MATCHING BY LUCK OR SEARCH? EMPIRICAL EVIDENCE FROM THE EXECUTIVE LABOR MARKET

KUN LI
TOULOUSE SCHOOL OF ECONOMICS, FRANCE
MICHAEL PETERS
UNIVERSITY OF BRITISH COLUMBIA, CANADA
PAI XU
UNIVERSITY OF HONG KONG, HONG KONG

Abstract. This paper provides a model of directed search in which workers have private information about type at the point where they make their applications to firms. Firms are able to observe these types once workers apply. The paper shows that for any smooth wage distribution there is a continuation equilibrium in which unemployed workers choose a reservation wage which is a strictly increasing function of their type, then apply with equal probability to all positions that offer more than that wage. We consider a case where matching occurs 'quickly', and show two main results. First, the wages at which workers are employed throughout their lives are correlated, but very imperfectly because of the fact that equilibrium involves a lot of mismatch. Second, the variance of future income of workers must be a decreasing function of the wage at which they are currently employed. In other words, high wage workers will enjoy more stable lifetime income.

These results make it possible to distinguish between the three main models of directed search empirically. The imperfect correlation - declining variance results in this paper contrast sharply with the classic directed search, where wages are uncorrelated over time, and models with assortative matching, in which wages are perfectly correlated over time.

The paper concludes with an analysis of data from the executive labor market from 1992 to 2009, where the imperfect correlation is supported.

We shall like to thank David Green, Espen Moen, Aloysius Siow, Sergei Severinov and participants in the Econometrics Lunch Workshop at UBC, Canadian theory workshop, and econometric society summer meetings for helpful comments.

1. Introduction

This paper provides a dynamic extension of the directed search model in (Peters 2010) in which workers and firms have private information about their characteristics that drive their search behavior. The original paper was designed to illustrate the connection between unemployment duration and exit wage. However, in the context we consider here, we are more interested in the matching that is supported. In particular we are interested in matching outcomes, and whether observable outcomes can be in any way understood using arguments from directed search. Directed search provides three different perspectives. In the most basic models of directed search (for example, (Peters 2000)), workers are identical but use mixed application strategies when they apply to firms, applying with highest probability at the firms who offer the highest prices. In the steady state of such a model, workers repeat this outcome in every period. Over their lifetimes, their matching outcomes will vary, but in a manner that exhibits no auto correlation at all. A worker who is lucky enough to land a high paying job in some period will mix again after he becomes unemployed. So his future outcome will tend to see his wages fall.

At the other extreme, models that support pure assortative matching (for example (Eeckhout and Kircher 2010)) will predict that workers who land high wage jobs in one period will do so again in future periods. Theoretically, outcomes are perfectly correlated over time. The same kind of outcome could be expected from wage-ladder like models (e.g., (Delacroix and Shi 2006)) in which homogeneous workers search on the job and implicitly use the current wage as a way of coordinating applications. Workers who are employed at some wage will apply to firms offering slightly higher wages until they are matched. Workers types are all the same, so there is no question about assortative matching. Yet the ladder like application behavior will mean that a worker's current outcome will be a good predictor of his future success.

In between these extremes is the model in (Peters 2010) in which workers have privately know types which matter to firms. The key difference is that workers use mixed application strategies. Nonetheless, these mixed strategies have considerable structure. Each worker adopts a 'reservation wage', which is an increasing function of the worker's type. Whenever the worker is unemployed, he or she makes applications to all the firms who offer wages above their reservation wage. Of course, the higher the worker's type, the more likely he is to be hired by the high wage firms. As a consequence, workers search outcomes are positively correlated with their type. As workers carry their types through time, search outcomes are positively correlated between periods. However, the correlation is far from perfect. Workers mixed strategies lead to mismatch. Though workers who become employed at low wage firms must have low enough types to make it sensible for them to apply to those firms, workers who find jobs at high wage firms may either have good types, or may just have gotten lucky. So good outcomes are very noisy signals of good types. So the model predicts that the correlation between present and future outcomes fall as outcomes improve.

Moreover, since workers who have low types (and are employed at low wages) do sometimes get lucky and land high wage jobs. It doesn't work the other way around for high type workers, high wage workers aren't likely to face a large decline in income unless they are low types. As a consequence, the variance of future income should be larger for workers who have low wage outcomes than it is for workers with high wage outcomes. These correlations seem plausible, yet they aren't consistent with either assortative matching, or homogeneous worker directed search. If anything, wage ladder models predict that variance of future outcome should be increasing with current wage. The reason is that workers whose current match is terminated apply at the lowest wages once they are unemployed. The decline in the wage rate for workers who make job transitions is highest for workers who are currently employed at the highest wage.

To illustrate these things, this paper begins with a simple dynamic model of worker job transitions. Matches are terminated randomly, after which workers adopt mixed application strategies similar to those described in (Peters 2010). This section illustrates the properties of equilibrium that drive the main predictions. The dynamic arguments will also serve to illustrate how comparable arguments would work with the better known variants of the directed search model.

We then turn to a data set taken from the executive labor market with the objective of trying to figure out which of the three directed search models best fits the matching data for that market. The executive market has a number of advantages from our perspective. First, executives' talent is a key input in the production (or, profit-generating) process. However executive talent is not captured by the number of MBA's or law degrees that an executive has. Rather, these talents seem largely interpreted as unobservables, such as connections with other executives, leadership ability, etc. Though these skills are unobservable to an outsider, firms seem to know them when they see them. Presumably reputation, reference letters, participation in successful projects are signals of managerial skill. At the same time, it is impossible to write a wage contract that conditions on these unobservables. This is the sort of environment which our theory fits.

Second, for executives in general, there are no segmented labor markets across industries, leadership skill is valuable in all industries. Thus we could effectively consider an integrated labor market for all executives.

Our work can be related to at least two strands of literature. One is the random matching literature. Equilibrium search with random matching ties wage dispersion to worker heterogeneity and firm heterogeneity.[1] Empricial work with directed search models is less common.

The other is structural analyses on executive compensation, which has exclusively focused on the assignment models, assuming executives and firms are matched assortatively.[2] Assignment models enable researchers to analytically solve and calibrate empirical models, but assume that markets are frictionless and efficient. Our intention is to consider whether market frictions can be used to get additional insight into these models.

