Estimation of a Roy/Search/Compensating Differential Model of the Labor Market

Christopher Taber

University of Wisconsin-Madison

Rune Vejlin

Aarhus University

August 16, 2011

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 of the other three sources of inequality. 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 earnings. 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 talent can earn different salaries. Finally, workers may have had poor luck in finding their ideal job. This type of search friction can also lead to heterogeneity in earnings as some workers may work for higher wage firms. In short, one worker may earn more than another a) because he has more talent at labor market entry (Roy Model), b) because he has accumulated more human capital while working (human capital), 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 earnings inequality.

For now we omit human capital. We develop and estimate a structural econometric model of wage determination that contains elements of the other three models. We use the estimated parameters to decompose overall earnings inequality into the three components.

We briefly discuss the relationship between this work and the previous literature in Section 2. We then describe the basic version of the model and the decomposition in Section 3 and describe an equilibrium version of the model in Section 4. Identification is discussed in 5 and the specific econometric specification is presented in Section 6. 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 7. Section 8 presents the auxiliary model we use and Section 9 presents the results.

2 Relation to Other Work

We do not know of another paper that has estimated a model containing all three of these elements. However, clearly there is a huge amount of work on search models, on the Roy model, 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 papers are Abowd, Kramarz, and Margolis (1999) and Postel-Vinay and Robin (2002). Both of these other papers use the Declarations Annuelles des Donnees Sociates data set which includes 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. In some ways, our first approach (i.e. not modeling firms) is a natural extension of Abowd, Kramarz, and Margolis (1999) while the second (i.e. the equilibrium model) is a natural extension of Postel-Vinay and Robin (2002).

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

$$\log(W_{itj}) = X_{it}'\beta + \theta_i + \mu_j + \zeta_{itj}$$

where θ_i is treated as an individual fixed effect, μ_j is treated as a firm fixed effect, and ζ_{itj} is independent and identically distributed. We will 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 that we observe seems to occur from market inefficiencies and how much occurs as a result of workers choice. Furthermore, the idea that conditional on worker i working for firm j, the assumption that ζ_{itj} is independent of θ_i and μ_j seems highly suspect, and our model will be able to account for this selection problem. Finally, we have all of the advantages of a structural model which helps in interpretation of the model and allows us to use it for policy simulations.

Postel-Vinay and Robin (2002) also 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 nonpecuniary 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 firm specific effect (v_{ij}) meaning that some workers match better with some firms. Furthermore, our estimation and identification strategy is very different. They estimate assuming that the model is in steady states, 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.

Another important related paper is the work by 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 includes compensating differentials into a search models. 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.

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 den Berg (2007) provide a nice discussion of empirical search models.

3 The Basic Model and Decomposition

In this section we present a very simple version of our economic framework. The details are presented in the technical appendix. Suppose there are a finite number of firm types J and let j index a firm type with $j = 0$ denoting non-employment.

In this section, we consider a price taking version of the model in which workers will take the wages offered by firms as given and adjust relative to that. We will estimate the distribution of wages offered across firms, but we will not explicitly estimate firm behavior. Let W_{ij} be the wage that firm j would offer to worker i. This will vary across different firms as the worker will have both absolute advantage and relative advantage at different jobs when compared to other workers. One can motivate this as heterogeneity in ability, but also that as a result of search frictions firms will have different wage policies as in Burdett and Mortensen (1998).

We use a continuous time model and assume that agents live forever. Individuals receive an offer from firm j at rate λ_j^e when they are employed and λ_j^n when they are non-employed. When they receive an offer they choose whether to accept and move to the new job or to turn it down. When individual i works for firm $j > 0$, she receives flow utility

$$U_{ij}(w),$$

where U_{ij} incorporates individual specific tastes for a job and is strictly increasing in wages. The job match exogenously ends at rate δ. Further, let U_{i0} represent the flow utility from non-employment.

For each job we will define $V_{ij}(w)$ as the value function associated with that job at wage rate w.[1] As time is continuous, we do not need to worry about workers receiving more than one job offer at a time. When an employed worker receives a job offer she compares it to her current job and will take it when the value function on the new job is higher. Letting $\Lambda^e = \sum_{j=1}^{J} \lambda_j^e$ and $\Lambda^n = \sum_{j=1}^{J} \lambda_j^n$ be arrival rates of any job offer then for any $j > 0$,

$$(\rho + \delta + \Lambda^e)\, V_{ij}\left(W_{ij}\right) = U_{ij}(W_{ij}) + \sum_{\ell=1}^{J} \lambda_\ell^e \max\{V_{i\ell}(W_{i\ell}), V_{ij}(W_{ij})\} + \delta V_{i0}.$$

It is straightforward to see that in this simple model the preference across jobs (excluding non-employment) will be ordered by flow utility $U_{ij}(W_{ij})$.

The behavior of non-employed agents is even simpler. They receive an offer from firm j at rate λ_j^n. They take the job if $V_{ij} > V_{i0}$. For non-employment the Bellman equation is

$$(\rho + \Lambda^n)V_{i0} = U_{i0} + \sum_{\ell=1}^{J} \lambda_\ell^n \max\{V_{i\ell}(W_{i\ell}), V_{i0}\}.$$

Note that as long as $\lambda_\ell^n \neq \lambda_\ell^e$, comparing the flow utility is not enough. In what follows we do not need to worry about flow utility and just focus on the value functions.

It is important to recognize that not all workers will take all jobs. That is there may be quite a few jobs for which

$$V_{ij} < V_{i0}$$

and we could never observe this worker in this job. This leads to an important restriction on what can be identified that will be described below. The basic problem is that we can never hope to identify the wage that a nobel prize winning economists would make as a dishwasher because it would never happen, but for the same basic reason it is not a particularly interesting counterfactual. Thus the fact that it is not identified is inconvenient but not usually problematic.

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:

[1] In this section as wages are fixed across jobs we could avoid writing the value function as a function of the wage, but rather just take it as given. However, in the next section this will be important and we want to keep the notation as consistent as possible.

- Worker variation in potential wages $W_{i1}, ..., W_{iJ}$ leads to "Roy Model" inequality or human capital inequality

- variation in the function $U_{ij}(\cdot)$ across persons accommodates "compensating differentials" inequality (and the mean of $U_{ij}(\cdot)$ will vary across jobs)

- λ_j^e and λ_j^n incorporate search frictions

After estimating the parameters of the model, we will 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 earnings exist and we will use different ones to highlight different features of the model. The following simulations represent an example of what could be done. We sequentially take the following steps.

a) First simulate the cross section variance of wages in model using all parameters (which should be approximately the same as overall wage inequality in the data).

b) Set λ_j^e and λ_j^n arbitrarily large so that workers move to the preferred job immediately.

c) In addition set wages to be $W_{ij} = E(W_{ij} \mid j)$ 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 search frictions, the difference between b) and c) is due to Roy model inequality, and the remaining fraction in c) is due to compensating differentials.

4 Model with Bertrand Competition for Wages

Next consider a model in which wages are determined by Bertrand competition following Postel-Vinay and Robin (2002). The details are presented in the appendix. We let π_{ij} be the productivity of worker i at firm j and assume is the maximum wage the firm would be willing to pay the worker.[2] The key idea of this model is that when a worker receives an outside offer the wage and outcome are determined as in a second price auction. That is we will see efficient turnover and the winning firm will pay a wage that makes the worker exactly indifferent between the new firm and the maximum bid at the old firm. This wage is kept fixed until it is renogotiated when the worker receives a better outside offer, or when the job spell ends.

For example suppose that worker i currently works at firm j and then gets an offer from firm ℓ. Suppose the workers current wage is w and the outside offer satisfies

$$V_{ij}(w) < V_{i\ell}(\pi_{ij\ell}) < V_{ij}(\pi_{ij}).$$

That is, the maximum offer on the new firm would dominate the current offer, but that the current firm would respond with a dominating offer. The Postel-Vinay Robin (2002) framework in this case would lead the worker to stay at the current firm, but with a wage w' determined as

$$V_{ij}(w') = V_{i\ell}(\pi_{i\ell}).$$

Without loss of generality we define the index across jobs in order of priority from lowest to highest so that when $j > \ell$,

$$U_{ij}(\pi_{ij}) > U_{i\ell}(\pi_{i\ell}).$$

It is straightforward to show that the continuous time value function evaluated at wage w takes the form

$$\begin{aligned} &\left(\rho + \delta + \Lambda^e - \Lambda^e_{\ell(w)}\right) V_{ij}(w) \\ &= U_{ij}(w) + \sum_{\ell=\ell(w)+1}^{j} \lambda^e_\ell V_{i\ell}(\pi_{i\ell}) + \left(\Lambda^e - \Lambda^e_j\right) V_{ij}(\pi_{ij}) + \delta V_{i0}. \end{aligned} \tag{1}$$

where we have defined $\ell(w)$ be the largest value of ℓ such that $V_{i\ell}(\pi_{i\ell}) \leq V_{i\ell}(w)$ and we generically define

$$\Lambda^e_\ell = \sum_{j=1}^{\ell} \lambda^e_j.$$

In deriving this result it is important to recognize that there are three different types of outside offers. When the outside offer satisfies $\ell \leq \ell(w)$, then it must be the case that

[2] We are implicitly assuming the value of a vacancy is zero. One could easily relax this assumption and just redefine π_{ij} to be the maximum wage a firm would ever offer.

$V_{i\ell}(\pi_{i\ell}) \leq V_{i\ell}(w)$ and the current firm has no incentive to respond. In the second case, if $\ell(w) < \ell \leq j$, then the current firm will respond to the outside offer with a wage raise, but the worker will remain at firm j. This corresponds to the part of the expression $\sum_{\ell=\ell(w)+1}^{j} \lambda_\ell^e V_{i\ell}(\pi_{i\ell})$. The firm responds so that the worker is indifferent between staying or going at the maximum wage the alternative firm would be willing to pay. That wage would yield the value function $V_{i\ell}(\pi_{i\ell})$. The third case is when the outside firm beats the inside firm. Such offers occur at rate $\left(\Lambda^e - \Lambda_j^e\right)$ and the worker moves to the new firm at a wage which gives value function $V_{ij}(\pi_{ij})$.

