 - p)(A + B)) + \gamma(1 - p).
\]

This is a constrained maximization problem of a weakly concave function on a convex set. The maximum exists and it is easy to see that there is a unique solution. The Kuhn-Tucker conditions for this program are

\[
1 - c'(e^*_h) - \lambda(1 - \theta c'(e^*_h)) = 0, \quad (3.49)
\]

\[
A - \lambda(A + B) - \gamma = 0, \quad (3.50)
\]

\[
\lambda \geq 0, \gamma \geq 0, \quad (3.51)
\]

\[
\lambda(U_0^i - Y^h - (e^*_h - \theta c(e^*_h)) + (1 - p)(A + B)) = 0, \quad (3.52)
\]

\[
\gamma(1 - p^*_h) = 0. \quad (3.53)
\]

According to the sign of the multipliers, there are three cases to consider for the solution to this program. First, \(\lambda = 0\), which corresponds to the case where the
incentive constraint is slack at the equilibrium effort and promotion offer. Second, \( \lambda > 0 \) and \( \gamma > 0 \), which corresponds to the case where the incentive constraint is binding at the equilibrium effort and promotion offer. In this case, \( \gamma > 0 \) implies that \( p_h^* = 1 \) by equation (3.53) so the promotion offer is efficient. In other words, only the period 1 effort will be distorted in equilibrium. Third, \( \lambda > 0 \) and \( \gamma = 0 \), which corresponds to the case where the incentive constraint is binding at the equilibrium effort and promotion offer. Moreover, both the equilibrium period 1 effort and promotion are distorted.

First, suppose \( \lambda = 0 \), so the low-types do not want to mimic the high types in equilibrium by some margin. Equation (3.53) implies that \( \gamma > 0 \), so we must have \( p_h^* = 1 \) by (3.53), i.e. the promotion level is efficient. Moreover, equation (3.53) implies that \( c'(e_{1h}^*) = 1 \), so the effort level required is efficient in this case. Finally, we can use the zero profit condition of the firms to calculate that the period 1 wage offer equals \( w_{1h}^* = e_h^* + \frac{K}{4} \). This gives the characterization of the equilibrium contract under \( \lambda = 0 \).

We also need to check that under the equilibrium contract, the low-types do not want to mimic the high types in equilibrium by a margin. This is given by

\[
U_0^l - Y^h - (e_{1h}^* - \theta c(e_h^*)) > 0,
\]

which follows from the IC constraint using \( e_{1h}^* = e_h^* \) and \( p_h^* = 1 \). In summary, when \( U_0^l - Y^h - (e_{1h}^* - \theta c(e_h^*)) > 0 \), we have

\[
e_{1h}^* = e_h^*, \ p_h^* = 1, \ w_{1h}^* = e_h^* + \frac{K}{4}.
\]

Second, suppose that \( \lambda > 0 \) and \( \gamma > 0 \) so the IC constraint is binding. Moreover, \( \gamma > 0 \) implies that \( p_h^* = 1 \) by equation ((3.53)). Since the IC constraint is binding, the equilibrium effort level must satisfy \( e_{1h}^* = e^* \). where \( e^* \geq e_h^* \) is the unique effort
level that satisfies
\[ U_0^l - Y^h - (e^{**} - \theta c(e^{**})) = 0. \] (3.54)

Note that we cannot have \( e_{1h}^* < e_h^* \) in equilibrium because it is strictly dominated by a contract that requires \( e_{1h}^* = e_h^* \), which increases the expected output and loosens the IC. Therefore, we must have \( e_{1h}^* \geq e_h^* \). Moreover, \( e^{**} \) is unique because \( e - \theta c(e) \) is strictly decreasing for \( e > e_h^* \). Finally, under the equilibrium level \( e_{1h}^* = e^{**} \), we can solve for the equilibrium wage offer in period one by using the zero profit condition. We obtain \( w_{1h}^* = e^{**} + \frac{K}{4} \).

