

Estimation of a Roy/Search/Compensating Differential  
Model of the Labor Market  
(Preliminary and Incomplete)

Christopher Taber

University of Wisconsin-Madison

Rune Vejlin

Aarhus University

May 2, 2013

# 1 Introduction

The four most important models of post-schooling wage determination in economics are almost certainly human capital, the Roy model, the compensating differentials model, and the search model. All four lead to wage heterogeneity. While separating human capital accumulation from the others is quite common, we know remarkably little about the relative importance and interactions of the other three sources of inequality. The goal of this paper is to quantify the effect of each of these explanations on overall wage inequality and also to investigate how they interact.

Human capital in general takes many forms. We will take human capital to mean post-schooling productivity. The main idea behind our notion of human capital is that this will change while the worker is on the job. This could happen for different reasons by we think of it as learning-by-doing. Increasing human capital increases productivity of the worker thereby creating wage heterogeneity.

The key aspect of the Roy model is comparative advantage in which some workers earn more than others as a result of different skill levels at labor market entry. Workers choose the job for which they achieve the highest level of wages. By contrast, in a compensating wage differentials model a worker is willing to be paid less in order to work on a job that they enjoy more. Thus, workers with identical skills and job opportunities can earn different salaries. Finally, workers may just have had poor luck in finding their ideal job. This type of search friction can also lead to heterogeneity in wages as some workers may work for higher wage firms. In short, one worker may earn more than another a) because he has accumulated more human capital while working (human capital), b) because he has more talent at labor market entry (Roy model), c) because he has chosen more unpleasant job (compensating differentials), or d) because he has had better luck in finding a good job (search frictions). The goal of this work is to uncover the contribution of these different components to overall wages inequality.

We develop and estimate a structural model of wage determination that contains elements of all four models. The model is estimated on Danish matched employer-employee data. We use the estimated parameters to decompose overall wage inequality into the four components in various ways. We find that while all four models are important contributors to overall earnings inequality, the Roy model inequality is the most important. The precise way in which the components matter depends on the way we measure them as there are interesting interactions between the components.

We briefly discuss the relationship between this work and the previous literature in Section 2. We then describe the model and the decomposition in Section 3. Identification is discussed in 4 and the specific econometric specification is presented in Section 5. Obtaining the right data is crucial to this exercise. Ideally one needs matched employer/employee data as well as a long panel on workers. We describe the data in section 6. Section 7 presents the auxiliary model that we use and Section 8 presents the results.

## 2 Relation to Other Work

Clearly there is a huge amount of work on search models, on the Roy model, on human capital acquired on the job and on compensating differentials. A full review of all of these literatures is beyond the scope of this paper. However, we discuss the relationship between our work and a few other papers. Two important related literatures were started by Abowd, Kramarz, and Margolis (1999) and by Postel-Vinay and Robin (2002). Both of these other papers use the Declarations Annuelles des Donnees Sociales (DADS) data set which is panel data on both firms and workers from France. Abowd, Kramarz, and Margolis (1999) use a fixed effect approach to estimate firm effects and worker effects while Postel-Vinay and Robin (2002) estimate a structural equilibrium search model.

Abowd, Kramarz, and Margolis (1999) estimate a model analogous to

$$\log(W_{itj}) = X'_{it}\beta + \theta_i + \mu_j + \zeta_{ijt} \quad (1)$$

where  $\theta_i$  is an individual fixed effect,  $\mu_j$  is a firm fixed effect, and  $\zeta_{ijt}$  is independent and identically distributed. We use this as a motivation for our auxiliary model. One major way in which we will build on their work is the inclusion of compensating differentials. In estimating both the firm specific effect in wages and the firm specific effect in utility, we can simulate how much of the differences in firms wages that we observe seems to occur from market inefficiencies (search) and how much occurs as a result of workers choice (compensating differentials). Furthermore, we write down a sample selection model for our analogue of  $\zeta_{itj}$  which makes interpretation of the model easier. Finally, we have all of the advantages of a structural model which further helps in interpretation of the model and allows us to use it for policy simulations.

Postel-Vinay and Robin (2002) decompose wage inequality into a search component and an ability related component. They find that for skilled workers the individual component is moderately important (close to 40%), but that for low skill workers virtually all of the

inequality can be assigned to search frictions. The main components that we add is that we allow for non pecuniary benefits and comparative advantage in jobs. Postel-Vinay and Robin allow for absolute advantage only - ability is one dimension and the relative productivity of two workers does not vary across firms. By contrast, we allow for a match specific effect ( $v_{ij}$ ) meaning that some workers match better with some firms. Furthermore, our estimation and identification strategy are very different. They estimate assuming that the model is in steady state, so the information we get from looking at the rate at which people switch jobs and from the revealed preference argument is not a source of identification their model.

Bagger, Fontaine, Postel-Vinay, and Robin (2011) extends Postel-Vinay and Robin (2002) in order to incorporate general human capital accumulated while working. Their goal is to separate the concave life-cycle wage profile into search and human capital. Human capital accumulation is found to be the most important source of wage growth for workers early in their careers. However, this is soon surpassed by search-induced wage growth. This is especially true for low educated workers. In general high educated workers have a higher return to experience than low educated workers.

Another important related literature is one started by the work of Keane and Wolpin (1997) and the many papers building on their model. They estimate a model that includes compensating differentials, human capital, and Roy model inequality. They do not explicitly incorporate search frictions and do not make use of firm level data. Becker (2009) uses a framework similar to ours in that it incorporates compensating differentials into a search models. However, it does not allow for as much Roy flexibility as we do and is not estimated using firm level data and focuses more on unemployment insurance. Another nice paper that has aspects of the four models we discuss above is Sullivan (2010). This paper includes elements of all four of our models above, though this is not the main focus (and while there are search frictions it is not a standard search model). From a modeling stand point the main difference between our paper and his is the determination of wages which he specifies using a reduced form. The second main difference is the type of data we use. We show in our identification section that matched worker/firm data is essential to perform our exercise. The combination of these two components is essential for our model in distinguishing the extent to which firm/worker effects are due to Roy model inequality or compensating differentials. It is also important for distinguishing between search frictions and compensating differentials in explaining firm wage premiums.

Dey and Flinn (2005) and Dey and Flinn (2008) estimate search models with a particular

type of non-wage characteristics: health insurance. Other than that, our models are quite different. Sullivan and Too (2011) is more similar in that they estimate a job search model with a general form of non-wage job characteristics. Thus, the model include search and compensating differentials. However, many differences exist between their paper and ours. First, they specify output as being only match specific, while we allow workers and firms to have constant ability and productivity across matches as well as match specific. The reason for this choice by Sullivan and Too (2011) is likely motivated by the use of NLSY data in the estimation. The NLSY follows only workers, so it is not suited for dealing with firms. Secondly, their model is only a partial equilibrium model in the sense that workers draw a wage and a non-wage component, but there is no negotiation between firms and workers. E.g. a firm does not try to negotiate the wage down in a match where the worker have a high value of the non-wage component. Finally, their model does not include human capital.

For a good discussion of the Roy model see Roy (1951), Heckman and Sedlacek (1985), Heckman and Honoré (1990), or Heckman and Taber (2008). Rosen (1987) provides an excellent discussion of compensating differentials models and Eckstein and Van den Berg (2007) provide a nice discussion of empirical search models.

### 3 The Model and Decomposition

We present a continuous time model in which agents are infinitely lived. Wages are determined similarly to Cahuc, Postel-Vinay, and Robin (2006), Dey and Flinn (2005), and Bagger, Fontaine, Postel-Vinay, and Robin (2011).<sup>1</sup> A substantive difference between our paper and most of the search literature is that we assume that there are finite types of establishments indexed  $j = 1, \dots, J$  with  $j = 0$  denoting nonemployment. Job offers from each type of establishment arrive at rate  $\lambda_j^n$  for nonemployed workers and  $\lambda_j^e$  for employed workers. Human capital is completely general and takes on a discrete set of values  $\psi_0, \dots, \psi_H$ . When individuals are employed human capital appreciates randomly to the next level ( $\psi_h$  to  $\psi_{h+1}$ ) at rate  $\lambda_h$ .<sup>2</sup> We let  $\pi_{ij}\psi_h$  be the productivity of worker  $i$  at establishment  $j$  when the worker has human capital  $\psi_h$ . In the bargaining protocol presented later the object of negotiation is the human capital rental rate. That is the employer and worker agree on a

---

<sup>1</sup>Note, that we assume bargaining over wages as opposed to wage posting. Hall and Krueger (2012) show that there is mixed evidence regarding the wage determination process. In their survey around one third of all workers report having bargained over their wage. Another third reports that they had precise information about the wage before meeting the employer, which is a sign of wage posting.

<sup>2</sup>And does not accumulate when people are not working.

rental rate  $R$  that is fixed until the next negotiation. This means that when human capital is augmented the wage is not renegotiated but automatically rises from  $R\psi_h$  to  $R\psi_{h+1}$ . Thus, the maximum rental rate that a firm is willing to pay is  $\pi_{ij}$ .<sup>3</sup> Notice, that as human capital augments, wages increase proportionally. Worker  $i$ 's flow utility from working at job  $j$  with wage  $W$  is  $u_{ij}(W)$ .<sup>4</sup>

Following Bagger, Fontaine, Postel-Vinay, and Robin (2011) a key aspect of this model is that when a worker receives an outside offer the human capital rental rate is determined by a form of generalized Nash Bargaining between the two firms. This form of wage setting leads to efficient turnover. The rental rate is kept fixed until both parties agree to renegotiate it, or when the job spell ends.<sup>5</sup>

Define  $V_{ijh}(R)$  to be the value function for worker  $i$  with the rental rate  $R$  working in job  $j$  and having human capital level  $h$ . Workers who are nonemployed will have flow utility  $U_{i0h}$  and value function  $V_{i0h}$ . If a non-employed worker  $i$  with human capital  $\psi_h$  receives an offer from firm  $j$ , they will take the job if  $V_{ijh}(\pi_{ij}) > V_{i0h}$ . In general the rental rate is denoted  $R_{ij\ell h}$ , since it depends on worker  $i$ , the current establishment  $j$ , the best outside option  $\ell$ , and units of human capital,  $h$ , *at the negotiation time*.

We assume that the negotiated rental rate,  $R_{ij0h}$ , for a nonemployed worker meeting firm  $j$  is defined by

$$V_{ijh}(R_{ij0h}) = \beta V_{ijh}(\pi_{ij}) + (1 - \beta)V_{i0h} \quad (2)$$

where  $\beta$  is the worker's bargaining power.<sup>6</sup>

Now suppose that worker  $i$  is working at establishment  $j$  and receives an outside offer from establishment  $\ell$ . One of three things can happen. First, the new job offer could dominate the old one,  $V_{i\ell h}(\pi_{i\ell}) > V_{ijh}(\pi_{ij})$ . In this case the worker will switch to the new establishment. Her new rental rate  $R_{i\ell jh}$  will be determined by

$$V_{i\ell h}(R_{i\ell jh}) = \beta V_{i\ell h}(\pi_{i\ell}) + (1 - \beta)V_{ijh}(\pi_{ij}) \quad (3)$$

---

<sup>3</sup>We are implicitly assuming the value of a vacancy is zero. One could easily relax this assumption and just redefine  $\pi_{ij}$  to be the maximum rental rate a firm would ever offer.

