

NBER WORKING PAPER SERIES

ESTIMATION OF A ROY/SEARCH/COMPENSATING DIFFERENTIAL MODEL
OF THE LABOR MARKET

Christopher Taber
Rune Vejlin

Working Paper 22439
<http://www.nber.org/papers/w22439>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2016

We would like to thank seminar participants at a number of places for helpful comments and suggestions. We especially thank Joe Altonji, Jesper Bagger, Jim Heckman, John Kennan, Rasmus Lentz, and Greg Veramendi for comments and Carl Sanders for comments and research assistance. Taber thanks the NSF for support. Vejlin acknowledges financial support from the Danish Social Sciences Research Council (grant no. 09 - 066745). All remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Christopher Taber and Rune Vejlin. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Estimation of a Roy/Search/Compensating Differential Model of the Labor Market
Christopher Taber and Rune Vejlin
NBER Working Paper No. 22439
July 2016
JEL No. J3

ABSTRACT

In this paper we develop a model capturing key features of the Roy model, a search model, compensating differentials, and human capital accumulation on-the-job. We establish which features of the model can be non-parametrically identified and which can not. We estimate the model and use it to assess the relative contribution of the different factors for overall wage inequality. We find that Roy model inequality is the most important component accounting for the majority of wage variation. We also demonstrate that there is substantial interaction between the other features - most notably the importance of the job match obtained by search frictions varies from around 9% to around 29% depending on how we account for other features. Compensating differentials and search are both very important for explaining other features of the data such as the variation in utility. Search is important for turnover, but so is compensating differentials: 1/3 of all choices between two jobs would have resulted in a different outcome if the worker only cared about wages.

Christopher Taber
Department of Economics
University of Wisconsin -Madison
1180 Observatory Dr
Social Sciences Building #6448
Madison, WI 53706-1320
and NBER
ctaber@ssc.wisc.edu

Rune Vejlin
Aarhus University
and IZA
rvejlin@econ.au.dk

1 Introduction

It is well-known that variation in wages across observably similar workers is high. There are several competing theories as to why this is the case. Four of the most important models of post-schooling wage determination are 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 to investigate how they interact.

In a human capital model, workers who have accumulated more human capital while working earn more money than less experienced workers. The key feature 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. In the canonical Roy model, 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 wages. Finally, workers may just have had poor luck in finding their preferred job. Labor market frictions can lead to heterogeneity in wages for two different reasons. First, some workers may work for firms for which they are a better match and earn more money. The second is monopsony. In the framework we use, the bargaining power of the worker depends on their outside option so two equally skilled workers at the same firm may earn different wages. In short, one worker may have higher wages than another because a) he has accumulated more human capital while working (human capital), b) he has more talent at labor market entry (Roy model), c) he has chosen more an unpleasant job (compensating differentials), or d) he has had better luck in finding a good job and receiving outside offers (search frictions). The goal of this work is to uncover the contribution of these different components and how they interact to produce overall wage inequality.

Combining these four prominent theories into one coherent framework and investigating the relative importance in various dimensions is important not only from a theoretical point of view but also from a practical standpoint. An understanding of the underlying causes of inequality is essential for policymakers who wish to decrease wage inequality. Perhaps it is even important for deciding whether or not decreasing wage inequality is a major concern. If the primary driver behind wage inequality is compensating differentials inequality might be less problematic since it simply reflects different preferences for jobs. If the primary driver is Roy model heterogene-

ity (productivity differences), search frictions, or human capital then decreasing wage inequality might seem like a more appropriate action. However, the best way to do it depends on the relative importance of the different channels - if the primary source is Roy model inequality that would suggest focusing on pre-market factors (like education), but for the other two post labor market policies would be more effective.

In this paper we develop a structural model of wage determination that contains elements of all four models. The model is estimated on Danish matched employer-employee data from 1985-2003. 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 contributors to overall wage inequality, the Roy model inequality is the most important accounting for between 64% and 83% of wage inequality. The magnitude of the effects depend on the way we implement the decomposition as there are important interactions between the different components of the model. Most striking is search. The monopsony component of search frictions explains roughly 8% of wage variation. The importance of the direct source of search frictions (workers work at different firms due to frictions) varies from around 1% to around 21%.

While search and compensating differentials are less important for wage inequality than the Roy model, we show that they are both essential for explaining other features of the data. For example, they strongly impact the level of wages as we show that without them workers would get 0.20 and 0.17 log points higher wages, respectively. As another example, rather than looking in variation in wages we look at variation in utility. Together, compensating differentials and search frictions explain the bulk of of the variation in utility.

Also, compensating differentials are important in determining job choices. Roughly 1/3 of all choices would be different if workers only cared about wages and not the non-pecuniary aspect of a job. We conclude that all four of these features are important aspects of the labor market in Denmark and should be included for a full understanding of wage inequality and job turnover.

The paper proceeds as follows. 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 and detailed information on job-to-job transitions. We describe the data and institutional features of the Danish labor market in section 6. Section 7 presents the auxiliary model that we use. Section 8 presents the results and section 9 concludes.

2 Relation to Other Work

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. See Roy (1951), Heckman and Sedlacek (1985), Heckman and Honoré (1990), or Heckman and Taber (2008) for a discussion about the Roy model. Rosen (1987) provides an excellent discussion of compensating differentials models and Eckstein and Van den Berg (2007), Rogerson, Shimer, and Wright (2005), and Roberson and Shimer (2011) provide a nice discussion of empirical search models. For human capital models see Weiss (1986) or Heckman, Lochner, and Todd (2006).

Rather than a broad discussion, we focus on the relationship between our work and a few other important key papers from the different literatures.

Two of the most important related literatures were started by Abowd, Kramarz, and Margolis (1999) and Postel-Vinay and Robin (2002). Both of these papers use 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

$$\log(W_{it}) = X'_{it}\beta + \theta_i + \mu_{j(i,t)} + \zeta_{it} \quad (1)$$

where i indexes an individual, $j(i, t)$ indexes a firm, and t indexes time. θ_i is an individual fixed effect, μ_j is a firm fixed effect, and ζ_{ijt} is independent and identically distributed. We use this as a motivation for our auxiliary model.

Abowd, Kramarz, and Margolis (1999) and the literature that has arisen around this methodology all find that variance in firm effects play an important role for wages. Important examples are Sørensen and Vejlin (2013), Andrews, Gill, Schank, and Upward (2008), Gruetter and Lalive (2004), Card, Heining, and Kline (2013), and Card, Cardoso, and Kline (2014).

The fact that some firms are able to survive while paying lower wages is often attributed to search frictions. However, in our model both search frictions and compensating differentials could result in differences in average wages across firms. Since we do not model the job formation process from the firm's point of view, we can not answer this question explicitly, but we can shed light on the nature of firm wage components from the perspective of the workers.

Postel-Vinay and Robin (2002) decompose wage inequality into a search component, a firm productivity component, and an ability related component. As such their research question is very

much related to ours. They find that for skilled workers, the individual component is moderately important (explains up to 40% of wage inequality), but that for low skill workers virtually all of the inequality can be assigned to search frictions and differences in firm productivity. The main components that are in our model but not in theirs are human capital, compensating differentials and comparative advantage in jobs. Postel-Vinay and Robin allow for absolute advantage only in the sense that ability is one dimension and the relative productivity of two workers does not vary across firms. By contrast, we allow for a match specific productivity meaning that some workers match better with some firms (the Roy component of our model). 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.

Several papers build on the framework developed by Postel-Vinay and Robin (2002). Cahuc, Postel-Vinay, and Robin (2006) extend it to account for a more general bargaining framework and Bagger, Fontaine, Postel-Vinay, and Robin (2014) allows for the accumulation of general human capital while working, see also Lentz (2010), Lise, Meghir, and Robin (2013), Robin (2011) for other applications. Our paper can most clearly be seen as building on this literature by adding non-pecuniary aspects of jobs and much more heterogeneity in skills.

The idea about compensating differentials date all the way back to Smith (1776). More recently, another important related literature was started by Keane and Wolpin (1997). They estimate a model where they introduce compensating differentials into a model of human capital, and Roy model inequality. Similar to us, they find that initial conditions on labor market entry are the main drivers of earnings inequality. However, their model is very different in that they do not explicitly incorporate search frictions and do not make use of matched employer-employee data as their focus is on occupations rather than firms. Becker (2009) uses a framework similar to ours in that it incorporates compensating differentials into a search models. We build on this work by incorporating more Roy flexibility and by using matched employer-employee data, which allows us to perform the decompositions which are the focus of our paper. Becker (2009) focuses more on unemployment insurance than wage inequality. Dey and Flinn (2005) and Dey and Flinn (2008) estimate search models with a particular type of non-wage characteristics: health insurance.

There is a vast literature on human capital models. Most related to our work is that by Huggett, Ventura, and Yaron (2011). They estimate a life-cycle model with idiosyncratic shocks to human capital, heterogeneity in ability to learn, initial human capital, and initial wealth. Their focus is quiet different from our in that they are interested in what the driver is behind differences in life-

time earnings, utility, and wealth. They find that most of the variation is due to differences in initial conditions. This is in accordance with our finding that most of the differences in instantaneous wage variation is due to differences in worker ability.

Closer to our setup is Sullivan (2010). This paper includes elements of all four components of our model above, though this is not the main focus. We build on this work by including a more complete search model with endogenous wage formation and by estimating it on matched employer-employee data (which, as said above, is essential to identify our counterfactuals). Other than that, our models are quite different. Sullivan and To (2014) 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 To (2014) 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. Also their model does not include human capital. Finally, Sorkin (2015) distinguishes between search and compensating differentials using a similar type of revealed preference argument as we do. Other than that, our models are very different as he specifies the wage function in a more reduced form manner and focuses on the distinction between search frictions and compensating differentials on understanding the firm fixed effect in a wage regression and mostly abstracts from comparative advantage and human capital.¹ By contrast we use a firm random effect approach and focus on wage inequality.

3 The Model and Decomposition

The model is in continuous time and wages are determined similarly to Cahuc, Postel-Vinay, and Robin (2006), Dey and Flinn (2005), and Bagger, Fontaine, Postel-Vinay, and Robin (2014).² We formally treat agents in our model as infinitely lived though in practice we think of it as a life-cycle

¹He also uses LEHD data which is quarterly data. While his paper studies a larger country, this data makes distinguishing job-to-job transitions from job-to non-employment to-job transitions much more difficult than in our data.

²Note, 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.

model where the date of retirement is sufficiently far away that it can be ignored. A substantive difference between our paper and most of the search literature is that we assume that there are finite number of job types indexed $j = 1, \dots, J$ with $j = 0$ denoting non-employment. In the data, job types correspond to establishments. We assume the economy consists of a very large number of potential employers of each type. In estimation this is similar to the finite number of types of individuals as is common in much of the structural labor literature, (see e.g. Heckman and Singer, 1984, or Keane and Wolpin, 1997). Job offers from each type of job arrive at rate λ_j^n for non-employed workers and λ_j^e for employed workers. We allow for worker heterogeneity in the job destruction rate, δ_i .³

We do not allow for borrowing or lending. Worker i 's flow utility from working at job j with wage W is $u_{ij}(W)$. The fact that this depends on j is an important aspect of our model that accommodates compensating differentials: workers care about jobs above and beyond the wage that they earn or expectations about future wages.

Human capital is completely general and takes on a discrete set of values ψ_0, \dots, ψ_H . When individuals are employed, human capital appreciates randomly to the next level (ψ_h to ψ_{h+1}) at rate λ_h and does not accumulate when people are not working.⁴ We let $\pi_{ij}\psi_h$ be the productivity of worker i at job type j when the worker has human capital ψ_h . Notice, that unlike many models which assume either additive or multiplicative production functions we allow π_{ij} to be completely general. 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 rental rate R that is fixed until the next negotiation which only takes place by mutual consent. This means that when human capital is augmented the wage is not renegotiated but automatically rises from $R\psi_h$ to $R\psi_{h+1}$.⁵ We treat production on the job as if it is linear, so hiring worker i does not crowd out hiring of other workers (or affect the productivity of current workers), so the value of a vacancy is zero and the value of the match is $\pi_{ij}\psi_h$.

After a job is destroyed some individuals immediately receive an offer. Specifically, with probability P^* the worker immediately receives a job offer which is drawn from the same distribution of jobs as for non-employed workers. The worker can either accept it or reject it and enter non-employment. Including this into the model reconciles involuntary job-to-job transitions which are

³We discuss below why we allow arrival rates to depend only on j and destruction rates to depend only on i .

⁴Allowing human capital to depreciate when out of the labor force could easily be imbedded into this framework though we think it would change our counterfactuals very little.

⁵Another natural way to do this would be to assume that the absolute wage is agreed upon and that the return to human capital only accrues when workers receive outside offers. This would be similar in nature as Postel-Vinay and Turon (2010). Human capital does not play a major role in this analysis and we expect the alternative contracting to give very similar results though be somewhat messier to implement.

seen in the data.

Following Bagger, Fontaine, Postel-Vinay, and Robin (2014), a key aspect of this model is that when a worker receives an outside offer then 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. Renegotiation requires mutual consent, so as long as the employment relationship continues, the rental rate is kept fixed until both parties agree to renegotiate it.⁶

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 non-employed have flow utility U_{i0h} and value function V_{i0h} . We let V_{i0h}^* denote the value function immediately after a match is destroyed. The difference between V_{i0h}^* and V_{i0h} is that the former incorporates the possibility of receiving an offer immediately.

If a non-employed worker i with human capital ψ_h receives an offer from job type j a match is formed if $V_{ijh}(\pi_{ij}) > V_{i0h}$. Note, that this is efficient in the sense that joint surplus between workers and firms are maximized. A major issue is the determination of the wage in the match. Both the worker and the firm would prefer the potential rental rate R to not forming the match as long as $V_{i0h} \leq V_{ijh}(R) \leq V_{ijh}(\pi_{ij})$. The issue is that there are many such values of R . We denote the equilibrium rental rate as $R_{ij\ell h_0}$ where it depends on worker i , the current establishment j , the best outside option ℓ , and units of human capital, h_0 , at the negotiation time. We assume that it is defined by the solution to

$$V_{ijh_0}(R_{ij\ell h_0}) = \beta V_{ijh_0}(\pi_{ij}) + (1 - \beta) V_{i0h_0} \quad (2)$$

where β is the worker's bargaining power.⁷ Since the bargaining position for a worker who has just been laid off is non-employment, this will also be the wage for workers who experience a match destruction, but are then immediately hired by a new firm. Note that when $\beta = 1$ the worker has all of the bargaining power and extracts full rent $R_{ij\ell h_0} = \pi_{ij}$. When $\beta = 0$ the firm has all of the bargaining power and pays the value of $R_{ij\ell h_0}$ that makes the worker indifferent between accepting the offer or staying non-employed.

Now suppose that worker i is working in job type j and receives an outside offer from job type ℓ . As in Postel-Vinay and Robin (2002), one of three things can happen. First, the new job offer

⁶In 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 non-employment to the current job under some circumstances. However, this will not happen in our empirical specification so we abstract from it here.

⁷We do not explicitly derive this from a bargaining game as do Cahuc, Postel-Vinay, and Robin (2006), but treat it 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.

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 job and the new rental rate $R_{ij\ell h_0}$ will be determined by

$$V_{i\ell h_0}(R_{ij\ell h_0}) = \beta V_{i\ell h_0}(\pi_{i\ell}) + (1 - \beta)V_{ijh_0}(\pi_{ij}). \quad (3)$$

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

$$V_{ijh}(R_{ij\ell h_0}) = \beta V_{ijh_0}(\pi_{ij}) + (1 - \beta)V_{i\ell h_0}(\pi_{i\ell}). \quad (4)$$

Since renegotiation only happens when both parties agree, the wage is renegotiated when $R_{ij\ell h_0}$ is higher than the workers current rental rate. If it is lower the worker will ignore the offer. This condition is relatively straight forward to check. Let $R_{ij\ell h_0}$ be the current rental rate. If $V_{ijh}(R_{ij\ell h_0}) < V_{i\ell h}(\pi_{i\ell})$ the worker will want to renegotiate. Note that we have been a bit sloppy with notation as we use the notation $R_{ij\ell h_0}$ to denote the rental rate that worker i with human capital ψ_{h_0} at the time of negotiation would receive from job type j when their outside option was job type ℓ . As one can see from equations (3) and (4), it will be the same regardless of whether they started at job type ℓ and moved to j or if they started at j and then used an outside offer from job type ℓ to renegotiate their wage.

To solve the model we need to calculate the value functions $V_{ijh}(R)$ and V_{i0t} as there are no easy closed form solutions for the wage as in Cahuc, Postel-Vinay, and Robin (2006) or Bagger, Fontaine, Postel-Vinay, and Robin (2014).

It is convenient to define

$$\Lambda_{ijh}^e(R) \equiv \sum_{\{\ell: V_{ijh}(R) < V_{i\ell h}(\pi_{i\ell})\}} \lambda_{\ell}^e$$

$$\Lambda_{i0h}^n \equiv \sum_{\{\ell: V_{i\ell h}(\pi_{i\ell}) > V_{i0h}\}} \lambda_{\ell}^n$$

these are the sums of arrival rates that will lead to some reaction, either renegotiation or switching job, for employed workers and non-employed workers, respectively. Thus, for worker i with human capital h who is currently employed at job type j with rental rate R this is the arrival rate of some outside offer that will change behavior.

We can write the value function for worker i with human capital h who is currently employed at job j with rental rate R as

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

Consider the different components on the right hand side of this equation. The first, $u_{ij}(R\psi_h)$, is the flow utility that the worker receives until something happens. The second component denotes outside offers that will lead the worker to renegotiate their wage but ultimately stay at the current job. If $V_{i\ell h}(\pi_{i\ell}) \leq V_{ijh}(R)$ the outside offer will not be useful for renegotiating and if $V_{ijh}(\pi_{ij}) < V_{i\ell h}(\pi_{i\ell})$ then the worker will leave to the next job. The component in brackets represents the value function of the renegotiated wage as described in equation (4). The next term denotes outside offers that lead the worker to leave the current job. Again, the term in brackets denotes the value function under the negotiated wage as described in equation (3). The next two terms $\delta_i V_{i0h}^*$ and $\lambda_h V_{ijh+1}(R)$ represent respectively the events in which the worker is laid off and in which human capital augments.

When $h = H$ we get an expression that is identical except that it no longer contains the possibility of human capital augmenting

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

There is only one way for the status to change following a non-employment spell - the worker can take a job. This leads to the simpler formulation

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

The first term is the flow utility and the second denotes the outcome in which an offer is received that dominates non-employment. The term in brackets represents the value function under the renegotiated rate.

Finally, the value function for workers immediately after their match is destroyed is

$$V_{i0}^*(h) = P^* \frac{\sum_{\{\ell: V_{i\ell}(\pi_{i\ell}) > V_{i0h}\}} \lambda_{\ell}^n V_{i\ell h}(R_{i\ell 0h})}{\sum_{\ell} \lambda_{\ell}^n} + \left(1 - P^* \frac{\Lambda_{i0t}^n}{\sum_{\ell} \lambda_{\ell}^n}\right) V_{i0h}$$

The first term is result of an acceptable offer while the second is the result from either no offer or an unacceptable offer.

This is the full model. Notice, that unlike many other search models, job ladders in this model are individual specific. This is both due to comparative advantages and tastes for non-pecuniary aspects of the job. This makes the model computationally harder to solve than e.g. Bagger, Fontaine, Postel-Vinay, and Robin (2014). Obviously, 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.

Next consider a decomposition that allows us to understand the various components. We can choose any measure of wage inequality (we later use 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. The mean of $u_{ij}(\cdot)$ will vary across jobs, so some jobs on average offer a higher utility level.
- λ_j^e and λ_j^n incorporate search frictions - note that these affect wage inequality in two different ways: directly through the job at which one takes, and indirectly through the bargaining process.
- Variation in ψ_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 the order in which we perform the decomposition matters. This is not surprising because the different sources interact.

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 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) Eliminate variation from human capital. We can do this by setting $\lambda_h = \infty$ so that workers obtain their maximum human capital immediately.
- c) Eliminate variation coming from heterogeneity in monopsony power through renegotiated wages. We can do this by setting $\beta = 1$ so that workers get paid their productivity immediately.
- d) Eliminate heterogeneity coming from worker luck in the job offers they have received. We do this by setting λ_j^e and λ_j^n arbitrarily high so that workers move to the preferred job immediately.
- e) To eliminate Roy model inequality we set rental rates to be $R_{ij} = E(R_{ij} | j)$ for all j eliminating variation due to skill differences so that job types pay constant wages (but holding workers ordering of jobs constant).
- f) Eliminate inequality based on compensating differentials. We do this by assuming individuals choose jobs to maximize wages only.