Now evaluate expression (1) at $w = \pi_{ij}$ (at this case $\ell(w) = j$),

$$\left(\rho + \delta + \Lambda^e - \Lambda_j^e\right) V_{ij}(\pi_{ij}) = U_{ij}(\pi_{ij}) + \left(\Lambda^e - \Lambda_j^e\right) V_{ij}(\pi_{ij}) + \delta V_{i0}$$

or[3]

$$V_{ij}(\pi_{ij}) = \frac{U_{ij}(\pi_{ij}) + \delta V_{i0}}{\rho + \delta}.$$

Now suppose a worker at firm j gets an offer from a firm ℓ^* that will be matched. Let the matched log wage be $W_{ij\ell^*}$, it must satisfy

$$\begin{aligned} U_{ij}(W_{ij\ell^*}) =& U_{i\ell^*}(\pi_{i\ell^*}) - \sum_{\ell=\ell^*+1}^{j} \lambda_\ell^e \frac{U_{i\ell}(\pi_{i\ell}) - U_{i\ell^*}(\pi_{i\ell^*})}{\rho + \delta} \\ & - \left[\Lambda^e - \Lambda_j^e\right] \frac{U_{ij}(\pi_{ij}) - U_{i\ell^*}(\pi_{i\ell^*})}{\rho + \delta}. \end{aligned}$$

This is also the new wage for a worker who is currently at firm ℓ^* but then accepts an offer from and outside firm j with $j > \ell^*$. This expression is quite intuitive. In order to keep a worker indifferent between working at firm j at wage $W_{ij\ell^*}$ and working at firm ℓ^* the firm must compensate for the flow utility $u(\pi_{i\ell^*})$ net of the change in the bargaining power. The change in the bargaining power takes to parts, the first is due to the difference in the response to an outside offer that firm j would match, which firm ℓ^* would not. The second is that if the worker switches to a firm superior to j, they will be in a better bargaining position if they are working for firm j than firm ℓ^*.

The next object to calculate V_{i0} which is pretty much identical to Postel-Vinay and Robin (2002). It is straight forward to show that

$$V_{i0} = \frac{U_{i0}}{\rho}.$$

[3] Postel-Vinay and Robin (2002) get the exact same expression.

When a firm hires a worker from non-employment it will set the wage so that the worker is indifferent between no-employment and working at the firm. Combining this with equation (1) yields

$$U_{ij}(W_{ij0}) = U_{i0} - \sum_{\ell=\ell_0}^{j} \lambda_\ell^e \frac{U_{i\ell}(\pi_{i\ell}) - U_{i0}}{\rho + \delta} - \left[\Lambda^e - \Lambda_j^e\right] \frac{U_{ij}(\pi_{ij}) - U_{i0}}{\rho + \delta}.$$

where ℓ_0 is the highest value job that the worker would decline to remain non-employed.

5 Identification

In this section we discuss identification of our model. We start with the basic model and then discuss the Bertrand version. We show which aspects of the model can and cannot be identified. We view both parts as important. In constructing the counterfactuals it is important to keep in mind that we can not credibly simulate counterfactuals that are not identified from the data. In particular, we will show that one can not hope to identify W_{ij} 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 to readers, but it is important to keep in mind that this limits the type of counterfactuals which can be simulated. We discuss this issue at the end of this section.

We define the original parameters in terms of flow utilities as a function of wage $U_{ij}(w)$, but since there is no variation in w within jobs, all we can hope to identify about tastes is $V_{ij}(W_{ij})$. The parameters and random variables in our model are $W_{ij}, V_{ij}, \lambda_j^e$, λ_j^n and δ. Proving general nonparametric identification of the remaining features when J is very large seems overly tedious, so instead we focus on a simpler case to illustrate how our model can be identified. We fully expect our result to generalize to larger (but finite) J.

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) by using a revealed preference argument to suggest 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 non-employment-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 nonpecuniary 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 suspect that this problem is unlikely to drive the results, but it is something we will address in future drafts as best we can.

A second key aspect of the model is that we show that the arrival rates (λ_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 very important component of inequality, then the rate of job switching should be fairly slow. However, if search frictions are relatively small (arrival rates are high) workers will receive an offer from their preferred firm and switch to a new firm rapidly.[4] Thus it is important to have high quality panel data on job switching and also matched firm/worker data.

To sketch the intuition for identification, we model a two firm labor market. We think it is clear that the logic of this argument will follow for any finite number of jobs but we have not explicitly shown this. First consider what can be identified without data on wages.

We assume that there are two types of jobs which we label as A and B. Let λ_A^e and λ_A^n be the arrival rates at job A and λ_B^e and λ_B^n the arrival rates at job B. We have essentially five different groups of people, those who would work only for firm B, those who would work only for firm A, those that work for either but prefer B, those who would work for either but prefer A, and those that never work. In this order, the five groups with their population proportions are

$$\begin{array}{cc} V_{iA}(W_{iA}) < V_{i0} < V_{iB}(W_{iB}) & P_{B-} \\ V_{iB}(W_{iB}) < V_{i0} < V_{iA}(W_{iA}) & P_{A-} \\ V_{i0} < V_{iB}(W_{iB}) < V_{iA}(W_{iA}) & P_{AB} \\ V_{i0} < V_{iA}(W_{iA}) < V_{iB}(W_{iB}) & P_{BA} \\ V_{iA}(W_{iA}) < V_{i0}, V_{iB}(W_{ib}) < V_{i0} & P_0 \end{array}$$

Note that we have assumed no ties.[5]

While our access to repeated spells simplifies identification considerably, we ignore that at this point and assume that we observe a single chain of spells from non-employment to employment back to non-employment. In addition, for workers who would never work we

[4]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 the 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).

[5]For the actual specifications we plan to use, ties will be a zero probability event.

observe this directly (and thus P_0 is trivially identified). For workers who ever work we observe the duration of non-employment followed by the duration of employment. During the employment spell we observe the duration of each job and we observe wages on the jobs which are collected periodically. It is important to distinguish an "employment spell" from a "job spell." By a job spell, we mean the amount of time that a worker spends on a particular job. By an employment spell we mean the amount of time a worker spends working between periods of non-employment. Thus an employment spell may contain several job spells.

First consider what can be identified without wage data. Ignoring wages, we have 10 objects in the model to identify $\lambda_A^e, \lambda_A^n, \lambda_B^e, \lambda_B^n, \delta, P_{B-}, P_{A-}, P_{AB}, P_{BA}$, and P_0. As mentioned above P_0 is trivially identified. Furthermore, the probabilities of types in the population must sum to one, so we really have 8 parameters left.

To see how δ is identified we can simply look at the distribution of employment spells. Thus the time from hiring from non-employment to layoff has an exponential distribution with parameter δ. Under the assumptions given we can condition on individuals whose employment durations are arbitrarily long. This is not necessary for identification, but is formally justifiable and simplifies the exposition.[6]

Given this assumption, identifying λ_A^e is similar to δ. We must first condition on individuals who start at B and move to A. However, once we have done this the amount of time at job B has an exponential distribution with parameter λ_A^e. Identification of λ_B^e is analogous. This confirms the intuition above that the search frictions can be identified from the rate of relative switching probabilities. One can see that it may turn out to be important to control for job or occupation specific human capital because one reason why transition rates out of j may be small relative to is because of specific human capital accumulated during the second period.

Now consider identification of the remaining objects. Abusing notation somewhat, let the event $J_i = A$ denote that the individual only worked at job A during the spell (and analogously for $J_i = B$). Let $J_i = AB$ denote that the individual started at A and then moved to B (and let $J_i = BA$ be defined analogously).

Let T_{iA}^n and T_{iB}^n be the duration of non-employment before the first job offer from firm A and B, respectively. Now let T_i^n be the duration of non-employment before the first job. For example for types $A-$, $T_i^n = T_{iA}^n$ and for types AB, $T_i^n = min(T_{iA}^n, T_{iB}^n)$. Now consider

[6] That is the notion of identification is what we could identify given the full joint distribution of the data. If we know the full joint distribution, there is no reason why we could not condition on extremely long employment spells.

very long employment spells where the worker only worked in firm A then[7]

$$
\begin{aligned}
Pr(T^n_{iA} > \tau \mid j = A) &= \frac{Pr(T^n_{iA} > \tau \mid A-)P_{A-} + Pr(T^n_{iB} > T^n_{iA} > \tau \mid AB)P_{AB}}{P_{A-} + P_{AB}Pr(T^n_{iB} > T^n_{iA})} \\
&= \frac{e^{-\lambda^n_A \tau}P_{A-} + \frac{\lambda^n_A}{\lambda^n_A+\lambda^n_B}e^{-\left(\lambda^n_A+\lambda^n_B\right)\tau}P_{AB}}{P_{A-} + P_{AB}\frac{\lambda^n_A}{\lambda^n_A+\lambda^n_B}}
\end{aligned}
$$

Now notice that

$$
\lim_{\tau\to\infty} e^{\lambda^n_A \tau} Pr(t_n > \tau \mid j = A) = \frac{P_{A-}}{P_{A-} + P_{AB}\frac{\lambda^n_A}{\lambda^n_A+\lambda^n_B}}
$$

so λ^n_A is identified since this expression would go to ∞ or 0 for any other value. Similarly λ^n_B is identified.

Finally condition on individuals for which the duration of employment is arbitrarily large (this is not important as the argument would go through with smaller levels, but the algebra is simplest in this case). Then

$$
\begin{aligned}
Pr(j = A) &= P_{A-} + P_{AB}\frac{\lambda^n_A}{\lambda^n_A + \lambda^n_B} \\
Pr(j = B) &= P_{B-} + P_{BA}\frac{\lambda^n_B}{\lambda^n_A + \lambda^n_B} \\
Pr(j = AB) &= P_{BA}\frac{\lambda^n_A}{\lambda^n_A + \lambda^n_B} \\
Pr(j = BA) &= P_{AB}\frac{\lambda^n_B}{\lambda^n_A + \lambda^n_B}.
\end{aligned}
$$

This is four linear equations in four linear unknowns so is identified.[8] Thus the model is identified. From here one can see the importance of panel data on job-to-job transitions.