<sup>4</sup>Note that by assuming people have utility for wages we are not allowing for borrowing and lending.

<sup>5</sup>In the presentation of this model we assume that workers would never want to renegotiate when their human capital augments. This does not have to be true as a worker may prefer nonemployment to the current job under some circumstances. However, this will not happen in our empirical specification so we abstract from it here.

<sup>6</sup>Given that workers are bargaining over utility and firms care about profit this does not fit in the standard Nash bargaining framework. We take this more as a functional form assumption: wages are indeterminate and this will give a wage that both parties will agree to. It has the nice property that if there is surplus in the match, when  $\beta = 1$  the worker get all of the surplus, when  $\beta = 0$  the firm gets all of the surplus, and when  $0 < \beta < 1$  the surplus between them is split.

If  $V_{i\ell h}(\pi_{i\ell}) < V_{ijh}(\pi_{ij})$  then the worker has the option to renegotiate their wage. If they choose to renegotiate, their new rental rate will be determined by

$$V_{ijh}(R_{ij\ell h}) = \beta V_{ijh}(\pi_{ij}) + (1 - \beta)V_{i\ell h}(\pi_{i\ell}). \quad (4)$$

Thus the worker's decision to renegotiate will just depend on whether  $R_{ij\ell h}$  is higher than her current rental rate. If so, she will renegotiate the contract, if not she will ignore the offer. This condition is relatively straight forward to check. Let  $R_{ij\ell h}$  be the current rental rate. If  $V_{ijh}(R_{ij\ell h}) < V_{i\ell h}(\pi_{i\ell})$ , then the worker will want to renegotiate. Otherwise she won't. Note that we have been a bit sloppy with notation as we use the notation  $R_{ij\ell h}$  to denote the rental rate that worker  $i$  with human capital  $\psi_h$  at the time of negotiation would receive from firm  $j$  when their outside option was firm  $\ell$ . As one can see from equations (3) and (4), it will be the same regardless of whether they started at firm  $\ell$  and moved to  $j$  or if they started at  $j$  and then used an outside offer from firm  $\ell$  to renegotiate their wage, but this is a result rather than an assumption.

To solve the model we need to be able to calculate the value functions  $V_{ijt}(r)$  and  $V_{i0t}$ .

It is convenient to define

$$\begin{aligned} \Lambda_{ijh}(R) &\equiv \sum_{\{\ell: V_{ijh}(R) < V_{i\ell h}(\pi_{i\ell})\}} \lambda_{\ell}^e \\ \Lambda_{i0t}^n &\equiv \sum_{\{\ell: V_{i\ell h}(\pi_{i\ell}) > V_{i0h}\}} \lambda_{\ell}^n \end{aligned}$$

The first is the arrival rate of any job that leads to either renegotiation or switching employers and the second is the arrival rate of any job that the worker would take from nonemployment.

First consider the value function for an employed worker. With an exogenous job destruction rate of  $\delta$  and a discount rate of  $\rho$ , we can solve for  $V_{ijh}(r)$  for  $h < H$  implicitly by

$$\begin{aligned} &(\rho + \delta + \lambda_h + \Lambda_{ijh}^e(R)) V_{ijh}(R) \\ &= U_{ij}(r\psi_h) + \sum_{\{\ell: V_{ijh}(R) < V_{i\ell h}(\pi_{i\ell}) \leq V_{ijh}(\pi_{ij})\}} \lambda_{\ell}^e [\beta V_{ijh}(\pi_{ij}) + (1 - \beta)V_{i\ell h}(\pi_{i\ell})] \\ &+ \sum_{\{\ell: V_{i\ell h}(\pi_{i\ell}) < V_{ijh}(\pi_{ij})\}} \lambda_{\ell}^e [\beta V_{i\ell h}(\pi_{i\ell}) + (1 - \beta)V_{ijh}(\pi_{ij})] + \delta V_{i0h} + \lambda_h V_{ijh+1}(R). \end{aligned}$$

When  $h = H$  we get

$$\begin{aligned}
& (\rho + \delta + \Lambda_{ijh}^e(R)) V_{ijH}(R) \\
&= U_{ij}(r\psi_H) + \sum_{\{\ell: V_{ijH}(R) < V_{i\ell H}(\pi_{i\ell}) \leq V_{ijH}(\pi_{ij})\}} \lambda_\ell^e [\beta V_{ijH}(\pi_{ij}) + (1 - \beta) V_{i\ell H}(\pi_{i\ell})] \\
&+ \sum_{\{\ell: V_{i\ell H}(\pi_{i\ell}) < V_{ijH}(\pi_{ij})\}} \lambda_\ell^e [\beta V_{i\ell H}(\pi_{i\ell}) + (1 - \beta) V_{ijH}(\pi_{ij})] + \delta V_{i0H}.
\end{aligned}$$

Since human capital does not augment when workers are nonemployed we can write the value function for nonemployed workers as

$$(\rho + \Lambda_{i0h}^n) V_{i0h} = U_{i0h} + \sum_{\{\ell: V_{i\ell h}(\pi_{i\ell}) > V_{i0h}\}} \lambda_j^n [\beta V_{i\ell h}(\pi_{i\ell}) + (1 - \beta) V_{i0h}].$$

This is the full model. There are many other features in the labor market that we have abstracted from. This is intentional. Our goal here is not to write down the most complicated model that is computationally feasible, but rather to write down the simplest model that captures the essence of our four models and allows us to distinguish between them. To see this, next consider a decomposition that allows us to understand the various components. We can choose any measure of wage inequality that we want (for example the variance of log wages). In the context of the model we have written down, one can see the different sources of wage inequality:

- Worker variation in potential rental rates  $\pi_{i1}, \dots, \pi_{iJ}$  leads to “Roy model” inequality
- Variation in the function  $U_{ij}(\cdot)$  across workers accommodate “compensating differentials” inequality (and the mean of  $U_{ij}(\cdot)$  will vary across jobs)
- $\lambda_j^e$  and  $\lambda_j^n$  incorporate search frictions - note that these effect wage inequality in two different ways: directly through the job at which one takes, and indirectly through the negotiation process
- Variation in  $\psi_h$  incorporates human capital

After estimating the parameters of the model, we use it to decompose overall post-schooling wage inequality into the different components. An orthogonal decomposition does not exist, so one can perform this decomposition in a number of different way. More generally the different sources interact. Perhaps most importantly, the fact that workers have comparative advantage in some jobs rather than others interacts with search frictions since search frictions

restrict not only access to firms with generous wage policies but also to firms that are good matches.

Thus, many possible ways to decompose wages exist and we use different ones to highlight different features of the model. The following simulations represent one example of a decomposition we take. We sequentially take the following steps.

- a) First simulate the cross section variance of wages using all parameters (which should be approximately the same as overall wage inequality in the data).
- b) set  $\lambda_h$  arbitrarily high so that workers obtain their maximum human capital immediately
- c) set  $\beta = 1$  so to eliminate variation coming from renegotiated wages
- d) Set  $\lambda_j^e$  arbitrarily large so that employed workers move to the preferred job immediately.
- e) In addition set rental rates to be  $R_{ij} = E(R_{ij} | j)$  for all  $j$  eliminating variation due to skill differences so that firms pay constant wages (but holding preferences across jobs constant).

The difference between a) and b) is due to human capital, the difference between b) and d) is due search frictions, the difference between e) and d) is due to Roy model inequality, and the remaining fraction in e) is due to compensating differentials.

## 4 Identification

In this section we discuss non-parametric identification of our model. We show which aspects of the model can and cannot be identified. We view both parts as important. This is empirically relevant in that we can not credibly simulate counterfactuals that are not identified from the data. We will respect this in our counterfactual exercises below by only simulating counterfactuals that can be identified.

Specifically, it will turn out that two aspects of our model are not identified. First, we show that one can not hope to identify  $R_{ij}$  (the rental rate that worker  $i$  would get at job  $j$ ) for a worker  $i$  on a job  $j$  that they would never take (i.e.  $V_{ij} < V_{i0}$ ). We suspect that this will not come as a great surprise, but it is important to keep in mind that this limits the type of counterfactuals which can be simulated. We view this not as a limit of our model or

particular data, but rather as a fundamental identification problem that will be an issue for non-parametric identification with any model and data set. It is essentially the identification problem discussed in (Flinn and Heckman 1982).

Second, we can not identify the bargaining parameter  $\beta$ . The reason is that we can use a revealed preference argument to identify the preference workers have across jobs, but we can not identify the levels of utility. For that reason we can not identify the bargaining parameter  $\beta$  as we can not separate out its level from the magnitude taste for jobs. We can, however, test whether  $\beta = 1$  and simulate a model without bargaining in which the worker receives a competitive wage offer ( $\beta = 1$ ).

Identification crucially depends on two types of arguments that are somewhat nonstandard and also depends on some special aspects of our data. The first is that we follow Villanueva (2007) (and others) by using a revealed preference argument that a worker has shown a preference for one job over another if they directly leave the first job to start the second. As a result it will be very important for us to distinguish job-to-job transitions (in which we will use our revealed preference argument) from job-to nonemployment-to job transitions (where we are not willing to use this argument). Intuitively, if we consistently observed that workers were willing to take wage cuts to go to a certain firm, this would indicate that this firm had high non-pecuniary benefits. A limitation of this approach is that in practice not every job-to-job transition is voluntary. In the Survey of Income and Program Participation (SIPP) approximately 15% of job to job transitions result from layoffs. We have also approximated this in a Danish context by linking the Danish Work Environment Cohort Study from 1995 and 2000 with information from the spell data used in this study. Of those interviewed that we have labeled as a job to job transitions 16 and 24 % say that they left on the employers initiative, respectively. We suspect that this problem is unlikely to drive the results, but it is something we will address in a future draft of the paper.

A second key aspect of the model is that we show that the arrival rates ( $\lambda_j^e$ ) can be identified by the rate at which workers switch jobs. If the reason that all workers do not work for the highest paying firm is because of search frictions, then eventually they should match with the highest paying firm. Thus, if search frictions are a very important component of inequality, the rate of job switching should be fairly slow. However, if search frictions are relatively unimportant (arrival rates are high) workers will quickly receive an offer from their preferred firm.<sup>7</sup> Thus it is important to have high quality panel data on job switching and

---

<sup>7</sup>It is important to point out that this does not necessarily incorporate all forms of search frictions. If the worker's first job restricts all jobs they can subsequently obtain, this will look identical to what we call

also matched employer-employee data.

Proving general nonparametric identification of the model when  $J$  is very large seems overly tedious, so instead we focus on a simpler case to illustrate how our model can be identified. There are two different types of jobs workers can get  $A$  and  $B$  and human capital only takes on two values,  $h = \{0, 1\}$ . We fully expect our result to generalize to larger (but finite)  $J$ . First consider what can be identified without data on wages.

We assume that we observe workers from time 0 to  $T$  and that all workers begin their working life nonemployed.<sup>8</sup> We observe labor force status during a worker's entire life - that is whether the worker is working and if so, the type of firm for which they work. We only observe wages at a finite number of times. For simplicity assume it is at the integers ( $t = 1.0, 2.0, \dots$ ). Let  $w_{it}^m$  denote the log of the wage measured at time  $t$  if the worker is working at time  $t$ .