The difference between a and b is due to human capital, the difference between b and d is due to search frictions (with b-c the monopsony part and c-d the match part), the difference between d and e is due to Roy model inequality, and the remaining fraction in e is due to compensating differentials. After implementing step f, all workers earn identical wages.

Notice, that we have estimated a model that tends to overstate the importance of search frictions when we take it to the data. First, preferences are stable. This means that job-to-job transitions only result from search frictions. Secondly, human capital is completely general. Specific human capital will play a very similar role as bargaining in our model in terms of increasing wages and in fact we use features of the returns to tenure in order to identify it.

One can also see the importance of matched employer-employee data. When we eliminate Roy model inequality we eliminate the match specific component of wages but allow for a job type level component.

Notice, that even though we have endogenous wage determination, the model does not endogenize job offer arrival rates nor the distribution of job types. This is important to keep in mind later when we do the decompositions - they are essentially partial equilibrium simulations. Allowing for these equilibrium effects would be an interesting and important extension of our work.

4 Identification

In this section we discuss non-parametric identification of our model. The details are given in Appendix A. We show which aspects of the model can and cannot be identified. Both parts are important. This is relevant in that we can not credibly simulate counterfactuals that are not identified from the data. We respect this in our main counterfactual exercises below by simulating counterfactuals that we have shown are identified (except in a couple of cases where we note it explicitly).

Proving general non-parametric identification of the model when the number of job types 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 . We fully expect our result to generalize to larger (but finite) J .⁸ One caveat to this argument is that the fact that we know there are only two job types and can observe the type proves useful for identification of some terms. Clearly, we can not literally operationalize this when we go to the data but rather take the finite job types as an approximation. In the empirical specification below we will recognize this issue and normalize the model appropriately. For example, with two job types we can identify λ_A^n , but we know from Flinn and Heckman (1982) that it is not non-parametrically identified from the reservation utility with a continuum of job types so we normalize the reservation utility.⁹ We also simplify to the case in which human capital only takes on two values, $h = \{0, 1\}$.

We first consider what can be identified without data on wages. Throughout this section we assume that we observe workers from the beginning of their working career in which they start non-employed.¹⁰ In this setting we can partition people into types depending on their preferences for jobs and non-employment. In a more general model the preference order would change with human capital accumulation, but we assume it does not to simplify the already complicated specification.¹¹ For example, one type prefers A to B and B to non-employment. This yields a total of 5 types: $A \succ B \succ 0$, $B \succ A \succ 0$, $A \succ 0 \succ B$, $B \succ 0 \succ A$, and non-workers: $0 \succ \max(A, B)$, where, for example, $A \succ B \succ 0$ denotes A preferred to B , B preferred to non-employment.

In Appendix A we show that from the observed job choices and the timing of transitions we can identify the transition parameters,

$$\lambda_A^n, \lambda_B^n, \lambda_A^e, \lambda_B^e, P^*$$

⁸Though of course the length of the panel we would need for identification would also increase.

⁹Barlevy (2008) provides a nice example of how to identify this type of model in the more complicated setting.

¹⁰In the data this is not true for all workers. However this is not relevant for identification in that we can focus only on those workers who we observe at the beginning of their careers.

¹¹This will also be imposed in our empirical specification.

the distribution of δ_i , and the population proportions of each of the five different types.

Identification crucially depends on two types of arguments 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 voluntarily leave the first job to start the second.¹² As a result it will be very important for us to distinguish job-to-job transitions (in which we can sometimes use our revealed preference argument) from job-to non-employment to-job transitions where we can not use this argument. Intuitively, if we consistently observed that workers were willing to take wage cuts to go to a certain establishment, this would indicate that this establishment had high non-pecuniary benefits. A limitation of this approach is that we do not observe whether a job-to-job transition is voluntary. In the model we allow for some job-to-job transitions to be involuntary through P^* . We show that P^* can actually be identified directly from standard data by the rates at which people experience job-to-job transitions. However in the estimation we will augment our standard data with additional data in which we directly observe the fraction of job-to-job transitions that are voluntary. We prefer to use the direct evidence over the more model specific identification shown in the appendix.

A second key aspect of the model is that we show that the arrival rates for employed (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 job is because of search frictions, then eventually they should match with the highest paying job. 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 job.¹³ Thus, it is important to have high quality panel data on job switching and also matched employer-employee data.

We next incorporate information from wages. Let w_{it} denote the log of the wage measured at time t if the worker is working. To incorporate the main features of the Danish data, we assume that we only observe wages a finite number of times. For simplicity assume it is at the integers ($t = 1.0, 2.0, \dots$). Without loss of generality we can normalize the initial level of human capital $\psi_0 = 1$. In the appendix we show that we can identify the wage distribution for each type in each state of the world in which we might observe that type.

As an example focus on those individuals who prefer job A to job B and job B to non-employment

¹²Other recent papers using job-to-job transitions as a revealed preference are Sorokin (2015) and Bagger and Lentz (2015).

¹³It 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 Roy model heterogeneity. However, the type of search friction we have modeled in this paper is the most common type in equilibrium search models such as Burdett and Mortensen (1998) or Postel-Vinay and Robin (2002).

($A \succ B \succ 0$). For each level of human capital there are seventeen possible labor market statuses depending upon current employment, current human capital, the workers best outside offer, and the level of human capital when the wage was negotiated. The wages for these workers depend on the human capital arrival rate, λ_h , and ψ_1 , the distribution of measurement error, and the eight dimensional set of wages

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

where $R_{ij\ell h_0}$ is the rental rate for worker i in the current job j with the best outside option ℓ who had h_0 units of human capital at the negotiation time. We show that we can identify the joint distribution of the wages for this type of worker. This requires panel data on wages for which we can observe the same worker with different employers. The employment status is not known with certainty but we can still identify this distribution by varying the timing of transitions. We show that similar types of expressions can be identified for the other types of workers. This is much simpler for workers that only work at one type of firm.

There are three important parts of our model that are not identified. We can not hope to identify the wage a worker would receive on a job they would never take. For example, consider the workers who prefer A to non-employment to B . They would never take a B type job so we can not identify the wage they would have earned in such a job. For these workers we can only identify the joint distribution of

$$(R_{iA00}, \pi_{iA}, R_{iA01}).$$

While obvious at some level, 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 at infinite problem discussed in Heckman (1990) and Heckman and Honoré (1990). Without an exclusion restriction moving the conditional probability of doing each job to one, we can not identify the full unconditional distribution of wages in that job.

Second, we explicitly show in the appendix that we can not non-parametrically identify the bargaining parameter β . The basic intuition is that we use a revealed preference argument to identify the preference workers have across jobs (i.e. the population distribution of the preference types) which identifies ordinal utility but not cardinal utility. We can not distinguish the magnitude of β from the magnitude of the strengths of preferences for jobs. As an example, consider the case above for workers that would only take job A and assume no human capital. There are only

two relevant labor market statuses for this group: a) they have received one offer from a type A job and b) they have received 2 or more offers from type A jobs within the employment spell. In the second state of the world, since two A firms have bid for them, they will receive a wage equal to their productivity at job A which is π_{iA} . In the other state of the world we can identify their wage and the difference between it and their productivity (π_{iA}). As long as they are different we know that $\beta < 1$. However, that wage depends on two different things; the bargaining parameter, β , and the difference in utility between not working and working at firm A . For any β , we can find some set of preferences that can reconcile the wage. We show in the appendix that this identification problem extends to more complicated cases so that β is not identified. In the empirical model, we fix this issue not by setting β ex-ante but by setting the intensity of preferences by assuming that the variance of the preference error terms is unity. A different value would yield a different β .

Third, we can identify the wages individuals would earn at each job and their preference ordering across jobs. However, we cannot identify how the preference ordering would change if productivity changed. This is closely related to the non-identification result for β . Since revealed preferences can not identify the relative importance of wages versus non-pecuniary aspects of jobs, we cannot identify how job orderings would change if we change productivity and/or wages. Presumably one could identify this by using wage changes at different establishments as an exclusion restriction, but this information is not needed for the counterfactuals we estimate.

It is important to point out that even though we can not identify β non-parametrically, we can perform two interesting counterfactuals. Completely eliminating search friction will lead people to be paid their productivity at their preferred job - this is non-parametrically identified. We will also simulate a counterfactual where, we get rid of inequality arising from the bargaining process (the difference between the wage in status a and status b above). We implement this by setting $\beta = 1$ and this counterfactual is identified because the worker will be paid π_{ij} in that case.

5 Econometric Specification/Parameterization

Even though the model is mostly non-parametrically identified, estimating it non-parametrically is not an option. In this section we present our empirical specification, where we try to be flexible. We assume that log productivity of individual i at job type j is specified as

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

where θ_i is the same for individuals at all jobs, μ_j^p is the same for all individuals at job j and v_{ij}^p is the match specific component. Thus, we allow for worker and firm heterogeneity in productivity as well as match (worker-firm) specific productivity.

The flow utility for individual i at job type j with human capital rental rate R is

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

where α is the weight workers put on log consumption compared to the non-pecuniary aspects of a job. What matters for occupational choice will be the utility evaluated at $\pi_{ij}\psi_h$. We can rewrite this flow utility as the sum of three terms

$$\begin{aligned} u_{ij}(\pi_{ij}\psi_h) &= \alpha \left(\theta_i + \mu_j^p + v_{ij}^p \right) + \alpha \log(\psi_h) + \mu_j^u + v_{ij}^u \\ &= \alpha(\theta_i + \log(\psi_h)) + \left(\alpha\mu_j^p + \mu_j^u \right) + \left(\alpha v_{ij}^p + v_{ij}^u \right). \end{aligned}$$

The first term is common across jobs so all that matters for job-to-job turnover are the latter two terms. The second term pertains to firm specific factors and the last term pertains to match specific factors. Log utility has a special feature that when human capital increases the income and substitution effects balance so that the taste across jobs does not change. This greatly simplifies the computation and given that human capital takes a relatively minor role, likely makes little difference to the ultimate results.¹⁴

We assume that θ_i is normally distributed with mean E_θ and variance σ_θ^2 and that we observe log wages with classical i.i.d. normally distributed measurement error ξ_{it} with mean zero and variance σ_ξ^2 .

The joint distributions (μ_j^p, μ_j^u) and (v_{ij}^p, v_{ij}^u) are ex-ante independent of each other.¹⁵ We set $var(v_{ij}^u) = 1$.¹⁶ Moreover, we assume that (v_{ij}^u, v_{ij}^p) are jointly normal distributed which gives us two parameters, $cov(v_{ij}^u, v_{ij}^p)$ and $var(v_{ij}^p)$. However, note from the equation above that the part of the idiosyncratic error term that is relevant for decisions is $(\alpha v_{ij}^p + v_{ij}^u)$. Thus all that matters for the model in determining wages is the joint distribution of v_{ij}^p and $(\alpha v_{ij}^p + v_{ij}^u)$. As a result, we can not separately identify $cov(v_{ij}^u, v_{ij}^p)$ from α , so we normalize $cov(v_{ij}^u, v_{ij}^p) = 0$ and estimate α along with σ_{v^p} , the standard deviation of v_{ij}^p .¹⁷

¹⁴We are able to solve the model without the assumption of log utility, but it complicates the numerical problem. Also the income effect associated with human capital would be quite different if we allowed workers to borrow against future human capital growth. It also might be quite different if human capital was not perfectly general. For the question at hand this does not seem particularly important so we keep it simple, but allowing for a richer model of human capital is an important avenue for future research.

¹⁵As a result of selection they will be correlated conditional on the jobs actually chosen.

¹⁶Normally, we would think of this as a normalization, but with its interaction with β and the other functional forms it is not technically a normalization.

¹⁷To see this notice that even after a scale normalization, what we can hope to identify is $cov(\alpha v_{ij}^p + v_{ij}^u, v_{ij}^p) = \alpha var(v_{ij}^p) + cov(v_{ij}^u, v_{ij}^p)$. We can not separate the first component from the second. Thus, setting $cov(v_{ij}^u, v_{ij}^p) = 0$ is an innocuous normalization for us, though there are some counterfactuals for which it would not be innocuous.

The same is true for the joint distribution of μ_j^u and $(\alpha\mu_j^p + \mu_j^u)$. We can not separate α from the joint distribution of μ_j^u and μ_j^p . However, we only get one normalization. Normalizing $cov(v_{ij}^u, v_{ij}^p) = 0$ delivers identification of α from the joint distribution of $cov(v_{ij}^u, v_{ij}^p)$ and $var(v_{ij}^p)$. Knowledge of this α takes the place of the normalization for the joint distribution of μ_j^u and $(\alpha\mu_j^p + \mu_j^u)$. Thus covariance of (μ_j^p, μ_j^u) is left unrestricted but its estimated value fundamentally depends on the $cov(v_{ij}^u, v_{ij}^p) = 0$ normalization.

Human capital evolves as

$$\log(\psi_h) = b_1h + b_2h^2 + b_3h^3$$

In most specifications¹⁸ we use a cubic spline so there are two free parameters and we choose the third to impose that

$$\frac{\partial \log(\psi_H)}{\partial h} = 0.$$

The time period is assumed to be one year, so we fix $\lambda_h = 1$. In Bagger, Fontaine, Postel-Vinay, and Robin (2014) human capital evolves deterministic. We do not view this difference as important, but if anything our specification tends to make human capital more important in explaining wage variation.

As mentioned above we allow the job destruction rate to vary across individuals but not across jobs. We specify it as

$$\log(\delta_i) \sim N(\mu_\delta, \sigma_\delta^2).$$

Allowing δ_i to vary across establishments in a way that is correlated with job types is feasible, but our goal is to keep the model simple and allowing the job destruction rate to vary across jobs does not add much to the model.

In terms of identification of the arrival rates of jobs λ^e and λ^n , there are two important considerations. The first is the identification issue noticed by Flinn and Heckman (1982). With a continuous number of jobs, we can not separately identify the arrival rate of jobs from the reservation utility. We address this issue by estimating the λ 's but restricting the location of the utility of non-employment. We take a simple specification for the value of non-employment by assuming

$$u_{i0t} = \alpha [E(\theta_i) + \gamma_\theta (\theta_i - E(\theta_i)) + v_{i0}^n]$$

with $v_{i0}^n \sim N(0, \sigma_v^2)$. When $\theta_i = E(\theta_i)$ and $v_{i0}^n = 0$ then $u_{i0t} = \alpha E(\theta_i)$. This means that when they have no human capital, average workers acceptance rate of jobs from non-employment would

¹⁸We use this specification for the base model and for low education men and women. This led to a strange shape for highly educated men and women so in that case we set $b_3 = 0$ and just estimated an unrestricted quadratic.

be roughly 50%. We think of this as a normalization; choosing a different value would lead to different estimates of λ .

For exactly the same reason as we can not separately identify the level of λ^e and λ^n from the level of u_{i0t} , we can not separate individual heterogeneity in λ^e and λ^n from heterogeneity in u_{i0t} . As can be seen we incorporate this as heterogeneity in u_{i0t} and do not allow for individual specific heterogeneity in the λ 's which again we think of as a normalization.

The second issue is allowing for job specific arrival rates, i.e. j subscripts on λ_j^e and λ_j^n . We parameterize the model somewhat differently. We estimate λ^e and λ^n which are the arrival rate of *any* job. We then estimate the distribution from which those jobs are drawn (and assume this is the same regardless of whether it comes from employment or non-employment). Thus, instead of allowing for differences in arrival rates across job types (and assume one of each type) we allow the mass of job types to differ. We tried to choose a relatively parsimonious functional form for the distribution of (μ_j^p, μ_j^u) which is a discrete distribution. With no obvious parametric alternative we use the following one:

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

where $U_1(j)$ and $U_2(j)$ are distributed as discrete uniform 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 giving us one hundred different firm types. Essentially f_1 governs the variance of μ_j^u , f_2 governs the variance of μ_j^p , and f_3 governs their correlation.

We fix the discount rate at $\rho = 0.05$. This leaves a total of 18 parameters to be estimated:

$$[\mu_\delta, \lambda^n, \lambda^e, E_\theta, \sigma_\theta, \sigma_\xi, \sigma_{v^p}, \alpha, f_1, f_2, f_3, b_1, b_2, \beta, P^*, \sigma_\delta^2, \sigma_v^2, \gamma_\theta]$$

Recall that even though we showed that β is not identified in the non-parametric model, we estimate it here. We fixed this problem by setting $var(v_{ij}^u) = 1$. This defines the scale of utility in which case β can be identified relative to this scale. The issue is that it is impossible to distinguish bargaining power from the intensity of preferences because all we could identify was ordinal utility not cardinal utility. Once we have normalized $var(v_{ij}^u) = 1$, we no longer have this problem and β is identified essentially from the importance of the bargaining process. The larger is β the smaller will be the importance of this process. Note as well that since β is identified from this normalization it can not be interpreted literally and can not be compared across different samples or the bargaining power estimated in related models.

6 Data and Danish Institutional Features

We use Danish matched employer-employee data. The data consists of two types. The first is weekly spell data that covers all individuals aged 15-74 in Denmark from 1985 to 2003. The data period stops in 2003 because this is as far as the spell data has been constructed.¹⁹ The version of the spell data used in this paper is generated from various raw registers maintained by Statistics Denmark. It consists of a worker identifier, firm and establishment identifiers, start and end date of the spell, and a state variable. The states are employed, unemployed, self-employed, retired, and non-participation. Unemployment is defined by receiving either unemployment insurance benefits or social assistance. 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 and other annual data sets.²⁰ IDA contains socioeconomic information on workers and background information on employers, and covers the entire Danish population age 15 to 74.²¹ Thus, the unit of observation in our data set is a worker-week-labor market state-establishment (if employed).

We choose to have our empirical analogue of a job type be at the establishment level and not the firm level. We differ on this point from most of the empirical search literature which uses firms as the employer unit. Using establishments have at least three advantages in the current setting. The firm identifier in the Danish data is not well defined over time, since it is based on a mixture of tax reporting numbers and legal units. Firms might change both the tax reporting number and the legal unit (and hence firm identifier) without changing anything else. The establishment identifier is consistent over time.²² Secondly, when thinking about compensating differentials, which is an integral part of the model, the most appropriate unit seems to be the establishment and not the legal firm. Third, it provides a more convenient way to think about government jobs. Treating the government as one firm is problematic. Since different government establishments often have very different responsibilities and do not coordinate with each other, thinking about them as if

¹⁹See Bunzel (2010) and Bobbio (2010) for a longer description. The research group at Aarhus University responsible for creating the spell data have made several attempts at extending the period that it covers. This has proven a great difficulty due to several data breaks in the raw registers in the mid-2000's. At the current moment there is an effort being made trying to update the spell data, but the data is not in a state such that it can be used for research as it is still being debugged.

²⁰Both IDA (Integreret Database for Arbejdsmarkedsforskning/Integrated Database for Labor Market Research) and the other annual data sets are constructed and maintained by Statistics Denmark.

²¹IDA contains the annual average hourly wage for the job occupied in the last week of November. However, as shown in Lund and Vejlin (2015) the hourly wage measure developed by Statistics Denmark have several drawbacks. Instead we use the hourly wage measure suggested by Lund and Vejlin (2015), which improves the original measure in multiple dimensions.

²²The establishment is constructed by Statistics Denmark and is the same across years if one of three criteria is met: same owner and industry; same owner and workforce; same workforce and either same address or same industry.

they are separate seems like the best approximation.

We aggregate unemployment and non-participation into non-employment. It is an open question how to think about non-participation. Looking at the data, it is clear that non-participation in the data is not an absorbing state in the sense that the hazard rate into employment from non-participation is around half of the hazard from unemployment. Our model allows for workers who do not take any job, so including workers who are not active in the labor market is not a big issue, since it will be captured by a high flow utility from non-employment.

We categorize a job-to-job transition as a transition between two establishments with a very short intermediate non-employment spell (details below). From the identification strategy it is clear that we need to distinguish between voluntary job-to-job transitions and involuntary job-to-job transitions in order to interpret them as a revealed preference. It is obvious that not all job-to-job transitions in the data are indeed voluntary. In Denmark, an average notice period is 2-3 months.²³ This means that on average a fired worker has around 2-3 months in order to find a new job before the old job stops. If the worker finds a job in this period and start before the old one stops it will look like a job-to-job transition in the data. Using only population-wide register data we are not able to credibly distinguish such a case from a true voluntary job-to-job transition. As shown above we can identify P^* (immediate offer after job destruction) without direct information on the status of the job-to-job transition. However, we chose to augment the standard data from Statistics Denmark with survey data from two representative samples of workers in 1995 and 2000.²⁴ Both surveys contain questions on whether the job spell in November 5 years ago terminated due to the employer or the employee. This is matched with the register data and used later to form a moment that identifies the parameter, P^* .