We also need to check that under the assumptions of this case, the equilibrium contract derived implies that \( \gamma > 0 \), or that promotion is efficient. To do this, we use the first-order-condition of effort in equation (3.49) to determine

\[ \lambda = \frac{c'((e^{**}) - 1}{\theta c'((e^{**}) - 1}. \] (3.55)

Then we plug this expression into equation (3.50) to derive the necessary and sufficient condition for \( \gamma > 0 \): we have \( \gamma > 0 \) if and only if

\[ \frac{A}{A + B} > \frac{c'((e^{**}) - 1}{\theta c'((e^{**}) - 1}, \] (3.56)

where recall that \( A = \frac{3\delta}{\delta}(1 + 2P_n^*), B = P_n^*(\theta - 1)c(e_h^*), \) and \( P_n^* = \frac{1}{2} - \frac{\delta}{2K} \).

In summary, if \( A > \frac{c'((e^{**}) - 1}{\theta c'((e^{**}) - 1}(A + B) \), where \( e^{**} \) is the unique effort level (greater than \( e_h^* \)) that satisfies \( U_0^l - Y^h = (e^{**} - \theta c(e^{**})) \), then in the equilibrium contract in period 1, we have

\[ e_{1h}^* = e^{**}, \ p_h^* = 1, \ w_{1h}^* = e^{**} + \frac{K}{4}. \] (3.57)

Third, suppose that \( \lambda > 0 \) and \( \gamma = 0 \) so the IC constraint is binding. In this case, we note that \( \lambda \) can be calculated by the primitives according to equation (3.50), so

\[ \lambda = \frac{A}{A + B} < 1, \] (3.58)
where recall that \( \frac{2K}{5} (1 + 2P^*_n), B = P^*_n (\theta - 1)c(e^*_h) \). Using the expression for \( \lambda \), we see that equation (3.49) implies that the equilibrium effort \( e^*_1 = \hat{e} \), where \( \hat{e} \) is the unique effort level (greater than \( e^*_h \)) that satisfies

\[
\frac{A}{A + B} = \frac{c'(\hat{e}) - 1}{\theta c'(\hat{e}) - 1}.
\]  

(3.59)

Note that we have \( e > e^*_h \) because of the same reasoning as in the previous case. Moreover, \( \hat{e} \) is unique since \( \frac{c'(e)-1}{\theta c'(e)-1} \) is strictly increasing when \( e \geq e^*_h \).

The equilibrium promotion probability follows directly from the equilibrium effort level using the condition that IC binds. In particular, we obtain

\[
p^*_h = 1 - \frac{\hat{e} - \theta c(\hat{e}) + Y^h - U^q_0}{A + B}.
\]  

(3.60)

Finally, we calculate the equilibrium wage in a high-type contract according to the zero expected profit condition. We obtain

\[
w^*_1h = \hat{e} + p^*_h \pi^*_p + (1 - p^*_h) \pi^*_n,
\]  

(3.61)

where recall that \( \pi^*_p = \frac{K}{4} \) is the firm’s expected profit from a promoted worker in period 2 and \( \pi^*_n = \frac{K}{4} (1 - \frac{\delta}{K})^2 \) is the firm’s expected profit from a nonpromoted worker in period 2.

In summary, when the conditions in the two cases above are not satisfied, we have

\[
e^*_1h = \hat{e}, p^*_h = 1 - \frac{\hat{e} - \theta c(\hat{e}) + Y^h - U^q_0}{A + B}, w^*_1h = \hat{e} + p^*_h \pi^*_p + (1 - p^*_h) \pi^*_n.
\]  

(3.62)

The combination of the three cases shows that there is an essentially unique separating equilibrium of the model.