The easiest parameter to identify is  $\delta$  which is directly identified from the data as the hazard rate out of employment.

Next consider identification of  $\lambda_A^n$ . Let  $L_{iA}^n$  be the length of a nonemployment spell for any worker whose first job was  $A$ . For any  $\tau_1 < \tau_2 < T$ , the conditional density of  $L_{iA}^n$  evaluated at  $\tau_1$  when we condition on  $L_{iA}^n < \tau_2$  is

$$\frac{\lambda_A^n e^{-\lambda_A^n \tau_1}}{1 - e^{-\lambda_A^n \tau_2}}.$$

Since this expression only depends on  $\lambda_A^n$ , it is identified.<sup>9</sup> Similarly  $\lambda_B^n$  is identified.

The arrival rates of offers on the job can be identified in the same way. That is for a worker who starts at  $B$  and moves to  $A$ , let  $L_{iA}^e$  be their length of employment at job  $B$  before taking  $A$ . For any  $\tau_1 < \tau_2 < T$ , the conditional density of  $L_{iA}^e$  evaluated at  $\tau_1$  when we condition on  $L_{iA}^e < \tau_2$  is

$$\frac{\lambda_A^e e^{-\lambda_A^e \tau_1}}{1 - e^{-\lambda_A^e \tau_2}}.$$

One can get  $\lambda_B^e$  analogously.

In terms of workers choice we can define a "type" by their preferences over jobs. For

---

Roy model heterogeneity. However, the type of search friction we have identified in the text of this paper is the most common type in equilibrium search models such as Burdett and Mortensen (1998) or Postel-Vinay and Robin (2002).

<sup>8</sup>In the data this is not true for all workers. However in terms of identification we can focus only on those workers who we observe at the beginning of their careers.

<sup>9</sup>To see this one can get a closed form solution for  $\lambda_A^n$  by taking the log of the ratio of this density evaluated at two different  $\tau_1$  points but the same  $\tau_2$  point.

example one type is:

$$V_{iA0}(\pi_{iA}) < V_{i00} < V_{iB0}(\pi_{iB}), V_{i01} < V_{iA1}(\pi_{iA}) < V_{iB1}(\pi_{iB}).$$

We can define all of these different combinations of types. However, to keep this already complicated section tractable we will assume that there are only 5 different types. Given our panel data, we fully expect the ideas here to extend beyond this case. The five types are:

Type	$h = 0$	$h = 1$
0	$V_{iA0}(\pi_{iA}) < V_{i00}, V_{iB0}(\pi_{iB}) < V_{i00}$	NA
B0	$V_{iA0}(\pi_{iA}) < V_{i00} < V_{iB0}(\pi_{iB})$	$V_{iA1}(\pi_{iA}) < V_{i01} < V_{iB1}(\pi_{iB})$
A0	$V_{iB0}(\pi_{iB}) < V_{i00} < V_{iA0}(\pi_{iA})$	$V_{iB1}(\pi_{iB}) < V_{i01} < V_{iA1}(\pi_{iA})$
BA	$V_{i00} < V_{iA0}(\pi_{iA}) < V_{iB0}(\pi_{iB})$	$V_{i01} < V_{iA1}(\pi_{iA}) < V_{iB1}(\pi_{iB})$
AB	$V_{i00} < V_{iB0}(\pi_{iB}) < V_{iA0}(\pi_{iA})$	$V_{i01} < V_{iB1}(\pi_{iB}) < V_{iA1}(\pi_{iA})$

What we have disallowed for is workers changing ordering after human capital accumulates. Again, we see no reason why this would be an issue, but focus on this case for simplicity.

Let  $D_{it}$  denote the labor force status of worker  $i$  at date  $t$ . This can take three different values:  $A$  if working for firm  $A$ ,  $B$  if working for firm  $B$ , or  $N$  for being nonemployed. Now choose any sequence of  $K$  time periods,  $\tau_1, \dots, \tau_K$ . Dente the labor market statuses at these time periods  $d_{\tau_1}, \dots, d_{\tau_K}$ . Let *type* index the workers type, then since we know the model for turnover, for each type we can calculate

$$Pr(D_{i\tau_1} = d_{\tau_1}, \dots, D_{i\tau_K} = d_{\tau_K} \mid type)$$

from the set of parameters  $(\delta, \lambda_A^n, \lambda_B^n, \lambda_A^e, \lambda_B^e)$ . Note that this typically will not have a closed form solution, but in principle it is known and in practice it can be approximated arbitrarily well. Moreover since different types make different decisions this function differs across them in nonlinear ways. Then from the data we can identify

$$Pr(D_{i\tau_1} = d_{\tau_1}, \dots, D_{i\tau_K} = d_{\tau_K}) = \sum_{type} Pr(D_{i\tau_1} = d_{\tau_1}, \dots, D_{i\tau_K} = d_{\tau_K} \mid type) Pr(type).$$

Note that this is a linear function of  $Pr(type)$  where the other components are identified. Thus, we can generate an uncountable number of these equations using different permutations  $(d_1, \dots, d_T)$  at different time periods. Since these conditional probabilities are nonlinear in type probability, with enough of these linear equations we can identify  $Pr(type)$  for all of the different types.

We now incorporate information from wages. Without loss of generality we can normalize  $\psi_0 = 1$ . We need to define some notation that is cumbersome, but necessary in order for us to be precise. For any worker who is currently working, there are four different states which are relevant for their wages: their current employer, their current human capital, the outside option when their current wage was negotiated, and the level of human capital when the current wage was negotiated. We denote these as functions of the individual and time as  $j(i, t)$ ,  $h(i, t)$ ,  $\ell(i, t)$  and  $h_0(i, t)$ , respectively. Then for each integer  $t$  at which time the agent is working, we observe

$$w_{it}^m \equiv \log(R_{ij(i,t)\ell(i,t)h_0(i,t)}) + \log(\psi_{[h(i,t)-h_0(i,t)]}) + \xi_{it}$$

where  $\xi_{it}$  is i.i.d. measurement error. One can see here the distinction between  $h(i, t)$  which is the current level of human capital and  $h_0(i, t)$  which is the level of human capital when the current rental rate was negotiated. If human capital was 0 when it was negotiated but 1 at time  $t$  then the term  $\log(\psi_{[h(i,t)-h_0(i,t)]})$  will show up.

To begin we condition on  $BA$  types. A random subset of them are easy to identify in the data as they constitute anyone who ever makes a job to job transition from  $A$  to  $B$  sometime before time  $T$ . We condition on this group, but do not make this conditioning explicit in the following for expositional simplicity.

Further condition on  $BA$  individuals who

- Are non-employed until time  $1 - d_1$
- Start working on job A at time  $1 - d_1$ , leave to nonemployment at  $1 + d_2$
- Are nonemployed until time  $2 - d_3$  when they start again at a type A firm and they stay through period 2

From this we can identify the joint distribution of

$$(w_{i1}^m, w_{i2}^m)$$

conditional on  $d_1, d_2$ , and  $d_3$ .

Take limits of this conditional distribution as  $d_1 \downarrow 0, d_2 \downarrow 0$ , and  $d_3 \downarrow 0$  and we get the distribution of

$$(\log(R_{iA00}) + \xi_{i1}, \log(R_{iA00}) + \xi_{i2})$$

for the  $BA$  types. Both of these wages will be  $\log(R_{iA00})$  because the people have not had enough time to accumulate human capital or get an outside offer. Notice, that since  $\psi_0 = 1$  then  $R_{iA00} = W_{iA00}$ , i.e. the rental rate is equal to the wage paid. Using Kotlarski's lemma (Kotlarski 1967) we can identify the the marginal distributions of both the measurement error and  $\log(R_{iA00})$ .

Consider the same group and continue to take  $\delta_1 \downarrow 0$  and  $\delta_3 \downarrow 0$ , but allow  $d_2$  to vary then let  $\phi_\xi(s)$  be the characteristic function of the measurement error which is identified. Assuming that the characteristic function of the measurement error does not vanish, we can identify

$$\frac{E(e^{isw_{i2}^m})}{\phi_\xi(s)} \approx e^{-\lambda_h \delta_2} E(e^{is \log(R_{iA00})}) + (1 - e^{-\lambda_h \delta_2}) E(e^{is \log(R_{iA01})}).$$

By varying  $\delta_2$  we can identify  $\lambda_h$  and  $E(e^{is \log(R_{iA01})})$ .<sup>10</sup>

Now consider individuals whose first job is  $A$ , they stay at the job through  $t = 1$  and  $t = 2$ , but eventually move to  $B$ . For them at each time period  $t = 1$  and  $t = 2$  there are 4 potential labor statuses:

1. Never got outside offer,  $h = 0$
2. Outside offer from  $A$ ,  $h = 0$
3. Never got outside offer,  $h = 1$
4. Outside offer from  $A$ ,  $h = 1$

---

<sup>10</sup>To see how, take the ratio of the derivatives of this function in terms of  $\delta_2$  at two different values of  $\delta_2$  and it will be a known function of  $\lambda_h$ .

The derivative of that expression with respect to  $\delta_2$  is

$$-\lambda_h e^{-\lambda_h \delta_2} E(e^{is \log(R_{iA00})}) + \lambda_h e^{-\lambda_h \delta_2} E(e^{is \log(R_{iA01})}) = \lambda_h e^{-\lambda_h \delta_2} [E(e^{is \log(R_{iA01})}) - E(e^{is \log(R_{iA00})})]$$

Now take the ratio of this at two different values of  $\delta_2$  say  $\delta_2^a$  and  $\delta_2^b$  then

$$\begin{aligned} \Delta(\delta_2^a, \delta_2^b) &\equiv \frac{\lambda_h e^{-\lambda_h \delta_2^a} [E(e^{is \log(R_{iA01})}) - E(e^{is \log(R_{iA00})})]}{\lambda_h e^{-\lambda_h \delta_2^b} [E(e^{is \log(R_{iA01})}) - E(e^{is \log(R_{iA00})})]} \\ &= e^{\lambda_h (\delta_2^b - \delta_2^a)} \end{aligned}$$

$\Delta(\delta_2^a, \delta_2^b)$  is directly identified from the data and

$$\lambda_h = \frac{\log(\Delta(\delta_2^a, \delta_2^b))}{\delta_2^b - \delta_2^a}$$