6.1 Sample Selection Criteria

In order to select an appropriate estimation sample, we use the following sample selection criteria.²⁵ First, we censor workers after age 55 in order to avoid retirement related issues. We also disregard spells before labor market entry (or age 19) defined as the time of highest completed education (and not observed in education later). This is done in order not to draw inference from jobs that are not a part of the worker's primary job market. We then proceed by disregarding workers with errors in educational information. Temporary non-employment spells with 12 weeks or less in between jobs at the same establishment are overwritten by the employment spell. Non-

²³The notice period typically depends positively on tenure and also varies across sectors.

²⁴The survey is the "The Danish Work Environment Cohort Study."

²⁵See Appendix B for a more thorough description.

employment spells that are 2 weeks or shorter are allocated to the first of the two employment spells. In the first case this is done because we believe that recalls are substantially different from being laid off permanently. In the second case we believe that for non-employment spells shorter than two weeks it is most likely the case that the job was obtained before starting non-employment and the two weeks are basically caused by a timing issue or short vacation in between jobs.²⁶ Since the model does not contain retirement or self-employment, we censor workers when they enter those two states. Finally, we also disregard workers who have gaps in their spell histories. This could arise if the worker for some reason have missing IDA data in a given year due to e.g. temporary emigration. This gives us our estimation sample.²⁷

Since the model is cast in steady state and cross-sectional wages therefore have no trend we detrend wages in logs by gender-educational groups conditional on experience. We condition on experience since the composition of workers change over the sample period due to an aging workforce.

Since job-to-job transitions play a vital role for the identification of our model, we ignore transitions from two types of establishments. The first are transitions for workers to or from establishments with missing ID (0.5 percent, cf. Table 2). We also ignore job-to-job transitions from closing establishments or establishments with mass layoffs.²⁸

6.2 Descriptive Statistics

In this section we present different descriptive statistics for the estimation sample. The number of years in sample and the number of establishments for each worker are important for the identification of the model. Table 2 show statistics for these measures together with other descriptive statistics. The worker is on average in the sample for 11 years and is 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 moderately changes over the sample period, since entering workers are better educated than those leaving the sample. The average cross-section age is 38. A total of 83 percent are employed in general, while 32 percent are employed in the public sector. The fact that only 83 percent are employed is intentional and is a result of the mild sample selection

²⁶The Danish labor market institutions give rise to the possibility of laying of workers temporary, so they actively support the possibility of recalls. Regarding the second case, it seems unlikely that a job should start just 1-2 weeks after entering non-employment. Usually, jobs are advertised at least 1-2 months before the starting date.

²⁷One worry about imposing strict sample selection criteria is that the sample will be unbalanced over the sample period. However, the sample selection results in an almost balanced sample in the sense that the sample consists of around 1.8 mill. workers in 1985 and 1.65 mill. in 2003.

²⁸A mass layoff is defined as the establishment having more than 15 workers in year t and less than 30 percent left in year $t + 1$. We define an establishment to be closing if we do not observe it in the data any longer.

that we impose. The average labor market experience is 13 years. Finally, in a given cross-section the establishment identifier is missing for 0.5 percent of all employment observations.

Figure 9 shows the estimates from a Kaplan-Meier estimator of the survival probability for employment spells (time between non-employment spells), job spells (a job is defined as an employment spell within an establishment), and non-employment spells. Notice, we only use non-left censored spells, so shorter spells are overrepresented compared to the moments presented later. The survival rate for non-employment is higher than what is normally observed as a consequence of the mild sample selection. The job survival rate is lower than the employment survival rate indicating significant job-to-job turnover between establishments. This is important for identification of job ladders.

The dips in both the survival rates for jobs and non-employment come from a “New Year” effect. In the reported data there is an over-representation of state changes at January 1st each year. Most of the state changes are coming from transitions the past year that have not been reported correctly.

6.3 Institutional Setting

The Danish labor market is characterized by having a so-called flexicurity system. The three main pillars in the flexicurity system are a generous social security system, flexible hiring and firing rules, and an active labor market policy with a focus on job search and re-employment possibilities, see Andersen and Svarer (2007) for a more thorough review. The generous social security system is a variant of the Ghent-model and consists of two main parts; unemployment insurance benefits and social assistance which is means-tested. The first is heavily subsidized by the government with members of unemployment insurance funds paying around one third of the total cost. If an individual is not eligible for unemployment insurance he or she can apply for social assistance. This is means-tested and eligibility requires among other things that the worker does not own much of value and is available for the labor market. The flexible hiring and firing rules and the low degree of employment protection results in a highly fluent labor market, see e.g. Bertheau, Bunzel, Hejlesen, and Vejlin (2016). These first two pillars were already in place in the 70’s. However, in the 70’s and 80’s Denmark did not do very well in international comparisons. A series of reforms in the early 90’s shifted focus from passive to active labor market policies, shortened the maximum unemployment duration, and increased the eligibility criteria. It is this combination that has been become known as the flexicurity model.

A full-time job in Denmark is around 1700 hours per year, which is lower than in the US. While hours are low the participation rate is rather high around 75-80 percent. Furthermore, as indicated

by the sample above the public sector is rather large, which is typical for a Scandinavian welfare state.

Clearly, the model presented makes some strong assumptions. Most notably regarding the wage setting mechanism. The fact that wages can and will be renegotiated when credible outside offers presents themselves is not something that necessarily happens in a labor market that is as highly unionized as the Danish. Even though unionization has been persistently high (around 70-80 % of workers are unionized) during the estimation period, Denmark underwent a transformation in the 80's and early 90's from having a centralized wage bargaining system to having a more and more decentralized wage setting system (see e.g. Fontaine and Vejlin, 2016 and Dahl, le Maire, and Munch, 2013). According to Boeri and Calmfors (2001) the centralization index of the Danish bargaining system has dropped from 0.64 for 1973–77 to 0.47 for 1983–87 and 0.34 for 1993–97.²⁹ This is also pointed out by Dahl, le Maire, and Munch (2013) who use the decentralization process to investigate how different wage bargaining systems affect wage dispersion. With this change in mind, centralized bargaining is still an important part of the Danish labor market especially for the public sector. It is clear that for most public sector jobs wage renegotiations does not take place when an outside offer appears. However, there are a wide use of tenure contracts in both the public and private sector in the sense that wages go up with tenure at a specific pace. This is especially true for contracts negotiated in the more centralized system. There are several papers starting with Burdett and Coles (2003) that study wage-tenure contracts theoretically. In these models tenure effects arise as a result of firms trying to reduce their quit rate by back-loading wages. This is essentially the same underlying mechanism as in our framework, where tenure effects also arise as a result of firms trying to retain workers. With obvious caveats tenure contracts could be interpreted as firms realizing that workers over time accumulate outside offers, but without being able or willing to renegotiate wages as they do. Thus, the same underlying economic mechanism is at play.

7 Auxiliary Model

We estimate our model using indirect inference following Gourieroux, Monfort, and Renault (1993). Our approach is to use the formal non-parametric identification results as a guide to which aspects of the data 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,

²⁹The index takes values from 0 to 1.

for each parameter we choose one auxiliary parameter that we think is useful for identifying it.³⁰ We find this approach to be highly beneficial for us in understanding the mapping between the parameters and the data. This is also relevant because identification of some of our parameters is more subtle than others. The most important example is β versus, for example, E_θ . We see that β is primarily identified by the tenure squared term in a fixed effect regression. This is a relatively imprecisely estimated parameter as opposed to mean wages which is very precisely measured. If we included a lot more moments in the data this feature would get very little weight and identification of β would come more from higher order moments of the wage distribution rather than the patterns of wage growth within a job which is why it is important.³¹

7.1 Notation

Let us define the variables used from the data

- $i = 1, \dots, N$ 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 ℓ 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
- $D_{i\ell j}$ the duration of time that the worker worked on job spell $i\ell j$
- $w_{i\ell jt}$ the t^{th} wage observation at job $i\ell j$
- $E_{i\ell jt}$ the t^{th} experience observation at job $i\ell j$
- $TE_{i\ell jt}$ the t^{th} tenure observation at job $i\ell j$. It is set to 0 at the first November cross-section in job $i\ell j$ and from that it increases with $E_{i\ell jt}$.
- $k = 1, \dots, K_i$ the number of non-employment spells for individual i
- D_{ik}^n the duration of non-employment spell ik

³⁰While 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.

³¹It is also not obvious which other features of the data we would use. Our main job is to decompose the variance of log wages and we target and fit the variance of log wages. We could look at higher order moments of the data and fit them - presumably if one really wanted to do that they would want to go beyond Gaussian assumptions about the error term. While adding this data and new parameters is feasible, we do not think it would add much to the current paper and it certainly would make the mapping between parameters and data much more opaque.

We assume that there are a large number of people in the economy, but a finite number of J job types. Our econometric specification differs from the identification section above in that we do not observe precisely which type an establishment is. Instead we infer the types of the establishments by using data from co-workers.³²

Given these variables from the data, we now define intermediate variables which we will use in the actual moments. First, we define intermediate variables related to transitions. As argued previously we use the revealed preference of a job-to-job transition to indicate preference for a job. With this in mind we define

$$S_{ilj} \equiv \begin{cases} 1 & \text{if spell } ilj \text{ starts with a job-to-job transition and does not end with one} \\ -1 & \text{if spell } ilj \text{ ends with a job-to-job transition and does not start with one} \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $S_{ilj}=1$ indicates that the worker has shown a preference for the job in the sense that he/she likely voluntarily chose it over another initially, but did not voluntarily choose another job when the employment relationship ended. Similarly $S_{ilj} = -1$ indicates that the worker did not particularly like the job the sense that they only entered when they had no other job offers and then left for another job. Since it is part of an auxiliary model it does not need to be a perfect measure of preference, just correlated with preference.

Notice, that if the spell is left censored we assume that it starts from non-employment. Likewise, if the job is right censored we assume that it ends in a job destruction. S_{ilj} should sum to zero for each individual when summing over jobs, but not when summing over establishments. S_{ilj} thus measures whether the worker has revealed a preference for working in establishment f_{ilj} . There could potentially be many different ways to construct a function related to revealed preferences. The above function only puts weight on the least preferred job (the first job in an employment spell) and the most preferred job (the last job in an employment spell) and assigns equal weight to the jobs in between. We prefer this functional form due to its simplicity.

As mentioned above, the other complicated aspect of identification is that we use co-workers as a proxy for job types. In particular, we use co-workers to estimate the covariance of workers preferences over jobs. For this we define the number of job-to-job separations, s_{-i}^q , and the number of job-to-job hires, h_{-i}^q , for each establishment q . Since we want to use this to estimate the correlation between person i 's preferences and their co-workers, we exclude individual i from this calculation.

³²We could use co-workers to estimate the posterior type probabilities for each establishment, but this is not central to anything we do.

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

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

From this define

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

Thus, $h_{-i\ell j}$ is a measure of the degree to which individual i 's co-workers at establishment $f_{i\ell j}$ like working there expressed by looking at job-to-job transition patterns. Since our goal is to construct a covariance we construct a residualized version of $h_{-i\ell j}$ by subtracting the mean value of $h_{-i\ell j}$ across jobs for individual i

$$\tilde{h}_{-i\ell j} \equiv \frac{h_{-i}^{f_{i\ell j}}}{h_{-i}^{f_{i\ell j}} + s_{-i}^{f_{i\ell j}}} - \frac{1}{\sum_{\ell^*=1}^{L_i} J_{i\ell^*}} \sum_{\ell^*=1}^{L_i} \sum_{k=1}^{J_{i\ell^*}} \frac{h_{-i}^{f_{i\ell^*k}}}{h_{-i}^{f_{i\ell^*k}} + s_{-i}^{f_{i\ell^*k}}} 1(N(f_{i\ell j}) > 50)$$

where $N(f_{i\ell j})$ is the number of individuals that contribute to $h_{-i}^{f_{i\ell j}}$.³³ Notice, that this does not include employment spells where the employee is employed by firm $f_{i\ell j}$ in the first job spell.

We also need some analogous intermediate variables regarding wages. First define \bar{w}_i as the mean worker wage over his working life

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

For the jobs where we observe a wage we also define the average job wage

$$\bar{w}_{i\ell j} = \frac{1}{T_{i\ell j}} \sum_{t=1}^{T_{i\ell j}} w_{i\ell j t}$$

If we do not observe any wages in the job we set $\bar{w}_{i\ell j}$ to zero.

³³We condition on the establishment having more than 50 observations, $N(f_{i\ell j}) > 50$, in order to be able to calculate the covariance matrix of the moments by bootstrap. The issue is that within a bootstrap repetition we might draw the same individual multiple times. This will generate an upward bias in the estimate of e.g. $\tilde{h}_{-i\ell j}$ as copies of the individual will have the same preference. However, this is only a problem for smaller establishments as the bias goes away when the establishment size increase. We will use this condition for moments involving co-workers.

For the jobs where we observe at least one wage define

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

$\tilde{w}_{i\ell j}$ has the nice feature that it will sum to zero across jobs for each individual.

When we look at co-workers, for the same reason as for $h_{i\ell j}$ sometimes need to take the individual out of the calculation, so define

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

$\tilde{w}_{-i\ell j}$ is measuring the average value of $\tilde{w}_{i^*\ell^*j^*}$ over jobs at establishment $f_{i^*\ell^*j^*}$ taking out the contribution from individual i .

7.2 Auxiliary Model

Using the definitions of the intermediate variables, we now define the moments used in the estimation. In general notice that many of the moments are calculated over different samples, since not all variables are defined for each job spell.

For expositional reasons we divide the parameters into two groups: those that are primarily important for turnover and those that are primarily important wages. There are 8 parameters that are important for turnover: $\sigma_{\zeta}^2, d_0, \lambda^e, \lambda^n, \sigma_v, f_1, P^*, \gamma_\theta$.

We first discuss which moments we chose to identify these variables.

- Level and variance of job destruction rate, $\mu_\delta, \sigma_\delta^2$: This is the rate at which people enter non-employment. We identify the mean and variance by using the duration of employment spells

$$\bar{L} = \frac{\sum_{i=1}^N \frac{\sum_{\ell=1}^{L_i} (\sum_{j=1}^{J_{i\ell}} D_{i\ell j})}{L_i}}{\sum_{i=1}^N 1[L_i > 0]}$$

and the variance

$$\frac{\sum_{i=1}^N \left(\frac{\sum_{\ell=1}^{L_i} (\sum_{j=1}^{J_{i\ell}} D_{i\ell j})}{L_i} - \bar{L} \right)^2}{\sum_{i=1}^N 1[L_i > 0]}$$

- Arrival rate of outside offers from employment, λ^e : This will be important for determining how long someone stays on a job before moving to the next job. We use the average length

of time on a job.³⁴

$$\bar{J} = \frac{\sum_{i=1}^N \frac{\sum_{\ell=1}^{L_i} (\sum_{j=1}^{J_{i\ell}^{-1}} D_{i\ell j})}{\sum_{\ell=1}^{L_i} J_{i\ell}^{-1}}}{\sum_{i=1}^N 1[\sum_{\ell=1}^{L_i} J_{i\ell}^{-1} > 0]}$$

- Arrival rate of jobs from non-employment, λ^n : Similar to the previous one this will be important for the rate at which people enter employment from non-employment and we use the average length of non-employment spells

$$\bar{K} = \frac{\sum_{i=1}^N \frac{\sum_{k=1}^{K_i} D_{ik}^n}{K_i}}{\sum_{i=1}^N 1[K_i > 0]}$$

- Variance in tastes for non-work, σ_v^2 : The way we have parameterized this, it will be related to the variance in non-employment spells

$$\frac{\sum_{i=1}^N \left(\frac{\sum_{k=1}^{K_i} D_{ik}^n}{K_i} - \bar{K} \right)^2}{\sum_{i=1}^N 1[K_i > 0]}$$

- Variation in non-pecuniary benefits at the establishment level, f_1 : This parameter picks up whether there is a lot of commonality across workers in the tastes for particular establishments. We use the covariance between co-workers in the preference for the job as measured through job-to-job transitions. Again, we condition on the size of the establishment, since this moment involves co-workers of individual i .

$$cov(\tilde{h}_{-i\ell j}, S_{i\ell j} | N(f_{i\ell j}, \tilde{h}_{-i\ell j}, S_{i\ell j}) > 50)$$

- Probability of immediate offer after job destruction, P^* : This will be important because it depends on what fraction of job-to-job transitions are voluntary. We have direct evidence on this from the survey data. We select those from the survey that are in our data and who made a job-to-job transition between the job asked about in the interview and the time of the interview. We then calculate the fraction that report that employment terminated on their employer's initiative. Formally, we match on

$$\frac{\sum_{i=1}^N 1[\text{Survey}_{1995,i} = 1 \wedge J_{1990,i} = 1 \wedge S_{1995,1990,i} = 1 \wedge R_{1995,i} = 1] + 1[\text{Survey}_{2000,i} = 1 \wedge J_{1995,i} = 1 \wedge S_{2000,1995,i} = 1 \wedge R_{2000,i} = 1]}{\sum_{i=1}^N 1[\text{Survey}_{1995,i} = 1 \wedge J_{1990,i} = 1 \wedge S_{1995,1990,i} = 1] + 1[\text{Survey}_{2000,i} = 1 \wedge J_{1995,i} = 1 \wedge S_{2000,1995,i} = 1]}$$

³⁴The distinction between this one and the first is that this one is the length of time working for a particular employer while the first is the length of time between non-employment spells. Thus, the first one includes more than one job if the movement was through a job-to-job transition.

where $Survey_{x,i}$ is an indicator variable for being in the survey in year x , $J_{y,i}$ is an indicator variable for being in a job in November year x , $S_{x,y,i}$ is an indicator for having a November cross-section job in year y ending in a job-to-job transition before year x , and $R_{y,i}$ is an indicator for having reported in year x that the employment terminated on the employers initiative.

- Importance of ability in tastes for non-employment, γ_θ : It is important for how the duration varies with general ability θ . We use the covariance between the length of a non-employment spell and average wages over the data (for people that ever work)

$$\frac{\sum_{i=1}^N \sum_{k=1}^{K_i} D_{ik}^n \bar{w}_i}{\sum_{i=1}^N K_i} - \left(\frac{\sum_{i=1}^N \sum_{k=1}^{K_i} D_{ik}^n}{\sum_{i=1}^N K_i} \right) \left(\frac{\sum_{i=1}^N \sum_{k=1}^{K_i} \bar{w}_i}{\sum_{i=1}^N K_i} \right)$$

For wages we have 10 parameters: $E_\theta, \sigma_\theta, \sigma_\xi, \sigma_{vp}^2, f_3, f_2, \alpha, \beta, b_1, b_2$

We use the following moments.

- Mean level of general ability, E_θ : Since all other variables are mean zero, this will be important for getting the overall wage in the economy, so we use the average wage across all time periods.

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

- The variance of general ability, the variance of measurement error, and the variance of the comparative advantage piece: $\sigma_\theta^2, \sigma_\xi^2, \sigma_{vp}^2$: Since these parameters will be important for variance at different levels we decompose the overall variance into three components: within job spell, between job spell/within worker, between worker. We think that σ_ξ will be important for the first, σ_{vp} will be important for the second, and σ_θ will be important for the third

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

- Importance of establishment for productivity, f_3 : Here we use the covariance of co-workers wages relative to the wages these workers have received at other firms. Again, we condition on the size of the establishment, since this moment involves co-workers of individual i .

$$cov(\tilde{w}_{i\ell j}, \tilde{w}_{-i\ell j} | N(f_{i\ell j}, \tilde{w}_{i\ell j}, \tilde{w}_{-i\ell j}) > 50)$$

- Correlation between establishment productivity and establishment non-pecuniary aspects, f_2 : Here we use the covariance between workers wages at the firm and tastes for the firm. Again, we condition on the size of the establishment, since this moment involves co-workers of individual i .

$$cov(\tilde{h}_{-i\ell j}, \tilde{w}_{i\ell j} | N(f_{i\ell j}, \tilde{h}_{-i\ell j}, \tilde{w}_{i\ell j}) > 50)$$

- Importance of wages as a component of utility, α : To identify this we use the fraction of wage losses that occur at job-to-job transitions.

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

The idea behind this is that the more important wages are as a determinant of job preference the less often we will see wage losses at job-to-job transitions.