Now consider using data on wages as well. For individual i, let W_{iA} be the wage at job A and W_{iB} be the wage at job B. Everything else is the same as before, but now we use the

[7]Let $f_{t^A_n}$ be the pdf of t^A_n then the probability $Pr(t^B_n > t^A_n > \tau)$ can be calculated like this

$$
\begin{aligned}
Pr(t^B_n > t^A_n > \tau) &= \int_\tau^\infty Pr(t^B_n > t^A_n \mid t^A_n = u) f_{t^A_n}(u)du = \int_\tau^\infty Pr(t^B_n > u) f_{t^A_n}(u)du \\
&= \int_\tau^\infty e^{-\lambda^n_B u}\lambda^n_A e^{-\lambda^n_A u}du = \lambda^n_A \int_\tau^\infty e^{-(\lambda^n_A+\lambda^n_B)u}du \\
&= \frac{\lambda^n_A}{\lambda^n_A + \lambda^n_B}e^{-(\lambda^n_A+\lambda^n_B)\tau}
\end{aligned}
$$

[8]Actually given that P_0 is known it is really 3 equations in three unknowns because both the equations and population types must sum to $1 - P_0$.

data on wages. For any two cutoffs w_A and w_B, we can identify

$$\Pr(W_{iA} \leq w_A, W_B \leq w_B \mid J = AB) = F_{BA}(w_A, w_B)$$

$$\Pr(W_{iA} \leq w_A, W_B \leq w_B \mid J = BA) = F_{AB}(w_A, w_B)$$

$$Pr(W_{iA} \leq w_A \mid j = A) = \frac{F_{AB}(w_A, \infty) P_{AB} \frac{\lambda_A^n}{\lambda_A^n + \lambda_B^n} + F_{A-}(w_A) P_{A-}}{P_{A-} + P_{AB} \frac{\lambda_A^n}{\lambda_A^n + \lambda_B^n}}$$

$$Pr(W_B \leq w_B \mid j = B) = \frac{F_{BA}(\infty, w_B) P_{BA} \frac{\lambda_B^n}{\lambda_A^n + \lambda_B^n} + F_{B-}(w_B) P_{B-}}{P_{B-} + P_{BA} \frac{\lambda_B^n}{\lambda_A^n + \lambda_B^n}}$$

where F_{BA} and F_{AB} are the cumulative distribution functions of (W_A, W_B) conditional on being in groups AB and BA respectively and F_{A-} and F_{B-} are the cumulative distribution functions of W_{iA} and W_{iB} conditional on being in groups $A-$ and $B-$ respectively. Since everything else in these equations is identified as shown above, F_{BA} and F_{AB} are identified from the two equations. Given these, one can then show that F_{A-} and F_{B-} are identified from the second equations given the other parameters which have been shown to be identified.

We have assumed that there is no measurement error in wages. In the example below we allow for an i.i.d. measurement error component in log wages. As long as one can observe multiple wages at the same time for at least some workers, a standard deconvolution argument will allow one to separately identify the joint distribution of wages from the measurement error.

It is important to consider not only what can be identified, but what is not identified as well. We have shown that F_{BA} and F_{AB} are identified, but we can not identify the joint distribution $F_{A-}(w_A, w_B)$ completely since we will never observe W_B for these types of workers. This is an issue that needs to be dealt with in estimation and simulation. Once we go to many jobs, clearly will not observe most workers at most jobs so the full joint distribution will not be identified for most workers. This is not problematic as long as the simulations that we use do not require knowing these counterfactuals. Since our goal is to understand the sources of wage inequality for jobs that we see, this is not fundamentally problematic. When we relax search frictions we would continue to assume that workers would not work on jobs on which they currently do not work. The only place where this is a bit of an issue when we relax compensating differentials versus Roy inequality. Since we do not observe W_B for the $A-$ workers, we can not say whether they do not choose those jobs because they have high disutility or because they have low wages. We plan to simply simulate the model continuing to assume that $A-$ workers would never work for firm B.

While doing this is somewhat unsatisfactory, it is clear that the distribution of W_B for those workers is fundamentally nonparametrically unidentified and we think the interest of the exercise is only slightly diminished when we condition on this.

We now have sufficient information to perform the decomposition described above. That is, once this model is identified we can simulate the variance of log wages under the various counterfactuals. That is we can followed the strategy outlined in Section 3 where in this particular model each step would correspond to:

a) First simulate the cross section variance of wages in model using all parameters.

b) To get rid of search frictions we would assume that type AB and $A-$workers would immediately move into job A and type BA and $B-$ workers would immediately move into job B.

c) To get rid of Roy inequality, set $W_{iA} = E(W_{iA}$ |workers of type $AB, A-$, and BA) and $W_{iB} = E(W_{iB}$ |workers of type $AB, B-$, and BA) for all i eliminating variation due to skill differences.

We could also get rid of compensating differential by allowing workers to purely maximize wages. However, we are restricted in how we can do that as we can not observe the wage a type $A-$ would earn at firm B.

5.1 Bertrand Model

Extension to the Bertrand model is relatively straightforward. First notice that as long as we define the types by $V_{iA}(\pi_{iA})$ and $V_{iB}(\pi_{iB})$ rather than $V_{iA}(W_{iA})$ and $V_{iB}(W_{iB})$ the model determining assignment to jobs is completely identical to the basic model so $\lambda_A^e, \lambda_A^n, \lambda_B^e, \lambda_B^n, \delta$, $P_{B-}, P_{A-}, P_{AB}, P_{BA}$, and P_0 are all identified.

To see how the rest of identification for this model works, consider workers that prefer A to B and prefer B to non-employment. Here it will be important that A and B represent types of jobs, so a worker could receive an offer from another firm with exactly the same type of job.

Thus in this model for any of these types of workers there will be five potential wages:

1. W_{iB0} is the wage at job B when hired directly from non-employment.

2. W_{iBB} is the wage at job B following an offer from another type B firm.

3. W_{iA0} is the wage at job A when hired directly from non-employment.

4. W_{iAB} is the wage at job A following an offer from a type B firm.

5. W_{iAA} is the wage at job A following an offer from another type A firm.

We know that:

$$W_{iAA} = \pi_{iA}$$
$$W_{iBB} = \pi_{iB}.$$

Note that this means already with a long enough panel we can identify the joint distribution of (π_{iA}, π_{iB}) for these workers.

For the rest of the parameters we must put more structure on the model and write the preferences as

$$U_{ij}(w) = \log(w) + \nu_{ij}$$

then

$$\log(W_{iB0}) + \nu_{iB} = \frac{(\rho + \delta + \lambda_A^e + \lambda_B^e)}{\rho + \delta} U_{i0} - [\lambda_B^e + \lambda_A^e] \frac{\pi_{iB} + \nu_{iB}}{\rho + \delta}$$

$$\log(W_{iA0}) + \nu_{iA} = \frac{(\rho + \delta + \lambda_A^e + \lambda_B^e)}{\rho + \delta} U_{i0} - \lambda_B^e \frac{\pi_{iB} + \nu_{iB}}{\rho + \delta} - \lambda_A^e \frac{\pi_{iA} + \nu_{iA}}{\rho + \delta}$$

$$\log(W_{iAB}) + \nu_{iA} = (\rho + \delta + \lambda_A^e) \frac{\pi_{iB} + \nu_{iB}}{\rho + \delta} - \lambda_A^e \frac{\pi_{iA} + \nu_{iA}}{\rho + \delta}$$

We know that preferences are only defined up to monotonic transformations, so without loss of generality we could increase the flow utilities by any number for all states of the world and this will not change preference. Thus like in discrete choice models we need a location normalization. The most natural is to set $U_{i0} = 0$, which leaves the following system of equations:

$$(\rho + \delta) \log(W_{iB0}) + (\rho + \delta) \nu_{iB} = [\lambda_B^e + \lambda_A^e] \pi_{iB} + [\lambda_B^e + \lambda_A^e] \nu_{iB}$$
$$(\rho + \delta) \log(W_{iA0}) + (\rho + \delta) \nu_{iA} = -\lambda_B^e \pi_{iB} - \lambda_A^e \pi_{iA} - \lambda_B^e \nu_{iB} - \lambda_A^e \nu_{iA}$$
$$(\rho + \delta) \log(W_{iAB}) + (\rho + \delta) \nu_{iA} = -\lambda_A^e \pi_{iA} - \lambda_A^e \nu_{iA.}$$

Taking expectations we have three equations in the three unknowns ρ, $E(\nu_{iB} \mid AB)$, and $E(\nu_{iA} \mid AB)$. One can solve out for the model and show it is a cubic and thus either has one solution or three solutions. At the very least ρ takes on three values. Given data on other moments and on the BA group, in the vast majority of cases ρ should be identified. However, in some very special cases there might be three different values of ρ and estimation would be done conditional on these three values.

Once ρ is identified, consider those workers who first obtain a job from B, then a counter offer from another B type, then move to an A type firm, and then receive a counter offer from another type A firm. From these workers we can identify the joint distribution of $(W_{iB0,}W_{iBB}, W_{iAB}, W_{iAA})$ from which one can identify the joint distribution of $(\pi_{iA}, \pi_{iB}, \nu_{iA}, \nu_{iB})$ for the AB types. Clearly the model is symmetric for the BA types so the conditional joint distribution of $(\pi_{iA}, \pi_{iB}, \nu_{iA}, \nu_{iB})$ can be identified as well.

Next consider the $A-$ types. For them only two wages are possible:

$$\begin{aligned}\log(W_{iA0}) + \nu_{iA} &= \frac{\rho\left(\rho+\delta+\lambda_A^e\right)V_{i0}}{\rho+\delta} - \lambda_A^e \frac{\pi_{iA}+\nu_{iA}}{\rho+\delta} \\ \log(W_{iAA}) &= \pi_{iA}.\end{aligned}$$

Given knowledge of everything else we can identify the joint distribution of (π_{iA}, ν_{iA}) for this group. Identification of the distribution of (π_{iB}, ν_{iB}) for group $B-$ is analogous.

6 Econometric Specification/Parameterization

6.1 Model 1

We assume that there are a large number of people in the economy but a finite number of J firm types. Multiple firms in the data will be of the same type. For each firm type, there are individual firms which have idiosyncratic properties. We will let $j_i(t)$ denote the job held by worker i at time t.

The transition parameters δ, λ^e, and λ^n take the same form as in the earlier sections.

True offered wages are

$$\log(W_{ij}) \quad = \quad \theta_i^w + \mu_j^w + \upsilon_{ij}^w$$

but we only get to observed this variable measured with error

$$\log(W_{ijt}^m) = \log(W_{ij}) + \xi_{it}$$

where all of these variables are independent of each other.