**Theorem 1:** The following strategies constitute a separating equilibrium. First, the strategies of the workers and firms in period 2 are described in Lemmas 1 and 3.
Second, the period 1 equilibrium decisions of the workers are summarized in Lemma 4. Third, at the beginning of period 1, each firm offers a low-type contract with

\[ (w_{i1}^*, e_{i1}^*, 0) = (e_i^*, e_i^* + \frac{K}{4}, 0). \]

In addition, each firm offers a high-type contract

\[ (w_{ih}^*, e_{ih}^*, p_h^*) = (e_h^*, 1, e_h^* + \frac{K}{4}) \quad \text{if } U_0^l - Y_h - (e_h^* - \theta c(e_h^*)) > 0, \]

where \( U_0^l = 2e_i^* - 2\theta c(e_i^*) \).

If \( U_0^l - Y_h - (e_h^* - \theta c(e_h^*)) \leq 0 \) and if \( A > \frac{c'(e_{**})-1}{\theta c'(e_{**})-1} (A + B) \), the equilibrium high-type contract satisfies

\[ (w_{ih}^*, e_{ih}^*, p_h^*) = (e_{**}, 1, e_{**} + \frac{K}{4}). \quad (3.63) \]

If \( U_0^l - Y_h - (e_h^* - \theta c(e_h^*)) \leq 0 \) and if \( A \leq \frac{c'(e_{**})-1}{\theta c'(e_{**})-1} (A + B) \), the equilibrium high-type contract satisfies

\[ (w_{ih}^*, e_{ih}^*, p_h^*) = (\tilde{e}, 1 - \frac{\tilde{e} - \theta c(\tilde{e}) + Y_h - U_0^l}{A + B}, \tilde{e} + p_h^* \pi_p^* + (1 - p_h^*) \pi_n^*), \quad (3.64) \]

where \( e_{**} \) satisfies \( e_{**} - \theta c(e_{**}) = U_0^l - Y_h \), \( A = \frac{3\delta}{\delta - \delta} (1 + 2P_n^*), B = P_n^*(\theta - 1)c(e_h^*), \) and \( \tilde{e} \) satisfies \( \frac{c'(\tilde{e})-1}{\theta c'(\tilde{e})-1} = \frac{A}{A + B} \).

### 3.4 Application of the Model

This section explores the implications of the model. We first show that the model delivers a reasonable prediction for the behavior of turnover. We then examine how the separating PBE's contract for high types varies with the exogenous parameters \( K \) and \( Y_h \); we find that while \( e \) is unambiguously decreasing in \( K \) and increasing in \( Y_h \), the behavior of \( p \) is not necessarily monotone. Large increases in \( K \) shift the separating PBE from a \( p = 1 \) to a \( p < 1 \) regime, but it is possible for \( p \) to be locally
increasing in $K$ once $p < 1$. The comparative statics of $Y_h$ on $p$ are also ambiguous, and we cannot generally say whether increases in $Y_h$ will move the PBE in or out of the $p = 1$ regime.

**Proposition 1:** In equilibrium, workers are more likely to stay with the incumbent if promoted.

**Proof.** Promoted workers leave the firm with probability $\frac{1}{2}$. Nonpromoted workers leave the firm with probability $1 - P^*_n = \frac{1}{2} + \frac{\delta}{2K} > \frac{1}{2}$. So this proposition is immediate.

We view Proposition 1 less as a substantive result and more as a check that the model delivers sensible predictions. Workers who take a high type contract and fail to be promoted are inefficiently assigned ex post, which we expect to increase the probability of turnover.

**Theorem 2:** Suppose $U^*_0 - Y^h - (e^*_h - \theta c(e^*_h)) < 0$. There exists a $K^*$ such that

$$e^*_1 = e^{**}, \quad p^*_h = 1 \quad \text{for } K < K^*, \quad (3.65)$$

$$\frac{de^*_1}{dK} < 0, \quad p^*_h < 1 \quad \text{for } K > K^*. \quad (3.66)$$