They can not have gotten an offer from firm B or they would have accepted it. When a worker at A gets an outside offer from another A firm, the bargained wage satisfies

$$V_{iAh}(R_{iAAh}) = \beta V_{iAh}(\pi_{iA}) + (1 - \beta)V_{iAh}(\pi_{iA}),$$

so  $R_{iAAh} = \pi_{iA}$ . Thus for status 4, the log wage will be  $\pi_{iA} + h_1$  regardless of whether the outside offer or human capital innovation came first. Let  $p_{r_1 r_2}(d_1)$  represent the probability that the status is  $r_1$  in the first period and  $r_2$  in the second for someone who started the job at time  $1 - d_1$ . Then let  $\phi_{w_{i1}^m, w_{i2}^m}(s_1, s_2; d_1)$  be the characteristic function of the joint distribution of  $(w_{i1}^m, w_{i2}^m)$  conditional on BA people who started job A at time  $1 - d_1$  and still employed there through time  $t = 2$ . Then we can identify

$$\begin{aligned} \frac{\phi_{w_{i1}^m, w_{i2}^m}(s_1, s_2; d_1)}{\phi_\varepsilon(s_1)\phi_\varepsilon(s_2)} = & p_{11}(d_1) E(e^{\nu(s_1+s_2)\log(R_{iA00})}) + p_{12}(d_1) E(e^{\nu(s_1\log(R_{iA00})+s_2\log(\pi_{iA}))}) \\ & + p_{13}(d_1) E(e^{\nu[(s_1+s_2)\log(R_{iA00})+s_2\log(\psi_1)]}) + p_{14}(d_1) E(e^{\nu(s_1\log(R_{iA00})+s_2\log(\pi_{iA}\psi_1))}) \\ & + p_{22}(d_1) E(e^{\nu(s_1+s_2)\log(\pi_{iA})}) + p_{24}(d_1) E(e^{\nu[(s_1+s_2)\pi_{iA}+s_2\log(\psi_1)]}) \\ & + p_{33}(d_1) E(e^{\nu(s_1+s_2)(\log(R_{iA00})+\log(\psi_1))}) + p_{34}(d_1) E(e^{\nu[s_1(\log(R_{iA00})+\log(\psi_1))+s_2(\log(\pi_{iA})+\log(\psi_1))]}) \\ & + p_{44}(d_1) E(e^{\nu(s_1+s_2)(\log(\pi_{iA})+\log(\psi_1))}) \end{aligned}$$

We have shown that the transition parameters are identified from which we can identify  $p_{j\ell}(d_1)$ . We can vary  $d_1$  continuously. Given that the form of  $p_{j\ell}(d_1)$  is nonlinear that we can solve for each of the expected values so by varying it we can get a system of linear equations and invert it to obtain each of the expected values in the equation above. From this it is clear that  $\psi_1$  is identified as well as the characteristic function (and thus joint distribution) of  $\log(R_{iA00})$  and  $\pi_{iA}$ .

Using similar logic and 8 periods of wage data we can identify the joint distribution of

$$R_{iA00}, \pi_{iA}, R_{iBA0}, R_{iB00}, \pi_{iB}, R_{iA01}, R_{iBA1}, R_{iB01}.$$

Given the number of different combinations of labor market statuses we do not write this out completely but the form will be similar to the example above.

We can use the exact same strategy to identify

$$(R_{iA00}, \pi_{iA}, R_{iAB0}, R_{iB00}, \pi_{iB}, R_{iA01}, R_{iAB1}, R_{iB01})$$

for the AB group.

Next consider the A0 type. This is more complicated as we can not directly distinguish an A0 type from an AB or BA type that never worked for a type B firm during our sample

period. For the A0 type, there are only three relevant wages  $R_{iA00}$ ,  $R_{iA01}$ , and  $\pi_{iA}$ . To see how to identify this joint distribution consider workers who begin their first job at A at  $1 - \delta_1$ , end that job at  $1 + \delta_2$ , begin a new job at A at  $2 - \delta_3$  and are still working at that job at time 3. Send  $\delta_1 \downarrow 0$  and  $\delta_3 \downarrow 0$  as above. We know that at the first period the worker has no human capital and no outside offers. At period 2 they have no outside offer, but may have human capital. At period three they can be in any of the four states listed above. Being someone loose with notation let  $P_i$  be a labor market profile for the three periods and let  $p(\delta_2)$  be the particular profile we are conditioning on (i.e. being unemployed until  $1 - \delta_1$ , working at A from  $1 - \delta_1$  until  $1 + \delta_2$ , being unemployed until  $1 - \delta_3$ , then working at job A through period 3). Then we can identify

$$\begin{aligned} & \phi_{w_{i1}^m, w_{i2}^m, w_{i3}^m}(s_1, s_2, s_3; \delta_2) \\ = & Pr(\text{type} = A0 \mid P_i = p(\delta_2)) E \left( e^{\lambda(s_1 w_{i1}^m + s_2 w_{i2}^m + s_3 w_{i3}^m)} \mid \text{type} = A0, P_i = p(\delta_2) \right) \\ & + Pr(\text{type} = AB \mid P_i = p(\delta_2)) E \left( e^{\lambda(s_1 w_{i1}^m + s_2 w_{i2}^m + s_3 w_{i3}^m)} \mid \text{type} = AB, P_i = p(\delta_2) \right) \\ & + Pr(\text{type} = BA \mid P_i = p(\delta_2)) E \left( e^{\lambda(s_1 w_{i1}^m + s_2 w_{i2}^m + s_3 w_{i3}^m)} \mid \text{type} = BA, P_i = p(\delta_2) \right). \end{aligned}$$

We have previously shown that everything in the above expression other than the conditional characteristic function  $E \left( e^{\lambda(s_1 w_{i1}^m + s_2 w_{i2}^m + s_3 w_{i3}^m)} \mid \text{type} = A0, P_i = p \right)$  is identified, thus it is identified from this expression.

We now continue to condition on the case above with  $\delta_1 \approx \delta_3 \approx 0$  and let  $p_{r_2 r_3}(\delta_1)$  represent the probability that the worker has labor market status  $r_2$  in the second period and status  $r_3$  in period 3 (statuses are listed above). From this we can identify (where we condition on  $\text{type}=A0$  but do not make it explicit)

$$\begin{aligned} & \frac{E \left( e^{\lambda(s_1 w_{i1}^m + s_2 w_{i2}^m + s_3 w_{i3}^m)} \mid \text{type} = A0, P_i = p \right)}{\phi_\varepsilon(s_1) \phi_\varepsilon(s_2) \phi_\varepsilon(s_3)} \\ = & p_{11}(\delta_2) E \left( e^{\lambda(s_1 + s_2 + s_3) \log(R_{iA00})} \right) + p_{12}(\delta_2) E \left( e^{\lambda((s_1 + s_2) \log(R_{iA00}) + s_3 \log(\pi_{iA}))} \right) \\ & + p_{13}(\delta_2) E \left( e^{\lambda[(s_1 + s_2 + s_3) \log(R_{iA00}) + s_3 \log(\psi_1)]} \right) + p_{14}(\delta_2) E \left( e^{\lambda((s_1 + s_2) \log(R_{iA00}) + s_3 (\log(\pi_{iA}) + \log(\psi_1)))} \right) \\ & + p_{33}(\delta_2) E \left( e^{\lambda(s_1 \log(R_{iA00}) + (s_2 + s_3) \log(R_{iA01}))} \right) + p_{34}(\delta_2) E \left( e^{\lambda(s_1 \log(R_{iA00}) + s_2 \log(R_{iA01}) + s_3 \log(\pi_{iA} \psi_1))} \right). \end{aligned}$$

Using the same logic as above we can identify all of the expected value terms on the right hand side including  $E \left( e^{\lambda(s_1 \log(R_{iA00}) + s_2 \log(R_{iA01}) + s_3 (\log(\pi_{iA}) + \log(\psi_1)))} \right)$ . From this we know the joint distribution of  $(R_{iA00}, R_{iA01}, \pi_{iA})$  for the A0 types.

An analogous argument shows identification of  $(R_{iB00}, R_{iB01}, \pi_{iB})$  for the B0 types.

We have shown that we can identify wages and revealed preference about job choices. The next question is whether we can identify the individual utility functions and in particular the bargaining parameter  $\beta$ . The answer turns out to be no. To see why, suppose that we could observe all of the wages for a particular BA worker (which is similar to observing the distribution of wages after filtering out the measurement error). We could observe the 8 wages  $(W_{iA00}, \pi_{iA}, W_{iBA0}, W_{iB00}, \pi_{iB}, W_{iA01}, W_{iBA1}, W_{iB01})$ . Assume even further that the utility function takes the form

$$U_{ij}(W) = \log(w) + v_{ij}.$$

Even in this restrictive case,  $\beta$  is not identified. Writing down the wage equations for the 6 wages we can get for human capital level 1:

$$\begin{aligned} (\rho + \delta) V_{iB1}(\pi_{iB}) &= \log(\pi_{iB}) + \log(\psi_1) + v_{iB} + \delta V_{i01} \\ (\rho + \delta + \lambda_B^e) V_{iA1}(\pi_{iA}) &= \log(\pi_{iA}) + \log(\psi_1) + v_{iA} \\ &\quad + \lambda_B^e [\beta V_{iB1}(\pi_{iB}) + (1 - \beta) V_{iA1}(\pi_{iA})] + \delta V_{i01} \\ (\rho + \lambda_A^n + \lambda_B^n) V_{i01} &= u_{i01} + \lambda_A^n [\beta V_{iA1}(\pi_{iA}) + (1 - \beta) V_{i01}] \\ &\quad + \lambda_B^n [\beta V_{iB1}(\pi_{iB}) + (1 - \beta) V_{i01}] \\ (\rho + \delta + \lambda_B^e + \lambda_A^e) [\beta V_{iA1}(\pi_{iA}) + (1 - \beta) V_{i01}] \\ &= \log(w_{iA01}) + v_{iA} + \lambda_A^e V_{iA1}(\pi_{iA}) \\ &\quad + \lambda_B^e [\beta V_{iB1}(\pi_{iB}) + (1 - \beta) V_{iA1}(\pi_{iA})] + \delta V_{i01} \\ (\rho + \delta + \lambda_B^e + \lambda_A^e) [\beta V_{iB1}(\pi_{iB}) + (1 - \beta) V_{i01}] \\ &= \log(w_{iB01}) + v_{iB} + \lambda_A^e [\beta V_{iB1}(\pi_{iB}) + (1 - \beta) V_{iA1}(\pi_{iA})] \\ &\quad + \lambda_B^e V_{iB1}(\pi_{iB}) + \delta V_{i01} \\ (\rho + \delta + \lambda_B^e) [\beta V_{iB1}(\pi_{iB}) + (1 - \beta) V_{iA1}(\pi_{iA})] \\ &= \log(w_{iBA1}) + v_{iB} + \lambda_B^e V_{iB1}(\pi_{iB}) + \delta V_{i01} \end{aligned}$$