Rather than model flow utilities we model the value functions directly as

$$\begin{aligned} V_{ij}(W_{ij}) &= \alpha \log(W_{ij}) + \mu_j^u + v_{ij}^u. \\ &= \alpha\theta_i + \left(\alpha\mu_j^w + \mu_j^u\right) + \left(\alpha v_{ij}^w + v_{ij}^u\right) \\ &\equiv \alpha\theta_i + \widetilde{\mu}_j^u + \widetilde{v}_{ij}^u \end{aligned}$$

It is important that we allow μ_j^w to be correlated with μ_j^u because of equilibrium considerations. Note that all that matters for job to job turnover is the reduced form. One can see that from this data α will not be identified, but we do not need to know it in order to perform our counterfactual simulations.

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 θ_i, and the distribution of ξ_{it} are all independent of each other. Furthermore we assume that ξ_{it} is i.i.d. across time.

We take a very simple specification for the value of non-employment by assuming

$$V_{i0} = \alpha\theta_i + V_0$$

where V_0 is constant across people. In future versions of this paper we will consider more complicated specifications. Given that the paper is not fundamentally about the transition from non-employment to employment, we suspect this will make little difference.

As with standard discrete choice models we need a location normalization, so we will choose to set $var(\widetilde{v}_{ij}^u) = 1$. Moreover, we assume that $\left(\widetilde{v}_{ij}^u, v_{ij}^w\right)$are jointly normal which gives us two parameters, $cov\left(\widetilde{v}_{ij}^u, v_{ij}^w\right)$ and $var\left(v_{ij}^w\right)$. We will also assume that θ_i is normal and that ξ_{it} is normal with mean zero.

We tried to choose a relatively parsimonious way to approximate the distribution of $\left(\widetilde{\mu}_j^u, \mu_j^w\right)$. Note that we want to keep this a discrete distribution. With no obvious parametric alternative we decided upon the following one:

$$\begin{aligned} \widetilde{\mu}_j^u &= f_1\left[U_1(j) + 0.5U_2(j)\right] \\ \mu_j^w &= f_2 U_1(j) + f_3 U_2(j) \end{aligned}$$

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 five different values and assume these are unrelated to each-other giving us 25 different firm types. Essentially f_1 governs the variance of $\widetilde{\mu}$, while f_1 and f_2 govern the covariance of $\left(\widetilde{\mu}_j^u, \mu_j^w\right)$ as well as the variance of μ_j^w.

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 $t = t_1, ..., t_{M_i}$.

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 no work.

We assume that we get to observe $w_i(\tau)$ for $\tau = t_1, ..., t_{M_i}$.

Note that a firm in the data is different than a job type in the model. The key thing, though, is that workers who work for the same firm also work for the same firm type.

Note that given our parametric restrictions, V_0 is identified, but in practice for the reasons discussed in Flinn and Heckman (1982), it is really not because we only focus on individuals in the data that work. Given the functional forms we can identify, but this is unsatisfactory. Furthermore, we do not think this is important for the purpose of our model which is to focus on wages conditional on working. What we will do is fix V_0 and estimate the rest of our model. We will then test the sensitivity of the results to different assumptions. This leaves a total of 11 parameters. Four of them dictate turnover patterns:

$$\delta, \lambda^n, \lambda^e, f_1.$$

These can be estimated separately from the seven parameters that determine wages conditional on turnover:

$$E(\theta_i), var(\theta_i), cov\left(\widetilde{v}_{ij}^u, v_{ij}^w\right), var\left(v_{ij}^w\right), var(\xi_{it}), f_2, f_3.$$

6.2 Model 2

We use a completely analogous specification for the Bertrand model. Now we define

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

and

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

We also define the reduced form as

$$\begin{aligned} U_{ij}(\pi_{ij}) &= \alpha \log(\pi_{ij}) + \mu_j^u + v_{ij}^u. \\ &= \alpha\theta_i + \left(\alpha\mu_j^w + \mu_j^u\right) + \left(\alpha v_{ij}^w + v_{ij}^u\right) \\ &\equiv \alpha\theta_i + \widetilde{\mu}_j^u + \widetilde{v}_{ij}^u \end{aligned}$$

and

$$U_{i0} = \alpha\theta_i + V_0.$$

The form of this model is that the decision between firms will simply depend on $U_{ij}(\pi_{ij})$ with the decision about non-employment also just comparing U_{i0} with $U_{ij}(\pi_{ij})$.[9]

We parameterize the turnover part of the model exactly in the way we do for the first model, so at this point there is no difference between them. The estimates of $\delta, \lambda^n, \lambda^e$, and f_1 will be identical.

In order to identify the offered wages we will use the implications of the model to determine wages. However, there are two extra elements of the model we need in order to do this. The first is the discount rate parameter, ρ. The second is that we can no longer use the reduced form, but must be able to identify the structural parameter α. In the identification section we showed that these two parameters could, in theory, be identified from the data. However, given the simplicity in which we model wage dynamics within the firm we are somewhat uncomfortable doing this. Secondly, our goal for now is to keep the same number of estimated parameters in the two models. Rather than estimating ρ, it seems reasonable to use other things we know about discount rates to choose ρ (although they aren't directly comparable). We choose $\rho = 0.05$, but will experiment with alternative values.

In terms of identifying α we choose to identify it by assume that $cov\left(v_j^u, v_j^w\right) = 0$. While taking literally this seems like a strong identifying assumption, but in practice it boils down to assuming that the positive covariance of $\left(\widetilde{v}_j^u, v_j^w\right)$ is due to the direct effect that individuals prefer higher wages. This seems quite reasonable to us.

An alternative way of identifying α is to use wage dynamics at the firm level. That is as firms change their wage policies we should see changes in the rates at which workers come and leave. This gives us a much more direct way of estimating α.

For this draft with use the assumption that $cov\left(v_j^u, v_j^w\right) = 0$, which gives us seven parameters to estimate:

$$E(\theta_i), var(\theta_i), \alpha, var\left(v_{ij}^w\right), var(\xi_{it}), f_2, f_3.$$

[9] This last aspect is actually non-optimal. If $\lambda^e \neq \lambda^n$ the planner would take the arrival rate into account when choosing between taking a job or not. However, since potential employers take all of the surplus from the better match, the worker has no incentive to take this into account.

7 Data

This section provides a description of the data used, the sample selection process, and some descriptive statistics on the final sample.

7.1 Description of Data Manipulation

The data used in this study consists of two types of data. The first type of data is yearly data from the Danish register-based matched employer-employee data set IDA covering the period 1980 to 2003.[10] IDA contains annual socioeconomic information on workers and background information on employers, and covers the entire Danish population. Although not all information pertains to November each year (some information is registered ultimo each year, i.e. by the 31st of December), we shall treat the data as providing repeated cross-sections taken in November. Besides the worker and firm identifiers, the most important piece of information for the current study is the earnings information, which consists of the annual average hourly wage in the job occupied in the last week of November.

The second type of data is spell data. The spell data is on a weekly basis and follows all individuals in Denmark. 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 id, an employer id, a yearly establishment id, start and end date of the spell, and a variable describing the state that the worker is in. We currently have spell data from 1985 to 2003. The data set contains all workers who at some point during the sampling period are in one of the registers and are between 15 and 70 years old.

To make the data more suited for this study we manipulate it in the follow ways. There are sixteen states that the worker can occupy in the generated spell data, these are aggregated into five states; employed (E), unemployed (U), nonparticipating (N), self-employed (S), and retirement (O). The yearly information from the IDA data sets are merge onto the spell data. Hence, the structure of the data set is such that a worker who occupies, say, three different labor market states during a given calendar year will have three observations associated with that calendar year. The information will be the type of state, a start date and an end date for the spell along with socio-economic information. As this latter piece of information is obtained from the IDA data, it is constant over the three observations relating to that given worker for the given calendar year. The unit of observation is going

[10] Integreret Database for Arbejdsmarkedsforskning (Integrated Database for Labor Market Research) is constructed and maintained by Statistics Denmark.

to be a worker, 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 id is not well defined over time since firms might change the legal unit without changing anything else. The establishment id constructed by Statistics Denmark is going to be consistent over time.[11] Also, when thinking about compensating differentials the most appropriate unit seems 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 the alternative. However, using establishments instead of firms introduces another problem. Before 1991 Statistics Denmark assigned workers that did not have a particular workplace (called fictitious workplaces) such as sailors, sales representatives, temporary workers, etc. to the largest establishment within the firm.[12] From 1991 this was coded separately. This poses two problems. First, it generates a time inconsistency in the data. Second, these workers have to be assigned to some workplace. We solve the problem by assigning fictitious workplaces to the largest establishment within the firm from 1991 and onward.

We define labor market entry to be the month of graduation from the highest completed education recorded.[13] 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,

[11] The establishment is considered to be the same if one of three criteria is met: same owner and industry; same owner and workforce; same workforce and either same address or same industry.

[12] Approximately 4-5 percent of a cross-section of workers are working in these fictitious workplaces

[13] 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.

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 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.

In Table 1 the effects of the data steps taken above are described for each step. The merged data set consists of both spell and merged yearly socio-economic variables such as education. Deleting spells before the highest attained education deletes no workers but deletes 16 percent of all spells and diminishes the number of workers in a cross-section by 13 percent. Disregarding workers who are under education deletes a total of 17 percent of all workers, but only 8 percent in a cross-section. The difference arises since workers under education are not in the sample for that many years. Deleting workers with missing educational information, changing information about the highest education, or where age minus education length is less than 5 years deletes 17 percent of the workers in the sample, but only 4 percent of the workers in a cross-section. Almost all of the workers deleted in this step are caused by missing educational information. This group consists of permanent immigrants for whom we have no educational information, workers who are just in Denmark for a shorter work visit and then returns to their home country, and finally some workers for whom we simply do not have educational information. A minor group are those who are not yet done with their compulsory schooling, but have turned 15 years. Censoring workers at age 55 deletes 13 percent of all workers, but 25 percent of a cross-section. Cleaning for temporary non-employment spells deletes no workers, but we reduce the number of spells by 18 percent in the sample. Censoring workers who enter self-employment deletes 6 percent of all workers, and 9 percent of all spells. Disregarding workers with gaps in their spell history deletes 2 percent. Finally, deleting the final year of observation, i.e. 2003, leaves us with the final sample of 2,852,996 workers.

In the numerical simulations of the model we replicate the labor market entry, sample entry, and sample exit of the data. All workers start their lives as unemployed. In order to calculate the joint distribution in the data we need graduation times for all workers. We only have the graduation dates for highest completed educations back to 1970. So for workers completing their highest education before this date we need to approximate it. For all workers we have length of the education so for all workers we can calculate the potential labor market entry if the worker had gone the direct route through the educational system. We use two cohorts born in 1959 and 1960 to get the distribution of the difference between real and potential labor market entry by educational length. We then use this distribution for workers completing their education before 1970. We suspect that this approximation will not matter much since the workers enter the sample in 1985 and therefore have been at least 15 years in the labor market.