- Bargaining parameter and human capital production parameters, β, b_1, b_2 : Estimation of these parameters is quite subtle and distinguishing between them is difficult. Choosing which moments to use was not straight forward and we explored many options ultimately deciding the following approach was intuitively better and more robust than the alternatives. In our model, β essentially picks up the importance of the bargaining process as a determinant of wages (smaller values of β imply that it is more common). Thus, β can be identified by measuring the importance of tenure. We estimate the following regression using those observations with no left censored tenure (meaning job spells where we do not observe the beginning of the spell)

$$w_{i\ell jt} = \beta_{i\ell j} + \beta_1 E_{i\ell jt} + \beta_2 E_{i\ell jt}^2 + \beta_3 TE_{i\ell jt}^2 + \epsilon_{i\ell jt}$$

where $\beta_{i\ell j}$ is a job spell fixed effect. We match on the estimates of β_1, β_2 , and β_3 . Note that since $\beta_{i\ell j}$ is a job spell fixed effect, as pointed out by Altonji and Shakotko (1987) and

Topel (1991) and others, experience and tenure are perfectly correlated within a spell. As a result we can not identify the coefficient on tenure directly. Instead use the coefficient on tenure square to pick up the importance of bargaining. Intuitively, this works because the rate at which people switch jobs is identified from other moments so for any level of bargaining power we know the rate at which that bargaining power should decline. The squared term should pick up both the magnitude of the tenure effect and the rate at which it stops increasing, and since other components of the model pick up the rate the magnitude of this parameter should identify the magnitude of the bargaining effect.

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 absolute value of the estimated parameters.³⁵ The covariance matrix of the data moments have been calculated by bootstrapping the sample 200 times.³⁶ We estimate first using the full sample and then by dividing into four different demographic groups on the basis of gender and education.

8.1 Fit and Estimates

The results for the full sample are presented in Tables 3 and 4. Table 3 shows the auxiliary parameters from the sample and model. One can see that the fit is excellent. This is perhaps not surprising because we have as many free parameters as we do auxiliary parameters to match. However, the model is non-linear, so there is no guarantee of a match.³⁷ The structural parameters of the model are presented in Table 4. We do not view these estimates as interesting in their own right as they can only be interpreted in the context of the other parameters. We judge them in the context of their contribution to the counterfactuals.

One thing to note is our estimate of λ^e relative to λ^n as we find a much higher value of the former than the latter. This is largely due to the fact that we are looking at non-employment rather

³⁵We do this instead of the most common approach which is to normalize weight by the inverse of the standard errors (which is related to the efficient weighting matrix). Our goal is to match all the moments, but the standard errors are very different for the different auxiliary parameters and often very small. We worry that the standard approach might put essentially no weight on some auxiliary parameters. Given the fit of the model, the weighting matrix should not matter very much.

³⁶More specifically, in each bootstrap repetition we draw N (recall that N is the sample size) individuals with replacement from the data and recalculate the moments on the bootstrap sample. After having done 200 repetitions we calculate the covariance matrix of the estimated moments.

³⁷It is also possible that there are multiple solutions of the model that all fit the data. While one can never guarantee a global optima, we have found no evidence of this.

than unemployment. We see in the raw data that non-employment spells tend to be similar in length to job spells, but switching jobs from employment should happen at a lower rate since workers with jobs are presumably much pickier. Whether this is due to heterogeneity in arrival rates or reservation values is hard to identify and we have made a certain normalization through heterogeneity in reservation utility rather than heterogeneity in arrival rates. The primary goal of this project is to explain wages rather than employment, so we do not view this issue as first order for this paper but worth exploring for other papers that are more concerned with explaining unemployment.

We also make some remarks about the raw data. We have talked about the importance of the coefficient on tenure squared in identifying β . One can see in the raw data that this coefficient is much larger in absolute value than the coefficient on experience squared. This results in the bargaining process being quite important for earnings inequality. Secondly, note that wage drops at job-to-job transitions are not at all uncommon, roughly 40%. Some of this will be explained by the fact that many of these transitions, 20%, are involuntarily. The rest will be explained by compensating differentials and the fact that workers will take a hit to wages in the short run for more wage growth in the long run.

8.2 Decomposition of Wage Variation

We focus on the decomposition of total log wage variance which is presented in Table 5. We sequentially eliminate the different sources of wage inequality and document their effect on inequality. Prior to the decomposition in the table, we eliminate measurement error. The total variance of log wages in the model and in the raw data is 0.124, but it falls to 0.106 without the measurement error.

We begin with the variance of 0.106 and try to determine which factors contribute to it. Recall that given the issues with the non-employment, eliminating compensating differentials makes less sense than the others. We simulate four different sequences of decompositions (A-D) in Table 5. It is important for the interpretation to understand that we eliminate things sequentially. For example in column (A) the first row presents results in which we eliminate human capital. The second row presents results where we eliminate human capital and monopsony. In the third we eliminate human capital, monopsony, and Roy model inequality, etc.

In all columns, our first two steps are the same. We eliminate human capital first by setting it to the maximum level, ψ_H . We do this since 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), so we view this as well known and less interesting for the purpose of explaining

variation in wages. In addition, given the structure of the model, human capital largely operates exogenously and separately from other aspects of the model, so it does not interact as much. We show that the small explanatory power is true here as well as human capital explains about 5% of wage variation.

The second step is to eliminate the variation in monopsony powers that firms have over workers reflected in the bargaining process by setting $\beta = 1$. We set $\beta = 1$ prior to the other counterfactuals for the reasons discussed above; the level of β is set by normalizations on other parameters (mostly the scale of preferences), so it does not make sense to hold it fixed while changing the other parameters. Setting $\beta = 1$ eliminates the variation in the model that comes from negotiation by giving all of the bargaining power to the worker. This lowers the variance of wages by about another 8%. The fundamental source in the model that leads to this heterogeneity is search frictions, i.e., in a perfectly competitive environment firms would have no monopsony power. Thus in an accounting sense, this 8% should be attributed to search frictions.

We next eliminate the remaining part of search frictions, compensating differentials, and Roy model inequality. As discussed above, the selection problem makes eliminating compensating differentials is the most tenuous of these. Thus, the most reliable simulations are (A) and (B) in which we eliminate the other two first. More specifically we do the following

- We eliminate search frictions by allowing workers to find the most preferred job immediately (i.e. $\lambda^e, \lambda^n \rightarrow \infty$)
- We eliminate Roy inequality by eliminating variation in wages within job type, but we hold the preference ordering across jobs exactly the same, but set $\sigma_\theta = \sigma_{v^p} = 0$ in the wage setting. We hold the preference ordering for jobs constant since we cannot non-parametrically identify how these would change when we change productivity as discussed in section 4 on identification. Note, that for this counterfactual one can see the necessity of matched employer-employee data. We still allow for μ_j^p , the variation across job types that is common to all workers.
- We eliminate compensating differentials by assuming workers choose among acceptable jobs only by comparing wages (i.e. $v_{ij}^u = \mu_j^u = 0$). However, for the reasons discussed above we condition on acceptable jobs by which we mean any job the worker would accept from non-employment in the real world. That is, any job rejected from non-employment in the full model would always be rejected in this counterfactual.

The first thing to note from Table 5 is that in all four simulations Roy model inequality is the

most important accounting for the vast majority of the variation in every decomposition. Non-parametrically we can not separately identify how much of this comes from the common component, θ_i , versus the match component, v_{ij}^p as this separability is a functional form restriction to make estimation feasible. Given the parametric form we find that roughly 60% of the Roy component comes from the common component.

The relative importance of search frictions and compensating differentials varies considerably across the four simulations so this is clearly not an orthogonal decomposition. We do not view this feature as a weakness of our decomposition but rather as a way of demonstrating how the aspects of the model interacts. Most interesting is search frictions. Recall that the monopsony aspect explains 8% of the variation in every case and that monopsony power arise in the this model because of search frictions. The remaining amount varies considerably across the specifications. It is about 1% in (A), 3% in (B), 21% in (C) and 2% in (D). What leads to these large differences in the importance of frictions? The order of the decomposition fundamentally alters the aspects of jobs that workers are searching for. In the base case, which corresponds to (B), workers are searching for good matches in four different dimensions - firm specific productivity (μ_j^p), firm specific non-pecuniary aspects (μ_j^u), match (individual \times firm type) productivity (v_{ij}^p), and match (individual \times firm type) utility (v_{ij}^u). When workers are searching for all four of these aspects this leads to 3% variation in wages (0.093-0.090 in column (B)). In experiment (A) we first eliminate Roy inequality. This means that we have eliminated the individual \times firm type productivity match (v_{ij}^p), which is a very important source of wage inequality. In this case workers are searching for a good match in terms of the non-pecuniary aspects (both match specific and firm specific) and firm productivity only. Perhaps not surprising these aspects are not particularly important for wage inequality and search only explains 1% of the inequality. In (C) we do the opposite, we first get rid of compensating differentials which means we are getting rid of both the firm specific and individual \times firm specific non-pecuniary characteristics (μ_j^u and v_{ij}^u). In this case workers are searching only for pecuniary aspects of the job and search frictions turn out to be very important. They explain 21% of the variation (29% total if one includes the monopsony part). In (D) we eliminate both the search for non-pecuniary aspects and the individual \times firm type productivity match, so that all is being searched for is the firm type productivity. This explains 2% of the variation.

8.3 Decomposition of Utility Variation

Table 5 quantifies the amount of variation in log wages. However, workers care about more than just wages. Another way of quantifying inequality is to look at variation in utility rather than just

wages which we do in Table 5. We should emphasize that unlike the main decomposition of Table 5, this one is not non-parametrically identified because we can only non-parametrically identify ordinal utility not cardinal utility. Thus, it should be viewed with some caution. For example, while we could get some bounds on utility (measured in units of foregone wages) by observing the wage cuts one could take, but the results would be sensitive to tail behavior at the most favored job which can not be bounded non-parametrically. Despite the limitations, the results show how potentially important the other two channels might be. For comparison we only use employed workers and we normalize utility to log wage equivalent units. That is, recall that flow utility is defined as

$$U_{ij}(W\psi_h) = \alpha \log(W\psi_h) + \mu_j^n + v_{ij}^n.$$

To put utility in the same units as log wages we can rescale simply by re-normalizing by dividing by α

$$\tilde{U}_{ij}(W\psi_h) \equiv \log(W\psi_h) + \left(\frac{\mu_j^n + v_{ij}^n}{\alpha} \right).$$

The results of this decomposition are presented in Table 6 and they are very different. This can be seen purely from looking at the variance in the first row - variance in tastes for non-pecuniary aspects contributes more to the variance than does variance in wages. First of all, from the overall variance one can see that the variance in wages accounts for only about 1/3 of the variation in utility. Unlike the wage variance, where the Roy model explained most of the variation, most of the variation in utility is explained by the interaction between compensating differentials and search. In this case it does not matter all that much how we do it, but getting rid of either search or compensating differentials eliminates most of the variation in flow utility.

8.4 Life-Time Profiles

So far we have investigated how the different components matter for cross-sectional variation in wages and utility. Here we turn our attention to how they affect average wage and utility over the life-cycle.³⁸ Figure 9 shows average wages for different combinations of counterfactuals. In the base model all four components of wage variation are present. The wage profile is increasing and concave, which we would also expect given our moments. The first step is to eliminate human capital by setting it to the maximum level. One can see that not surprisingly a large part

³⁸Once again, there is a large literature looking at this and a full literature review is beyond the scope of this paper. Weiss (1986) and Sanders and Taber (2012) provide nice surveys. Recent examples include Bagger, Fontaine, Postel-Vinay, and Robin (2014), Sanders (2014), Lu (2016), and Han (2016).

the increase in average wages over the life-cycle is due to increases in human capital. This is in accordance with the results from Bagger, Fontaine, Postel-Vinay, and Robin (2014). The next step is to eliminate monopsony power by setting $\beta = 1$, so workers get their full productivity. This eliminates any effect of outside wage offers. We can see that this is more important than general human capital and also relatively more important early in workers careers. However, it is also interesting that it still has a large effect on the level of wages after 20 years, which basically tells us as evidence that it takes a long time for the worker to move to his/her preferred firm match and to then generate outside offers for rent extraction. Next, we delete search frictions by letting employed workers move to their preferred job immediately. We can see that this increase wages and essentially makes the profile flat.³⁹ Finally, we delete Roy model heterogeneity. This decrease average wages by about 0.2 log points. This is due to the considerable variance in selection on match productivity which no longer exists.

Figure 9 shows average utility with the same decomposition as the one above for wages. Eliminating human capital and monopsony power has about the same effect as before. Contrary to before, eliminating search increase utility by much more than search and monopsony power combined. This is a result of workers moving to their preferred job, which does not only have higher wages on average, but also have a higher non-pecuniary return. It is clear from this that search frictions and compensating differentials are not only important for explaining the variation in utility, but also for explaining the level. Eliminating Roy model heterogeneity has the exact same effect as on wages.

8.5 Other Aspects of the labor market

The fact that search frictions and compensating differentials are not that important for wage inequality does not mean that they are not important for the labor market more generally. We quantify their importance in a few other ways. Compensating differentials are important for turnover. In roughly 1/3 of the competing offers we see in the simulation of our full model, the workers would have made a different choice if they only cared about wages. The consequences are large as workers earn about 0.20 log points lower wage as a result of these choices. In our model search frictions are obviously important in explaining turnover as there would be no turnover without them. To quantify, wages would be about 0.17 higher in the absence of search frictions. Of these 0.17, roughly 0.07 is due to the negotiation and 0.10 is due to the actual job match.

³⁹The reason it is not completely flat is because all workers start as non-employed. High ability individuals have lower flow utility from unemployment, so they are willing to take more jobs out of non-employment and since they are high ability individuals they have high wages, so the profile is slightly decreasing.

8.6 Sub-groups

As mentioned previously, we also estimate the model for different subgroups. In particular, we divide individuals by gender and whether or not they have more education than high school. In Tables 7 and 8 we present the results for the different demographic groups in terms of moment fit and parameter estimates. First, as is clear from Table 7 the fit is very good for each of the four groups. For the most part the parameters are quite similar across the different groups and the differences make sense. However, there are two notable differences about college men as compared to the other groups: they have a substantially higher value of α and a substantially lower value of β . This will lead to differences in the simulations between college men and the other groups that we discuss below.

Table 9 presents the decomposition results for all four groups. To fit the information in a single table, rather than present the results in four different types of decompositions as in Table 5, we present the results of the different counterfactuals that can be used to produce those results. That is while the decompositions in Table 5 are sequential, the counterfactuals presented in Table 9 are more complicated. The first two rows are analogous to Table 5. The remaining rows continue to assume no human capital and no monopsony, but otherwise are not sequential. The main result holds for all four groups. Roy model inequality is clearly the most important factor for all four demographic groups. Furthermore, the basic results above regarding the interaction between search and the other components are generally true as well.

Despite the similarities, there are quite a few factors that are different across the different groups. As suggested by the parameter estimates above college men differ from the other groups in three important ways. First, the level of inequality is very different: the variance is more than twice as large for college men as it is for high school women. Second, compensating differentials are less important for college men than for the other groups. Third, the importance of monopsony is also quite different: it is a very important factor for college men explaining almost 30% of the variation while it only explains 12% for high school women. As discussed above this comes directly from the fact that the coefficient on tenure squared is much larger for this group which leads to a smaller value of β . With the exception of college men, the largest drops for search occur when we drop compensating differentials first. Second, the bargaining process is much more important for college men than it is for the other groups.

These two results can be anticipated from the difference in parameter estimates. The α dictates the importance of wages relative to the idiosyncratic taste since the variance of v_{ij}^p is normalized to one. The higher value suggests that college men place a relatively higher value on pecuniary

aspects of the job than non-pecuniary aspects. The lower value of β suggests that these men have lower bargaining power leading to a greater role for monopsony. While a smaller bargaining parameter for college men seems surprising on its face, it makes more sense from a broader perspective. First, one should keep in mind that it is only identified given the normalization on the variance of v_{ij}^p so we can not literally compare it across groups. It could just be picking up the fact that the true variance of v_{ij}^p is lower for this group of people. However, as we see in the simulations it does lead to a larger importance of bargaining. As a practical matter, this occurs because the coefficient on tenure squared is considerably larger for college men than the other groups as can be seen in Table 7. The implication is that the bargaining process is more important for this group than the others. Intuitively this makes sense to us. While it is not incorporated in our model, one could interpret β as picking up the relative importance of jobs that post wages versus those that bargain with employers. That is, if a firm commits to wage posting our model will interpret that as the worker having all the bargaining power, since there will be no return to tenure. One can see that the order of the magnitude of β perfectly corresponds with the order of the coefficient on tenure squared, so one can see what factor in the data are driving these results. Intuitively, it seems plausible to us that college educated men is the group that will encounter wage posting offer the least. There exists a small literature on the choice of wage posting versus sequential auction from the firms point of view. Postel-Vinay and Robin (2004) propose a random search model with on-the-job search. Workers can choose search intensity thereby potentially generating a moral hazard problem in the sequential auction setting. Firms can commit to either matching (sequential auction) or not matching (wage posting) outside offers. In this environment high productive firms chose to match, while low productive firms chose to post wages.⁴⁰ If there is positive assortative matching as suggested by Bagger and Lentz (2015) and Lise, Meghir, and Robin (2016) then college men would be more likely to be employed in a high productive firm which chooses to match outside offers. Distinguishing a model with both wage posting and bargaining from one with bargaining but allowing for a flexible β seems extremely difficult and as a practical manner we are not sure how to distinguish them. We leave this as an open research question for future work.

The relative importance of search frictions depends quite a bit on the experiment we do. For example, if we eliminate compensating differentials first, then the impact of getting rid of search frictions next leads to a large decline for all groups except college educated men, but if we just eliminate search friction the largest decline happens for college educated men.

⁴⁰Doniger (2015) use a similar setup but in a more general framework.

9 Conclusions

The goal of this paper is to estimate the primary drivers behind wage variation. For this purpose we have developed and estimated a general labor market model. The model includes features from the Roy model (productivity differences across workers), search frictions, general human capital acquired on the job, and compensating differentials. The model includes endogenous wage determination.

We add to the literature by investigating non-parametric identification of the model given revealed preferences from job-to-job transitions and wages. We show that almost all aspects of the model are identified. Two important exceptions are bargaining power and wages in jobs that workers would never take. The later is expected, while the former was unanticipated by us, but comes from the fact that revealed preferences can identify the preference order, but not the differences in utility.

We estimate the model on Danish matched employer-employee data using indirect inference. The model fit is very good. Using the estimated model we show that Roy model heterogeneity is the most important driver behind wage variation explaining between 64% to 83% of the total. Search friction, human capital, and compensating differentials matter to a lesser extent. These factors also interact in important ways.

Simulations show that compensating differentials and search friction are both important for other aspects of the data. One third of all encounters between two firms and a worker would have changed transition path if workers only cared about wages and not the non-pecuniary aspect of a job. Worker would on average earn 0.20 log point high wages if they did not care about non-pecuniary aspects of the job but only about wages and 0.17 log points higher if there were no search frictions. We show that variance in utility is primary driven by search frictions and compensating differentials and not Roy model heterogeneity. These conclusions in general hold across education-gender subgroups.

We should also emphasize that we have intentionally kept the model very simple. Computation of this model is quite fast so adding other features into this framework is straight forward computationally (although likely difficult to identify). Two that come to mind are job specific human capital and non-pecuniary costs of switching jobs. We conclude that future research that wants to understand wage determination and labor market transitions should worry about all four features.

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.
- Andersen, T. and M. Svarer (2007). Flexicurity - Labour Market Performance in Denmark. CE-Sifo Working Paper Series 2108, CESifo Group Munich.
- Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008). High wage workers and low wage firms: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society A*(2008) 171(Part 3), 673–697.
- Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics.
- Bagger, J. and R. Lentz (2015). An equilibrium model of wage dispersion and sorting. Working Paper.
- Barlevy, G. (2008, January). Identification of search models using record statistics. *Review of Economic Studies* 75(1).
- Becker, D. (2009). Non-wage job characteristics and the case of the missing margin. Federal Trade Commission.
- Bertheau, A., H. Bunzel, M. Hejlesen, and R. Vejlin (2016). Labor market transitions over the business cycle. Unpublished Working Paper, Department of Economics, Aarhus University.
- Bobbio, E. (2010). The Danish Matched Employer-Employee Data. *Unpublished Manuscript*.
- Boeri, Tito, A. B. and L. Calmfors (2001). *The role of unions in the twenty-first century*. Oxford University Press.
- Bunzel, H. (2010). The LMDG Data Sets. *Unpublished Manuscript*.
- Burdett, K. and M. Coles (2003). Equilibrium wage-tenure contracts. *Econometrica* 71(5), 1377–1404.
- 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.