For human capital level 0

$$(\rho + \delta + \lambda_h) V_{iB0}(\pi_{iB}) = \log(\pi_{iB}) + v_{iB} + \delta V_{i0} + \lambda_h V_{iB1}(\pi_{iB})$$

$$\begin{aligned} (\rho + \delta + \lambda_B^e + \lambda_h) V_{iA0}(\pi_{iA}) &= \log(\pi_{iA}) + v_{iA} + \lambda_B^e [\beta V_{iB0}(\pi_{iB}) + (1 - \beta) V_{iA0}(\pi_{iA})] \\ &\quad + \delta V_{i0} + \lambda_h V_{iA1}(\pi_{iA}) \end{aligned}$$

$$\begin{aligned} (\rho + \lambda_A^n + \lambda_B^n) V_{i00} &= u_{i00} + \lambda_A^n [\beta V_{iA0}(\pi_{iA}) + (1 - \beta) V_{i00}] \\ &\quad + \lambda_B^n [\beta V_{iB0}(\pi_{iB}) + (1 - \beta) V_{i00}] \end{aligned}$$

$$\begin{aligned} (\rho + \delta + \lambda_B^e + \lambda_A^e + \lambda_h) [\beta V_{iA0}(\pi_{iA}) + (1 - \beta) V_{i00}] & \\ = \log(w_{iA00}) + v_{iA} + \lambda_A^e V_{iA0}(\pi_{iA}) & \\ + \lambda_B^e [\beta V_{iB0}(\pi_{iB}) + (1 - \beta) V_{iA0}(\pi_{iA})] + \delta V_{i00} + \lambda_h V_{iA1}(w_{iA00}) & \\ + \lambda_h \frac{\log(w_{iA00}) + \log(\psi_1) + v_{iA} + \delta V_{i01}}{\rho + \delta + \lambda_B^e + \lambda_A^e} & \\ + \lambda_h \frac{\lambda_A^e V_{iA1}(\pi_{iA}) + \lambda_B^e [\beta V_{iB1}(\pi_{iB}) + (1 - \beta) V_{iA0}(\pi_{iA})]}{\rho + \delta + \lambda_B^e + \lambda_A^e} & \\ (\rho + \delta + \lambda_B^e + \lambda_A^e + \lambda_h) [\beta V_{iB0}(\pi_{iB}) + (1 - \beta) V_{i00}] & \\ = \log(w_{iB00}) + v_{iB} + \lambda_A^e [\beta V_{iB0}(\pi_{iB}) + (1 - \beta) V_{iA0}(\pi_{iA})] & \\ + \lambda_B^e V_{iB0}(\pi_{iB}) + \delta V_{i00} + \lambda_h V_{iB1}(w_{iB00}) & \\ + \lambda_h \frac{\log(w_{iB00}) + \log(\psi_1) + v_{iB} + \delta V_{i01}}{\rho + \delta + \lambda_B^e + \lambda_A^e} & \\ + \lambda_h \frac{\lambda_A^e [\beta V_{iB0}(\pi_{iB}) + (1 - \beta) V_{iA0}(\pi_{iA})] + \lambda_B^e V_{iB0}(\pi_{iB})}{\rho + \delta + \lambda_B^e + \lambda_A^e} & \\ (\rho + \delta + \lambda_B^e + \lambda_h) [\beta V_{iB0}(\pi_{iB}) + (1 - \beta) V_{iA0}(\pi_{iA})] & \\ = \log(w_{iBA0}) + v_{iB} + \lambda_B^e V_{iB0}(\pi_{iB}) + \delta V_{i00} & \\ \lambda_h \frac{\log(w_{iBA0}) + \log(\psi_1) + v_{iB} + \lambda_B^e V_{iB1}(\pi_{iB}) + \delta V_{i01}}{\rho + \delta + \lambda_B^e} & \end{aligned}$$

This is 12 equations in 11 unknowns

$$V_{i00}, V_{i01}, V_{iB0}(\pi_{iB}), V_{iB1}(\pi_{iB}), V_{iA0}(\pi_{iA}), V_{iA1}(\pi_{iA}), u_{i0}, u_{i1}, v_{iA}, v_{iB},$$

and  $\beta$ . Beyond this we can clearly get a normalization on flow utility, say normalizing  $u_{i00}$  to zero, so this is 12 equations in 10 unknowns. However, three of the equations are linearly dependent so we actually have only 9 separate equations and 10 unknowns. Specifically, it

is straightforward to show that

$$\begin{aligned}\log(w_{iBA1}) - (\log(w_{iB01}) - \log(w_{iA01})) &= \log(\pi_{iA}) + \log(\psi_1) \\ \log(w_{iBA0}) - (\log(w_{iB00}) - \log(w_{iA00})) &= \log(\pi_{iA}) \\ \log(w_{iBA0}) &= \log(w_{iBA1}) - \log(\psi_1)\end{aligned}$$

Thus, the model is not identified and in particular as long as  $\beta < 1$  we could choose it to any value and solve all equations. However  $\beta = 1$  can be tested as it would imply, for example,  $w_{iA00} = \pi_{iA}$ .

## 5 Econometric Specification/Parameterization

We assume that there are a large number of people in the economy but a finite number of  $J$  employer types. Our econometric specification differs from the identification section above in that we do not observe establishment types. We will assume that multiple establishments in the data will be of the same type but we will not know exactly which those are. The key aspect from the data is that workers who work for the same establishment also work for the same establishment type. We will let  $j_i(t)$  denote the job (establishment) held by worker  $i$  at time  $t$ .

The transition parameters  $\delta$ ,  $\lambda^e$ , and  $\lambda^n$  take the same form as in the earlier sections though they do not vary across jobs.

The maximum offered rental rate is specified as

$$\log(\pi_{ij}) = \theta_i + \mu_j^w + v_{ij}^w.$$

Again the easiest way to think about this is as the marginal productivity of a worker and that firms have constant returns to scale and no capacity constraints. This implies that hiring this worker does not restrict hiring of any other workers, nor does it affect their marginal productivities. We only get to observe wages with normal measurement error  $\xi_{it}$ , with mean zero and variance  $\sigma_\xi^2$ .

The tastes for jobs are

$$u_{ij}(W) = \alpha \log(W) + \mu_j^u + v_{ij}^u.$$

It is important that we allow  $\mu_j^w$  to be correlated with  $\mu_j^u$  because of equilibrium considera-

tions. Note that all that matters for job to job turnover is the reduced form

$$\begin{aligned} u_{ij}(\pi_{ij}\psi_t) &= \alpha (\theta_i + \mu_j^w + v_{ij}^w) + \alpha \log(h_t) + \mu_j^u + v_{ij}^u \\ &= \alpha \theta_i + \alpha \log(h_t) + (\alpha \mu_j^w + \mu_j^u) + (\alpha v_{ij}^w + v_{ij}^u). \end{aligned}$$

We assume that the joint distribution of  $(u_j^w, u_j^u)$ , the joint distribution of  $(v_{ij}^w, v_{ij}^u)$ , the distribution of  $\theta_i$ , and the distribution of  $\xi_{it}$  are all independent of each other. Thus,  $(v_{ij}^w$  and  $v_{ij}^u)$  can be interpreted as match-specific variation around the mean. Furthermore we assume that  $\xi_{it}$  is i.i.d. across time. As with standard discrete choice models we need a location normalization, so we will choose to set  $var(v_{ij}^u) = 1$ . Moreover, we assume that  $(v_{ij}^u, v_{ij}^w)$  are jointly normal which gives us two parameters,  $cov(v_{ij}^u, v_{ij}^w)$  and  $var(v_{ij}^w)$ . However, as should be clear from the reduced form equation above we can not separately identify  $cov(v_{ij}^u, v_{ij}^w)$  from  $\alpha$ , so we normalize  $cov(v_{ij}^u, v_{ij}^w) = 0$  and estimate  $\alpha$  along with  $\sigma_{vw}$ , the standard deviation of  $v_{ij}^w$ .<sup>11</sup> This assumes that the reason why workers tend to prefer to work at a high productive matches is because they receive high wages at those jobs whether than because they tend to like the non-wage characteristics of jobs on which they are particularly productive. This seems like a very reasonable assumption to us. Notice, that the covariance of  $(u_j^w, u_j^u)$  is left unrestricted, so the common part of productivity and non-pecuniary returns are allowed to be correlated.

We also assume that  $\theta_i$  is normal with mean  $E_\theta$  and variance  $\sigma_\theta^e$ .

We take a very simple specification for the value of nonemployment by assuming

$$u_{i0t} = \alpha E(\theta_i).$$

We tried to choose a relatively parsimonious way to approximate the distribution of  $(\mu_j^u, \mu_j^w)$  which is a discrete distribution. With no obvious parametric alternative we decided upon the following one:

$$\begin{aligned} \mu_j^u &= f_1 [U_1(j) + f_3 U_2(j)] \\ \mu_j^w &= f_2 [f_3 U_1(j) + U_2(j)] \end{aligned}$$

where  $U_1(j)$  and  $U_2(j)$  are uniformly distributed across  $[-1,1]$ . In our specification we allow each of  $U_1$  and  $U_2$  to take ten different values and assume these are unrelated to each-other

---

<sup>11</sup>This is a normalization to the extent that it is innocuous in terms of the model we estimated since choices across jobs depends only upon the overall  $u_{ij}(\pi_{ij}h_t)$  not on its components. However, it is not a normalization in that in counterfactuals in which  $\pi_{ij}$  changes, the preference across jobs depends on this assumption.

giving us one hundred different firm types. Essentially  $f_1$  governs the variance of  $\mu_j^u$ ,  $f_2$  governs the covariance of  $\mu_j^2$ , and  $f_3$  governs their correlation.

We let  $i$  denote workers and assume that for each worker we get to observe them from when they enter the labor market until time period  $T_i$ . We also observe  $M_i$  different wages for workers at times  $t = t_1, \dots, t_{M_i}$ .

To keep the idea of the data simple assume that every moment  $t \in [0, T_i]$  we observe the firm for which the worker worked (call this  $j_i(t)$ ) and without loss of generality let  $j_i(t) = 0$  denote non-employment.

We fix  $\rho = 0.05$ . This leaves a total of 14 parameters:

$$[\delta, \lambda^n, \lambda^e, E_\theta, \sigma_\theta, \sigma_\xi, \sigma_{vw}, \alpha, f_1, f_2, f_3, b_1, b_2, \beta].$$

## 6 Data

We use Danish register data of two types. The first is weekly spell data that covers all individuals aged 15-70 in Denmark from 1985 to 2003. The data is constructed using various sources of register data, see Bunzel (2010) and Bobbio (2010) for a longer description. The data generated from these sources consists of a worker identifier, a firm and establishment identifier, start and end date of the spell, and a state variable. The states are employed, unemployed, self-employed, retirement, and non-participation. Non-participation is a residual state in the sense that it means that we do not observe the worker in any of the available registers. The second type of data is annual cross-section data from the Danish register-based matched employer-employee data set IDA.<sup>12</sup> IDA contains socioeconomic information on workers and background information on employers, and covers the entire Danish population age 15 to 70. IDA contains the annual average hourly wage for the job occupied in the last week of November.

We choose to identify employers at the establishment level. Thus, the unit of observation is a worker, year, state, establishment. We differ on this point from most of the empirical search literature which uses firms as the employer unit. However, using establishments have three advantages in the current setting. The employer identifier is not well defined over time since firms might change the legal unit (and hence firm identifier) without changing anything else. The establishment identifier is consistent over time.<sup>13</sup> Secondly, when thinking about

<sup>12</sup>Integreteret Database for Arbejdsmarkedsforskning (Integrated Database for Labor Market Research) is constructed and maintained by Statistics Denmark.

<sup>13</sup>The establishment is constructed by Statistics Denmark and is the same across years if one of three

compensating differentials the most appropriate unit seems to us to be the establishment and not the legal firm. The third advantage is that we are able to break up larger firms into separate establishments. Treating all workers within the same large firm the same seems to be inappropriate. This is especially important in the government sector, where firms tend to be large and potentially cover many different types of establishments. While this is still not a completely satisfactory way of dealing with the government sector, we think it is much better than treating the government sector as a single firm or dropping it from the data.