7.2 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. From the distribution of education we see that there are a relatively large group with only preparatory education, meaning 10 years of schooling or less. Remember that the averages are over cross-sections from 1985 to 2002, so the older generations in the beginning of the sample are low educated. Also the group with vocational educations are rather large, but this group also encompasses a lot of different educations from carpenters to social workers. The average cross-section age is 39 years old with the earliest labor market entry at age 15 and the highest age in the sample being age 55. 27 percent are employed in the public sector, while in total 78 percent are employed. Finally, in a given cross-section we miss establishment id for 0.8 per cent of all employment observations. We will deal with this when constructing each moment later.

Figure 1 displays the distribution over workers over number of years in the sample. As expected the number of workers declines a little but increase at the maximum number of years. However, the relative flatness of the graph indicates that we do not censor many individuals. Figure 2 shows the distribution of number of establishments per worker. This declines asymptotically toward zero. However, as can be seen we observe more than two thirds of the workers in more than one establishment.

Figure 3 displays the estimate from a Kaplan Meier estimator of the survival probability for employed and non-employed. The non-employment spells have a lot of short durations, but after approximately three to four years of being non-employed the probability of exiting the state is very small. The survival rate does not seem to approach zero. This could be due to the fact that the sample also include a relatively large amount of workers that are actually non-participants. However, we do not view this as a problem since we in the model 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. Notice, that this number is probably affected to a relatively large extent by the choice of deletion of temporary non-employment spells. I.e. if a larger fraction of temporary spells was deleted we would see a higher employment survival curve. The dips in both the survival rates comes from a "New Year" effect, i.e. there is an overrepresentation 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, but we are still working on resolving the issue.

Next, we present hiring and separation rates for percentiles of establishment fixed effects. First, we use the AKM decomposition to calculate establishment fixed effects. Each establishment is then weight by the number of November cross-section workers. Given this weighting we calculate the percentiles of the establishment effects. Finally, the number of job to job separations and job to job hires within each percentile are calculated. These are then divided by the number of cross-section observations within the percentile. These rates are displayed in Figure 4 where each point is a percentile of the establishment fixed effect distribution. The final graphs show the fraction of job to job hires to the sum of job to job separations and job to job hires (i.e. how large a fraction of the total number of job to job transitions were hires).

The results displayed on the graphs are essential for the later model that we are going to estimate. It shows that job to job transitions in and out of the establishment are not largely effected by establishment wages. However, the job to job separation rate seems to be slightly decreasing in wages, while the job to job hire rate seems to slightly increase. Also, which is essential, the two graphs seem to cross each other, which is something that any job ladder model would predict. The relative rate of job to job rates to job to job transitions seems to be a little increasing in establishment fixed effect, but still relatively flat.

8 Auxiliary Model

We will 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 map between the parameters and the data.

8.1 Notation

The data available to us is very rich. While that is very useful for identifying the model, it does mean that we have to use a large amount of notation to describe it.

- $i = 1, ..., N$ index individuals
- $\ell = 1, ..., L_i$ index employment spells-a spell is continuous employment with no non-employment in between (but may involve multiple establishments)

- $j = 1, ..., J_{i\ell}$ index an establishment spell that occurs within employment spell ℓ for individual i
- $t = 1, ..., T_{i\ell j}$ index the set of wage observations on job spell $i\ell j$.
- $f_{i\ell j}$ the firm associated with this job spell
- $1, ..., Q$ be the number of establishments in the data
- $D^w_{i\ell j}$ the duration of time that the worker worked on job spell $i\ell j$
- $w^m_{i\ell jt}$ the tth wage observation at job $i\ell j$
- $k = 1, ...K_i$ the number of non-employment spells for individual i
- D^n_{ik} the duration of non-employment spell ik

8.2 Definition of transition variables

We define a variable to pick up revealed preference on a job as

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

where JJ transition means a Job-to-Job transition. Note that there are four possibilities. If the worker enters this job from another and then leaves to non-employment (or is right truncated at the end of the sample) we will find $S_{i\ell j} = -1$ which will indicate that this worker likes this job. If they enter from non-employment, but then go directly to another job we get $S_{i\ell j} = 1$ which indicates that they did not like the job. For workers who either entered via JJ and then left via JJ, or workers who entered through non-employment and left through non-employment, there is no revealed preference information and $S_{i\ell j} = 0$.

It will also be useful to define

$$\widetilde{S}_{i\ell j} \equiv S_{i\ell j} - \frac{1}{\sum_{\ell=1}^{L_i} J_{i\ell}} \sum_{\ell=1}^{L_i} \sum_{k=1}^{J_{i\ell}} S_{i\ell k}$$

This way $\widetilde{S}_{i\ell j}$ should sum to zero for each individual when summing over jobs (or establishments, we use these interchangeably).

We define the number of JJ separations, s^q_{-i}, and JJ hires, h^q_{-i}, for each firm where individual i do not contribute.

For individual i at job j we will also want to get a sense of

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

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

Then we will use

$$\frac{h^{f_{i\ell j}}_{-i}}{h^{f_{i\ell j}}_{-i} + s^{f_{i\ell j}}_{-i}}$$

as an indicator of individual i's coworkers tastes for the job. Essentially when we obtain revealed preference information on a firm either workers are picking firm j or leaving firm j. This is the amount of time the firm wins the head to head matchup. We exclude individual i for consistency reasons. It will also be useful to define the deviation of this fraction from the workers mean value

$$\widetilde{h}_{-i\ell j} \equiv \frac{h^{f_{i\ell j}}_{-i}}{h^{f_{i\ell j}}_{-i} + s^{f_{i\ell j}}_{-i}} - \frac{1}{\sum_{\ell^*=1}^{L_i} J_{i\ell^*}} \sum_{\ell^*=1}^{L_i} \sum_{k=1}^{J_{i\ell}} \frac{h^{f_{i\ell^* k}}_{-i}}{h^{f_{i\ell^* k}}_{-i} + s^{f_{i\ell^* k}}_{-i}}.$$

8.3 Definition of wage variables

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

$$\overline{w_i} = \frac{\sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} \sum_{t=1}^{T_{i\ell j}} w^m_{i\ell jt}}{\sum_{\ell=1}^{L_i} \sum_{j=1}^{J_{i\ell}} T_{i\ell j}}$$

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

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

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

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

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

Analogous to $\widetilde{h}_{-i\ell j}$define:

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

8.4 Auxiliary Parameters

In general notice that a lot of the moments are calculated over different samples, since not all variables are defined for each job spell.

There are four parameters that are important for turnover: $\delta, \lambda^e, \lambda^n, f_1$

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

- δ : Average length of employment spell:

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

- λ^n Similarly we use the average length of the non-employment spell:

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

- λ^e : Average length of a job spell:

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

- f_1 : Use only spells where $s_{-i}^{q} + h_{-i}^{q} > 0$

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

  One can see the importance of excluding i from the calculation of $\widetilde{h}_{-ij}$ from this moment. As long as the number of workers at a firm is finite, there would be a correlation between the individual's revealed preference and the overall revealed preference of the firm. Since we do not model firm size we do not want to account for this. The parameter we used here has the property that one will get the same expected value of this parameter regardless of the number of coworkers.

For Wages we have 7 parameters: $E(\theta_i), var(\theta_i), cov\left(\widetilde{v}^u_{ij}, v^w_{ij}\right), var\left(v^w_{ij}\right), var(\xi_{it}), f_2, f_3$.

We use the moments:

- $E(\theta)$: Just use

$$\overline{w} \equiv \frac{\sum_{i=1}^{N}\sum_{\ell=1}^{L_i}\sum_{j=1}^{J_{i\ell}}\sum_{t=1}^{T_{i\ell j}} w^m_{i\ell jt}}{\sum_{i=1}^{N}\sum_{\ell=1}^{L_i}\sum_{j=1}^{J_{i\ell}} T_{i\ell j}}$$

- $var(\theta_i), var(\xi_i), var(v^w_{ij})$: We use the decomposition

$$\frac{\sum_{i=1}^{N}\sum_{\ell=1}^{L_i}\sum_{j=1}^{J_{i\ell}}\sum_{t=1}^{T_{i\ell j}} \left(w^m_{it} - \overline{w}\right)^2}{\sum_{i=1}^{N}\sum_{\ell=1}^{L_i}\sum_{j=1}^{J_{i\ell}} T_{i\ell j}} = \frac{\sum_{i=1}^{N}\sum_{\ell=1}^{L_i}\sum_{j=1}^{J_{i\ell}}\sum_{t=1}^{T_{i\ell j}} \left(w^m_{i\ell jt} - \overline{w_{i\ell j}}\right)^2}{\sum_{i=1}^{N}\sum_{\ell=1}^{L_i}\sum_{j=1}^{J_{i\ell}} T_{i\ell j}}$$
$$+ \frac{\sum_{i=1}^{N}\sum_{\ell=1}^{L_i}\sum_{j=1}^{J_{i\ell}}\sum_{t=1}^{T_{i\ell j}} \left(\overline{w_{i\ell j}} - \overline{w_i}\right)^2}{\sum_{i=1}^{N}\sum_{\ell=1}^{L_i}\sum_{j=1}^{J_{i\ell}} T_{i\ell j}} + \frac{\sum_{i=1}^{N}\sum_{\ell=1}^{L_i}\sum_{j=1}^{J_{i\ell}}\sum_{t=1}^{T_{i\ell j}} \left(\overline{w_i} - \overline{w}\right)^2}{\sum_{i=1}^{N}\sum_{\ell=1}^{L_i}\sum_{j=1}^{J_{i\ell}} T_{i\ell j}}$$

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

- f_2, f_3 : These two parameters govern $cov\left(\widetilde{\mu}^u_j, \mu^w_j\right)$ and $var(\mu^w_j)$. To pick up $var(\mu^w_j)$ we only use spells where $\widetilde{w}_{-ilj}$ are defined and use the moment

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

  where the means are taken over the appropriate sample. In the data it matters a little if we take the covariance or the product, since the mean over $\widetilde{w}_{i\ell j}$ is not zero when taking it over spells where $\widetilde{w}_{-i\ell j}$ is defined so we normalize things appropriately so it is like a covariance.