Though our results are somewhat in contrast to the assortative matching that occurs in executive compensation models, our intention isn't to compete with this alternative literature, but to compare the performance of different directed search models.

2. Fundamentals

We begin with a description of equilibrium in a labor market in which worker types are privately known to them.

A labor market consists of measurable sets of positions and workers. A measurable set of workers are assumed to be unemployed each instant. They search for unfilled positions by making applications to specific positions. Firms controlling these positions hire one of the workers who applies. We assume that these empty positions and searching workers are the result of matches that fail for exogenous reasons. We focus on the case where the measures of sets of searching workers and unfilled positions are constant over time, and are equal to one another.

Workers are parameterized by their *type* which is an element of some compact subset Y of $\mathbb{R}_+$. The measure of the set of searching workers with types less than or equal to y is given by $F(y)$, where F is some

[1] See, for example, (Postel-Vinay and Robin 2002) who used French panel data to decompose the cross-employee wage variance into market imperfection and person effects.

[2] (Tervio 2008) shows that the observed joint distribution of CEO pay and market value can be used to infer the economic value of underlying ability differences. (Gabaix and Landier 2008) find quantitative explanation for the rise of CEO pay.

distribution function that is monotonically increasing and differentiable with support equal to some interval $Y = [\underline{y}, \overline{y}]$. Unfilled positions are parameterized by some characteristic $x \in X$, where X is a compact subset of $\mathbb{R}$. The measure of position types is given by some distribution function H. Workers' types are private information when they apply for jobs, though we assume that workers can *show* their types to firms when they apply. Position types are assumed to be public information.

Matching and production occur over an infinite number of periods and all participants in the matching process are assumed to be infinitely lived and perfectly patient. At the beginning of each period a measurable set of matches is exogenously terminated. The subsequent re-matching occurs very quickly so that all workers are re-matched by the end of the period. Workers and firms don't discount future payoffs,. However, there is an exogenous matching cost that each party bears when they search for partners. This cost is assumed to be a proportion $(1-\gamma)$ of their future earnings or profits. Workers care only about their total wage payments net of these matching costs. Firms care about their total profits net of matching costs. Let w be the expected payments made to a worker during a match with a firm. The expected profit earned by a firm with a position of type x who hires a worker of type y and pays him or her this wage, is given by some function $v(w,x,y)$ which is assumed decreasing in w and weakly increasing in y. To maintain an order on position types, it is assumed that for any pair (w,y) and (w',y') with $(w,y) \geq (w',y')$, if $v(w,x,y) \geq v(w',x,y')$ for some type x, then $v(w,x',y) \geq v(w',x',y')$ for any higher type $x' \geq x$. In words, this *single crossing condition* says that higher type positions generate more profits from higher type workers than lower type positions do. We assume wages are chosen from a compact interval W.

At the beginning of each period, a worker is either employed, or unemployed. If she is unemployed, the worker simply chooses where to apply with full knowledge of the expected payments she will receive from any firm who hires her. If she is hired, she expects to receive the promised payments until her employment with the firm is terminated. If she isn't hired by the firm where she applies, she tries again, but bears the cost of unemployment. Her application strategy is chosen to balance unemployment costs against future wage payments.

Payoffs are such that firms always hire the worker who applies who has the highest type. Since this paper is primarily concerned with workers application behavior, we make the simplifying assumption the firm doesn't have the option of refusing its best application in order to search for a worker of higher quality. The underlying presumption is that all the workers who are searching for work have the same observable qualifications like education and experience.

3. The Market

The payoffs that players receive depend on their own actions, and on the distributions of actions taken by the other players. We specify these payoffs using standard arguments from directed search. Let G be the steady state wage *offer* distribution. We'll assume throughout that G is monotonic, differentiable, and has interval support $\overline{G} = [\underline{w}, \overline{w}]$. Generally, the wage offer distribution should differ from the accepted wage distribution. However, in a steady state in which matches are exogenously broken up in a manner that is independent of the wage, then reformed quickly, these two distributions will be the same. So we don't distinguish between them in what follows.

Let P represent the steady state distribution of applications, where $P(w, y)$ is understood to be the measure of the set of workers who have type y or less who apply at wage w or less. From the perspective of an individual position offering wage w, we will be interested in the measure of the set of searching workers of type y or less who apply at wage w. Denote this *conditional distribution* by $p_w(y)$ and observe that the relationship between P and p_w is given by

$$P(w, y) = \int_{\underline{w}}^{w} p_{w'}(y)\, dG(w'). \tag{3.1}$$

Since P must be absolutely continuous with respect to the distribution G, it should be apparent that p_w is the Radon-Nikodym derivative of $P(w, y)$ with respect to G.

To understand the payoffs of workers, observe that a worker who has current type y is always hired before workers with lower types. As a consequence, he is concerned not with the total number of applicants expected to apply at the firm where he applies (the 'queue size'), but with the measure of the set of applicants who apply whose type is as least as large as his. When he applies at wage w, this number is given by

$$\int_{y}^{\overline{y}} dp_w(\tilde{y}).$$

We use the familiar formula $e^{-\int_y^{\overline{y}} dp_w(\tilde{y})}$ to give the probability that the worker will be hired if he applies at wage w.

From this it is straightforward to write down the payoff to a worker of type y who is searching for a job

$$U(y) =$$

$$\max_{w' \in \overline{W}} \left(\left(w' + \gamma U(y) \right) e^{-\int_y^{\overline{y}} dp_{w'}(\tilde{y})} + \left(1 - e^{-\int_y^{\overline{y}} dp_{w'}(\tilde{y})} \right) \gamma U(y) \right) \Bigg] . \tag{3.2}$$

For firms, an unfilled position then has value

$$V(x) =$$

$$\max_{w} \left[\int_{\underline{y}}^{\overline{y}} \left(v(w,x,y) + \gamma V(x) \right) e^{-\int_y^{\overline{y}} dp_w(\tilde{y})} dp_w(y) + \left(1 - \int_{\underline{y}}^{\overline{y}} e^{-\int_y^{\overline{y}} dp_w(\tilde{y})} dp_w(y) \right) \gamma V(x) \right] . \tag{3.3}$$

Each expression contains an expected profit or wage term that applies to the duration of the match, then the value to the firm or worker of finding a new match once the existing one terminates discounted to reflect the costs search. A steady state equilibrium for this model is a pair of distributions (G, P) having two properties: (i) at every wage w, G(w) coincides with the measure of the set of unfilled positions which maximize expected payoff by offering a wage w or less; and (ii) for each pair (w, y), there is a set of unemployed workers of measure $P(w, y)$ whose types are less than or equal to y and who maximize expected payoff by applying at wage w or less. In other words, the distribution of best replies to the distributions G and P are G and P themselves.