## 6.1 Sample Selection Criteria

We use the following sample selection criteria<sup>14</sup>

- We censor workers after age 55 and disregard spells before labor market entry (or age 19) defined as the time of highest completed education (and not observed in education later).
- We disregard workers with errors in educational information.
- Temporary non-employment spells shorter than 13 weeks are deleted and non-employment shorter than 3 weeks are allocated to the last of the two employment spells.
- We censor workers when they enter self-employment or retirement.
- We delete workers that have gaps in their spell histories. This could arise if the worker for some reason have missing IDA data in a given year.

In Table 1 the effects of the data steps taken above are described for each step.

Since the model is cast in steady state and cross-sectional wages therefore have no trend we detrended wages in logs by gender-educational groups conditional on experience. We do this since the composition of workers changes over the sample period. Therefore it is important not to impose e.g. average wages to be constant over all years.

### Voluntary Transitions

Since job-to-job transitions play a vital role for the identification of our model we will ignore transitions from two types of establishments. The first are transitions for workers

---

criteria is met: same owner and industry; same owner and workforce; same workforce and either same address or same industry.

<sup>14</sup>See Appendix for a more thorough description.

to or from establishments with missing ID (0.5 percent, cf. Table 2). We will also ignore job-to-job transitions from closing establishments or establishments with mass layoffs.<sup>15</sup>

## 6.2 Descriptive Statistics

In this section we present different descriptive statistics for the sample used in this study. The number of years and the number of establishments are important for the identification of the model. Table 2 show statistics for these measures using repeated cross-sections.

The worker is on average 11 years in the sample and are employed in 2.7 different establishments. There are almost as many women as men in the sample. This is because we are not censoring or deleting public employees of which many are women. The workers have on average twelve years of education. However, this changes some over the sample period, since entering workers are better educated than those leaving the sample. The average cross-section age is 38 with the earliest labor market entry at age 19 and the highest age in the sample being age 55. A total of 84 percent are employed in general, while 32 percent are employed in the public sector. Finally, in a given cross-section we miss the establishment identifier for 0.5 percent of all employment observations. The average labor market experience is 13 years.

Figure 1 displays the estimate from a Kaplan Meier estimator of the survival probability for employed, unemployed, and non-employed. The unemployment spells have a much shorter duration than both employment and non-employment in general. Notice, that the survival rate for non-employment does not seem to approach zero. This could be due to the fact that the sample also include workers that are actually not in the labor force, i.e. not actively looking for a job. However, we do not view this as a problem, since in the model we allow for workers who simply choose not to take a job. Turning to the employment spells we can see that there is a slower decline in the survival probability. The probability of working in the same establishment after the initial two years is 60 percent. The dips in both the survival rates for employment and non-employment comes from a “New Year” effect. In the reported data there is an over representation of state changes at January 1st each year. We suspect that some the state changes are coming from transitions the past year that have not been reported correctly.

---

<sup>15</sup>A mass layoff is defined as the establishment having more than 15 workers and the next year only has 30 percent or less left.

## 7 Auxiliary Model

We estimate our model using Indirect Inference (Gourieroux, Monfort, and Renault 1993). Our approach is to use the argument in the identification section as a guide to which aspects of the data we should be using to identify the different parameters. To keep the relationship between the parameters and the data as transparent as possible we focus on the exactly identified case. In particular, for each parameter we choose one auxiliary parameter that we think is useful for identifying it. While we use this language, it is not precisely how the estimation works. In practice, all the auxiliary parameters are useful for identifying all of the structural parameters. However, we find that this approach to be highly beneficial for us in understanding the mapping between the parameters and the data.

### 7.1 Notation

Let

- $i = 1, \dots, N$  index individuals
- $\ell = 1, \dots, L_i$  index employment spells—a spell of consistent employment with no non-employment in between
- $j = 1, \dots, J_{i\ell}$  index a job spell that occurs within employment spell  $\ell$  for individual  $i$
- $t = 1, \dots, T_{i\ell j}$  index the set of wage observations on job spell  $i\ell j$ .
- $f_{i\ell j}$  the establishment associated with this job spell
- $1, \dots, Q$  be the number of establishments
- $D_{i\ell j}$  the duration of time that the worker worked on job spell  $i\ell j$
- $w_{i\ell jt}^m$  the  $t^{\text{th}}$  wage observation at job  $i\ell j$
- $E_{i\ell jt}^m$  the  $t^{\text{th}}$  experience observation at job  $i\ell j$
- $T_{i\ell jt}^m$  the  $t^{\text{th}}$  tenure observation at job  $i\ell j$
- $\ell^m = 1, \dots, L_i^m$  index employment spells that contains a November cross-section
- $w_{i\ell 11}^m$  be the first wage observation at employment spell  $\ell^m$  for individual  $i$

- $E_{i\ell_{11}}^m$  is the corresponding first experience observation at employment spell  $\ell^m$  for individual  $i$
- $k = 1, \dots, K_i$  the number of nonemployment spells for individual  $i$
- $D_{ik}^n$  the duration of nonemployment spell  $ik$
- $k^n = 1, \dots, K_i^u$  the number of nonemployment spells for individual  $i$  where experience at the first November cross-section following the end of nonemployment spell is available.
- $E_{ik}^n$  is experience at the first November cross-section following the end of nonemployment spell  $ik$

## 7.2 Definition of Transition Variables

Define

$$S_{ilj} \equiv \begin{cases} 1 & \text{if spell } ilj \text{ starts with a JJ transition and do not end with a JJ transition} \\ -1 & \text{if spell } ilj \text{ ends with a JJ transition and did not start with a JJ transition} \\ 0 & \text{otherwise} \end{cases}$$

where JJ transition means a job to job transition. Notice, if the spell is left censored then for the purpose of defining  $S_{ilj}$  we assume that it starts from non-employment. Likewise, if the job is right censored we assume that it ends in a firing. I.e. the sum of  $S_{ilj}$  across jobs for an individual is therefore always 0, but not when summing over firms.

For consistency reasons we will sometimes take the individual out of the calculation, so define:

$$S_{-ilj} = \frac{\sum_{i^*=1}^N \sum_{\ell^*=1}^{L_{i^*}} \sum_{j^*=1}^{J_{i^*\ell^*j^*}} S_{i^*\ell^*j^*} 1[i^* \neq i, f_{i^*\ell^*j^*} = f_{ilj}]}{\sum_{i^*=1}^N \sum_{\ell^*=1}^{L_{i^*}} \sum_{j^*=1}^{J_{i^*\ell^*j^*}} 1[i^* \neq i, f_{i^*\ell^*j^*} = f_{ilj}]}$$

where  $1[\cdot]$  is the indicator function. In other words  $S_{-ilj}$  is just the average value of  $S_{i^*\ell^*j^*}$  for people who work at establishment  $f_{ilj}$  excluding individual  $i$ .

Define the number of JJ separations,  $s_{-i}^q$ , and JJ hires,  $h_{-i}^q$ , for each firm where individual  $i$  do not contribute. Again we take transitions from closing establishments out.

$$s_{-i}^q = \sum_{i^*=1}^N \sum_{\ell=1}^L \sum_{j=1}^{J_{i\ell j}-1} 1[q = f_{i^*\ell j}, i \neq i^*]$$

$$h_{-i}^q = \sum_{i^*=1}^N \sum_{\ell=1}^L \sum_{j=2}^{J_{i\ell j}} 1[q = f_{i^*\ell j}, i \neq i^*]$$

Also define

$$h_{-ilj} \equiv \frac{h_{-i}^{f_{ilj}}}{h_{-i}^{f_{ilj}} + s_{-i}^{f_{ilj}}}$$

$$\tilde{h}_{-ilj} \equiv \frac{h_{-i}^{f_{ilj}}}{h_{-i}^{f_{ilj}} + s_{-i}^{f_{ilj}}} - \frac{1}{\sum_{\ell^*=1}^{L_i} J_{i\ell^*}} \sum_{\ell^*=1}^{L_i} \sum_{k=1}^{J_{i\ell^*}} \frac{h_{-i}^{f_{i\ell^*k}}}{h_{-i}^{f_{i\ell^*k}} + s_{-i}^{f_{i\ell^*k}}}.$$

### 7.3 Definition of Wage Variables

The worker effect is just mean worker wage over his working life

$$\bar{w}_i = \frac{\sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} \sum_{t=1}^{T_{i\ell j}} w_{i\ell j t}^m}{\sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} T_{i\ell j}}$$

When  $T_{i\ell j} > 0$ , define

$$\overline{w_{i\ell j}} = \frac{1}{T_{i\ell j}} \sum_{t=1}^{T_{i\ell j}} w_{i\ell j t}^m,$$

and define it to be zero otherwise. The distinction between  $\bar{w}_i$  and  $\overline{w_{i\ell j}}$  is intentional - we will use them at different points. Then when  $T_{i\ell j} > 0$  we define

$$\tilde{w}_{i\ell j} \equiv \overline{w_{i\ell j}} - \frac{\sum_{\ell^*=1}^{L_i} \sum_{j^*=1}^{J_{i\ell^*}} \overline{w_{i\ell^*j^*}}}{\sum_{\ell^*=1}^{L_i} \sum_{j^*=1}^{J_{i\ell^*}} 1 [T_{i\ell^*j^*} > 0]}$$

This now has the nice feature that we think of as standard - it will sum to zero across jobs for each individual.

Analogous to  $\tilde{S}_{-ilj}$  define:

$$\tilde{w}_{-ilj} \equiv \frac{\sum_{i^*=1}^N \sum_{\ell^*=1}^{L_{i^*}} \sum_{j^*=1}^{J_{i^*\ell^*j^*}} \tilde{w}_{i^*\ell^*j^*} 1 [i^* \neq i, f_{i^*\ell^*j^*} = f_{ilj}]}{\sum_{i^*=1}^N \sum_{\ell^*=1}^{L_{i^*}} \sum_{j^*=1}^{J_{i^*\ell^*j^*}} 1 [i^* \neq i, f_{i^*\ell^*j^*} = f_{ilj}]}$$

### 7.4 Definition of Experience Variables

When  $T_{i\ell j} > 0$ , define

$$\overline{E_{i\ell j}} = \frac{1}{T_{i\ell j}} \sum_{t=1}^{T_{i\ell j}} E_{i\ell j t}^m,$$

## 7.5 Auxiliary Model

There are 4 parameters that we “identify” using turnover data:  $\delta, \lambda^e, \lambda^n, f_1$

Here are a set of moments we could use to identify them:

- $\delta$  : Average length of employment spell:

$$\frac{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \left( \sum_{j=1}^{J_{i\ell}} D_{i\ell j} \right)}{\sum_{i=1}^N L_i}$$

- $\lambda^n$  : Average length of nonemployment spell.

$$\frac{\sum_{i=1}^N \sum_{k=1}^{K_i} D_{ik}^n}{\sum_{i=1}^N K_i}$$

- $\lambda^e$  : Average length of a job spell:

$$\frac{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} D_{i\ell j}}{\sum_{i=1}^N \sum_{\ell=1}^{L_i} J_{i\ell}}$$

- $f_1$  :

$$\frac{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} \tilde{h}_{-i\ell j} S_{i\ell j}}{\sum_{i=1}^N \sum_{\ell=1}^{L_i} J_{i\ell}}$$

If  $\tilde{h}_{-i\ell j}$  can be defined for all firm-worker combinations this should be the same as the covariance of  $S_{i\ell j}$  and  $\tilde{h}_{-i\ell j}$ . Notice, that for some firm-worker combinations  $\tilde{h}_{-i\ell j}$  cannot be defined in the data. In the data we use the covariance between  $\tilde{h}_{-i\ell j}$  and  $S_{i\ell j}$ .