- Card, D., A. Cardoso, and P. Kline (2014). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women.
- Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality. *Quartely Journal of Economics* 128(3), 967–1015.
- Dahl, C. M., D. le Maire, and J. R. Munch (2013). Wage dispersion and decentralization of wage bargaining. *Journal of Labor Economics* 31(3), 501–533.
- 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.
- Doniger, C. L. (2015, August). Wage dispersion with heterogeneous wage contracts. Finance and economics discussion series 2015-023. washington: Board of governors of the federal reserve system.
- 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.
- Fontaine, F. and R. Vejin (2016). Who benefits from the unemployment insurance?
- French, E. and C. Taber (2011). Identification of models of the labor market. In D. C. O. Ashenfelter (Ed.), *The Handbook of Labor Economics Volume 4A*.
- Gourieroux, C., A. Monfort, and E. Renault (1993, December). Indirect inference. *Journal of Applied Econometrics* 8, S85–S118.
- Gruetter, M. and R. Lalive (2004). The importance of firms in wage determination. IEW - Working Papers 207, Institute for Empirical Research in Economics - IEW.
- 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.
- Han, J. (2016, April). Multidimensional learning and earnings dynamics.
- Heckman, J. (1990). Varieties of selection bias. *American Economic Review, Papers and Proceedings* 80.
- Heckman, J. and B. Honore (1989). The identifiability of the competing risks mode. *Biometrika* 76, 325–330.

- 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. and B. Singer (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52, 271–320.
- Heckman, J. J. and B. E. Honoré (1990). The empirical content of the roy model. *Econometrica* 58(5), 1121–1149.
- Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). *Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond*, Volume 1, Chapter 7, pp. 307–458. Amsterdam: North Holland.
- Heckman, J. J. and C. R. Taber (2008). Chapter Roy model. Palgrave Macmillan.
- Huggett, M., G. Ventura, and A. Yaron (2011). Sources of lifetime inequality. *American Economic Review* 101(7), 2923–54.
- 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.
- Lentz, R. (2010). Sorting by search intensity. *Journal of Economic Theory* 145(4), 1436 – 1452. Search Theory and Applications.
- Lise, J., C. Meghir, and J.-M. Robin (2013, January). Mismatch, sorting and wage dynamics. Working Paper 18719, National Bureau of Economic Research.
- Lise, J., C. Meghir, and J.-M. Robin (2016). Matching, sorting and wages. *Review of Economic Dynamics* 19, 63 – 87. Special Issue in Honor of Dale Mortensen.
- Lu, C. (2016, January). Wages, career progress, and multidimensional skills: What do recessions tell us?
- Lund, C. and R. Vejlin (2015). Documenting and improving the hourly wage measure in the danish ida database.
- Postel-Vinay, F. and J.-M. Robin (2002, November). Equilibrium Wage Dispersion with Worker and Employer Heterogeneity. *Econometrica* 70(6), 2295–2350.
- Postel-Vinay, F. and J.-M. Robin (2004). To match or not to match?: Optimal wage policy with endogenous worker search intensity. *Review of Economic Dynamics* 7(2), 297 – 330.

- Postel-Vinay, F. and H. Turon (2010). On-the-job search, productivity shocks, and the individual earnings process*. *International Economic Review* 51(3), 599–629.
- Roberson, R. and R. Shimer (2011). Search in macroeconomic models of the labor market. In D. C. O. Ashenfelter (Ed.), *The Handbook of Labor Economics Volume 4A*.
- Robin, J.-M. (2011). On the dynamics of unemployment and wage distributions. *Econometrica* 79(5), 1327–1355.
- Rogerson, R., R. Shimer, and R. Wright (2005). Search-theoretic models of the labor market: A survey. *Journal of Economic Literature* XLIII, 959–988.
- 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.
- Sanders, C. (2014, August). Skill accumulation, skill uncertainty, and occupational choice.
- Sanders, C. and C. Taber (2012). Life-cycle wage growth and heterogeneous human capital. *Annual Review of Economics* 4.
- Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of the Nations*. W. Strahan and T. Cadell.
- Sørensen, T. and R. Vejlin (2013). The importance of worker, firm and match effects in the formation of wages. *Empirical Economics* 45(1), 435–464.
- Sorkin, I. (2015, October). Ranking firms using revealed preference.
- 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. To (2014). Search and nonwage job characteristics.
- Topel, R. (1991). Specific capital, mobility, and wages: wages rise with job seniority. *The Journal of Political Economy* 99(1), 145–176.
- Villanueva, E. (2007). Estimating compensating wage differentials using voluntary job changes: Evidence from germany. *Industrial and Labor Relations Review* 60(4), 544–561.
- Weiss, Y. (1986). The determination of lifecycle earnings: A survey. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 1, pp. 603–640. Amsterdam: North-Holland.

Appendix A: Identification

This section is composed of three different parts. We first consider what can be identified without information on wages. We then show what can be additionally identified with wage information. Finally, we show that the bargaining parameter, β , is not identified.

Identification of parameters determining turnover (without using wages)

As mentioned in the text, we assume that workers begin their working life non-employed and that we observe all data on workers from time 0 to T . To identifying the full distribution of δ_i non-parametrically we need to observe completed spells so we assume that we can let T go to some arbitrarily large number. Since we allow for the possibility of large T anyway, we use it for some other arguments. We expect (but have not showed explicitly) that if we eliminated or restricted heterogeneity in δ_i we could place an upper bound on the length of the panel needed. We observe labor force status during a worker's entire life - that is whether the worker is working and if so, the and establishment for which they work. We assume that we only observe wages at a finite number of times. For simplicity assume it is at the integers ($t = 1.0, 2.0, \dots$). Both the assumption on continuous observable labor market status and on wages being only observable at integers match the Danish data, where we have labor market status at a weekly basis and where wages are only observed for the job in the last week of November. Let w_{it} denote the log of the wage measured at time t for individual i working at time t . Human capital takes on only two values 0 and 1.

As mentioned in the text, we can classify workers into different "types" by their preferences over jobs. For 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 by not allowing workers to change their ordering after human capital accumulates. We see no reason why this would be an issue, but focus on this case for simplicity. It will be true for our empirical specification. The five types are:

Type	
0	$V_{iAh}(\pi_{iA}) < V_{i0h}, V_{iBh}(\pi_{iB}) < V_{i0h}$
B0	$V_{iAh}(\pi_{iA}) < V_{i0h} < V_{iBh}(\pi_{iB})$
A0	$V_{iBh}(\pi_{iB}) < V_{i0h} < V_{iAh}(\pi_{iA})$
BA	$V_{i0h} < V_{iAh}(\pi_{iA}) < V_{iBh}(\pi_{iB})$
AB	$V_{i0h} < V_{iBh}(\pi_{iB}) < V_{iAh}(\pi_{iA})$

for $h = \{0, 1\}$.

We start by showing that we can identify $\lambda_A^n, \lambda_B^n, \lambda_A^e, \lambda_B^e, P^*$, the sample probabilities of the five types above, and the distribution of δ_i (denote it F_δ) without using data on wages. The biggest complication is the addition of P^* because when we observe a job-to-job-transition we do not know whether it was voluntary or involuntary.

Identification of λ_A^n and λ_B^n

Condition on workers who are initially non-employed, start at an A type firm, become non-employed, start at a B type firm, become non-employed, and then start at another firm. We know these are either AB types or BA types. The probability that the third firm is a B type firm is

$$P_B \equiv \frac{\lambda_B^n}{\lambda_A^n + \lambda_B^n}.$$

We define P_A in an analogous manner. Let T_1 be the duration of the first non-employment spell, T_2 be the duration of the first employment spell, T_3 the duration of the second non-employment spell, T_4 the duration of the second employment spell, and T_5 the duration of the third non-employment spell. For any value $(t_1^a, t_1^b, t_2, t_3, t_4, t_5)$ we can identify

$$\frac{\Pr(T_1 > t_1^a, T_2 \leq t_2, T_3 \leq t_3, T_4 \leq t_4, T_5 \leq t_5)}{\Pr(T_1 > t_1^b, T_2 \leq t_2, T_3 \leq t_3, T_4 \leq t_4, T_5 \leq t_5)} = \frac{e^{-(\lambda_A^n + \lambda_B^n)t_1^a}}{e^{-(\lambda_A^n + \lambda_B^n)t_1^b}}.$$

Clearly, $\lambda_A^n + \lambda_B^n$ is identified as long as we pick values such that $t_1^a \neq t_1^b$. From this sum and P_B we can identify λ_A^n and λ_B^n .

We should point out that the fact that we observe the identity of the firm facilitates identification. As mentioned in the text, in general we know from Flinn and Heckman (1982) that separating arrival rates from reservation wages can not be done non-parametrically. In the actual empirical application we do not use the identity of the firms, so we identify this only after normalizing the location of the taste for non-employment.

Identification of P^*

Next, we establish identification of P^* . This is the probability of getting an offer from a firm immediately after a spell is terminated conditional on being terminated. As a practical manner in

the estimation we use additional information as we observe the fraction of job-to-job transitions that are voluntary directly from survey data. Here we show that we can identify P^* without that knowledge.⁴¹

This is simplified by conditioning on people who we know are either AB types or BA types. We do this by conditioning on individuals who start at a type B job, leave to non-employment and then start a type A job at t_1 . We will derive the hazards of those moving from A to B and from A to non-employment at time $t_1 + \tau$. Since they have accepted offers from both jobs we know that they are either members of the AB group and the BA group. The events that put one into this conditioning set are independent of type, so the relative proportion of AB types to BA types will be the same as it is in the population. We have assumed that the distribution of δ_i is independent of types, so it is the same for both the AB types and the BA types. Since we have conditioned on being at a B job beforehand and then leaving to non-employment, we have a selected sample of the distribution of δ_i which we define to be F_c .

Note that AB types (preferring A to B and B to non-employment) could move directly from A to B only if their job was destroyed and they immediately got an offer from a B type firm. However, for the BA types there are two ways that this transition could happen, either they are laid off and rehired or they receive an offer directly. The other key difference is that the only way that the BA types would still be working at the original type A firm at time $t_1 + \tau$ is if they had not received an outside offer from a type B firm. Putting these together the hazard rate out of job A and into job B is

$$\frac{P(AB) \int e^{-\delta_i \tau} \delta_i P^* P_B dF_c(\delta_i) + P(BA) e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} [\delta_i P^* P_B + \lambda_B^e] dF_c(\delta_i)}{P(AB) \int e^{-\delta_i \tau} dF_c(\delta_i) + P(BA) e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} dF_c(\delta_i)}$$

We can also identify the conditional hazard of movements to non-employment at time $t_1 + \tau$. This is

$$\frac{P(AB) \int e^{-\delta_i \tau} \delta_i [1 - P^*] dF_c(\delta_i) + P(BA) e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} \delta_i [1 - P^*] dF_c(\delta_i)}{P(AB) \int e^{-\delta_i \tau} dF_c(\delta_i) + P(BA) e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} dF_c(\delta_i)}$$

where $P(AB)$ and $P(BA)$ are the probabilities of being an AB or BA type worker, respectively. Intuitively, the model is identified from these equations because our only parameters are essentially $P^*, \lambda_B, P(AB)$ and $P(BA)$ (because F_c is identified if P^* is known), but we can move τ continuously. To show this formally we assume that δ_i is bounded from below by δ^ℓ to avoid strange behavior of F_δ near $\delta_i = 0$.

To see how to identify P^* we take the ratio of these two hazards and take the limit at $\tau \rightarrow \infty$

⁴¹Doing it this way is easier than using the other information as the expression is very complicated.

$$\begin{aligned}
& \lim_{\tau \rightarrow \infty} \frac{P(AB) \int e^{-\delta_i \tau} \delta_i P^* P_B dF_c(\delta_i) + P(BA) e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} [\delta_i P^* P_B + \lambda_B^e] dF_c(\delta_i)}{P(AB) \int e^{-\delta_i \tau} \delta_i [1 - P^*] dF_c(\delta_i) + P(BA) e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} \delta_i [1 - P^*] dF_c(\delta_i)} \\
&= \lim_{\tau \rightarrow \infty} \frac{P(AB) P^* P_B + P(BA) P^* P_B e^{-\lambda_B^e \tau} + P(BA) \lambda_B^e \frac{e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} dF_c(\delta_i)}{\int e^{-\delta_i \tau} \delta_i dF_c(\delta_i)}}{P(AB) [1 - P^*] + P(BA) [1 - P^*] e^{-\lambda_B^e \tau}} \\
&= \frac{P^* P_B}{[1 - P^*]}
\end{aligned}$$

Since P_B is identified, P^* is identified.⁴²

The intuition for this result is that as τ gets large all of the BA types should have gotten a B offer by then so that all that is left is AB types and the only way that they make a transition to a B firm is through a displacement. Comparing that to their exit rate to non-employment gives P^* . As mentioned above, since we need to assume that we do not have right truncated spells for identification of δ_i we use it here for convenience, but the intuition of the argument is that as τ gets larger the proportion of AB in the sample shrinks making involuntary switches relatively more important which identifies P^* . It does not need to shrink all the way to zero - it just makes the expression simpler if it does.

Identification of $P(AB)/P(BA)$, λ_A^e , and λ_B^e

Using the same conditioning as above, once P^* has been identified it is trivial to identify F_c because the hazard rate to non-employment for this group is $\delta_i [1 - P^*]$. Here it is important that we have no right truncated spells in which case we can identify the distribution of the log of length of the employment spell. This distribution is the convolution of the distribution of $\delta_i [1 - P^*]$ and an extreme value distribution (see e.g. Heckman and Honore, 1989, or French and Taber, 2011), so the distribution of $\delta_i [1 - P^*]$ and thus F_c is identified.

To identify λ_B^e and $P(AB)/P(BA)$ note that we can identify the probability that someone is still employed at the type A firm at time $t_1 + \tau$ (all conditional on having worked at a B firm). This is

$$\frac{P(AB) \int e^{-\delta_i \tau} dF_c(\delta_i) + P(BA) e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} dF_c(\delta_i)}{P(AB) + P(BA)}.$$

Since F_c is identified,

$$\frac{P(AB) + P(BA) e^{-\lambda_B^e \tau}}{P(AB) + P(BA)}$$

is as well.

⁴²We used the fact that δ_i is bounded from below for $0 < \lim_{\tau \rightarrow \infty} \frac{e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} dF_c(\delta_i)}{\int e^{-\delta_i \tau} \delta_i dF_c(\delta_i)} < \lim_{\tau \rightarrow \infty} \frac{e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} dF_c(\delta_i)}{\delta^\ell \int e^{-\delta_i \tau} dF_c(\delta_i)} = 0$.

Evaluate this letting τ go to ∞ and $P(AB) / (P(AB) + P(BA))$ is identified and thus so is $P(BA) / (P(AB) + P(BA))$. Given this λ_B^e is identified at any other value of τ .

We can identify λ_A^e by using an analogous argument.

Identification of $P(0)$, $P(A0)$, and $P(B0)$

Note that $P(0)$ is identified directly from the data as those that never work.

To identify the $P(A0)$ group we use the same argument we used for identification of P^* except that we no longer condition on having a B spell prior to the A spell. We condition on all individuals who's first spell is a type A spell which starts at t_1 and condition on how it ends. Now three types can experience the A spell.

The hazard rate of moving to B at time $t_1 + \tau$ conditional on starting at an A type firm and remaining at that firm is⁴³

$$\frac{P(AB) P_A \int e^{-\delta_i \tau} \delta_i P^* P_B dF(\delta_i) + P(BA) P_A e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} [\delta_i P^* P_B + \lambda_B^e] dF(\delta_i)}{P(AB) P_A \int e^{-\delta_i \tau} dF(\delta_i) + P(BA) P_A e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} dF_c(\delta_i) + P(A0) \int e^{-\delta_i \tau} dF_c(\delta_i)}$$

We can also identify the conditional hazard of movements to non-employment at time $t_1 + \tau$.

This is

$$\frac{[P(AB) + P(BA)e^{-\lambda_B^e \tau}] P_A \int e^{-\delta_i \tau} \delta_i [1 - P^*] dF(\delta_i) + P(A0) \int e^{-\delta_i \tau} \delta_i [1 - P^* P_A] dF(\delta_i)}{[P(AB) + P(BA)e^{-\lambda_B^e \tau}] P_A \int e^{-\delta_i \tau} dF_c(\delta_i) + P(A0) \int e^{-\delta_i \tau} dF_c(\delta_i)}$$

As above we take the ratio between these and send $\tau \rightarrow \infty$

$$\begin{aligned} & \lim_{\tau \rightarrow \infty} \frac{P(AB) P_A \int e^{-\delta_i \tau} \delta_i P^* P_B dF(\delta_i) + P(BA) P_A e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} [\delta_i P^* P_B + \lambda_B^e] dF(\delta_i)}{[P(AB) + P(BA)e^{-\lambda_B^e \tau}] P_A \int e^{-\delta_i \tau} \delta_i [1 - P^*] dF(\delta_i) + P(A0) \int e^{-\delta_i \tau} \delta_i [1 - P^* P_A] dF(\delta_i)} \\ &= \lim_{\tau \rightarrow \infty} \frac{P(AB) P_A P^* P_B + P(BA) P_A e^{-\lambda_B^e \tau} + P(BA) P_A \frac{e^{-\lambda_B^e \tau} \int e^{-\delta_i \tau} \lambda_B^e dF(\delta_i)}{\int e^{-\delta_i \tau} \delta_i dF(\delta_i)}}{[P(AB) + P(BA)e^{-\lambda_B^e \tau}] P_A [1 - P^*] e^{-\lambda_B \tau} + P(A0) [1 - P^* P_A]} \\ &= \frac{P(AB) P_A P^* P_B}{P(AB) P_A [1 - P^*] + P(A0) [1 - P^* P_A]} \end{aligned}$$

We have shown that everything in this expression apart from $P(A0)/P(AB)$ is identified, thus $P(A0)/P(AB)$ is identified.

An analogous argument gives $P(B0)/P(AB)$. Knowledge of $P(0)$, $P(AB)/P(BA)$, $P(A0)/P(AB)$, and $P(B0)/P(AB)$ is sufficient to identify all of the pieces.

⁴³Note that we have used the fact that the probability of being an AB type conditional on an original A spell is $P(AB) P_A / (P(AB) P_A + P(BA) P_A + P(A0))$ with similar expressions for the conditional probability of being a BA type. The denominator in this expression cancels.

Identification of distribution of unconditional distribution of δ_i

Continue to condition on individuals whose first spell is A . We showed how to identify F_c above, but we need to show identification of the unconditional distribution of δ_i using a similar argument. We can identify the distribution of the hazard rates to non-employment and thus the distribution of the logs of the hazard rates to non-employment.⁴⁴ Thus, we can identify the distribution of

$$\begin{aligned} & \frac{P(A0)}{P(A0) + P_A(P(AB) + P(BA))} [\log(\delta_i) + \log(1 - P^*P_A)] \\ & + \frac{P_A(P(AB) + P(BA))}{P(A0) + P_A(P(AB) + P(BA))} [\log(\delta_i) + \log(1 - P^*)] \end{aligned}$$

Since everything except δ_i in this expression is identified, we can identify the distribution of δ_i .

So far we have used only information about turnover and no information about wages. We have shown that with this information we are able to identify transition rates and type probabilities.

Identification of Wage Components

We now incorporate information from wages. As a reminder of the notation, 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. We will also use the notation $r_{ij\ell h} = \log(R_{ij\ell h}\psi_h)$. Then for each integer t at which time the agent is working, we observe

$$w_{it} \equiv r_{ij(i,t)\ell(i,t)h_0(i,t)} + \log\left(\frac{\psi_{h(i,t)}}{\psi_{h_0(i,t)}}\right) + \zeta_{it}$$

where ζ_{it} is i.i.d. measurement error. One can see here the distinction between $h(i, t)$ and $h_0(i, t)$.

Identification of Distribution of Measurement Error (ζ_{it})

First, we identify the distribution of measurement error and then the arrival rate of human capital, λ_h . We condition on a group who

- Are non-employed until time $1 - d_1$
- Start working in job A at time $1 - d_1$ and leaves to non-employment at $1 + d_2$

⁴⁴Note that in this case we are using the distribution of the employment spell only, not the job spell. It is only the hazard rate to non-employment.

- Are non-employed until time $2 - d_3$ when they start again at a type A firm and they stay through period 2

We assume that the d_j 's are sufficiently small, so spells do not overlap. From this we can identify the joint distribution of

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

conditional on the events above for alternative values of d_1, d_2 , and d_3 .

Taking limits of the above object as $d_1 \downarrow 0, d_2 \downarrow 0$, and $d_3 \downarrow 0$, we can identify the conditional distribution of

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

for our conditioning group. Without the measurement error, both of these wages will correspond to r_{iA00} because the workers 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 r_{iA00} . We need to additionally assume that the characteristic function of the measurement error, $\phi_\xi(s)$, does not vanish.