4. Continuation Strategies

Since the workers make their application decisions conditional on the distribution of wage offers, we can begin to characterize the equilibrium by describing the equilibrium value functions conditional on distributions G and P. The continuation equilibrium we are about to describe follows (Peters 2010). The utility function $U(y)$ in the theorem that follows should be interpreted as the *market payoff function* since it describes the payoff to a worker of type y who follows his equilibrium strategy. The reservation wage $\omega(y)$ is the wage that would yield the worker his market payoff if he expected to be hired at that wage for sure.

In the continuation equilibrium we are about to describe, each worker applies to every wage above his reservation wage with equal probability unless he is already employed in a position that pays more than his reservation wage. That means, that for any interval of wages, the probability that the worker applies to a wage in that interval is equal to the measure of wage offers in that interval divided by the measure of wage offers above the worker's reservation wage.

Theorem 1. *For any differentiable wage offer distribution G, there is a continuation equilibrium characterized by a monotonically increasing reservation wage strategy $\omega(y)$ in which each worker applies with equal probability at every wage at or above* $\max[\underline{w}, \omega(y)]$. *The reservation wage is characterized by the solution to the differential equation*

$$\omega'(y) = \frac{\omega(y) F'(y)}{G(\overline{w}) - G(\omega(y))} \tag{4.1}$$

through the point $(\overline{y}, \overline{w})$. The market payoff is given by $U(y) = \frac{\omega(y)}{1-\gamma}$ when $\omega(y) > \underline{w}$, and by

$$U(y) = \frac{\underline{w} e^{-\int_y^{\omega^{-1}(\underline{w})} \frac{1}{G(\overline{w}) - G(\underline{w})} dF(y')}}{1-\delta}$$

otherwise. Finally for every wage w in the support of G, the queue size faced by a worker who applies for a position offering wage w is

$$\int_y^{\overline{y}} dp_w(\tilde{y}) = \int_y^{\omega^{-1}(w)} \frac{1}{G(\overline{w}) - G(\omega(y'))} dF(y') \tag{4.2}$$

The proof of this theorem is given in the appendix. From the point of view of the empirics discussed below, the important part of this theorem is the assertion that there is a symmetric equilibrium in which every worker applies to every wage above his reservation wage with equal probability. Before we discuss this, we detour slightly to explain what the overall equilibrium looks like.

4.1. **Firms' Strategies.** To complete the description of equilibrium, we need to describe the determination of firms' equilibrium strategies. In all the discussion below, the wage offers made by firms are simply taken to be fixed. The types of firms are assumed to be whatever they need to be to support the wage distribution that appears in the data. Readers who are only interested in the empirical implications of Theorem 1 can skip to the next subsection.

The connection between Theorem 1 and the wage distribution can be gleaned by substituting the continuation strategies for workers into firms' payoff functions. The corresponding expression is given by

$$V(x) =$$

$$(4.3)\quad \max_{w}\left[\int_{\underline{y}}^{\omega^{-1}(w)} v(w,x,y')e^{-\int_{y'}^{\omega^{-1}(w)}\frac{dF(\tilde{y})}{G(\overline{w})-G(\omega(\tilde{y}))}}\frac{dF\left(y'\right)}{G\left(\overline{w}\right)-G\left(\omega\left(y'\right)\right)}\right.$$

$$\left.+\gamma V\left(x\right)\right].$$

This expression has a very convenient interpretation. Once a firm chooses a wage, it will receive applications from workers whose types are such that their reservation wage in the continuation equilibrium does not exceed the wage the firm offers. A slightly simpler formulation is to think of the firm as choosing the highest worker type it wants to try to attract, then offering the reservation wage of that worker type to all workers. The profit function then has the slightly simpler form

$$V\left(x\right)=$$

$$(4.4)\qquad \max_{y^*}\left[\int_{\underline{y}}^{y^*} v(\omega\left(y^*\right),x,y')e^{-\int_{y'}^{y^*}\frac{dF(\tilde{y})}{G(\overline{w})-G(\omega(\tilde{y}))}}\frac{dF\left(y'\right)}{G\left(\overline{w}\right)-G\left(\omega\left(y'\right)\right)}\right.$$

$$\left.+\gamma V\left(x\right)\right].$$

In the latter formulation of profits, the firm is maximizing its profit when the iso-profit line (in (y,w) space) associated with the argument in the maximization above is tangent to the reservation price rule.

Of course, this only defines payoffs in the support of the wage offer distribution. We should define payoffs outside this support. To keep things simple, we just define payoffs outside the support to ensure that a tangency with $\omega\left(y\right)$ is sufficient for profit maximization.[3]

To deduce the firm type distribution needed to support the observed wage distribution, one would take the profit function inside the bracket in (4.4) and for each wage w, try to find the value for x, the firm's type, that would make the iso-profit lines associated with this profit function tangent to the reservation wage function $\omega\left(\cdot\right)$ evaluated at whatever y satisfies $\omega\left(y\right)=w$.

It may be apparent from the expression, that in order to do this, one needs to know the distribution F. From the perspective of this paper, we are not interested so much in how valuable firms think this unobservable type y is, we are only interested in whether they use it to select workers. One approach, then, is simply to assume that F is just a uniform distribution and that firms value only the rank of the worker within the distribution of worker types. This is the approach we use in the rest of the paper.

[3]This is equivalent to imposing the usual market utility assumption in directed search, though the payoff when an offer is made above the support of the distribution of G requires some subtle considerations. See (Peters 2010) for details.