**Proof.** From Theorem 1, we observe that $p^*_h = 1$ if and only if

$$\frac{A}{A + B} > \frac{c'(e^{**}) - 1}{\theta c'(e^{**}) - 1}, \quad (3.67)$$

where $U^*_0 - Y^h - (e^{**} - \theta c(e^{**})) = 0$. The right hand side is independent of $K$. Now notice that $\frac{A}{A + B}$ is strictly decreasing in $\frac{B}{A}$, which can be written as a function of $P^*_n$ as

$$\frac{B}{A} = \frac{8(\theta - 1)c(e^*_h)}{3\delta} \left( \frac{P^*_n}{1 + 2P^*_n} \right). \quad (3.68)$$

It is clear that $\frac{B}{A}$ is strictly increasing in $P^*_n = \frac{1}{2} - \frac{\delta}{2K}$, which is in turn strictly increasing in $K$. Therefore, $\frac{A}{A + B}$ is strictly decreasing in $K$, so there exists a threshold
Below which equation (3.67) is satisfied.

Now suppose $K > K^*$. It is clear that in this region $p_h^* < 1$ by Theorem 1. In addition, the equilibrium period 1 effort $c_h^* = \widehat{c}$, where $\widehat{c}$ satisfies

$$\frac{A}{A + B} = \frac{c'(\widehat{c}) - 1}{\theta c'(\widehat{c}) - 1}.$$  \hfill (3.69)

Note that $\frac{c'(e) - 1}{\theta c'(e) - 1}$ is strictly increasing in $e$ because $\theta > 1$. On the other hand, $\frac{A}{A + B}$ is strictly decreasing in $K$ so we have $\frac{dc_h^*}{dK} = \frac{dc^*}{dK} < 0$. $\blacksquare$

When job mobility costs decline, the promotion distortion mechanism works less well in the sense that the ratio of marginal benefits to marginal costs of $p$ go up. Holding the severity of the incentive compatibility constraint fixed, this relative price effect induces the firm to substitute distortions in $e$ for distortions in $p$. However, lower mobility costs may also tighten the incentive compatibility constraint, so that the firm must increase the total amount of distortion in order to keep separating workers by type. This latter effect comes from the fact that low mobility costs increase turnover among non-promoted (high-type contract) workers, and it is the net threat of ending up in job 1 that discourages low types from mimicking. When $K$ is small, a low type who mimics and fails to get promoted can very likely find a managerial job at another firm, and this threat is weak.

Because the relative price effect and the effect of $K$ on the tightness of the incentive compatibility constraint push $p$ in opposite directions, we cannot generally sign $\frac{\partial p_h^*}{\partial K}$. In the neighborhood of $p = 1$, however, only the relative price effect operates; "perverse" behavior of $p$ is more likely to occur in a region where job assignment is already very distorted.

Finally, while turnover rates conditional on non-promotion increase when $K$ goes down,$^{11}$ the behavior of aggregate turnover is unclear. If $p$ falls, then aggregate turnover must increase, since non-promoted workers are less likely to remain with the incumbent. However, $p$ must rise over many ranges, and in this case the net result is

$^{11}$Turnover conditional on promotion is fixed at $\frac{1}{2}$. 171
ambiguous: fewer workers fall into the high-turnover group, but turnover within that group increases.

**Theorem 3:** Suppose \( U_0^l - Y^h - (e_h^* - \theta c(e_h^*)) < 0 \), then \( \frac{d e_h^*}{d Y_h} > 0 \).

**Proof.** Recall that \( \hat{e} > e_h^* \) is the unique effort level such that

\[
\frac{A}{A + B} = \frac{c' \hat{e} - 1}{\theta c' \hat{e} - 1}.
\]  

(3.70)

Using the expressions in terms of the more primitive variables, we find that

\[
\frac{A}{B} = \frac{\delta}{12(\theta - 1)c(e_h^*)} \left(1 + \frac{K}{K - \delta}\right),
\]  

(3.71)

where \( \delta = Y_h - (e_h^* - c(e_h^*)) \). It is clear that \( \frac{A}{B} \) is strictly increasing in \( \delta \) and thus in \( Y_h \) because \( \delta \) and \( Y_h \) changes one-to-one. Since \( \frac{A}{B} \) moves with \( \frac{A}{A + B} \), this implies that \( \frac{A}{A + B} \) is strictly increasing in \( Y_h \). This implies that

\[
\frac{d \hat{e}}{d Y_h} > 0.
\]  