Identification of λ_h

Next, we show how to identify λ_h . To economize on notation we will use $E(\cdot | d)$ to denote the expectation conditional on the events described above at values of $d = (d_1, d_2, d_3)$. We use the same conditioning group as in the Measurement Error section and continue to send $d_1 \downarrow 0$ and $d_3 \downarrow 0$, but allow d_2 to vary. We can identify

$$\lim_{d_1, d_3 \downarrow 0} \frac{E(e^{i\zeta w_{i2}} | d)}{\phi_\zeta(s)} = \lim_{d_1, d_3 \downarrow 0} \left[e^{-\lambda_h d_2} E(e^{i\zeta r_{iA00}} | d) + (1 - e^{-\lambda_h d_2}) E(e^{i\zeta r_{iA01}} | d) \right].$$

By varying d_2 we can identify λ_h .⁴⁵ Intuitively, varying d_2 varies the time that the worker has to get a human capital shock.

⁴⁵To see how, take the ratio of the derivatives of this function in terms of d_2 at two different values of d_2 and it will be a known function of λ_h . First note that the derivative with respect to d_2 is

$$\begin{aligned} \lim_{d_1, d_3 \downarrow 0} \left[-\lambda_h e^{-\lambda_h d_2} E(e^{i\zeta r_{iA00}} | d) + \lambda_h e^{-\lambda_h d_2} E(e^{i\zeta r_{iA01}} | d) \right] &= \lambda_h e^{-\lambda_h d_2} [E(e^{i\zeta r_{iA00}} | A) - E(e^{i\zeta r_{iA01}} | A)] \\ &= \lambda_h e^{-\lambda_h d_2} [E(e^{i\zeta r_{iA00}} | A) - E(e^{i\zeta r_{iA01}} | A)] \end{aligned}$$

where the notation $E(\cdot | A)$ means the expected value conditional on taking an A job first. Now take the ratio of this at two different values of d_2 say d_2^a and d_2^b then

$$\begin{aligned} \Delta(d_2^a, d_2^b) &\equiv \frac{\lambda_h e^{-\lambda_h d_2^a} [E(e^{i\zeta r_{iA00}} | A) - E(e^{i\zeta r_{iA01}} | A)]}{\lambda_h e^{-\lambda_h d_2^b} [E(e^{i\zeta r_{iA00}} | A) - E(e^{i\zeta r_{iA01}} | A)]} \\ &= e^{\lambda_h (d_2^b - d_2^a)}. \end{aligned}$$

Identification of ψ_1 and the Distribution of Wages for the AB types

We now consider identification of ψ_1 and demonstrate identification of the full wage distribution for the AB type. Identification of the ladder is complicated, so we will do this stepwise by showing identification of expanding subsets of the full distribution.

For the AB types these are the seventeen different labor market statuses possible

Table AA1
Labor Market Statuses for AB workers

$j(i, t)$	$h(i, t)$	$\ell(i, t)$	$h_0(i, t)$	Wage
A	0	0	0	R_{iA00}
A	0	B	0	R_{iAB0}
A	0	A	0	π_{iA}
A	1	0	0	$R_{iA00}\psi_1$
A	1	B	0	$R_{iAB0}\psi_1$
A	1	A	0	$\pi_{iA}\psi_1$
A	1	0	1	$R_{iA01}\psi_1$
A	1	B	1	$R_{iAB1}\psi_1$
A	1	A	1	$\pi_{iA}\psi_1$
B	0	0	0	R_{iB00}
B	0	B	0	π_{iB}
B	1	0	0	$R_{iB00}\psi_1$
B	1	B	0	$\pi_{iB}\psi_1$
B	1	0	1	$R_{iB01}\psi_1$
B	1	B	1	$\pi_{iB}\psi_1$
0	0	NA	NA	NA
0	1	NA	NA	NA

From the table one can see that for an AB worker wages depend on the joint distribution of eight objects (in addition to ψ_1)

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

While there are multiple ways to show identification, we focus on a particular set of transitions. We assume that workers start their labor market career in non-employed and receive their first job at $1 - d_1$. The following table shows the transition path.

$\Delta(d_2^a, d_2^b)$ is directly identified from the data and

$$\lambda_h = \frac{\log(\Delta(d_2^a, d_2^b))}{d_2^b - d_2^a}.$$

Transition	Time
Start at A	$1 - d_1$
Move to non-employment	$1 + d_2$
Start at B	$2 - d_3$
Move to non-employment	$2 + d_4$
Start at B	$3 - d_5 - d_6$
Move to A	$3 - d_6$
Move to non-employment	$3 + d_7$
Start at A	$4 - d_8$
Move to non-employment	$4 + d_9$
Start at B	$6 - d_{10}$
Move to non-employment	$6 + d_{11}$
Start at B	$8 - d_{12} - d_{13}$
Move to A	$8 - d_{13}$
Move to B	$8 + d_{14}$

with $d_j \geq 0$ for $j = 1, \dots, 14$. We also assume that the d_j 's are sufficiently small such that the above spells do not overlap. The goal here will be to look at the joint distribution of wages conditional on the d_j 's. Analogous to above, we use the notation $E[\cdot | d]$ to mean the conditional expectation conditioning on events occurring at times denoted by $d_1 - d_{14}$.

In going forward we condition on functions of wages from the first eight periods $f(w_{i1}, \dots, w_{i8})$. For any function $f(w_{i1}, \dots, w_{i8})$ notice that

$$E(f(w_{i1}, \dots, w_{i8}) | d) = P[AB | d] E[f(w_{i1}, \dots, w_{i8}) | d, AB] \\ + P[BA | d] E[f(w_{i1}, \dots, w_{i8}) | d, BA].$$

The last value d_{14} will play a crucial roll in distinguishing between these expressions. As it is not realized until after period 8 it does not affect either $E[f(w_{i1}, \dots, w_{i8}) | d, AB]$ or $E[f(w_{i1}, \dots, w_{i8}) | d, BA]$. However, it does influence $P[AB | d]$ because a BA can move from A to B directly either because they got an outside offer from a B firm or because the they were laid off and got an immediate offer. For an AB type it can only be due to the latter event. It is straight forward to show that $P[AB | d]$ is a known increasing function of d_{14} .⁴⁶ Notice then for any distribution of wages, $f(\cdot)$, and any values of d_1, \dots, d_{13} , by moving d_{14} we can separately identify $E[f(w_{i1}, \dots, w_{i8}) | d, AB]$ from $E[f(w_{i1}, \dots, w_{i8}) | d, BA]$. We refrain from making this argument repeatedly but just condition on types implicitly assuming that $E[f(w_{i1}, \dots, w_{i8}) | d, AB]$ is identified. In what follows we will

⁴⁶It is

$$P[AB | d] = \frac{P(AB) E\left(e^{-(\lambda_A^e + \delta_i)(d_5 + d_{12}) - \delta_i(d_{13} + d_{14})} [\delta_i P^* P_A + \lambda_A^e]^2 \delta_i P^* P_B\right)}{P(AB) E\left(e^{-(\lambda_A^e + \delta_i)(d_5 + d_{12}) - \delta_i(d_{13} + d_{14})} [\delta_i P^* P_A + \lambda_A^e]^2 \delta_i P^* P_B\right) \\ + P(BA) E\left(e^{-\delta_i(d_5 + d_{12}) - (\lambda_B^e + \delta_i)(d_{13} + d_{14})} [\delta_i P^* P_A]^2 [\delta_i P^* P_B + \lambda_B^e]\right)}$$

start with a subset of the full joint distribution, $(R_{iA00}, R_{iAB0}, \pi_{iA}, R_{iB00}, \pi_{iB}, R_{iA01}, R_{iAB1}, R_{iB01})$, and show identification. We will then increase the subset until we reach the full joint distribution.

Identification of joint distribution of $(R_{iA00}, R_{iB00}, R_{iAB0})$ for the AB types

We start by sending $d_1 \dots d_{13} \downarrow 0$ and look at the joint distribution of (w_{i1}, w_{i2}, w_{i3}) . A complication is that in period 3, individuals who moved directly from B to A could have either have gotten an outside offer from an A firm or been laid off and found a new job immediate. The probability that it is a voluntary transition (the former case) is defined to be $\rho_3(d)$ which is a complicated expression involving transition parameters which we have shown are identified. Since $\rho_3(d)$ is a known expression of parameters we have shown it is identified.

Then for any values of $s_1 - s_3$ we can identify

$$\begin{aligned} & \lim_{d_1, \dots, d_{13} \downarrow 0} \frac{E \exp [i (s_1 w_{i1} + s_2 w_{i2} + s_3 w_{i3}) \mid d, AB]}{\phi_{\bar{\zeta}}(s_1) \phi_{\bar{\zeta}}(s_2) \phi_{\bar{\zeta}}(s_3)} \\ &= \left[\lim_{d_1, \dots, d_{13} \downarrow 0} \rho_3(d) \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} \mid AB)] \\ & \quad + \left[\lim_{d_1, \dots, d_{13} \downarrow 0} (1 - \rho_3(d)) \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} \mid AB)]. \end{aligned}$$

We will use the same basic argument for identification of the model throughout this section. We will be explicit about it here, but not as explicit in what follows (which will involve many more terms).

1. $\lim_{d_1, \dots, d_{13} \downarrow 0} \rho_3(d)$ is identified as it is a known function of parameters that we have shown are identified.
2. By setting $s_3 = 0$ we can identify $E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} \mid AB)]$ from the expression above.
3. Once this is identified, $E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} \mid AB)]$ is identified as we vary s_3 .
4. Everything in the expression above is identified except $E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} \mid AB)]$ so we can solve for this expression as well.
5. $E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} \mid AB)]$ is the characteristic function of $(r_{iA00}, r_{iB00}, r_{iAB0})$, so if this is identified, the joint distribution of $(r_{iA00}, r_{iB00}, r_{iAB0})$ is identified.

Identification of $(R_{iA00}, R_{iB00}, R_{iAB0}, R_{iA01}, R_{iB01}, R_{iAB1})$ for the AB types

Now we extend the argument to include the joint distribution of

$$(R_{iA00}, R_{iB00}, R_{iAB0}, R_{iA01}, R_{iB01}, R_{iAB1})$$

for the AB types by adding periods 4, 6, and 8. The key here is that we will now vary d_7 which will allow the possibility that human capital evolves between time 3 and $3 + d_7$. We keep $d_7 < 1$. Note that there are now 8 possible events that can occur in the data; (human capital evolves or does not evolve between period 3 and $3 + d_7$) \times (the job-to-job transition to job A at time 3 is voluntary/involuntary) \times (the job-to-job transition to job A at time 8 is voluntary/involuntary). Let $\rho_3(d)$ and $\rho_8(d)$ be the conditional probability that the job-to-job transitions are voluntary at time 3 and 8, respectively. These are identified as is their limit. To simplify the following expression define

$$\rho_3^{\ell_1} \equiv \lim_{d_1, \dots, d_6, d_8, \dots, d_{13} \downarrow 0} \rho_3(d)$$

$$\rho_8^{\ell_1} = \lim_{d_1, \dots, d_6, d_8, \dots, d_{13} \downarrow 0} \rho_8(d)$$

and note that these values are identified. For any value of $s_1 - s_6$ we can identify

$$\lim_{d_1, \dots, d_6, d_8, \dots, d_{13} \downarrow 0} \frac{E \exp [i (s_1 w_{i1} + s_2 w_{i2} + s_3 w_{i3} + s_4 w_{i4} + s_5 w_{i6} + s_6 w_{i8}) \mid d, AB]}{\phi_{\xi}(s_1) \phi_{\xi}(s_2) \phi_{\xi}(s_3) \phi_{\xi}(s_4) \phi_{\xi}(s_5) \phi_{\xi}(s_6)}$$

$$= e^{-\lambda_h d_7} \left[\rho_3^{\ell_1} \rho_8^{\ell_1} \right] E \exp [i ((s_1 + s_4) r_{iA00} + (s_2 + s_5) r_{iB00} + (s_3 + s_6) r_{iAB0}) \mid AB]$$

$$+ e^{-\lambda_h d_7} \left[\rho_3^{\ell_1} (1 - \rho_8^{\ell_1}) \right] E \exp [i ((s_1 + s_4 + s_6) r_{iA00} + (s_2 + s_5) r_{iB00} + s_3 r_{iAB0}) \mid AB]$$

$$+ e^{-\lambda_h d_7} \left[(1 - \rho_3^{\ell_1}) \rho_8^{\ell_1} \right] E \exp [i ((s_1 + s_3 + s_4) r_{iA00} + (s_2 + s_5) r_{iB00} + s_6 r_{iAB0}) \mid AB]$$

$$+ e^{-\lambda_h d_7} \left[(1 - \rho_3^{\ell_1}) (1 - \rho_8^{\ell_1}) \right] E \exp [i ((s_1 + s_3 + s_4 + s_6) r_{iA00} + (s_2 + s_5) r_{iB00}) \mid AB]$$

$$+ (1 - e^{-\lambda_h d_7}) \left[\rho_3^{\ell_1} \rho_8^{\ell_1} \right] \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + s_4 r_{iA01} + s_5 r_{iB01} + s_6 r_{iAB1}) \mid AB]$$

$$+ (1 - e^{-\lambda_h d_7}) \left[\rho_3^{\ell_1} (1 - \rho_8^{\ell_1}) \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_6) r_{iA01} + s_5 r_{iB01}) \mid AB]$$

$$+ (1 - e^{-\lambda_h d_7}) \left[(1 - \rho_3^{\ell_1}) \rho_8^{\ell_1} \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + s_4 r_{iA01} + s_5 r_{iB01} + s_6 r_{iAB1}) \mid AB]$$

$$+ (1 - e^{-\lambda_h d_7}) \left[(1 - \rho_3^{\ell_1}) (1 - \rho_8^{\ell_1}) \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_6) r_{iA01} + s_5 r_{iB01}) \mid AB]$$

We showed above that the first four expressions are identified. Thus we have four new expressions to identify

- (a) $E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + s_4 r_{iA01} + s_5 r_{iB01} + s_6 r_{iAB1}) \mid AB]$
- (b) $E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_6) r_{iA01} + s_5 r_{iB01}) \mid AB]$
- (c) $E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + s_4 r_{iA01} + s_5 r_{iB01} + s_6 r_{iAB1}) \mid AB]$
- (d) $E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_6) r_{iA01} + s_5 r_{iB01}) \mid AB]$

We use the same approach as above. If we evaluate at $s_3 = s_6 = 0$ these expressions are the same and thus $E [i (s_1 r_{iA00} + s_2 r_{iB00} + s_4 r_{iA01} + s_5 r_{iB01}) \mid AB]$ is identified. This identifies (d) for any values of $s_1 - s_6$.

Again using the same type of argument, given (d) , keeping $s_3 = 0$ but varying the other values of s_j identifies (c) and setting $s_6 = 0$ but varying the others gives (b) . Then everything in the large equation above is identified other than (a) , so it is identified by varying all values of s_j .

Identification of ψ_1

Next, we consider identification of ψ_1 which we can do from $E[\log(W_{i7}) | d, AB]$ alone. In order to do this we condition on $1 < d_{11} < 2$ so that we observe w_{i7} and we will vary d_2 but send the rest of the $d_j \downarrow 0$ (except d_{14} as usual). For this case there are three possibilities: human capital has not evolved before period 7, human capital evolves between time 1 and time $1 + d_2$, and human capital evolves between periods 6 and 7.

$$\begin{aligned}
& \lim_{d_1, d_3, \dots, d_{13} \downarrow 0} E[\log(w_{i7}) | d, AB] = \\
& e^{-\lambda_h[1+d_2]} \left(e^{-\lambda_B^c} E[r_{iB00} | AB] + \left(1 - e^{-\lambda_B^c}\right) E[\log(\pi_{iB}) | AB] \right) \\
& + \left(1 - e^{-\lambda_h d_2}\right) \left(e^{-\lambda_B^c} E[r_{iB01} | AB] + \left(1 - e^{-\lambda_B^c}\right) (E[\log(\pi_{iB}) | AB] + \log(\psi_1)) \right) \\
& + \left(e^{-\lambda_h d_2} - e^{-\lambda_h[1+d_2]} \right) \left(e^{-\lambda_B^c} E[r_{iB00} | AB] + \left(1 - e^{-\lambda_B^c}\right) E[\log(\pi_{iB}) | AB] + \log(\psi_1) \right) + E(\xi_{i7}) \\
& = e^{-\lambda_h d_2} e^{-\lambda_B^c} E[r_{iB00} | AB] + \left(1 - e^{-\lambda_h d_2}\right) e^{-\lambda_B^c} E[r_{iB01} | AB] + E(\xi_{i7}) \\
& + \left(1 - e^{-\lambda_A}\right) E[\log(\pi_{iB}) | AB] + \left[\left(1 - e^{-\lambda_h d_2}\right) \left(1 - e^{-\lambda_B^c}\right) + \left(e^{-\lambda_h d_2} - e^{-\lambda_h[1+d_2]} \right) \right] \log(\psi_1)
\end{aligned}$$

Everything is known in this expression except $E[\log(\pi_{iB}) | AB]$ and $\log(\psi_1)$, so by varying d_2 they can be separately identified.

Identification of $(R_{iA00}, R_{iAB0}, R_{iB00}, R_{iA01}, R_{iAB1}, R_{iB01}, \pi_{iA}, \pi_{iB})$ conditional on AB

Now we assume that $1 < d_{11} < 2$ and $1 < d_9 < 2$ such that we observe wages at time 5 and 7. Using the same argument as above by varying d_7 , we can identify the expected value of $f(w_{i1}, \dots, w_{i8})$ conditional on human capital arriving between time 3 and $3 + d_7$ (write this conditioning as $H_{i4} = 1$). To simplify the notation we define

$$\begin{aligned}
\omega_{iA} & \equiv \log(\pi_{iA}) \\
\omega_{iB} & \equiv \log(\pi_{iB}).
\end{aligned}$$

We will set the rest of the d_j 's close to zero (other than d_7, d_{11}, d_9 , and d_{14}). Since we condition on human capital arriving between period 3 and $3 + d_7$, we know that the wage in the first period will be approximately R_{iA00} , the second period R_{iB00} , the fourth R_{iA01} , and the sixth R_{iB01} . As before for the third and the eighth period the wage can take two values depending on whether the job-to-job transition was voluntary or not (R_{iA00} or R_{iAB0} in 3 and R_{iA01} or R_{iAB1} in 8). For period 5 the

wage can take 3 values depending on outside offers either R_{iA01} if no outside offers, R_{iAB1} if an offer from a B type only, or π_{iA} if an offer from an A type. Similarly in period 7 the wage can take 2 values depending on whether there was no outside offer (R_{iB01}) or an outside offer from a B firm (π_{iB}).⁴⁷ This gives a total of $2 \times 3 \times 2 \times 2 = 24$ different possibilities.

Analogous to above we define

$$\rho_3^{\ell_2} \equiv \lim_{d_1, \dots, d_6, d_8, d_{10}, d_{12}, d_{13} \downarrow 0} \rho_3(d)$$

$$\rho_8^{\ell_2} \equiv \lim_{d_1, \dots, d_6, d_8, d_{10}, d_{12}, d_{13} \downarrow 0} \rho_8(d)$$

Putting this together can identify the complicated expression with the relevant 24 terms.