  For $cov\left(\widetilde{\mu}^u_j, \mu^w_j\right)$ we only use spells where $s^q_{-i} + h^q_{-i} > 0$ and use the moment

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

Again here, as in the moment above it matters in the data whether or not we do the covariance or the product and we again normalize things appropriately

- $cov\left(\widetilde{v}^u_{ij}, v^w_{ij}\right)$:

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

That is the fraction of times that job to job changes lead wages to fall.

9 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. Instead we focus on the decomposition of the the amount of total earnings variance into its components. This is done for model 1 in Table 5. First note that we get rid of measurement error. The total variance of log wages in the model is 0.124 it falls to 0.101 after we get rid of the measurement error. The results of each of the decompositions is shown in Table 5. 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). One sees that these give very similar results. Roy heterogeneity accounts for 85-89% of wage variation, compensating differentials for 10%, and search for 1-5%. Despite the issue, columns (C) is very interesting as it gives a very different result about search frictions. In this case Search frictions account for 25% of variation. The reason is that in columns (A) and (B) workers searching for better jobs were not only worried about wages, but about the nonpecuniary aspects of the jobs as well. Since in column (C) we eliminate compensating differentials first, now workers are only worried about wages and search frictions are much more important. This suggests that they were likely really important in column (A) and (B) as well, but we were mismeasuring things as they were important for “utility inequality” rather than wage inequality. We have not figured out a good way to quantify this yet.

We next estimate Model 2 using the same moments and present the results in Tables 6-8. They are quite different from Model 1. First one can see that this model does not fit the data as well. In particular, unlike Model 1, the wage distribution can not be pinned down exactly. Since the turnover parameters affect wages, this leads us to fit this distribution somewhat worse as well. The parameter estimates are presented in Table 7. Note that relative to model 1, but the measurement error and the variance of v_{ij}^{w} are much lower than in Model 1 (which should not be surprising).

We present the simulations in Table 8. For the base model we argued that eliminating compensating differentials was a strange counterfactual given the selection issue with the wages we observe. This is even more problematic in Model 2 because the wage one receives immediately after a non-employment spells depends on the flow value of non-employment. Since this is measured in utility measures, its not at all clear how to translate this to something comparable to a wage. While one could try various alternatives, we instead just chose to focus on the decompositions in which compensating differentials is left by itself. The results in Table 8 are very different than Table 5 as search frictions are roughly as important as Roy inequality. The difference between these results and those in Table 5 is that search frictions have two roles in this model: they affect which firm that employs you, but through wage setting mechanism, they also affect the amount of monopsony power that the firm has. This second role is absent in the first model, but is the driving force in the second.

10 Next Steps

Clearly Model 1 and Model 2 give very different results about the importance of Search frictions. Our next goal is to estimate a hybrid model along the lines of Cahuc and Robin (2006). They estimate a model with Nash bargaining where they estimate the Nash bargaining power. One can think of our Model 1 as a special case in which the worker has all of the bargaining power while in the second model the firm has all of the power. The main difference in the model is the amount of wage growth within a firm. In model 1 there is no growth, while in model 2 there is quite a bit. We will use this as the moment we use to pin down the bargaining power. Of course, it would be non-sensical to include wage growth on the job without allowing for human capital, so we will also allow for human capital loosely following Bagger, Fontaine, Postel-Vinay, and Robin (2007). Other features we plan to add are additional wage dynamics, shocks to tastes, conditions on industry and/or occupation, conditioning on schooling and gender, and adding more flexibility to various aspects of the model.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High Wage Workers and High Wage Firms. *Econometrica 67*(2), 251–333.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2007, February). A tractable equilibrium search model of individual wage dynamics with experience accumulation.

Becker, D. (2009). Non-wage job characteristics and the case of the missing margin. Federal Trade Commision.

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., P.-V. F. and J.-M. Robin (2006, March). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica 74*(2), 323–364.

Eckstein, Z. and G. V. 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.

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.

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.

Villanueva, E. (2007). Estimating compensating wage differentials using voluntary job changes: Evidence from germany. *Industrial and Labor Relations Review 60*(4), 544–561.

Technical Appendix

Base Model

Following the notation in the text, we assume firm j would pay worker i the wage W_{ij}. Letting Δ be a change in time, when Δ is small the value function takes the form

$$V_{ij}(W_{ij}) = \Delta U_{ij}(W_{ij}) + \frac{1}{1+\rho\Delta}\left[\Delta \sum_{\ell=1}^{J} \lambda_\ell^e max\{V_{ij}(W_{ij}), V_{i\ell}(W_{i\ell}) + \Delta\delta V_{i0} + (1-\Delta\Lambda^e - \Delta\delta))V(W_{ij})\right]$$

Now some algebra:

$$(\rho\Delta + \Delta\Lambda^e + \Delta\delta)\, V_{ij}(W_{ij}) = (1+\rho\Delta)\, \Delta U_{ij}(W_{ij}) + \Delta \sum_{\ell=1}^{J} \lambda_\ell^e max\{V_{ij}(W_{ij}), V_{i\ell}(W_{i\ell}) + \Delta\delta V_{i0}$$

Dividing by Δ and taking limits as $\Delta \to 0$,

$$\begin{aligned} &(\rho + \delta + \Lambda^e)\, V_{ij}(W_{ij}) \\ &= U_{ij}(W_{ij}) + \sum_{\ell=1}^{J} \lambda_\ell^e max\{V_{ij}(W_{ij}), V_{i\ell}(W_{i\ell}) + \delta V_{i0} \end{aligned}$$

For non-employment,

$$V_{i0} = \Delta U_{i0} + \frac{1}{1+\rho\Delta}\left[\Delta \sum_{\ell=1}^{J} \lambda_\ell^n max\{V_{i0}, V_{i\ell}(W_{i\ell})\} + (1-\Delta\Lambda^n)\, V_{i0}\right]$$

gives

$$(\rho + \Lambda^n)\, \Delta V_{i0} = (1+\rho\Delta)\, \Delta U_{i0} + \Delta \sum_{\ell=1}^{J} \lambda_\ell^n max\{V_{i0}, V_{i\ell}(W_{i\ell})\}$$

taking limits

$$(\rho + \Lambda^n)\, V_{i0} = U_{i0} + \sum_{\ell=1}^{J} \lambda_\ell^n max\{V_{i0}, V_{i\ell}(W_{i\ell})\}$$

Bertrand Competition

Now consider the Bertrand competition version of the model. This is very similar to Postel-Vinay and Robin (2002) with two major differences: workers are now maximizing utility rather than wages and we consider a finite number of firms rather than a continuum. We index j so that if $j > \ell$, $U_{ij}(\pi_{ij}) > U_{i\ell}(\pi_{i\ell})$. The key here is that

- When a firm hires a worker from nonemployment they will pay them a wage so that the worker is just indifferent between taking the job or not. This means that W_{ij0} is defined so that

$$V_{ij}(W_{ij0}) = V_{i0}.$$

- When a firm hires a worker directly from another firm, the Bertrand competition assumption means that the winning firm pays a wage to make the worker indifferent between accepting the offer, or taking the maximum offer from the alternative firm and staying there. That is if $\ell > j$,

$$U_{i\ell}(W_{i\ell j}) = U_{ij}(\pi_{ij}).$$

- The model is symmetric when an outside offer is matched, thus if $j \geq \ell$,

$$U_{ij}(W_{ij\ell}) = U_{i\ell}(\pi_{i\ell}).$$

Following the notation in the text, we assume that the most a firm will ever pay as a wage will be $w = \pi_{ij}$ and use the same definition of $\ell(w)$. If the value of a vacancy were zero this would just be the productivity of the job, more generally it is the flow value of the match. We start with a more general model in which we do not explicitly model wage determination other than assuming than defining the wage for worker i at firm j whose best outside option is $W_{ij\ell}$. We then go on to consider the Bertrand competition. Again let Δ be a small value of time,

$$\begin{aligned} V_{ij}(w) = \Delta U_{ij}(w) + \frac{1}{1+\rho\Delta}[\Delta \sum_{\ell=\ell(w)+1}^{j} \lambda_\ell^e V_{ij}(W_{ij\ell}) + \\ + \Delta \sum_{\ell=j+1}^{J} \lambda_\ell^e V_{i\ell}(W_{i\ell j}) + \Delta\delta V_{i0} + \big(1 - \left[\Lambda^e - \Lambda^e_{\ell(w)}\right] - \Delta\delta\big)) V(w)]. \end{aligned}$$

After some algebra

$$\begin{aligned} \left(\rho\Delta + \Delta\left[\Lambda^e - \Lambda^e_{\ell(w)}\right] + \Delta\delta\right) V_{ij}(w) = (1+\rho\Delta)\,\Delta U_{ij}(w) + [\Delta \sum_{\ell=\ell(w)+1}^{j} \lambda_\ell^e V_{i\ell}(W_{ij\ell}) + \\ + \Delta \sum_{\ell=j+1}^{J} \lambda_\ell^e V_{i\ell}(W_{i\ell j}) + \Delta\delta V_{i0}] \end{aligned}$$

Dividing by Δ and taking limits as $\Delta \to 0$,

$$\begin{aligned}
&\left(\rho + \delta + \left[\Lambda^e - \Lambda^e_{\ell(w)}\right]\right) V_{ij}(w) \\
&= U_{ij}(w) + \left[\sum_{\ell=\ell(w)+1}^{j} \lambda^e_\ell V_{i\ell}(W_{ij\ell}) + \sum_{\ell=j+1}^{J} \lambda^e_\ell V_{i\ell}(W_{i\ell j}) + \delta V_{i0}\right]
\end{aligned}$$

For non-employment, let ℓ_0 denote the first job one would take if offered from nonemployment, then

$$V_{i0} = \Delta U_{i0} + \frac{1}{1+\rho\Delta}\left[\Delta \sum_{j=\ell_0}^{J} \lambda^n_j V_{ij}(W_{ij0}) + \left(1 - \Delta\left[\Lambda^n - \Lambda^n_{\ell_0 - 1}\right]\right) V_{i0}\right]$$

gives

$$\left(\rho + \left[\Lambda^n - \Lambda^n_{\ell_0-1}\right]\right) \Delta V_{i0} = (1+\rho\Delta)\,\Delta U_{i0} + \left[\Delta \sum_{j=\ell_0}^{J} \lambda^n_j V_{ij}(W_{ij0})\right]$$

taking limits

$$\left(\rho + \left[\Lambda^n - \Lambda^n_{\ell_0-1}\right]\right) V_{i0} = U_{i0} + \sum_{j=\ell_0}^{J} \lambda^n_j V_{ij}(W_{ij0})$$