For wages we have 10 parameters:  $E_\theta, \sigma_\theta, \sigma_\xi, \sigma_{v^w}, \alpha, f_2, f_3, b_1, b_2, \beta$

We use the moments:

- $E_\theta$  : Just use

$$\bar{w} \equiv \frac{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} \sum_{t=1}^{T_{i\ell j}} w_{i\ell j t}^m}{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} T_{i\ell j}}$$

- $\sigma_\theta^2, \sigma_\xi^2, \sigma_{vw}$  : We use the decomposition

$$\frac{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} \sum_{t=1}^{T_{i\ell j}} (w_{it}^m - \bar{w})^2}{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} T_{i\ell j}} = \frac{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} \sum_{t=1}^{T_{i\ell j}} (w_{i\ell j t}^m - \bar{w}_{i\ell j})^2}{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} T_{i\ell j}} + \frac{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} \sum_{t=1}^{T_{i\ell j}} (\bar{w}_{i\ell j} - \bar{w}_i)^2}{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} T_{i\ell j}} + \frac{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} \sum_{t=1}^{T_{i\ell j}} (\bar{w}_i - \bar{w})^2}{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} T_{i\ell j}}$$

That is we use each of the three expressions on the right hand side

- $f_2 : \text{Cov}(\tilde{w}_{i\ell j}, \tilde{w}_{-i\ell j})$ .

$$\frac{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} \tilde{w}_{i\ell j} \tilde{w}_{-i\ell j}}{\sum_{i=1}^N \sum_{\ell=1}^{L_i} J_{i\ell}}$$

If  $\tilde{w}_{-i\ell j}$  can be defined for all firm-worker combinations this should be the same as the covariance of  $\tilde{w}_{i\ell j}$  and  $\tilde{w}_{-i\ell j}$ . This is the case in the simulations. However, notice that for some firm-workers combinations  $\tilde{w}_{-i\ell j}$  cannot be defined in the data. In the data we use the covariance between  $\tilde{w}_{i\ell j}$  and  $\tilde{w}_{-i\ell j}$ .

- $f_3 :$

$$\frac{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} \tilde{w}_{i\ell j} \tilde{h}_{-i\ell j}}{\sum_{i=1}^N \sum_{\ell=1}^{L_i} J_{i\ell}}$$

If  $\tilde{h}_{-i\ell j}$  can be defined for all firm-worker combinations this should be the same as the covariance of  $\tilde{w}_{i\ell j}$  and  $\tilde{h}_{-i\ell j}$ . This is the case in the simulations. However, notice that for some firm-workers combinations  $\tilde{h}_{-i\ell j}$  cannot be defined in the data. In the data we use the covariance between  $\tilde{w}_{i\ell j}$  and  $\tilde{h}_{-i\ell j}$ .

- $\alpha :$

$$\text{Pr}(\overline{w}_{i\ell j+1} < \overline{w}_{i\ell j}) = \frac{\sum_{i=1}^N \sum_{\ell=1}^{L_i} \sum_{j=2}^{J_{i\ell}} 1[\overline{w}_{i\ell j} < \overline{w}_{i\ell j-1}]}{\sum_{i=1}^N \sum_{\ell=1}^{L_i} (J_{i\ell} - 1)}$$

- $\beta, b_1, b_2$  : For these we use moments inspired by Altonji and Shakotko (1987). Define tenure at a job,  $TE_{iljt}$  as

$$TE_{iljt} = E_{iljt} - E_{ilj1}$$

i.e. tenure is set to 0 at the first November cross-section (also for left-censored tenure). Now define  $dTE_{iljt}$  as

$$dTE_{iljt} = TE_{iljt} - \frac{\sum_{t=1}^{T_{ilj}} 1[E_{iljt} < 20]TE_{iljt}}{\sum_{t=1}^{T_{ilj}} 1[E_{iljt} < 20]}$$

Estimate by OLS the following regression for all observations with  $E_{iljt} < 20$ :

$$TE_{iljt} = \tilde{\alpha}_0 + \tilde{\alpha}_1 E_{iljt} + \tilde{\alpha}_2 1[E_{iljt} > 7](E_{iljt} - 7) + \tilde{\alpha}_3 dTE_{iljt}$$

Let the predicted values be given by  $\widehat{TE_{iljt}}$ . Now estimate by OLS the following regression for all observations with  $E_{iljt} < 20$ :

$$w_{iljt}^m = \tilde{b}_0 + \tilde{b}_1 E_{iljt} + \tilde{b}_2 1[E_{iljt} > 7](E_{iljt} - 7) + \tilde{b}_3 \widehat{TE_{iljt}}$$

We use the estimates  $\tilde{b}_1$ ,  $\tilde{b}_2$ , and  $\tilde{b}_3$  as auxiliary parameters.

## 8 Results

We estimate the model using indirect inference with the auxiliary model described above. Our objective function is the sum of the squared deviation between the simulated model and the data weighted by the inverse of the variance of the estimated parameter. In this version of the paper we do not use the full variance/covariance matrix because it allows us to more easily see where the model is missing.

The results of this procedure for Model 1 (the base model) are presented in Tables 3 and 4. Table 3 shows the model fit. One can see that the fit of the model is excellent. The structural parameters of the model are presented in Table 4. However, we do not view these as particularly interesting in their own right. We will focus on the decomposition of the amount of total wage variance into its components, but we make some brief comments first. This paper is primarily about wages rather than transitions into and out of unemployment. For that reason we have not tried to get our parameters to match exactly what is done in the

literature. For example, we do not make a distinction between workers who are unemployed and workers who are out of the labor force. Moreover, since workers do not accept all offers one can not interpret  $\lambda^n$  and  $\lambda^e$  directly because it is not separately identified from the probability of getting an acceptable job conditional on a wage offer (see (Flinn and Heckman 1982)).

The decomposition is presented in Table 5, where we sequentially eliminate the different sources of wage inequality. First note that we get rid of measurement error. The total variance of log wages in the model is 0.120, but it falls to 0.098 after we get rid of the measurement error. Recall that given the issues with the non-employment, eliminating compensating differentials makes less sense than the others. Thus, the most reliable simulations are (A) and (B). To do the simulations

- We eliminate human capital accumulation by allowing human capital to accumulate to its highest level immediately.
- We eliminate renegotiation of contracts by setting the workers' bargaining power to one ( $\beta = 1$ ).
- We eliminate search frictions by allowing people to find the most preferred job immediately.
- We eliminate compensating differentials by setting  $v_{ij}^u = \mu_j^u = 0$  for all acceptable jobs (i.e. those that would be taken from nonemployment).
- We eliminate Roy inequality by setting  $\sigma_\theta = \sigma_{v,w} = 0$ .

In all four cases we eliminate human capital first as it is well known to have little explanatory power (i.e. the  $R^2$  in a Mincer model does not change much when experience and experience squared are dropped). We show that this is true here as well as human capital explains about 6% of wage variation. Changing other characteristics and holding  $\beta$  fixed makes little sense so we set it to  $\beta = 1$ . This gets rid of variation in the model that comes from renegotiation by giving all of the bargaining power to the worker. This actually slightly increases wage inequality. We next eliminate Roy inequality, Search inequality, and compensating differentials in different orders. All four simulations indicate that the Roy inequality is the most important in determining wage inequality. However, the relative importance depends upon the order. This is not surprising as these aspects interact in different ways. For example in the base model workers are searching for good firm matches to wages ( $\mu_j^w$ ) and to utility

( $\mu_j^u$ ) as well as good idiosyncratic matches to wages ( $v_{ij}^w$ ) and utility ( $v_{ij}^u$ ). In column (B) since we first eliminate search we are picking up all four of these and this actually leads to an increase in total wage variance. When we eliminate Roy model inequality first in column (A) they are only searching for good firm matches and good utility matches. In column (C) we eliminate compensating differentials first so they are only searching for good wage matches (both firm and idiosyncratic). In this case the importance of search frictions is largest. In the final column we have eliminated both compensating differentials and search frictions so firms are only searching for good  $\mu_j^w$  matches.

One of the motivations of this paper was to answer why firms with different wage policies can coexist in the labor market. The two main explanations in our model was that it was either a result of search frictions or compensating differentials. We estimate that the covariance between the firm specific productivity component ( $\mu_j^w$ ) and the firm specific taste component ( $\mu_j^u$ ) is positive but very small. Thus, worker's weakly prefer to work at high productive firms ignoring the wage benefit. This implies that the reason that less productive firms can continue to stay in the market is due to search frictions and not compensating differentials.

## 9 Conclusions

Not for this draft.

## References

- Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High Wage Workers and High Wage Firms. *Econometrica* 67(2), 251–333.
- Altonji, J. G. and R. A. Shakotko (1987, July). Do wages rise with job seniority? *Review of Economic Studies* 54(3), 437–59.
- Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2011, August). Tenure, experience, human capital and wages: A tractable equilibrium search model of wage dynamics.
- Becker, D. (2009). Non-wage job characteristics and the case of the missing margin. Federal Trade Commission.
- Bobbio, E. (2010). The Danish Matched Employer-Employee Data. *Unpublished*

*Manuscript.*

- Bunzel, H. (2010). The LMDG Data Sets. *Unpublished Manuscript*.
- Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review* 39(2), 257–273.
- Cahuc, C., F. Postel-Vinay, and J.-M. Robin (2006, March). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica* 74(2), 323–364.
- Dey, M. S. and C. J. Flinn (2005, 03). An equilibrium model of health insurance provision and wage determination. *Econometrica* 73(2), 571–627.
- Dey, M. S. and C. J. Flinn (2008, January). Household search and health insurance coverage.
- Eckstein, Z. and G. Van den Berg (2007). Empirical labor search. *Journal of Econometrics* 136(2), 531–564.
- Flinn, C. and J. Heckman (1982). New methods for analyzing structural models of labor force dynamics. *Journal of Econometrics* 18, 115–168.
- Gourieroux, C., A. Monfort, and E. Renault (1993, December). Indirect inference. *Journal of Applied Econometrics* 8, S85–S118.
- Hall, R. E. and A. B. Krueger (2012, September). Evidence on the incidence of wage posting, wage bargaining, and on-the-job search. *American Economic Journal: Macroeconomics* 4(4), 56–67.
- Heckman, J. and G. Sedlacek (1985). Heterogeneity, aggregation, and market wage functions: An empirical model of self-selection in the labor market. *Journal of Political Economy* 93(6), 1077–1125.
- Heckman, J. J. and B. E. Honoré (1990). The empirical content of the Roy model. *Econometrica* 58(5), 1121–1149.
- Heckman, J. J. and C. R. Taber (2008). Chapter Roy model. Palgrave Macmillan.
- Keane, M. P. and K. I. Wolpin (1997). The career decisions of young men. *The Journal of Political Economy* 105(3), 473–522.
- Kotlarski, I. (1967). On characterizing the gamma and the normal distributions. *Pacific Journal of Mathematics* 20(1), 69–76.