⁴⁷Since we are considering AB types they could not have gotten an offer from an A firm or they would have left

$$\begin{aligned}
& \lim_{d_1, \dots, d_6, d_8, d_{10}, d_{12}, d_{13}, d_0} \frac{E \exp [i (s_1 w_{i1} + s_2 w_{i2} + s_3 w_{i3} + s_4 w_{i4} + s_5 w_{i5} + s_6 w_{i6} + s_7 w_{i7} + s_8 w_{i8})] | d, AB, H_{i4} = 1]}{\phi_\zeta (s_1) \phi_\zeta (s_2) \phi_\zeta (s_3) \phi_\zeta (s_4) \phi_\zeta (s_5) \phi_\zeta (s_6) \phi_\zeta (s_7) \phi_\zeta (s_8)} \\
&= e^{-(\lambda_A^c + \lambda_B^c)} e^{-\lambda_B^c} \left[\rho_3^{\ell_2} \rho_8^{\ell_2} \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_5) r_{iA01} + (s_6 + s_7) r_{iB01} + s_8 r_{iAB1})] | AB] \\
&+ e^{-(\lambda_A^c + \lambda_B^c)} \left(1 - e^{-\lambda_B^c} \right) \left[\rho_3^{\ell_2} \rho_8^{\ell_2} \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_5) r_{iA01} + s_6 r_{iB01} + s_7 \omega_{iB} + s_8 r_{iAB1})] | AB] \\
&+ \left(1 - e^{-\lambda_A^c} \right) e^{-\lambda_B^c} \left[\rho_3^{\ell_2} \rho_8^{\ell_2} \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + s_4 r_{iA01} + s_5 \omega_{iA} + (s_6 + s_7) r_{iB01} + s_8 r_{iAB1})] | AB] \\
&+ \left(1 - e^{-\lambda_A^c} \right) \left(1 - e^{-\lambda_B^c} \right) \left[\rho_3^{\ell_2} \rho_8^{\ell_2} \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + s_4 r_{iA01} + s_5 \omega_{iA} + s_6 r_{iB01} + s_7 \omega_{iB} + s_8 r_{iAB1})] | AB] \\
&+ e^{-\lambda_A^c} \left(1 - e^{-\lambda_B^c} \right) e^{-\lambda_B^c} \left[\rho_3^{\ell_2} \rho_8^{\ell_2} \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + s_4 r_{iA01} + (s_6 + s_7) r_{iB01})] | AB] \\
&+ e^{-\lambda_A^c} \left(1 - e^{-\lambda_B^c} \right) \left(1 - e^{-\lambda_B^c} \right) \left[\rho_3^{\ell_2} \rho_8^{\ell_2} \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + s_4 r_{iA01} + s_6 r_{iB01} + s_7 \omega_{iB} + (s_5 + s_8) r_{iAB1})] | AB] \\
&+ e^{-(\lambda_A^c + \lambda_B^c)} e^{-\lambda_B^c} \left[\rho_3^{\ell_2} \rho_8^{\ell_2} \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_5 + s_8) r_{iA01} + (s_6 + s_7) r_{iB01})] | AB] \\
&+ e^{-(\lambda_A^c + \lambda_B^c)} \left(1 - e^{-\lambda_B^c} \right) \left[\rho_3^{\ell_2} \rho_8^{\ell_2} \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_5 + s_8) r_{iA01} + s_7 \omega_{iB} + s_6 r_{iB01})] | AB] \\
&+ \left(1 - e^{-\lambda_A^c} \right) e^{-\lambda_B^c} \left[\rho_3^{\ell_2} \rho_8^{\ell_2} \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_8) r_{iA01} + s_5 \omega_{iA} + (s_6 + s_7) r_{iB01})] | AB] \\
&+ \left(1 - e^{-\lambda_A^c} \right) \left(1 - e^{-\lambda_B^c} \right) \left[\rho_3^{\ell_2} \rho_8^{\ell_2} \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_8) r_{iA01} + s_5 \omega_{iA} + s_6 r_{iB01} + s_7 \omega_{iB})] | AB] \\
&+ e^{-\lambda_A^c} \left(1 - e^{-\lambda_B^c} \right) e^{-\lambda_B^c} \left[\rho_3^{\ell_2} \rho_8^{\ell_2} \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_8) r_{iA01} + s_5 r_{iAB1} + (s_6 + s_7) r_{iB01})] | AB] \\
&+ e^{-\lambda_A^c} \left(1 - e^{-\lambda_B^c} \right) \left(1 - e^{-\lambda_B^c} \right) \left[\rho_3^{\ell_2} \rho_8^{\ell_2} \right] E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_8) r_{iA01} + s_5 r_{iAB1} + s_6 r_{iB01} + s_7 \omega_{iB})] | AB] \\
&+ e^{-(\lambda_A^c + \lambda_B^c)} e^{-\lambda_B^c} \left[\left(1 - \rho_3^{\ell_2} \right) \rho_8^{\ell_2} \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_5) r_{iA01} + (s_6 + s_7) r_{iB01} + s_8 r_{iAB1})] | AB] \\
&+ e^{-(\lambda_A^c + \lambda_B^c)} \left(1 - e^{-\lambda_B^c} \right) \left[\left(1 - \rho_3^{\ell_2} \right) \rho_8^{\ell_2} \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_5) r_{iA01} + s_6 r_{iB01} + s_7 \omega_{iB} + s_8 r_{iAB1})] | AB] \\
&+ \left(1 - e^{-\lambda_A^c} \right) e^{-\lambda_B^c} \left[\left(1 - \rho_3^{\ell_2} \right) \rho_8^{\ell_2} \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + s_4 r_{iA01} + s_5 \omega_{iA} + (s_6 + s_7) r_{iB01} + s_8 r_{iAB1})] | AB] \\
&+ \left(1 - e^{-\lambda_A^c} \right) \left(1 - e^{-\lambda_B^c} \right) \left[\left(1 - \rho_3^{\ell_2} \right) \rho_8^{\ell_2} \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + s_4 r_{iA01} + s_5 \omega_{iA} + s_6 r_{iB01} + s_7 \omega_{iB} + s_8 r_{iAB1})] | AB]
\end{aligned}$$

$$\begin{aligned}
& + e^{-\lambda_A^e} (1 - e^{-\lambda_B^e}) e^{-\lambda_B^e} \left[(1 - \rho_3^{\ell_2}) \rho_8^{\ell_2} \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + s_4 r_{iA01} + (s_5 + s_8) r_{iAB1} + (s_6 + s_7) r_{iB01}) \mid AB] \\
& + e^{-\lambda_A^e} (1 - e^{-\lambda_B^e}) (1 - e^{-\lambda_B^e}) \left[(1 - \rho_3^{\ell_2}) \rho_8^{\ell_2} \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + s_4 r_{iA01} + (s_5 + s_8) r_{iAB1} + s_6 r_{iB01} + s_7 \omega_{iB}) \mid AB] \\
& + (1 - e^{-\lambda_A^e}) e^{-\lambda_B^e} \left[(1 - \rho_3^{\ell_2}) (1 - \rho_8^{\ell_2}) \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_5 + s_8) r_{iA01} + (s_6 + s_7) r_{iB01}) \mid AB] \\
& + (1 - e^{-\lambda_B}) e^{-(\lambda_A + \lambda_B)} (1 - \rho_3^{\ell_2}) (1 - \rho_8^{\ell_2}) E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_5 + s_8) r_{iA01} + s_6 r_{iB01} + s_7 \omega_{iB}) \mid AB] \\
& + e^{-\lambda_B} (1 - e^{-\lambda_A}) \left[(1 - \rho_3^{\ell_2}) (1 - \rho_8^{\ell_2}) \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_8) r_{iA01} + s_5 \omega_{iA} + (s_6 + s_7) r_{iB01}) \mid AB] \\
& + (1 - e^{-\lambda_A^e}) (1 - e^{-\lambda_B^e}) \left[(1 - \rho_3^{\ell_2}) (1 - \rho_8^{\ell_2}) \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_8) r_{iA01} + s_5 \omega_{iA} + s_6 r_{iB01} + s_7 \omega_{iB}) \mid AB] \\
& + e^{-\lambda_A^e} (1 - e^{-\lambda_B^e}) e^{-\lambda_B^e} \left[(1 - \rho_3^{\ell_2}) (1 - \rho_8^{\ell_2}) \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_8) r_{iA01} + s_5 r_{iAB1} + (s_6 + s_7) r_{iB01}) \mid AB] \\
& + e^{-\lambda_A^e} (1 - e^{-\lambda_B^e}) (1 - e^{-\lambda_B^e}) \left[(1 - \rho_3^{\ell_2}) (1 - \rho_8^{\ell_2}) \right] E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_8) r_{iA01} + s_5 r_{iAB1} + s_6 r_{iB01} + s_7 \omega_{iB}) \mid AB]
\end{aligned}$$

We now have sixteen new terms that have not been previously identified.

- (a) $E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_5) r_{iA01} + s_6 r_{iB01} + s_7 \omega_{iB} + s_8 r_{iAB1}) | AB]$
- (b) $E \exp [i (s_1 R_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + s_4 r_{iA01} + s_5 \omega_{iA} + (s_6 + s_7) r_{iB01} + s_8 r_{iAB1}) | AB]$
- (c) $E \exp [i (s_1 R_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + s_4 r_{iA01} + s_5 \omega_{iA} + s_6 r_{iB01} + s_7 \omega_{iB} + s_8 r_{iAB1}) | AB]$
- (d) $E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + s_4 r_{iA01} + s_6 r_{iB01} + s_7 \omega_{iB} + (s_5 + s_8) r_{iAB1}) | AB]$
- (e) $E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_5 + s_8) r_{iA01} + s_7 \omega_{iB} + s_6 r_{iB01}) | AB]$
- (f) $E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_8) r_{iA01} + s_5 \omega_{iA} + (s_6 + s_7) r_{iB01}) | AB]$
- (g) $E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_8) r_{iA01} + s_5 \omega_{iA} + s_6 r_{iB01} + s_7 \omega_{iB}) | AB]$
- (h) $E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_3 r_{iAB0} + (s_4 + s_8) r_{iA01} + s_5 r_{iAB1} + s_6 r_{iB01} + s_7 \omega_{iB}) | AB]$
- (i) $E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_5) r_{iA01} + s_6 r_{iB01} + s_7 \omega_{iB} + s_8 r_{iAB1}) | AB]$
- (j) $E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + s_4 r_{iA01} + s_5 \omega_{iA} + (s_6 + s_7) r_{iB01} + s_8 r_{iAB1}) | AB]$
- (k) $E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + s_4 r_{iA01} + s_5 \omega_{iA} + s_6 r_{iB01} + s_7 \omega_{iB} + s_8 r_{iAB1}) | AB]$
- (l) $E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + s_4 r_{iA01} + (s_5 + s_8) r_{iAB1} + s_6 r_{iB01} + s_7 \omega_{iB}) | AB]$
- (m) $E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_5 + s_8) r_{iA01} + s_6 r_{iB01} + s_7 \omega_{iB}) | AB]$
- (n) $E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_8) r_{iA01} + s_5 \omega_{iA} + (s_6 + s_7) r_{iB01}) | AB]$
- (o) $E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_8) r_{iA01} + s_5 \omega_{iA} + s_6 r_{iB01} + s_7 \omega_{iB}) | AB]$
- (p) $E \exp [i ((s_1 + s_3) r_{iA00} + s_2 r_{iB00} + (s_4 + s_8) r_{iA01} + s_5 r_{iAB1} + s_6 r_{iB01} + s_7 \omega_{iB}) | AB]$

We use the same basic approach as above: we set various values of s_j to zero we can identify the components. To see how to identify all of these terms, setting $s_3 = s_8 = s_5 = 0$ all of the terms simplify to either

$$E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_4 r_{iA01} + s_6 r_{iB01} + s_7 \omega_{iB}) | AB]$$

or

$$E \exp [i (s_1 r_{iA00} + s_2 r_{iB00} + s_4 r_{iA01} + (s_6 + s_7) r_{iB01}) | AB]$$

However, we have already shown identification of latter of these terms which means the former is identified. Identification of this gives identification of term (m). Using a similar argument, setting $s_3 = s_8 = s_7 = 0$ we can identify term (n). Given these setting $s_3 = s_5 = 0$ we can show that (p), (i) and (l) are identified. Setting $s_3 = s_8 = 0$ we can identify (o), $s_3 = s_7 = 0$ gives (j), $s_5 = s_8 = 0$ gives (e), and $s_7 = s_8 = 0$ gives (f). Now with these setting $s_3 = 0$ gives (k), $s_8 = 0$ gives (g), $s_5 = 0$ gives (a), (d), and (h), and $s_7 = 0$ gives (b). This leaves only term (c) which is identified

by varying all 8 terms given knowledge of all the other terms. This is the characteristic function for the joint distribution. Thus, we have shown that the joint distribution of wages for type AB workers can be non-parametrically identified.

Identification of the Distribution of Wages for the Other Types

Using a symmetric argument reversing A and B we can show that the distribution of

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

for the BA types is also identified.

Next consider the $A0$ types. We will use an argument similar to above, though it will be much simpler as there are fewer labor market statuses to worry about.

Table AA2
Labor Market Statuses for $A0$ workers

$j(i, t)$	$h(i, t)$	$\ell(i, t)$	$h_0(i, t)$	Wage
A	0	0	0	R_{iA00}
A	0	A	0	π_{iA}
A	1	0	0	$R_{iA00}\psi_1$
A	1	A	0	$\pi_{iA}\psi_1$
A	1	0	1	R_{iA01}
A	1	A	1	$\pi_{iA}\psi_1$

From Table AA2 one can see that for an $A0$ worker wages depend on the joint distribution of just three objects (in addition to ψ_1)

$$(R_{iA00}, \pi_{iA}, R_{iA01}).$$

We consider the following the transition path. People begin non-employed at time zero and we will take $d_4 > 1$

Transition	Time
Start at A	$1 - d_1$
Move to non-employment	$1 + d_2$
Start at A	$2 - d_3$
Move to non-employment	$2 + d_4$

We can identify

$$\begin{aligned} E \exp [i (s_1 w_{i1} + s_2 w_{i2} + s_3 w_{i3}) | d] = & P(AB | d) E \exp [i (s_1 w_{i1} + s_2 w_{i2} + s_3 w_{i3}) | AB, d] \\ & + P(BA | d) E \exp [i (s_1 w_{i1} + s_2 w_{i2} + s_3 w_{i3}) | BA, d] \\ & + P(A0 | d) E \exp [i (s_1 w_{i1} + s_2 w_{i2} + s_3 w_{i3}) | A0, d] \end{aligned}$$

Since everything else in this expression is identified, we can identify $E \exp [i (s_1 w_{i1} + s_2 w_{i2} + s_3 w_{i3}) | A0, d]$.

Furthermore, analogous to the argument above using d_7 , now by varying d_2 , we can identify the expected value of $f(w_{i1}, \dots, w_{i3})$ conditional on human capital arriving between time 1 and $1 + d_2$ (write this conditioning as $H_{i2} = 1$). Then we can identify

$$E \exp [i (s_1 w_{i1} + s_2 w_{i2} + s_3 w_{i3}) | A0, d, H_{i2} = 1].$$

Thus taking $d_1 \downarrow 0$ and $d_3 \downarrow 0$ we can identify

$$\begin{aligned} & \lim_{d_1, d_3 \downarrow 0} \frac{E \exp [i (s_1 w_{i1} + s_2 w_{i2} + s_3 w_{i3}) | A0, d, H_{i2} = 1]}{\phi_{\xi}(s_1) \phi_{\xi}(s_2) \phi_{\xi}(s_3)} \\ &= e^{-\lambda A} E \exp [i (s_1 r_{iA00} + [s_2 + s_3] r_{iA01}) | A0] \\ &+ (1 - e^{-\lambda A}) E \exp [i (s_1 r_{iA00} + s_2 r_{iA01} + s_3 \omega_{iA}) | A0]. \end{aligned}$$

Set $s_3 = 0$ and we can identify $E \exp [i (s_1 r_{iA00} + s_2 r_{iA01}) | A0]$. Knowledge of this gives knowledge of $E \exp [i (s_1 r_{iA00} + [s_2 + s_3] r_{iA01}) | A0]$ and then allowing s_3 to vary means we can identify $E \exp [i (s_1 r_{iA00} + s_2 r_{iA01} + s_3 \omega_{iA}) | A0]$.

An analogous argument gives identification of $E \exp [i (s_1 r_{iB00} + s_2 r_{iB01} + s_3 \omega_{iB}) | B0]$.

Thus, we have shown that wages, turnover parameters, and type proportions are identified.

Non-Identification of β

The next question is whether we can identify the bargaining parameter β . The answer turns out to be no. To see why, suppose that we could observe all of the wages for a particular AB worker (which is similar to observing the distribution of wages after filtering out the measurement error). We could observe the 8 wages $(W_{iA00}, \pi_{iA}, W_{iAB0}, W_{iB00}, \pi_{iB}, W_{iA01}, W_{iAB1}, W_{iB01})$ and know their d_j 's. Assume even further that the utility function takes the form

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

Even in this restrictive case, β is not identified. Table AA1 showed that for the AB types, there are seventeen different possible labor market statuses. We can write the Bellman equation for each of the seventeen where we impose the restrictions from bargaining (e.g. $V_{iA0}(R_{iA00}) = \beta V_{iA0}(\pi_{iA}) + (1 - \beta) V_{i00}$). We present them in the exact same order as in the table.

The first nine are associated with employment at an A firm

$$(\rho + \delta + \lambda_B^e + \lambda_A^e + \lambda_h) [\beta V_{iA0}(\pi_{iA}) + (1 - \beta) V_{i00}] \quad (\text{A-1})$$

$$= r_{iA00} + v_{iA} + \lambda_A^e V_{iA0}(\pi_{iA}) \\ + \lambda_B^e [\beta V_{iA0}(\pi_{iA}) + (1 - \beta) V_{iB0}(\pi_{iB})] + \delta_i V_{i00}^* + \lambda_h V_{iA1}(R_{iA00})$$

$$(\rho + \delta + \lambda_A^e + \lambda_h) [\beta V_{iA0}(\pi_{iA}) + (1 - \beta) V_{iB0}(\pi_{iB})] \quad (\text{A-2})$$

$$= r_{iAB0} + v_{iA} + \lambda_A^e V_{iA0}(\pi_{iA}) + \delta_i V_{i00}^* + \lambda_h V_{iA1}(R_{iAB0})$$

$$(\rho + \delta + \lambda_h) V_{iA0}(\pi_{iA}) = \log(\pi_{iA}) + v_{iA} + \delta_i V_{i00}^* + \lambda_h V_{iA1}(\pi_{iA}) \quad (\text{A-3})$$

$$(\rho + \delta + \lambda_B^e + \lambda_A^e) V_{iA1}(R_{iA00}) \quad (\text{A-4})$$

$$= r_{iA00} + \log(\psi_1) + v_{iA} + \lambda_A^e V_{iA1}(\pi_{iA}) \\ + \lambda_B^e [\beta V_{iA1}(\pi_{iA}) + (1 - \beta) V_{iB1}(\pi_{iB})] + \delta_i V_{i01}^*$$

$$(\rho + \delta + \lambda_B^e + \lambda_A^e) V_{iA1}(R_{iAB0}) \quad (\text{A-5})$$

$$= r_{iAB0} + \log(\psi_1) + v_{iA} + \lambda_A^e V_{iA1}(\pi_{iA}) \\ + \lambda_B^e \max \{V_{iA1}(R_{iAB0}\psi_1), V_{iA1}(R_{iAB1})\} + \delta_i V_{i01}$$

$$(\rho + \delta) V_{iA1}(\pi_{iA}) = \log(\pi_{iA}) + \log(\psi_1) + v_{iA} + \delta_i V_{i01}^* \quad (\text{A-6})$$

$$(\rho + \delta + \lambda_B^e + \lambda_A^e) [\beta V_{iA1}(\pi_{iA}) + (1 - \beta) V_{i01}] \quad (\text{A-7})$$

$$= r_{iA01} + v_{iA} + \lambda_A^e V_{iA1}(\pi_{iA}) \\ + \lambda_B^e [\beta V_{iA1}(\pi_{iA}) + (1 - \beta) V_{iB1}(\pi_{iB})] + \delta_i V_{i01}^*$$

$$(\rho + \delta + \lambda_A^e) [\beta V_{iA1}(\pi_{iA}) + (1 - \beta) V_{iB1}(\pi_{iB})] \quad (\text{A-8})$$

$$= r_{iAB1} + v_{iA} + \lambda_A^e V_{iA1}(\pi_{iA}) + \delta_i V_{i01}^*$$

$$(\rho + \delta) V_{iA1}(\pi_{iA}) = \log(\pi_{iA}) + \log(\psi_1) + v_{iA} + \delta_i V_{i01}^* \quad (\text{A-9})$$

The next six are associated with employment at a *B* type firm.

$$(\rho + \delta + \lambda_B^e + \lambda_A^e + \lambda_h) [\beta V_{iB0}(\pi_{iA}) + (1 - \beta) V_{i00}] \quad (\text{A-10})$$

$$\begin{aligned} &= r_{iB00} + v_{iB} + \lambda_A^e [\beta V_{iA0}(\pi_{iA}) + (1 - \beta) V_{iB0}(\pi_{iB})] \\ &\quad + \lambda_B^e V_{iB0}(\pi_{iB}) + \delta_i V_{i00}^* + \lambda_h V_{iB1}(R_{iB00}) \end{aligned}$$

$$(\rho + \delta + \lambda_A + \lambda_h) V_{iB0}(\pi_{iB}) = \log(\pi_{iB}) + v_{iB} + \lambda_A^e [\beta V_{iA0}(\pi_{iA}) + (1 - \beta) V_{iB0}(\pi_{iB})] \quad (\text{A-11})$$

$$+ \delta_i V_{i00}^* + \lambda_h V_{iB1}(\pi_{iB})$$

$$(\rho + \delta + \lambda_B^e + \lambda_A^e) V_{iB1}(R_{iB00}) \quad (\text{A-12})$$

$$\begin{aligned} &= r_{iB00} + \log(\psi_1) + v_{iB} + \lambda_A^e [\beta V_{iA1}(\pi_{iA}) + (1 - \beta) V_{iB1}(\pi_{iB})] \\ &\quad + \lambda_B^e V_{iB1}(\pi_{iB}) + \delta_i V_{i01}^* \end{aligned}$$

$$(\rho + \delta + \lambda_A^e) V_{iB1}(\pi_{iB}) = \log(\pi_{iB}) + \log(\psi_1) + v_{iB} \quad (\text{A-13})$$

$$+ \lambda_A^e [\beta V_{iA1}(\pi_{iA}) + (1 - \beta) V_{iB1}(\pi_{iB})] + \delta_i V_{i01}^*$$

$$(\rho + \delta + \lambda_B^e + \lambda_A^e) [\beta V_{iB1}(\pi_{iA}) + (1 - \beta) V_{i01}] \quad (\text{A-14})$$

$$\begin{aligned} &= r_{iB01} + v_{iB} + \lambda_A^e [\beta V_{iA1}(\pi_{iA}) + (1 - \beta) V_{iB1}(\pi_{iB})] \\ &\quad + \lambda_B^e V_{iB1}(\pi_{iB}) + \delta_i V_{i01}^* \end{aligned}$$

$$(\rho + \delta + \lambda_A^e) V_{iB1}(\pi_{iB}) = \log(\pi_{iB}) + \log(\psi_1) + v_{iB} \quad (\text{A-15})$$

$$+ \lambda_A^e [\beta V_{iA1}(\pi_{iA}) + (1 - \beta) V_{iB1}(\pi_{iB})] + \delta_i V_{i01}^*$$