Now to figure out the Bertrand competition case, first evaluate this at $w = \pi_{ij}$ (and at this case $\ell(\pi_{ij}) = j$),

$$\left(\rho + \delta + \left[\Lambda^e - \Lambda^e_j\right]\right) V_{ij}(\pi_{ij}) = U_{ij}(\pi_{ij}) + \left[\Lambda^e - \Lambda^e_j\right] V_{ij}(\pi_{ij}) + \delta V_{i0}.$$

Note that we have made use of the assumption that when $\ell > j$, $V_{i\ell}(W_{i\ell j}) = V_{ij}(\pi_{ij})$. This expression simplifies to

$$V_{ij}(\pi_{ij}) = \frac{U(\pi_{ij}) + \delta V_{i0}}{\rho + \delta}$$

And for nonemployment:

$$\begin{aligned}
\left(\rho + \left[\Lambda^n - \Lambda^n_{\ell_0-1}\right]\right) V_{i0} =& U_{i0} + \sum_{j=\ell_0}^{J} \lambda^n_j V_{ij}(W_{ij0}) \\
=& U_{i0} + \left[\Lambda^n - \Lambda^n_{\ell_0-1}\right] V_{i0}
\end{aligned}$$

and thus

$$V_{i0} = \frac{U_{i0}}{\rho}.$$

Finally we need to solve for $W_{ij\ell}$. Bertrand competition implies that $W_{ij\ell^*}$is determined from

$$\left(\rho + \delta + \left[\Lambda^e - \Lambda^e_{\ell^*}\right]\right) V_{i\ell^*}(\pi_{\ell^*}) = \left(\rho + \delta + \left[\Lambda^e - \Lambda^e_{\ell^*}\right]\right) V_{ij}(W_{ij\ell^*})$$

So using our formulas for both sides

$$\begin{aligned}
&(\rho+\delta+[\Lambda^e-\Lambda^e_{\ell^*}])\,\frac{U_{i\ell^*}(\pi_{i\ell^*})+\delta V_{i0}}{\rho+\delta}\\
=&U_{ij}(W_{ij\ell^*})+\left[\sum_{\ell=\ell^*+1}^{j}\lambda^e_\ell V_{i\ell}(\pi_{i\ell})+\left[\Lambda^e-\Lambda^e_j\right]V_{ij}(\pi_{ij})+\delta V_{i0}\right]\\
=&U_{ij}(W_{ij\ell^*})+\left[\sum_{\ell=\ell^*+1}^{j}\lambda^e_\ell\frac{U_{i\ell}(\pi_{i\ell})+\delta V_{i0}}{\rho+\delta}+\left[\Lambda^e-\Lambda^e_j\right]\frac{U_{ij}(\pi_{ij})+\delta V_{i0}}{\rho+\delta}+\delta V_{i0}\right]\\
=&U_{ij}(W_{ij\ell^*})+\sum_{\ell=\ell^*+1}^{j}\lambda^e_\ell\frac{U_{i\ell}(\pi_{i\ell})}{\rho+\delta}+\left[\Lambda^e-\Lambda^e_j\right]\frac{U_{ij}(\pi_{ij})}{\rho+\delta}\\
&+V_{i0}\left[\frac{\delta\left[\Lambda^e_j-\Lambda^e_{\ell^*}\right]+\delta\left[\Lambda^e-\Lambda^e_j\right]+(\rho+\delta)\,\delta}{\rho+\delta}\right]\\
U_{ij}(W_{ij\ell^*})=&U_{i\ell^*}(\pi_{i\ell^*})-\sum_{\ell=\ell^*+1}^{j}\lambda^e_\ell\frac{U_{i\ell}(\pi_{i\ell})-U_{i\ell^*}(\pi_{i\ell^*})}{\rho+\delta}\\
&-\left[\Lambda^e-\Lambda^e_j\right]\frac{U_{ij}(\pi_{ij})-U_{i\ell^*}(\pi_{i\ell^*})}{\rho+\delta}
\end{aligned}$$

Similarly we can get W_{ij0} from

$$\begin{aligned}
\left(\rho+\delta+\left[\Lambda^e-\Lambda^e_{\ell_0-1}\right]\right)V_{i0}&=\left(\rho+\delta+\left[\Lambda^e-\Lambda^e_{\ell_0-1}\right]\right)V_{ij}(W_{ij0})\\
\left(\rho+\delta+\left[\Lambda^e-\Lambda^e_{\ell_0-1}\right]\right)\frac{U_{i0}}{\rho}&=U_{ij}(W_{ij0})+\sum_{\ell=\ell_0}^{j}\lambda^e_\ell\frac{U_{i\ell}(\pi_{i\ell})+\delta V_{i0}}{\rho+\delta}+\left[\Lambda^e-\Lambda^e_j\right]\frac{U_{ij}(\pi_{ij})+\delta V_{i0}}{\rho+\delta}+\delta V_{i0}
\end{aligned}$$

$$\begin{aligned}
U_{ij}(W_{ij0})&=\left(\rho+\delta+\left[\Lambda^e-\Lambda^e_{\ell_0-1}\right]-\left[\Lambda^e-\Lambda^e_{\ell_0-1}\right]\frac{\delta}{\rho+\delta}-\delta\right)\frac{U_{i0}}{\rho}-\sum_{\ell=\ell_0}^{j}\lambda^e_\ell\frac{U_{i\ell}(\pi_{i\ell})}{\rho+\delta}-\left[\Lambda^e-\Lambda^e_j\right]\frac{U_{ij}(\pi_{ij})}{\rho+\delta}\\
&=\left(\frac{\left(\rho+\delta+\left[\Lambda^e-\Lambda^e_{\ell_0-1}\right]\right)(\rho+\delta)}{\rho+\delta}-\frac{\left[\Lambda^e-\Lambda^e_{\ell_0-1}\right]\delta}{\rho+\delta}-\frac{\delta\,(\rho+\delta)}{\rho+\delta}\right)\frac{U_{i0}}{\rho}\\
&\quad-\sum_{\ell=\ell_0}^{j}\lambda^e_\ell\frac{U_{i\ell}(\pi_{i\ell})}{\rho+\delta}-\left[\Lambda^e-\Lambda^e_j\right]\frac{U_{ij}(\pi_{ij})}{\rho+\delta}\\
&=\left(\frac{\left[\rho+\delta+\Lambda^e-\Lambda^e_{\ell_0-1}\right]\rho}{\rho+\delta}\right)\frac{U_{i0}}{\rho}-\sum_{\ell=\ell_0}^{j}\lambda^e_\ell\frac{U_{i\ell}(\pi_{i\ell})}{\rho+\delta}-\left[\Lambda^e-\Lambda^e_j\right]\frac{U_{ij}(\pi_{ij})}{\rho+\delta}\\
&=U_{i0}-\sum_{\ell=\ell_0}^{j}\lambda^e_\ell\frac{U_{i\ell}(\pi_{i\ell})-U_{i0}}{\rho+\delta}-\left[\Lambda^e-\Lambda^e_j\right]\frac{U_{ij}(\pi_{ij})-U_{i0}}{\rho+\delta}
\end{aligned}$$

Table 1

Description of Sample Selection

	Number of workers	Number of Firms	Number of Spells	Number of Establishments	Fraction Employed
Merged data	4,954,649	454,395	59,057,162	496,154	
Labor market entry	4,954,649	420,383	49,379,610	459,154	
Delete under education	4,328,725	402,619	42,040,361	441,052	
Delete workers with missing educational variables	3,600,716	397,016	40,023,507	435,030	
Censoring at age 56	3,119,241	386,423	34,719,458	421,915	
Clean for temporary non-employment	3,119,241	386,423	28,363,330	421,915	
Censoring self-employment	2,927,279	370,264	25,778,040	404,184	
Delete workers with gabs	2,858,156	367,361	24,963,887	401,037	
Censor last year of observation	2,852,996	356,862	23,911,828	389,416	
In cross-section					
Merged data	3,576,263	129,384	3,576,263	165,361	0.58
Labor market entry	3,092,696	117,754	3,092,696	153,798	0.58
Delete under education	2,819,283	112,799	2,819,283	148,548	0.58
Delete workers with missing educational variables	2,691,062	111,608	2,691,062	147,282	0.59
Censoring at age 56	2,010,609	105,878	2,010,609	140,429	0.70
Clean for temporary non-employment	2,010,609	106,320	2,010,609	140,923	0.71
Censoring self-employment	1,767,064	100,246	1,767,064	134,519	0.77
Delete workers with gabs	1,726,538	99,460	1,726,538	133,623	0.77
Censor last year of observation	1,641,194	94,477	1,641,194	126,808	0.77

Table 2
Descriptive Statistics

	Mean	Std. Dev.	Min.	Max.
Number of years in sample	10.964	6.107	1	18
Number of firms per worker	2.658	1.831	1	17
Cross-section				
- Female	0.480		0	1
- Preparatory educations	0.341		0	1
- Highschool	0.039		0	1
- Vocational education	0.417		0	1
- Short further education	0.046		0	1
- Medium-length further education/Bachelor-degree	0.112		0	1
- Master-degree	0.042		0	1
- PhD degrees	0.002		0	1
- Age	38.576	9.860	15	55
- Public Employment	0.274		0	1
- Employed	0.776		0	1
- Missing Establishment ID	0.006		0	1

Table 3
Auxliary Model and Estimates for Model 1

Moment	Data	Model
Avg. Length Emp. Spell	259.9	260.0
Avg. Length Nonnemp. Spell	104.0	104.0
Avg. Length Job	116.0	116.0
$E(\widetilde{S}_{i\ell j}\widetilde{h}_{-i\ell j})$	-0.00292	-0.00292
Between Persons	0.0810	0.0810
Between Jobs	0.0277	0.0277
Within Job	0.0150	0.0150
Sample mean w_{it}	4.520	4.520
$\mathrm{E}(\widetilde{w}_{it}\widetilde{w}_{-it})$	0.00358	0.00357
$E(\widetilde{w}_{it}\widetilde{h}_{-it})$	0.00151	0.00151
Fraction Wage Drops	0.418	0.417