The main implications of this theory for the matching data stems from the fact that the type dependent application strategy imposes restrictions on what happens to workers as they move between jobs. The implications are quite straightforward. Workers in the model bear an unemployment cost that is proportional to their earnings whenever they look for a job. However, the matching process itself doesn't take them any time. They simply apply until they find a job. Since their application strategy has them making applications to every firm whose wage is above their reservation wage, the future wage of a worker of type y is a random variable whose distribution is just $\frac{G(w)-G(\omega(y))}{G(\overline{w})-G(\omega(y))}$. This is just the distribution G truncated below at the point $\omega(y)$. As $\omega(y)$ is increasing, workers with higher types are drawing their future wage from the same distribution but conditional on an event that is a strict subset of the conditioning event for workers of lower types. It then seems plausible that the variance of a workers future wage will fall as his type increases.

Of course, the workers' types are not observed directly, but they can be indirectly observed by looking at the wage at which a worker was last employed. Again, using the reservation wage strategy and the fact that the matching probability for a worker of type y at wage w is $\frac{\omega(y)}{w}$, the probability distribution of types employed at wage w is given by

$$\int_{\underline{y}}^{y} \frac{\omega(y')}{w} \frac{dF(y')}{G(\overline{w}) - G(\omega(y'))}$$

divided by the measure of the set of types who are employed at wage w (the formulas above with $y = \omega^{-1}(w)$). Since $\omega(y)$ is strictly increasing, an increase in w supports a new distribution the first order stochastically dominates the initial distribution. Putting this together with the fact that the variance of future income is falling with worker type, illustrates that the variance of future income is a declining function of the wage at which a worker is currently employed. This forms the basis of the empirical test in the data that follows below.

This leads to a very straightforward corollary of the main theorem:

Proposition 2. *If the variance of the random variable whose distribution is described by* $G_w(\cdot) = \frac{G(\cdot)-G(w)}{G(\overline{w})-G(w)}$ *is decreasing in* w*, then the variance of a workers future wage is a decreasing function of the wage at which he is currently employed.*

This Proposition makes it possible to describe the critical implications of this theorem. The following picture may help make the results in the empirical section, and the connections between the various search models clearer.

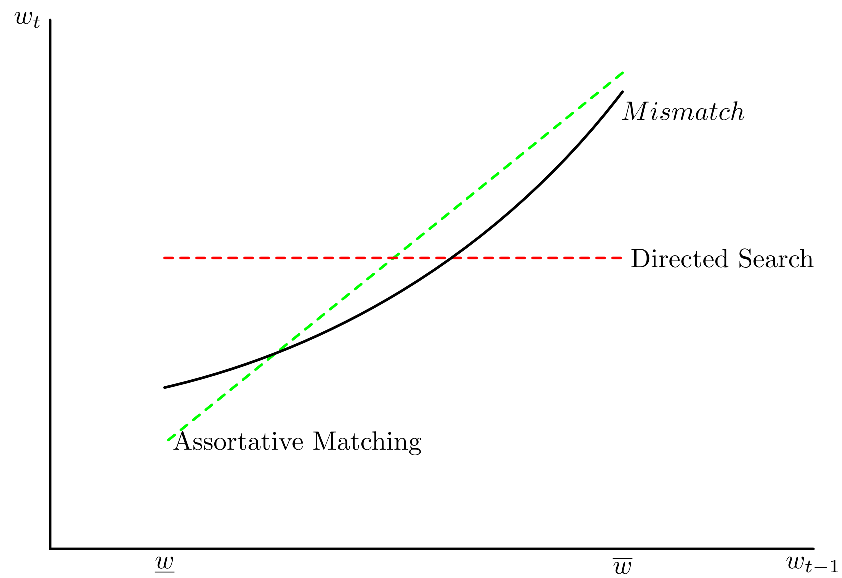

FIGURE 4.1. Comparison of Models

The horizontal axis in the picture represents the current wage of a worker, while the vertical axis represents the wage in the job that the worker moves to after a match is terminated. All the predictive consequences of the model we consider here emerge from studying the predictive content of current wage. To see how this should work, start with the most basic directed search model in which workers are all identical, but firms offer different wages. To support equilibrium, worker used mixed application strategies where they all apply with higher probability to high wage firms than low wage firms. This mixing will lead some workers to find employment in high wage firms. However, once such a high wage match is terminated, the worker will revert to the mixed application strategy, and could end up at any firm after a job transition. The important point is the there should be no correlation between his current wage, and the wage where he ends up after a transition.

The implication is that regressing wage outcomes of workers on their past wages should yield a flat relationship. like the one represented by the flat dashed line in the figure, labeled, "Directed Search". Furthermore, there will be a lot of variance in outcomes for workers at different wages, but this variance won't be connected in any way to their original wage.

Similarly, a directed search model of assortative matching ((Eeckhout and Kircher 2010)) or a wage ladder model ((Shi 2009)) would both predict that the current wage is a perfect predictor of future wage. Regressing future wages on current wage would yield a simple linear relationship with slope 1. Assortative matching would give no systematic relationship between past wage and variance of future wages, while the wage ladder model would give and increasing variance.

Finally, the model of mismatch discussed in this paper says that workers who leave a match will end up with a wage that is a random draw from the distribution of wages above their reservation wage. This has a couple of implications. First, workers' future wages will rise with their current wage but not as quickly because of mismatch. Furthermore, since low type workers apply and get jobs at all wage levels, the wage is a better predictor of worker type at low wages than it is at high wages. This suggests the relationship between past and future wages should be non-linear with the correlation being highest for low wages. Finally, the mis-match model we discuss suggests that the variance of future income should be falling as the current wage rises.

5. Empirical Application - The Executive Market

5.1. **Data.** To look for evidence on mismatch, we used a dataset on executive salaries collected from Compustat and Execucomp. The sample comprises the observed characteristics of both executives and firms for the period from 1992 to 2009. Especially, we observe executives' age, tenure, gender, turnover on his/her career path, basic salary and total compensation. On the firm side, we observe the total assets and their return, sales, employment, and the industry classification code. Table 2 in appendix lists the summary statistics about these observables that we used for our empirical analysis. All monetary terms are converted to the dollar value of year 1992.

To exploit this data, we started by breaking the executives up into various age categories, reasoning that older workers were verifiably more experienced, and would receive a wage premium as a result that would not be related to their type. Table 3 in appendix summarizes some basic information regarding these groups.

For each group, we took the logarithm of the observed wages, and then normalized them by the average of wages of that year and in that group. Doing so effectively made all the wage distributions comparable across groups and years. The normalized wages also eliminated impacts on wage levels due to any common shocks in the year. In essence, we would like to study the worker-firm mismatches and transitions between jobs. We therefore restrain our focus of the study on the moves in these normalized wage distributions, rather than the variations in wage levels. We by default refer to the normalized (logarithm of) wages whenever discussing wages in the empirical part of the paper.