Finally, workers can be non-employed with two different levels of human capital. This leads to four more equations because of the distinction between V_{i00} and V_{i00}^*

$$(\rho + \lambda_A^n + \lambda_B^n) V_{i00} = U_{i00} + \lambda_A^n [\beta V_{iA0}(\pi_{iA}) + (1 - \beta) V_{i00}] \quad (\text{A-16})$$

$$+ \lambda_B^n [\beta V_{iB0}(\pi_{iB}) + (1 - \beta) V_{i00}]$$

$$V_{i00}^* = P^* \frac{\lambda_A^n [\beta V_{iA0}(\pi_{iA}) + (1 - \beta) V_{i00}] + \lambda_B^n [\beta V_{iB0}(\pi_{iA}) + (1 - \beta) V_{i00}]}{\lambda_A^n + \lambda_B^n} + (1 - P^*) V_{i00}$$

$$(\text{A-17})$$

$$(\rho + \lambda_A^n + \lambda_B^n) V_{i01} = U_{i01} + \lambda_A^n [\beta V_{iA1}(\pi_{iA}) + (1 - \beta) V_{i01}] \quad (\text{A-18})$$

$$+ \lambda_B^n [\beta V_{iB1}(\pi_{iB}) + (1 - \beta) V_{i01}]$$

$$V_{i01}^* = P^* \frac{\lambda_A^n [\beta V_{iA1}(\pi_{iA}) + (1 - \beta) V_{i01}] + \lambda_B^n [\beta V_{iB1}(\pi_{iA}) + (1 - \beta) V_{i01}]}{\lambda_A^n + \lambda_B^n} + (1 - P^*) V_{i01}$$

$$(\text{A-19})$$

Note that equations (A-6) and (A-9) are identical as are (A-13) and (A-15). Thus we have 17

equations in 16 unknowns:

$$(V_{iA0}(\pi_{iA}), V_{i00}, V_{i00}^*, v_{iA}, V_{iB0}(\pi_{iB}), V_{iA1}(R_{iA00}), V_{iA1}(R_{iAB0}), V_{i01}, V_{i01}^*, V_{iA1}(\pi_{iA}), v_{iB}, V_{iB1}(R_{iB00}), U_{i00}, U_{i01}, \beta).$$

We can also normalize $U_{i00} = 0$ as a location normalization leaving 17 equations in 15 unknowns. However working through the algebra, one can show that these equations are linearly dependent. Specifically the following three linear relationships exists.

$$\begin{aligned} r_{iAB1} + (r_{iB01} - r_{iA01}) &= \log(\pi_{iB}) + \log(\psi_1) \\ r_{iAB0} + (r_{iB00} - r_{iA00}) &= \log(\pi_{iB}) \\ r_{iAB1} - r_{iAB0} &= \log(\psi_1) \end{aligned}$$

This leaves 12 linear equations and 13 unknowns, so β is not separately identified from v_{iA} and v_{iB} . The basic reason is that revealed preferences can only identify the ordering of jobs, but not the cardinal utility associated with each of them.

Appendix B: Data

Data Selection We define labor market entry to be the month of graduation from the highest completed education recorded.⁴⁸ We disregard spells that are before this date. We disregard workers who are observed in education after the date of highest completed education. For example, 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

⁴⁸We have information on highest completed education back to 1971, so highest completed education is missing for workers who took it before 1974. Also, immigrants and workers who never finished primary school have missing values. We keep these 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 1971 have entered the labor market.

in a given year. Wages are detrended in logs (but 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 likely to be 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. Likewise, we also ignore job-to-job transitions from mass layoffs. A mass lay-off is defined as the establishment having more than 15 workers in year t and less than 30 percent left in year $t + 1$. We define an establishment to be closing if we do not observe it in the data any longer. We calculate calculate establishment closings and mass lay-offs from the raw data without imposing any of our sample selection criteria.

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

In total	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 (hffsp)	4,147,463	445,298	54,714,582	594,070	
Delete workers with gaps in educational variables (hffsp)	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 hffsp 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 gabs	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 (hffsp)	2,894,716	124,084	2,894,716	161,285	0.69
Delete workers with gaps in educational variables (hffsp)	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 hffsp 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 gabs	1,722,319	99,802	1,722,319	135,375	0.83

Table 2: Summery Statistics: Pooled Cross-sections

	Mean	Std. Dev.
Number of years in sample	11.06	6.27
Number of Establishments per worker	2.70	1.85
Female	0.49	
Preparatory educations	0.30	
Highschool	0.04	
Vocational education	0.42	
Short further education	0.04	
Medium-length further education/Bachelor-degree	0.14	
Master-degree and PhD	0.05	
Average years of education	11.69	3.18
Age	38.31	9.63
Employed	0.83	
Public Employed	0.32	
Missing Establishment ID	0.01	
Real Experience in Years	13.46	9.47
Log Hourly Wages	4.50	0.35

Table 3: Auxiliary Model and Estimates: Full Sample

Moment	Data	Model
Avg. Length Emp. Spell	377 (0.202)	382
Avg. Length Nonnemp. Spell	91.4 (0.086)	91.5
Avg. Length Job	108 (0.101)	107
$E(\tilde{S}_{itj}\tilde{h}_{-itj}) \times 100$	1.53 (0.003)	1.51
Between Persons $\times 100$	8.03 (0.012)	8.02
Between Jobs $\times 100$	2.87 (0.006)	2.89
Within Job $\times 100$	1.49 (0.003)	1.49
Sample mean w_{it}	4.50 (0.000)	4.50
$E(\tilde{w}_{it}\tilde{w}_{-it}) \times 100$	0.284 (0.002)	0.284
$E(\tilde{w}_{it}\tilde{h}_{-it}) \times 100$	0.108 (0.001)	0.108
Fraction Wage Drops	0.400 (0.000)	0.408
Coeff Exper $\times 100$	2.48 (0.008)	2.44
Coeff Exper ² $\times 1000$	-0.291 (0.002)	-0.295
Coef Tenure ² $\times 1000$	-0.460 (0.005)	-0.462
Var(Nonemployment)	16000 (47.472)	15992
Cov(\bar{w}_i , Non-employment)	-3.42 (0.013)	-3.43
Var(Employment Dur)	102000 (71.434)	99688
Invol Job to Job	0.205 (0.011)	0.205

Table 4: Parameter Estimates: Full Sample

Parameter	Estimate
d_0	-2.89 (0.035)
λ^n	0.98 (0.007)
λ^e	2.07 (0.061)
E_θ	4.34 (0.013)
σ_θ	0.231 (0.002)
σ_ξ	0.134 (0.001)
f_1	2.207 (0.139)
f_2	0.122 (0.001)
f_3	-0.000 (0.009)
σ_{v^w}	0.197 (0.003)
α	2.82 (0.123)
b_1	0.019 (0.001)
b_2	-0.001 (0.000)
β	0.812 (0.008)
P^*	0.435 (0.019)
σ_{v_0}	0.434 (0.003)
γ_θ	-0.387 (0.025)
σ_d	1.99 (0.026)

Table 5: Model Decompositions: Variance of log wages, Full Sample

(A)		(B)		(C)		(D)	
Total	0.106	Total	0.106	Total	0.106	Total	0.106
No HC	0.101	No HC	0.101	No HC	0.101	No HC	0.101
No Monop	0.093	No Monop	0.093	No Monop	0.093	No Monop	0.093
No Roy	0.006	No Search	0.090	No Comp	0.089	No Comp	0.087
No Search	0.005	No Roy	0.005	No Search	0.068	No Roy	0.002

Table 6: Model Decompositions: Variance of flow utility, Full Sample

(A)		(B)		(C)		(D)	
Total	0.315	Total	0.315	Total	0.315	Total	0.315
No HC	0.303	No HC	0.303	No HC	0.303	No HC	0.303
No Monop	0.284	No Monop	0.284	No Monop	0.284	No Monop	0.284
No Roy	0.218	No Search	0.098	No Comp	0.089	No Comp	0.087
No Search	0.060	No Roy	0.060	No Search	0.068	No Roy	0.002

Table 7: Auxliary Model and Estimates: Demographic Groups

Moment	Col Men		HS Men		Col Women		HS Women	
	Data	Model	Data	Model	Data	Model	Data	Model
Avg. Length Emp. Spell	430 (0.173)	435	379 (0.200)	382	392 (0.165)	391	346 (0.175)	348
Avg. Length Nonnemp. Spell	60.1 (0.073)	60.3	80 (0.086)	80	63 (0.072)	63	119 (0.104)	119
Avg. Length Job	119 (0.087)	120	104 (0.091)	103	109 (0.081)	109	109 (0.098)	109
$E(\tilde{S}_{i\ell_j}\tilde{h}_{-i\ell_j}) \times 100$	1.59 (0.004)	1.61	1.45 (0.003)	1.44	1.08 (0.003)	1.08	1.68 (0.003)	1.67
Between Persons $\times 100$	9.88 (0.013)	9.92	6.20 (0.010)	6.23	5.05 (0.008)	5.04	4.49 (0.008)	4.50
Between Jobs $\times 100$	3.21 (0.007)	3.21	3.13 (0.006)	3.11	2.55 (0.006)	2.55	2.57 (0.006)	2.55
Within Job $\times 100$	1.83 (0.004)	1.83	1.39 (0.003)	1.39	1.44 (0.014)	1.44	1.50 (0.003)	1.50
Sample mean w_{it}	4.78 (0.000)	4.78	4.56 (0.000)	4.56	4.48 (0.000)	4.48	4.33 (0.000)	4.33
$E(\tilde{w}_{it}\tilde{w}_{-it}) \times 100$	4.60 (0.004)	4.61	0.407 (0.002)	0.407	0.203 (0.002)	0.203	1.81 (0.002)	1.81
$E(\tilde{w}_{it}\tilde{h}_{-it}) \times 100$	1.74 (0.001)	1.74	0.113 (0.001)	0.113	0.083 (0.001)	0.083	0.726 (0.001)	0.726
Fraction Wage Drops	0.335 (0.000)	0.342	0.408 (0.000)	0.410	0.409 (0.000)	0.404	0.414 (0.000)	0.494
Coeff Exper $\times 100$	4.26 (0.008)	4.38	2.34 (0.009)	2.40	2.62 (0.007)	2.57	1.83 (0.010)	1.79
Coeff Exper ² $\times 1000$	-0.640 (0.002)	-0.648	-0.259 (0.002)	-0.259	-0.352 (0.002)	-0.356	-0.207 (0.002)	-0.211
Coef Tenure ² $\times 1000$	-0.793 (0.005)	-0.759	-0.470 (0.006)	-0.468	-0.493 (0.005)	-0.498	-0.286 (0.006)	-0.287
Var(Nonemployment)	7830 (33.52)	7775	12057 (43.23)	12000	8571 (35.92)	8550	22800 (54.97)	22231
Cov(\bar{w}_i , Non-employment)	-1.96 (0.005)	-1.96	-3.02 (0.008)	-3.02	-1.38 (0.003)	-1.38	-1.43 (0.007)	-1.43
Var(Employment Dur)	102000 (54.24)	98663	106032 (71.39)	107000	96283 (65.26)	98100	94700 (69.65)	90903
Invol Job to Job	0.182 (0.010)	0.182	0.243 (0.011)	0.243	0.086 (0.007)	0.086	0.218 (0.012)	0.218

Table 8: Parameter Estimates By Group

Parameter	Col	HS	Col	HS
	Men	Men	Women	Women
d_0	-4.38 (0.131)	-2.76 (0.026)	-3.57 (0.018)	-2.41 (0.022)
λ^n	1.92 (0.011)	1.14 (0.016)	1.78 (0.036)	0.737 (0.006)
λ^e	2.59 (0.039)	2.35 (0.015)	2.34 (0.037)	1.57 (0.093)
E_θ	4.64 (0.009)	4.41 (0.009)	4.35 (0.006)	4.169 (0.012)
σ_θ	0.264 (0.005)	0.181 (0.003)	0.150 (0.001)	0.134 (0.001)
σ_ξ	0.095 (0.011)	0.128 (0.001)	0.124 (0.000)	0.143 (0.001)
f_1	5.067 (0.199)	2.85 (0.198)	0.91 (0.007)	2.92 (0.394)
f_2	0.202 (0.006)	0.145 (0.002)	0.121 (0.003)	0.103 (0.001)
f_3	-0.043 (0.088)	-0.015 (0.010)	0.095 (0.006)	0.044 (0.008)
σ_{v^w}	0.177 (0.014)	0.202 (0.002)	0.201 (0.002)	0.198 (0.003)
α	15.85 (1.00)	3.07 (0.067)	2.64 (0.123)	4.61 (0.312)
b_1	0.015 (0.033)	0.012 (0.001)	0.041 (0.007)	-0.001 (0.003)
$b_2 \times 10$	-0.021 (0.053)	-0.005 (0.001)	-0.006 (0.002)	0.008 (0.002)
β	0.310 (0.047)	0.791 (0.009)	0.665 (0.009)	0.845 (0.018)
P^*	0.691 (0.057)	0.451 (0.020)	0.258 (0.018)	0.409 (0.018)
σ_{v_0}	0.188 (0.015)	0.358 (0.053)	0.331 (0.024)	0.366 (0.013)
γ_θ	0.583 (0.021)	-1.221 (0.182)	-0.357 (0.050)	-0.595 (0.041)
σ_d	2.42 (0.053)	2.25 (0.014)	2.05 (0.013)	1.65 (0.024)

Table 9: Model Decompositions: Variance of log wages by Demographic Groups

	Col Men	HS Men	Col Women	HS Women
Total	0.141	0.091	0.075	0.065
No HC	0.140	0.087	0.073	0.061
No Monop	0.099	0.076	0.063	0.057
No Roy	0.011	0.008	0.006	0.004
No Search	0.086	0.071	0.060	0.047
No Comp	0.095	0.071	0.056	0.053
No Roy/Search	0.006	0.007	0.005	0.003
No Roy/Comp	0.004	0.003	0.002	0.002
No Search/Comp	0.080	0.047	0.036	0.029

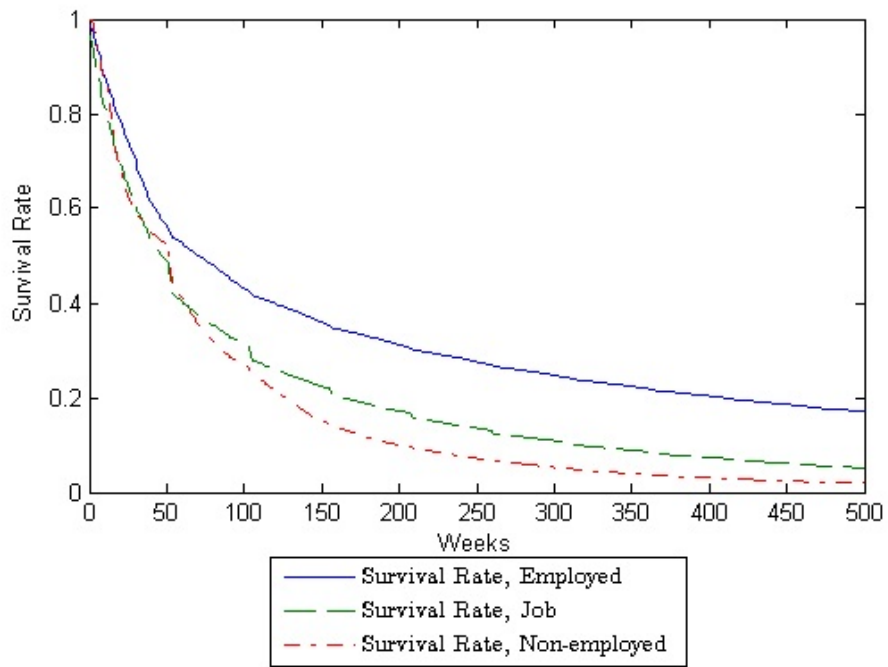


Figure 1: KAPLAN-MEIER SURVIVAL ESTIMATES

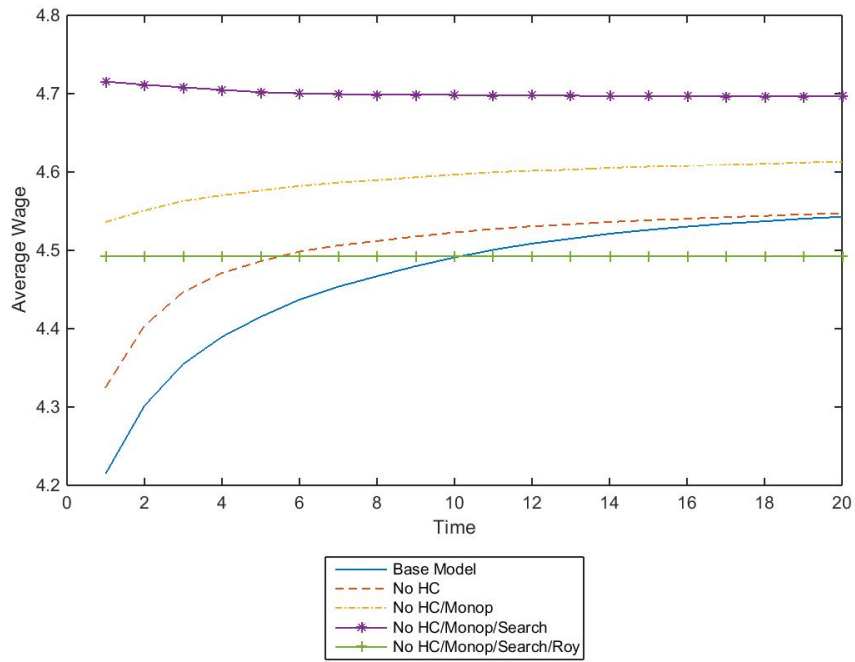


Figure 2: LIFE-TIME WAGE PROFILES

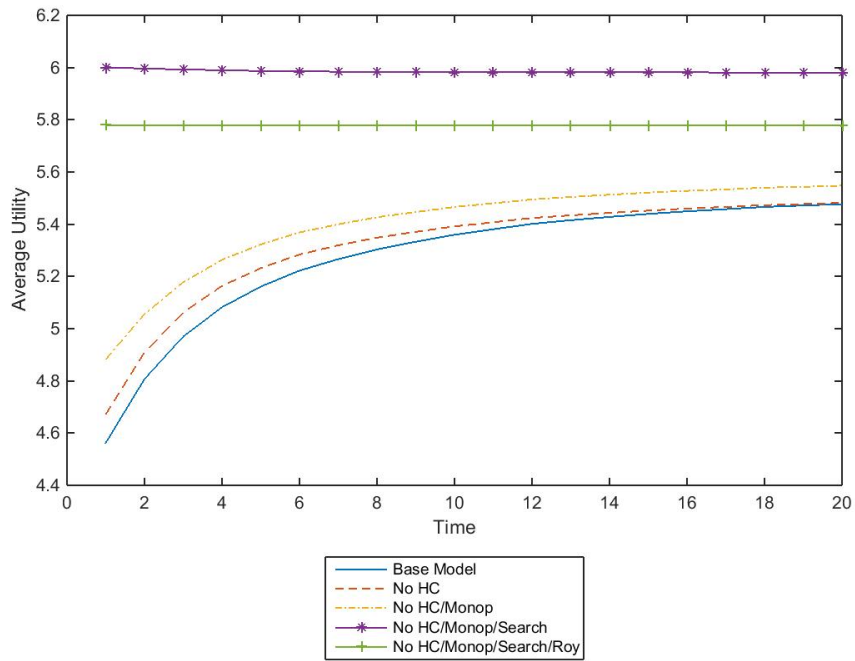


Figure 3: LIFE-TIME UTILITY PROFILES