Table 4
Parameter Estimates Model 1

Parameter	Estimate
δ	0.166
λ^n	0.843
λ^e	1.913
f_1	1.019
E_θ	4.377
σ_θ	0.229
σ_ξ	0.151
f_2	-0.0112
f_3	0.161
$\sqrt{var(\nu_{ij}^w)}$	0.206
α	2.422

Table 5
Model 1 Decompositions

(A)			(B)			(C)			(D)		
Total	0.101		Total	0.101		Total	0.101		Total	0.101	
No Roy	0.011	89%	No Sear.	0.096	5%	No CD	0.094	7%	No C.D.	0.094	7%
No Sear.	0.097	1%	No Roy	0.097	85%	No Sear.	0.068	25%	No Roy	0.007	86%
C.D.	-	10%	C.D.	-	10%	Roy	-	68%	Search	-	7%

Table 6

Auxliary Model and Estimates for Model 2

Moment	Data	Model
Avg. Length Emp. Spell	260.0	251.4
Avg. Length Nonnemp. Spell	104.0	104.3
Avg. Length Job	116.0	116.9
$E(\widetilde{S}_{i\ell j}\widetilde{h}_{-i\ell j})$	-0.00292	-0.00285
Between Persons	0.0810	0.0810
Between Jobs	0.0277	0.0277
Within Job	0.0150	0.0150
Sample mean w_{it}	4.520	4.520
$\mathrm{E}(\widetilde{w}_{it}\widetilde{w}_{-it})$	0.00357	0.00349
$E(\widetilde{w}_{it}\widetilde{h}_{-it})$	0.00151	0.00105
Fraction Wage Drops	0.417	0.394

Table 7

Parameter Estimates Model 2

Parameter	Estimate
δ	0.173
λ^n	0.838
λ^e	1.662
f_1	1.023
E_θ	4.723
σ_θ	0.239
σ_ξ	0.0506
f_2	0.0323
f_3	0.156
$\sqrt{var(\nu_{ij}^w)}$	0.010
α	7.454

Table 8

Model 2 Decompositions

(A)			(B)		
Total	0.121		Total	0.121	
No Roy	0.063	48%	No Search	0.068	44%
No Search	0.009	45%	No Roy	0.009	49%
No CD	-	7%	No CD	-	7%

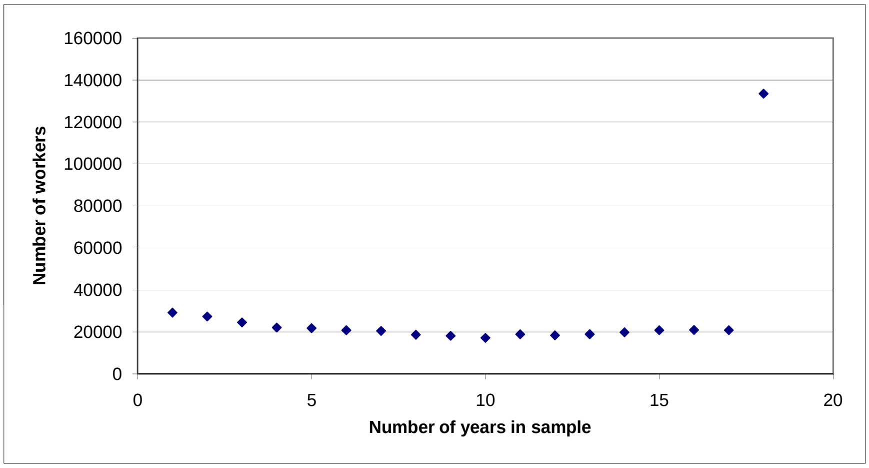

Figure 1: DISTRIBUTION OF NUMBER OF YEARS PER WORKER

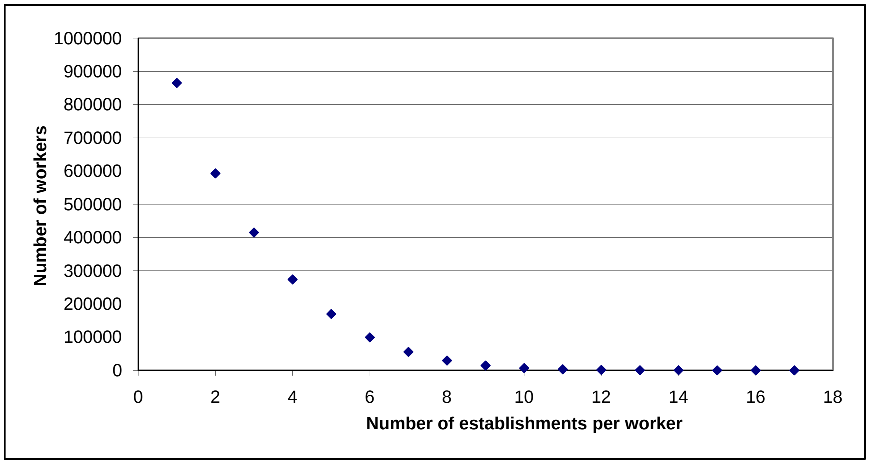

Figure 2: Distribution of Number of Establishments Per Worker

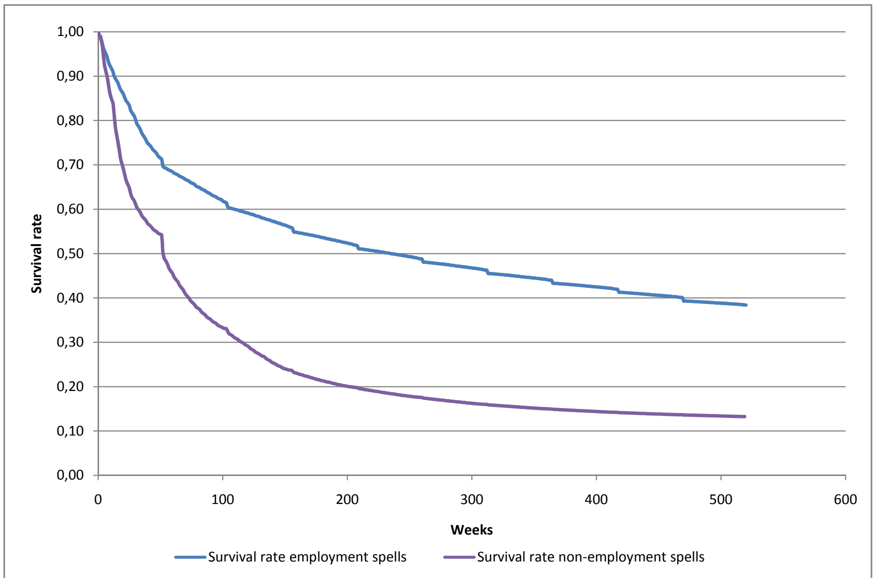

Figure 3: Survival Plots of Employment and Non-Employment Spells

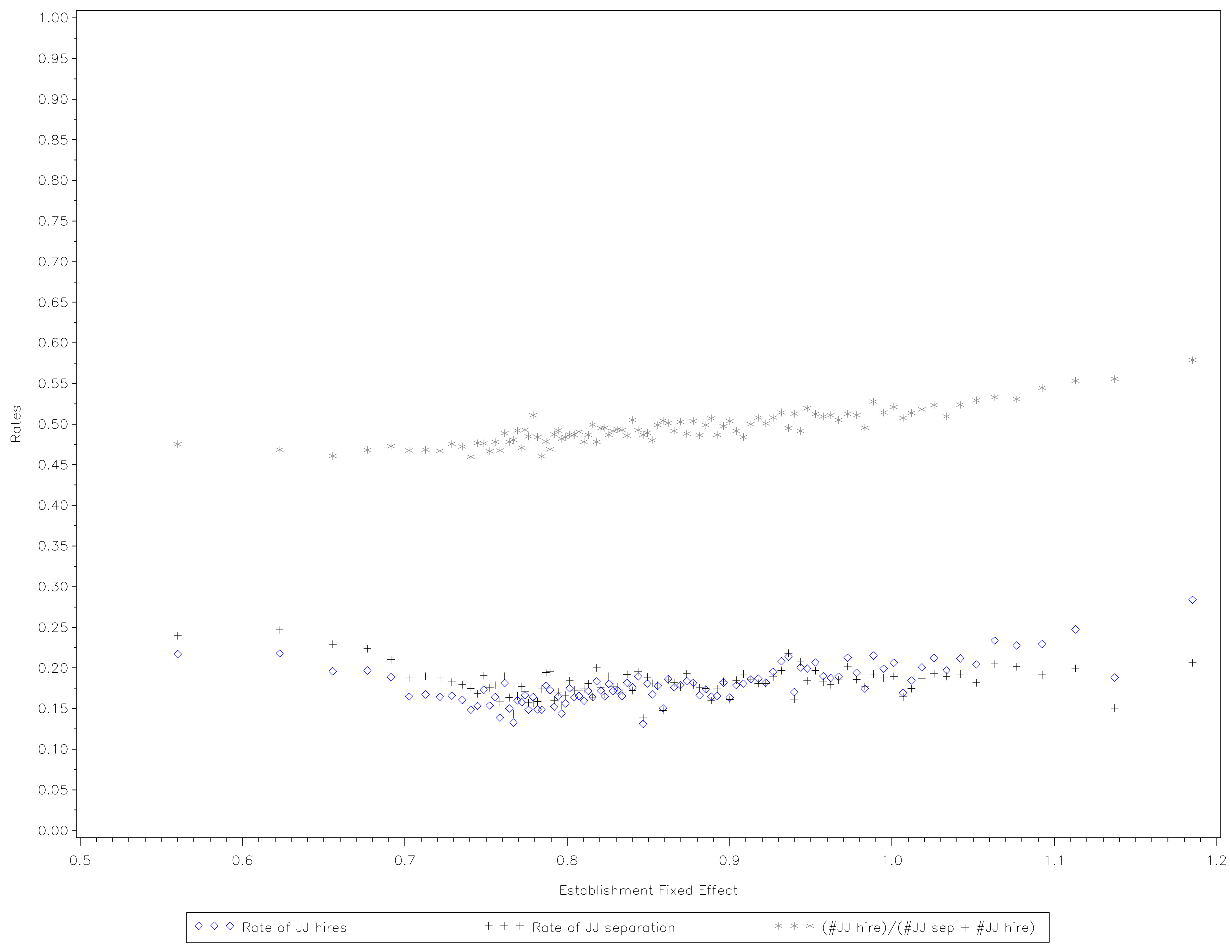

Figure 4: Hire and separation rates for percentiles of Firm Fixed effects