(3.72)

Also recall that \( e^{**} \geq e_h^* \) is the unique effort level such that

\[
U_0^l - Y^h - (e^{**} - \theta c(e^{**})) = 0.
\]  

(3.73)

Since \( e - \theta c(e) \) is decreasing in \( e > e_h^* \), it is easy to see that

\[
\frac{d e^{**}}{d Y_h} > 0.
\]  

(3.74)

Now equation (3.56) implies that \( p_h^{*} = 1 \) if

\[
\frac{c'(e^{**}) - 1}{\theta c'(e^{**}) - 1} < \frac{A}{A + B} = \frac{c'(\hat{e}) - 1}{\theta c'(\hat{e}) - 1}.
\]  

(3.75)
and since \( \frac{c'(e)}{c'(e)-1} \) is increasing in \( e \), this implies that

\[
e^*_{1h} = \min\{e^{**}, \hat{e}\}.
\]

Because we have both \( \frac{de}{dY_h} > 0 \) and \( \frac{de^{**}}{dY_h} > 0 \), this gives that

\[
\frac{de^*_{1h}}{dY_h} > 0.
\]

Increases in \( Y_h \) have the same basic effects as decreases in \( K \): the relative price of distorting \( p \) goes up, and the incentive compatibility constraint becomes more binding. Both of these effects push work effort upward. The difference between this case and Theorem 2 comes from the fact that the effect of \( Y_h \) on the incentive compatibility constraint does not disappear as \( p \to 1 \). Even when \( p \) is very close to 1, increases in \( Y_h \) can therefore have perverse effects, and we cannot sign \( \frac{\partial p_A^2}{\partial Y_h} \) at all. That is, the \( p = 1 \) regime may occur at low values of \( Y_h \), at high values, at both or at neither. When the price effect is small relative to the incentive compatibility effect, the firm will respond to a rise in \( Y_h \) by increasing distortions of both \( e \) and \( p \); when the price effect is relatively large, the firm will choose to reduce inefficient task assignment at the cost of even greater distortions to work hours.

As in Theorem 2, the response of aggregate turnover to changes in \( Y_h \) is ambiguous: high \( Y_h \) increases turnover among workers who fail to be promoted, but the proportion of such workers in the workforce may go either up or down.

### 3.5 Discussion and Conclusion

When jobs vary in the degree to which productive effort is observable, firms face a hard problem. Efficiency demands that hard workers (high types) be assigned to jobs that are difficult to monitor, but firms must prevent lazier workers (low types) from
mimicking their colleagues. The problem is made harder still by the fact that low types stand to gain more from a job where effort is non-contractable. We explore two instruments that firms can use to screen out low types. First, firms can require employees to work very hard at the beginning of their careers as a precondition of access to these jobs. Since low types find the effort more onerous, such measures deter them from seeking promotion. Secondly, firms can introduce noise into the promotion process. While this measure has the cost that some high types end up assigned to the wrong job, it helps to ensure that low types are assigned efficiently. Firms that want to screen workers trade off the costs and benefits of these two instruments.

More generally, we can think of firms varying career tracks along two broad dimensions: the characteristics of the entry level work within each track and the path of jobs the worker expects to receive in the future. Monopolistic firms would solve the second best problem and find the most efficient tradeoff between distorting entry level characteristics and future job paths. However, competition (smaller values of $K$ in the model) makes it difficult for firms to distort workers' future job paths. Once a worker has successfully signalled his type, rival firms will offer him the *ex post* efficient job path and the worker will accept as long as mobility costs are small. Thus competition inefficiently tilts the solution of the screening problem away from career path distortions in favor of distortions that can be introduced into entry level work (like work hours).