Another issue arose in this empirical application when we made attempts to study wage variations from job transitions - we have to identify the moves of executives in data, as they are not directly observable. To this end, we define a job in the following way. We traced the same worker for any two consecutive years. If she worked in a same firm, involving no changes in both the job title and wage rankings in the firm, we regarded her staying on one same job. Otherwise, we considered her changing jobs, either internally promoted or externally moved. Then, for any job, we computed the averages of earnings over the period that the worker stayed on the job. We used this average as the wage for the job in analysis that follows.

Our theoretical prediction requires the wage distributions be at steady state. Accordingly, we had to ensure the data we used in the regression analysis represented a steady state for each age group. For this purpose, we designed an algorithm to partition the years into two groups - stable years, and unstable years. The stable years are those in which the (normalized) wage offer distributions appeared to come from the same underlying distribution. Unstable years represent years in which the wage distribution is significantly different from the overall distribution in the stable years.

Prior to describing the algorithm, we first define stability in the following sense. A collection of years are referred to as *stable years* if - (i) the wage distributions of stable years are identical, in the sense that the wage distribution of a stable year is the same (in the sense of the Kolmogorov-Smirnov (KS) test) with the distribution of wages in the rest of the years in the stable group; and (ii) the wage distribution from each unstable year is different (according to the KS test) from the distribution of wages in the stable years. For example, suppose there are four different years with associated wage distributions A, B, C, D. If we think the first year, with distribution A is unstable, while the other three are unstable, then we require:

- wage distribution B appears identical (by the KS test) to the distribution generated by combining distributions C and D ;
- wage distribution C appears identical (by the KS test) to the distribution generated by combining distributions B and D;
- wage distribution D appears identical (by the KS test) to the distribution generated by combining distributions B and C;

- wage distribution A appears to be significantly different from the distribution generated by combining the other three distributions.

To find this partition, we started with the guess that 1992 was the only abnormal year, then ran the KS checks as above to see whether the partition where 1992 was unstable, while all other years were stable. If this partition does not satisfy the various checks as above, we move on to check whether 1993 might be unstable while all the others were stable. If all partitions involving a single unstable year failed the checks above, we went on to try groups containing two unstable years. The algorithm stopped the first time a partition satisfied all our checks above.

Effectively, our approach searches for the partition of set that contains the largest set of stable years.[4] We repeated this procedure until we found all the wage distributions justifying the steady state assumption for all age groups.[5]

The testing results are reported in the table 4. There, a 1 indicates that our algorithm found the year to be part of a stable set of years in the sense described above. For example, the table indicates that for the youngest executives, all the years between 1997 and 2009 but one (2002) had wage distributions that appeared to be the same. We only used the observations from the age groups and years that satisfying the steady state assumption for the following empirical analysis.

Before moving onto the regression analysis, we would like to examine two properties that the wage distribution has to satisfy in order for the strong properties predicted by the theory to hold.

The first key prediction of the theory is that the variance of future income should be declining in current income. This follows from the fact that workers make their application decisions by sampling randomly from a wage distribution that is truncated from below. The truncation point is higher the higher the worker's type. Intuitively, variance should fall as a random variable is sampled from a finer and finer event. This isn't true of all distributions, so we simply check whether it is true for the wage distribution available in our executive data.

To check this, we computed the conditional variances using cutoff values equal to quantile values (i.e., 0.1, 0.2, ... 0.9). In particular,

[4]Our approach shares the spirit of (Ruefli and Wiggins 2000), who also used a series of Kolmogorov-Smirnov tests for identifying stratification. The method of this kind is sensitive to all moments of the distributions involved, does not depend on any parametric assumptions, and appears superior than conventional cluster techniques.

[5]We chose critical values at 10% significant level for all KS tests.

we took the cutoff type being the specified quantile, then computed the standard deviations of wages above this threshold. The results are reported in the table 5. The shrinking variance patterns indeed hold for all age groups.

It might seem that wages must somehow be bounded above. Workers with high wages simply have lower variance in wage income when they transit between jobs simply because the keep bumping into this upper bound. A central part of our theory is that mismatch that occurs as a result of workers randomizing their applications will result in workers who move between jobs suffering a wage loss as a result. This potential for wage decline is supposed to represent a large chunk of the variance in future income. As a preliminary sanity check on the data, we checked the distribution of wage changes. The results are in table 6.

The first cell of the table 6 reports 12% of the workers in age group 1, found their ranking in the wage distribution fall by 0.2 or more when they moved between jobs. The 18% in the next cell to the right is the proportion of workers who found their ranking fall by more than .1 when they moved. Since these things add up, this means that 18% of workers found their rank fall by something less than .1. To the right of that, 41% of workers found their rank rise by at least .1, while 27% found their rank rise by at least .2. The other rows represent these changes for the other age (experience) groups.

The point of this table is that there was a lot of wage movement during job transitions. It isn't entirely clear what this table is supposed to look like under the various theories. Wage ladder models would seem to suggest lots of workers increasing their rank a little, with a lump of workers (those who go back to the bottom of the latter) reducing their ranking a lot. This table doesn not seem to reconcile with standard "on-the-job" random search models. One exception is (Cahuc, Postel-Vinay, and Robin 2006) who treated this as a strategic outcome from firm competitions and ex-post multi-lateral bargaining. In our model, we regard decreases in wage ranking as a sign for mismatch due to mixing of application strategy.

5.2. **Regression Analysis.** We then used observations involving job transitions between steady-state years for the next regression analysis. We first regressed the wage a worker earns at his new position ($w_{i,t}$) on the wage he earned at his last position ($w_{i,t-1}$) as in the following equation:

$$w_{i,t} = \alpha_0 + \alpha_1 w_{i,t-1} + \alpha_2 w_{i,t-1}^2 + \gamma_1' X_{i,t} + \mu_{i,t} \tag{5.1}$$

Here$X$ is the vector of observed characteristics of the worker. The squared term appears here because the past wage is a signal of the worker's type. The quality of this signal varies with the wage at which a worker was previously employed. Very low or very high wages are relatively good signals of type, while intermediate wages aren't. The relationship between past and future wages should be non-linear for this reason. The regression coefficients α_1 and α_2 are the basics for the first test as described in Figure 4.1.