- Postel-Vinay, F. and J.-M. Robin (2002, November). Equilibrium Wage Dispersion with Worker and Employer Heterogeneity. *Econometrica* 70(6), 2295–2350.
- Rosen, S. (1987). *Handbook of Labor Economics*, Volume 1, Chapter 12, The Theory of Equalizing Differences,, pp. 641–692.
- Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3(2), 135–146.
- Sullivan, P. (2010). A dynamic analysis of educational attainment, occupational choices, and job search. *International Economic Review* 51(1), 289–317.
- Sullivan, P. and T. Too (2011, September). Search and non-wage job characteristics.
- Villanueva, E. (2007). Estimating compensating wage differentials using voluntary job changes: Evidence from germany. *Industrial and Labor Relations Review* 60(4), 544–561.

## Appendix

**Data Selection** We define labor market entry to be the month of graduation from the highest completed education recorded.<sup>16</sup> We disregard spells that are before this date. If the worker after the date of highest completed education is observed in education the worker is disregarded. E.g. if the highest recorded education for a worker is high school and he graduated in 2001 and we later observe him in education, say in 2003 then we delete him. Workers with changing codes for highest completed education and where age minus education length is less than 5 years are also disregarded. We censor workers after age 55.

Temporary non-employment (unemployment and non-participation) spells shorter than 13 weeks where the previous and next establishment id are the same as one employment spell, i.e. unemployment and non-participation spells are treated as one type of spells. Short unemployment or non-participation spells between two employment spells shorter than 3 weeks are allocated to the last of the two employment spells.

We censor workers when they enter a self-employment state. We delete workers that have gaps in their spell histories. This could arise if the worker for some reason have missing IDA data in a given year. Wages are detrended in logs (but so far not trimmed). We label the states unemployment, retirement and non-participation as non-employment. For some of the employed workers in the final sample we do not observe the establishment ID. However, this is a relatively small fraction, see Table 2. In the calculations of the moments we will take this into account, and only use observations for which we do observe the establishment. E.g. if a transition happens from establishment 1 to an unknown establishment we will not count this as a transition. Likewise, if a transition from an unknown establishment to firm 1 will not be counted as a transition.

In the identification strategy we heavily rely on the fact that observed job to job transitions are actually voluntary. One might suspect that workers in closing establishments might move to a new establishment without actually preferring it compared to the old one. In order to avoid drawing inference from such observations we do not count job to job transitions from an establishment in the year that it closes. E.g. if a workers is employed in establishment 1 in week 1 to 40 in 1995 (and we do not observe the establishment in the data

---

<sup>16</sup>We have information on highest completed education back to 1969, so highest completed education is missing for workers who took it before 1969. Also, immigrants and workers who never finished primary school have missing values. We keep this workers since we suspect that the problems with immigrants and workers who never finished primary school are quit small, and workers who took there education before 1969 have entered the labor market.

after 1995) and in week 41 he is employment in establishment 2 then we do not count that as job to job separation for establishment 1 nor a job to job hire in establishment 2, although we will count them as a separation and a hire. However, if the worker had transitioned into non-employment we would have counted a job destruction. This limits the sample to 2002, since we cannot observe if 2003 was the last year of which an establishment was alive. This gives us the final sample.

**Estimating Labor Market Entry:** We observe graduation times from 1971 and forward. Our sample starts in 1985 which means that we observe some individuals with labor market entry before 1971 (around 1/3 of all workers). We therefore need to approximate the entry year. We use population data (not just those in our sample) graduating in 1971 and 1972 to derive the age distribution at graduation time by gender-education group. This gives us around 70 groups. However a few workers in the sample cannot be matched, so we use a more rough groups for those. We now use the gender-education specific graduation age distribution conditional on the fact that we know the individual did not graduate after 1970. If the minimum age in the estimated distribution implies entry after 1970 we set entry to 1970. This is the case for 1 percent of the workers we approximate.

**Estimating Experience** Experience is observed yearly from 1964 (although with different degrees of precision). We define experience to be experience accumulated from labor market entry. Given that we have approximated entry time we need to approximate experience up to 1970. To do this we again use those entering the labor market in 1971 and 1972. There are several ways to do this. However, one of the simplest is to calculate the yearly mean experience increase. Assuming that individuals either work full time or not at all we approximate experience up to 1970 using a binomial distribution with probability estimated by gender-education-time since entry groups. We thus divide workers into 4 groups based on time since entry. These are 1-5, 6-10, 11-15, and above 15 years since entry. An example of a group could be female Kindergarten teachers with 1 to 5 years in the labor market.

Table 1: Overview of data creation

	Number of workers	Number of Firms	Number of Spells	Number of Establishments	Fraction Employed
Merged data	5,116,625	455,054	60,914,366	610,130	
Censoring at age 56	4,348,157	446,957	56,090,810	596,777	
Delete workers with all missing educational variables (hifsp)	4,147,463	445,298	54,714,582	594,070	
Delete workers with gaps in educational variables (hifsp)	4,147,463	445,298	54,714,582	594,070	
Delete obs. below age 19	3,967,419	420,223	51,494,700	550,388	
Labor market entry	3,967,419	400,670	43,406,322	517,710	
Delete under education	3,461,332	386,894	37,007,709	496,792	
Changing hifsp codes	3,346,165	385,833	36,778,315	495,453	
To early labor market entry	3,322,994	385,289	36,540,726	494,727	
Clean for temporary non-employment	3,313,246	384,988	29,495,476	494,275	
Censoring self-employment	3,118,361	368,839	26,920,923	472,219	
Censoring retirement	3,041,336	368,747	26,462,593	472,095	
Delete workers with gaps	2,969,454	365,959	25,630,532	467,983	
Delete workers with missing experience	2,969,454	365,959	25,630,532	467,983	
Average over yearly cross-sections					
Merged data	3,717,692	129,506	3,717,692	167,882	0.60
Censoring at age 56	2,955,775	124,698	2,955,775	161,952	0.68
Delete workers with all missing educational variables (hifsp)	2,894,716	124,084	2,894,716	161,285	0.69
Delete workers with gaps in educational variables (hifsp)	2,894,716	124,084	2,894,716	161,285	0.69
Delete obs. below age 19	2,700,370	116,688	2,700,370	153,681	0.70
Labor market entry	2,375,046	110,346	2,375,046	146,898	0.71
Delete under education	2,143,752	106,593	2,143,752	142,682	0.72
Changing hifsp codes	2,137,499	106,470	2,137,499	142,548	0.72
To early labor market entry	2,127,965	106,353	2,127,965	142,423	0.72
Clean for temporary non-employment	2,120,247	106,635	2,120,247	142,777	0.72
Censoring self-employment	1,872,217	100,595	1,872,217	136,301	0.78
Censoring retirement	1,763,850	100,568	1,763,850	136,269	0.83
Delete workers with missing experience	1,722,319	99,802	1,722,319	135,375	0.83

Table 2: Summery Statistics: Pooled Cross-sections

	Mean	Std. Dev.	Min.	Max.
Number of years in sample	11.09	6.27	1	19
Number of Establishments per worker	2.74	1.87	1	18
Female	0.49			
Average years of education	11.66	3.18	7	20
Age	38.33	9.63	19	55
Employed	0.84			
Public Employed	0.32			
Missing Establishment ID	0.005			
Experience	13.18	9.22	0.00	40.92
Log Wages	4.54	0.34	0.04	9.07

Table 3: Auxliary Model and Estimates

Moment	Data	Model
Avg. Length Emp. Spell	270.8	272
Avg. Length Nonnemp. Spell	79.8	78.7
Avg. Length Job	118.3	119
$E(\tilde{S}_{ilj}\tilde{h}_{-ilj})$	0.0211	0.0212
Between Persons	0.0763	0.0762
Between Jobs	0.0279	0.0280
Within Job	0.0154	0.0154
Sample mean $w_{it}$	4.54	4.54
$E(\tilde{w}_{it}\tilde{w}_{-it})$	0.00362	0.00362
$E(\tilde{w}_{it}\tilde{h}_{-it})$	0.00166	0.00166
Fraction Wage Drops	0.404	0.397
Coeff Exper	0.0287	0.0283
Coeff Exper $\times 1(\text{Exper} > 7)$	-0.0248	-0.0249
Tenure	0.00271	0.00271

Table 4: Parameter Estimates

$\delta$	0.160
$\lambda^n$	0.879
$\lambda^e$	1.087
$E_\theta$	4.36
$\sigma_\theta$	0.240
$\sigma_\xi$	0.148
$f_1$	1.12
$f_2$	0.15
$f_3$	0.09
$\sigma_{vw}$	0.18
$\alpha$	2.168
$b_1$	0.036
$b_2$	-0.0022
$\beta$	0.89

Table 5: Model Decompositions: Variance of log wages

(A)		(B)		(C)		(D)	
Total	0.098	Total	0.098	Total	0.098	Total	0.098
No HC	0.092						
$\beta = 1$	0.094						
No Roy	0.009	No Search	0.095	No Comp	0.098	No Comp	0.098
No Search	0.007	No Roy	0.007	No Search	0.068	No Roy	0.007

# Figures

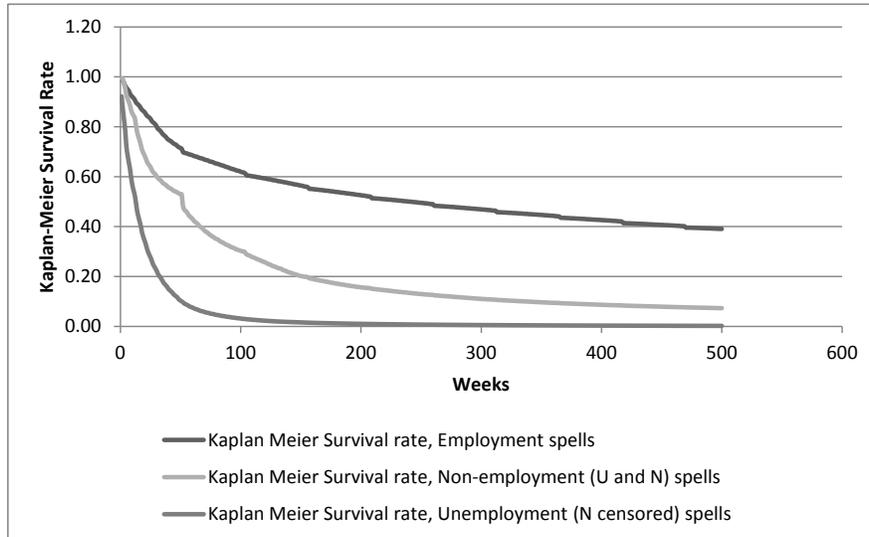


Figure 1: SURVIVAL PLOTS OF EMPLOYMENT, UNEMPLOYMENT, AND NON-EMPLOYMENT SPELLS