Viewed as a measure of competition, $K$ can be given at least two specific interpretations. First, $K$ may reflect the cost of job search or the rate at which new jobs open up. When workers can easily find alternative employers, career paths cannot deviate far from what we would see under spot markets. High mobility costs and limited job creation tie workers to their initial employer and allow more flexibility in how future career paths are designed. Note that short run fluctuations in mobility costs are not relevant; what matters are expectations about mobility over the course of a working life. However, $K$ can be interpreted as a measure of persistent institutional barriers to mobility. Secondly, $K$ may reflect the relative importance of firm-specific
human capital, or the firm-worker pair-specific surplus relative to the worker's next best match. When firm-specific human capital is important, there is a large surplus wrapped up in existing employment relationships. This discourages mobility and allows internal labor markets to deviate from external ones. As skills become more transferable, these internal labor markets will tend to break down.

This discussion points to an even more fundamental interpretation of $K$, as a measure of the ease with which long term contracts are sustainable. Greater competition in the labor market is one way for the viability of long term contracts to erode, but alternative mechanisms are possible as well. For example, if firms are expected to die at a faster rate, then workers know that they will need to find new employers in the future. The career practices of the current firm are now less relevant than what one is able to signal to the outside market, and low types will have greater incentives to mimic other workers. Similarly, a one time incentive to renege on long term arrangements may destroy expectations that they will be honored in the future, pushing the economy into an equilibrium based more on spot markets.

We interpret $Y_h$ as a measure of the relative productivity of jobs where worker inputs are hard to observe. To a large extent these will overlap with high-level managerial and professional positions. (Interpreting changes in $Y_h$ as changes in the relative productivity of high types is somewhat less precise, since all of the comparative statics would be identical if $Y_l$ changed in step with $Y_h$.) In an economy with distortionary taxation or labor market institutions, $Y_h$ will reflect the net surplus generated by these jobs for a firm paired with a hard worker. Thus greater progressive taxation, or institutional wage compression, is equivalent to a decrease in $Y_h$, and we should expect it to lower the use of inefficiently long work hours. This is not quite a standard labor supply effect: work hours decline in job 1, not in the job that is actually hit by the distortions. Instead, the intuition is that competition in the rat race is less attractive when the prize is exogenously made smaller.

In terms of welfare, both increases in $K$ and decreases in $Y_h$ slacken the incentive compatibility constraint and in that sense decrease the degree of inefficiency in the
economy. The exact welfare consequences of changes in $K$ and $Y_h$ will depend on how the changes arise, however. As we have constructed the model, the economy’s production possibilities frontier is shrinking in $K$ and expanding in $Y_h$; a more informative comparison is to consider fully rebated taxes on worker mobility that increase $K$ and progressive income taxes that decrease $Y_h$. While we do not solve for the welfare effects of such policies explicitly, they will tend to improve efficiency at the margin. Taxes on mobility tend to make firms *ex post* monopsonists, which helps them to enforce a contract closer to the second best tradeoff between $e$ and $p$. Similarly, small taxes on high incomes will slacken the incentive compatibility constraint, allowing firms to reduce work hours and possibly improve the efficiency of task assignment. Sufficiently large income taxes, of course, will discourage firms from promoting hard workers even in the absence of an incentive compatibility constraint and will be very inefficient. In a sense, however, our setup is rigged to make income taxation look too appealing: in a more general model where managerial jobs require effort (and firms use second best contracts to induce that effort), we would expect taxes to improve the adverse selection problem and worsen moral hazard problems simultaneously.

We conclude on an empirical note. In the introduction, we noted that comparisons between the U.S. and Europe, as well as recent trends within the U.S., are consistent with the predictions of our model. That is, very rough cuts of the data suggest work hours are positively correlated with both job mobility (or short term contracting) and the returns to high-level work. Of course, these patterns might not be causal: all three might be reflections of greater liberalism, for example. In future work, we hope to subject the model to a more telling empirical test that, at a minimum, holds political institutions constant. Differences in the rate at which job mobility has changed across subsets of the U.S. labor market (e.g., occupational groups) might provide one such feasible test.
Bibliography