Next, we took the absolute values of residuals from the regression (5.1) as conditional standard deviation of current wage, and regressed them on past wage in order to check for a declining variance effect. That is,

$$|\mu_{i,t}| = \beta_0 + \beta_1 w_{i,t-1} + \gamma_2' X_{i,t} + \epsilon_{i,t} \tag{5.2}$$

We present the main results of these regressions in the next table, while leaving out the conditioning variables for simplicity. The complete set of regression results is instead reported in the table 8 in appendix.

TABLE 1. Regression Results

Variables	(1) $w_{i,t}$	(2) $\lvert\mu_{i,t}\rvert$	(3) $w_{i,t}$
$w_{i,t-1}$	0.17	-0.05***	0.17***
	(0.136)	(0.015)	(0.011)
$w^2_{i,t-1}$	0.08		0.08***
	(0.064)		(0.006)
Constant	0.63***	0.09***	0.62***
	(0.063)	(0.015)	(0.008)
Covariates	Yes	Yes	Yes
Observations	34,088	34,088	34,088
R^2	0.854	0.114	0.855
	OLS	OLS	GLS

Column (1) presents the results of regression (5.1), regressing current wages on past wages. We don't read too much into the coefficients associated with $w_{i,t-1}$ and its square terms because we are expecting heterosckedasticity in the sense that the variance of the residuals of this equation might be declining with $w_{i,t-1}$.

To check for the presence of heterosckedasticity, we conducted several formal tests, whose results are reported in the table 7. We first implemented White's general test for heteroskedasticity. Then, we did the standard Breusch-Pagan test, which is designed to test only for the linear format of heteroskedasticity. Both test statistics are associated with $p-$values of 0.000, strongly rejecting the null of homoskedasticity. Another one is Goldfeldt-Quant (GQ) test, which can be used when it is strongly believed that the variance of the error term decreases (or increases) consistently as one of the regressors increases. The F-statistic for our GQ test rejects the null at 1% significance level, suggesting a clear decreasing pattern of variances with respect to $w_{i,t-1}$.

We then regressed the absolute value of residuals from (5.1) on past wages, Column (2) gives the results of regression (5.2). As can be seen, the coefficient is negative and significant, indicating that the variance of outcomes is declining with past wages, again consistent with the theory of mismatch, but inconsistent with either of the other theories of directed search.

The significant coefficient of β_1 suggests the presence of heteroskedasticity. Interpreting the significance of estimates on α_1 and α_2, we had to adjust the estimation procedure taking into account the fact of shrinking conditional variance. To this end, we further assume $Var(\mu_{i,t}|w_{i,t-1}) = \sigma^2/w_{i,t-1}^2$. This specification enables us to conduct a feasible GLS. The results are reported in Column (3) of table 1.

As is apparent current wages rise with past wages. Furthermore, the squared past wage term in the regression illustrates that this effect gets stronger as the wage rises, which is again consistent with our theory of mismatch. Quantitatively, this suggests that the marginal impact of past wages on current wage be between 0.13 and 0.40. [6] It is bounded well below one - apparently ruling out assortative matching, and wage ladders. Meanwhile, it is bounded well away from 0 which would seem to rule out simple directed search.

Prior to concluding remarks, we should like to comment on the complications from our data application. First, it has been well known that an increasing fraction of jobs in the US labor market explicitly pay workers for their performance using bonus pay, commissions, or piece-rate contracts. It has caught quite some attentions from labour economists when studying wage inequalities of this market. One possible view to different pay structures is that they are designed provide incentives for workers.

[6] Our normalized wages in the sample is mostly spread between 0.6 and 1.4, where 1 meaning the mean wage in the age group.

To be effective, incentive pay has to be variable, at least in the short run. So declining variance in future income may be due to the fact that higher wage workers receive less of their income as incentive pay. Since we can observe what proportion of an executives pay comes from basic salary, we attempt to control the variation of this kind from entire wage distribution. This measure is already included in the regression analysis for the table 1.

6. Conclusion

The results reported here are consistent with the theory of mismatch we have presented, but not consistent with either of the other three models of directed search. The model we use is about as simple as it could possibly be. Nonetheless it seems to get some of the basic empirical properties right.

On the theoretical side, there are at least three dimensions in which the model seems to be going too far. First, it assumes that workers' types don't change over the course of their life. This may be the least objectionable assumption. Workers will obviously acquire new skills as they age. Yet these skills are more often than not contractible. For example, a worker who acquires an MBA will probably be compensated for it. We don't interpret this as an improvement in the worker's type. In our regressions, we capture this by adding experience (measured by worker age) and assuming that this will explain much of the rise in income that workers experience as they move between jobs.

We also assume that match termination is independent of worker and firm type. Notice that this is different from the assumption that match termination doesn't depend on duration. It surely does. However, in our theory no one cares about duration per se, and wages represent expected income of the life of the match. The only important assumption is that matches are terminated in a way that maintains the distribution of worker and position types available on the market in any period.

Lastly, we do not allow firms to refuse to hire. If we did, high type firms would refuse to employ very low type workers making current wage a much better signal of worker quality. Pursuing this modification goes well beyond the scope of the current paper.

This brings us to one of the implications of the results here. The results are consistent with a model in which firms use uncontractible information to screen candidates when they hire. A natural empirical question this suggests is whether firms actually value this uncontractible information, or whether it is simply a way of coordinating search. These are independent questions. For example, it could well be that a firm hires a worker because the worker knows one of the bosses relatives. That isn't the same as saying that the firm is willing to pay more to hire someone who knows a relative, nor that the firm is more profitable when it hires someone who knows a relative instead of someone who doesn't. The model here seems to fit well enough to move to a structural approach which tries to estimate the distribution of firm types.

References

Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006): "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 74(2), 323–364.

Delacroix, A., and S. Shi (2006): "Directed Search on the Job and the Wage Ladder," *International Economic Review*, 47, 651–699.

Eeckhout, J., and P. Kircher (2010): "Sorting and Decentralized Price Competion," *Econometrica*, 78(2), 539–574.

Gabaix, X., and A. Landier (2008): "Why Has CEO Pay Increased So Much?," *The Quarterly Journal of Economics*, 123(1), 49–100.

Peters, M. (2000): "Limits of Exact Equilibria for Capacity Constrained Sellers with Costly Search," *Journal of Economic Theory*, 95(2), 139–168.

——— (2010): "Unobservable Heterogeneity in Directed Search," *Econometrica*, 78(4), 1173–1200.

Postel-Vinay, F., and J.-M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70(6), 2295–2350.

Ruefli, T. W., and R. R. Wiggins (2000): "Technical Note: Longitudinal Performance Stratification—An Iterative Kolmogorov-Smirnov Approach," *Management Science*, 46(5), 685–692.

Shi, S. (2009): "Directed Search for Equilibrium Wage Tenure Contracts," *Econometrica*, 77, 561–584.

Tervio, M. (2008): "The Difference That CEOs Make: An Assignment Model Approach," *American Economic Review*, 98(3), 642–68.

7. Appendix

7.1. **(Proof of Theorem 1).**

Proof. Fix a continuous non-decreasing rule $\omega : Y \to W$. Notice that ω is not required in this definition to have range contained in $\overline{G}$, so the proper interpretation is that $\omega(y)$ is the wage that yields the worker his market payoff if he is hired for sure at that wage. If all searching workers apply to all wages at or above their reservation wage, then

$$P(w,y) = \int_{\underline{y}}^{\min\left[\omega^{-1}(w),y\right]} \frac{G(w) - G(\omega(y'))}{G(\overline{w}) - G(\omega(y'))} dF(y').$$

The 'queue size' $p_w(y)$ has to satisfy (3.1), so

$$p_w(y) = \int_{\underline{y}}^{\min\left[\omega^{-1}(w),y\right]} \frac{1}{G(\overline{w}) - G(\omega(y'))} dF(y').$$

To see this observe that for any w,

$$\int_{\underline{w}}^{w} p_{\tilde{w}}(y)\, dG(w) = \int_{\underline{w}}^{w} \int_{\underline{y}}^{\min\left[\omega^{-1}(\tilde{w}),y\right]} \frac{1}{G(\overline{w}) - G(\omega(y'))} dF(y')\, dG(\tilde{w})$$

$$= \int_{\underline{y}}^{\min\left[\omega^{-1}(w),y\right]} \int_{\omega(y')}^{w} \frac{dG(\tilde{w})}{G(\overline{w}) - G(\omega(y'))} dF(y') =$$

$$\int_{\underline{y}}^{\min\left[\omega^{-1}(w),y\right]} \frac{G(w) - G(\omega(y'))}{G(\overline{w}) - G(\omega(y'))} dF(y').$$

This implies that

$$\int_{y}^{\overline{y}} dp_w(\tilde{y}) = \int_{y}^{\omega^{-1}(w)} \frac{1}{G(\overline{w}) - G(\omega(y'))} dF(y'). \tag{7.1}$$

So hiring probabilities will be given by (4.2) provided that workers all use the application strategy described. Given this matching probability we can now describe the condition that $\omega(y)$ has to satisfy in order for them to be willing to follow this strategy. In order for a searching worker of type $\omega(y) > \underline{w}$ to be indifferent between all wages above his reservation wage, it should be that for each $w' > \omega(y)$

$$(w' + \gamma U(y))\, e^{-\int_y^{\overline{y}} dp_{w'}(\tilde{y})} + \left(1 - e^{-\int_y^{\overline{y}} dp_{w'}(\tilde{y})}\right) \gamma U(y)$$
$$= \omega(y) + \gamma U(y),$$

or

$$w' e^{-\int_y^{\overline{y}} dp_{w'}(\tilde{y})} = \omega(y).$$

Taking logs yields

$$\int_{y}^{\overline{y}} dp_{w'}(\tilde{y}) = \log(w') - \log(\omega(y)).$$

By the fundamental theorem of calculus this implies

$$\int_{\omega(y)}^{w'} \frac{1}{\tilde{w}} d\tilde{w} = \int_{y}^{\overline{y}} dp_{w'}(\tilde{y}). \tag{7.2}$$

Substituting (7.1), then gives the identity

$$\int_{\omega(y)}^{w} \frac{1}{\tilde{w}} d\tilde{w} = \int_{y}^{\omega^{-1}(w)} \frac{1}{G(\overline{w}) - G(\omega(y'))} d\mathcal{F}(y')$$

is satisfied for all y. Differentiating both sides with respect to w gives the differential equation

$$\omega'(y) = \frac{\omega(y) F'(y)}{G(\overline{w}) - G(\omega(y))}. \tag{7.3}$$

The reservation wage function ω will support the continuation equilibrium if it has a solution with $\omega(\overline{y}) = \overline{w}$. This is not immediate since the right hand side does not have a continuous derivative around the point $(\overline{y}, \overline{w})$.

However it does have a solution through the point $(\overline{y}, \overline{w} - \epsilon)$ for any $\epsilon > 0$. Denote the solution for $\epsilon > 0$ as $\omega^{\epsilon}(y)$. Observe that each ω^{ϵ} is strictly increasing and that ω^{ϵ} and $\omega^{\epsilon'}$ cannot cross, therefore the sequence $\{\omega^{\epsilon}\}_{\epsilon \to 0}$ is an increasing sequence of increasing functions. As the sequence $\omega^{\epsilon}(y)$ is a bounded increasing sequence of real numbers, ω^{ϵ} converges point-wise, therefore uniformly (Dini's Theorem) to some function ω. If (7.3) fails at some point y, then by uniform convergence, it must fail for small ϵ. So ω is a solution to (7.3).

The remaining bits of the theorem then follow by using (3.2) along with the reservation wage. □

Table 2. Summary statistics

Variable	**NOB**	**Mean**	**Std. Dev.**
ln(Total Compensation)	34088	7.08	1.05
Normalized ln(Total Compensation)	34088	1.01	.15
ln(Salary)	34024	5.82	.61
SP500	34088	.32	.47
ln(Total Asset)	34088	7.42	1.79
Return of Asset	34088	.13	.1
Fama-French 12 Industry Classification	34088	7.39	3.38
ln(Sales)	34088	.98	.23
ln(Employment)	34088	.19	.53
Productivity	34088	3.00	15.40
Age	34088	49.99	7.36
CEO	34088	.13	.33
CFO	34088	.06	.24
Female	34088	.06	.24
Age Group	34088	3.48	1.4
Incentive Share	34088	.64	.23
Tenure	34088	1.38	1.43

7.2. **Additional Tables.**

Table 3. Summary statistics on job transition in groups

Group (Age)	NOB	Freq.	Ave. Wage	# Firms	# Workers
group 1 (30-40)	2567	7.5	6.5	929	1395
group 2 (40-44)	6464	19.0	6.8	1799	3697
group 3 (45-49)	9062	26.6	7.0	2114	5146
group 4 (50-54)	7426	21.8	7.2	1880	4263
group 5 (55-59)	5299	15.5	7.3	1701	3360
group 6 (60-70)	3270	9.6	7.4	1161	1755
Overall	34088	100.0	7.0	2661	13898

Table 4. Testing results on steady-state distributions

Year	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
1992	0	0	0	0	0	0
1993	0	0	0	0	0	0
1994	0	0	0	0	0	0
1995	0	0	0	0	0	0
1996	0	1	0	0	0	0
1997	1	1	1	0	0	0
1998	1	1	1	0	1	0
1999	1	1	1	0	1	0
2000	1	1	1	1	0	1
2001	1	1	1	1	1	1
2002	0	0	1	1	0	1
2003	1	1	1	1	1	1
2004	1	1	1	1	1	1
2005	1	1	1	1	1	1
2006	1	1	1	1	1	1
2007	1	1	0	1	1	1
2008	1	1	1	1	1	0
2009	1	1	1	0	0	0

Table 5. Shrinking conditional standard deviations of wage distributions

Group (Age)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
group 1 (30-40)	0.15	0.15	0.14	0.14	0.13	0.13	0.13	0.12	0.11
group 2 (40-44)	0.14	0.13	0.13	0.12	0.12	0.11	0.11	0.10	0.10
group 3 (45-49)	0.13	0.13	0.12	0.11	0.11	0.10	0.10	0.09	0.09
group 4 (50-54)	0.13	0.12	0.11	0.11	0.10	0.10	0.09	0.08	0.08
group 5 (55-59)	0.13	0.12	0.11	0.11	0.10	0.09	0.08	0.08	0.06
group 6 (60-70)	0.14	0.13	0.12	0.11	0.10	0.10	0.09	0.08	0.07

Table 6. Distribution of wage rank changes

Group (Age)	**-0.2**	**-0.1**	**0**	**+0.1**	**+0.2**
group 1 (30-40)	0.12	0.18	0.36	0.41	0.27
group 2 (40-44)	0.12	0.20	0.40	0.35	0.21
group 3 (45-49)	0.11	0.20	0.43	0.30	0.18
group 4 (50-54)	0.09	0.18	0.42	0.29	0.15
group 5 (55-59)	0.11	0.19	0.42	0.26	0.13
group 6 (60-70)	0.13	0.23	0.47	0.22	0.11
Overall	0.11	0.20	0.42	0.30	0.17

Table 7. Heteroskedasticity test

Test	**Test Statistics**	**p-value**
White	196.5	0.000
Breusch-Pagan	182.8	0.000
Goldfeld-Quandt	1.053	0.003

Table 8: Regression results with full sample

VARIABLES	$w_{i,t}$	$\|\mu_{i,t}\|$	$w_{i,t}$
$w_{i,t-1}$	0.17	-0.05***	0.17***
	(0.136)	(0.015)	(0.011)
$w_{i,t-1}^2$	0.08		0.08***
	(0.064)		(0.006)
incentive share	-0.01	-0.15***	-0.01**
	(0.016)	(0.010)	(0.006)
incentive share^2	0.34***	0.17***	0.34***
	(0.015)	(0.011)	(0.006)
last tenure	-0.00**	-0.00	-0.00***
	(0.001)	(0.001)	(0.001)
last tenure^2	0.00***	0.00	0.00***
	(0.000)	(0.000)	(0.000)
external move	-0.01	0.04***	-0.01***
	(0.005)	(0.004)	(0.002)
firm size	0.01**	0.01***	0.01***
	(0.003)	(0.002)	(0.001)
firm size^2	0.00	-0.00***	0.00***
	(0.000)	(0.000)	(0.000)
return of assets	-0.07***	-0.01	-0.07***
	(0.023)	(0.018)	(0.011)
return of assets^2	0.07**	0.08***	0.07***
	(0.028)	(0.020)	(0.015)
productivity	-0.00	-0.00	-0.00
	(0.000)	(0.000)	(0.000)
productivity^2	-0.00***	0.00	-0.00***
	(0.000)	(0.000)	(0.000)
size × roa	0.01***	0.00	0.01***
	(0.004)	(0.003)	(0.002)
size × productivity	0.00	0.00	0.00
	(0.000)	(0.000)	(0.000)
size × ceo	0.00*	-0.00	0.00***
	(0.001)	(0.000)	(0.000)
size × cfo	-0.00***	-0.00***	-0.00***
	(0.001)	(0.000)	(0.001)
roa × productivity	0.00	0.00	0.00*

Continued on next page

VARIABLES	$w_{i,t}$	$\|\mu_{i,t}\|$	$w_{i,t}$
	(0.000)	(0.000)	(0.000)
roa × ceo	0.02*	-0.02**	0.02**
	(0.011)	(0.007)	(0.008)
roa × cfo	-0.01	-0.00	-0.01
	(0.011)	(0.008)	(0.012)
size × incentive share	0.00	-0.00	0.00
	(0.001)	(0.000)	(0.000)
roa × incentive share	-0.05*	-0.02	-0.05***
	(0.024)	(0.016)	(0.011)
female	-0.00*	-0.00*	-0.00**
	(0.001)	(0.001)	(0.001)
sp500	0.00	-0.00	0.00***
	(0.002)	(0.001)	(0.001)
ceo	0.03***	0.01***	0.03***
	(0.006)	(0.004)	(0.004)
cfo	0.02***	0.00	0.02***
	(0.005)	(0.003)	(0.006)
Constant	0.63***	0.09***	0.62***
	(0.063)	(0.015)	(0.008)
fixed effects	YES	YES	YES
Observations	34,088	34,088	34,088
R^2	0.854	0.114	0.855
	OLS	OLS	GLS

Continued on next page