

# Estimating the Employer Switching Costs and Wage Responses of Forward-Looking Engineers

Jeremy T. Fox\*

University of Chicago and NBER

January 2009

## Abstract

I estimate the relative magnitudes of worker switching costs and whether the employer switching of experienced engineers responds to outside wage offers. Institutional features imply that voluntary turnover dominates switching in the market for Swedish engineers from 1970–1990. I use data on the allocation of engineers across a large fraction of Swedish private sector firms to estimate the relative importance of employer wage policies and switching costs in a dynamic programming, discrete choice model of voluntary employer choice. The differentiated firms are modeled in employer characteristic space and each firm has its own age-wage profile. I find that a majority of engineers have moderately high switching costs and that a minority of experienced workers are responsive to outside wage offers. Younger workers are more sensitive to outside wage offers than older workers.

---

\*This paper is part of the project Pay and Promotion in Swedish Private Industries 1970–1990, which is under the leadership of Eva Meyerson–Milgrom, with financial support from the Swedish Council for Research in the Humanities and Social Sciences. Thanks to Arie Hietasalo, Åke Kempe and Svenskt Näringsliv for generously providing the SAF data and guidance on their use. The Kapnick Foundation supported me through a grant to the Stanford Institute for Economic Policy Research. An earlier version of this paper is part of my dissertation for Stanford University. Thanks go to my dissertation advisers Patrick Bajari, Edward Lazear, John Pencavel, and Frank Wolak, as well as the other faculty at Stanford, particularly Lanier Benkard, Kenneth Judd, Paul Oyer and Peter Reiss. Thanks for helpful comments from seminar participants at Boston University, the BYU / Princeton Monopsony Conference, Chicago, Irvine, Northwestern, U. Pennsylvania, the Society for Economic Dynamics, Stanford, Texas, and the Aarhus Labor Market Models conference, which was in honor of Dale Mortensen. Thanks also to the comments from my discussants, Alan Manning and Frank McIntyre. My email address is fox@uchicago.edu.

# 1 Introduction

Moderately low levels of voluntary worker turnover between employers often characterize labor markets for skilled workers, such as the market for Swedish engineers studied in this paper. This empirical fact opens the possibility that enough high-skilled workers do not switch employers in response to outside wage offers in a way that would lead to voluntary reallocation of workers across firms and sectors in response to the varying product market successes of firms. Indeed, in a survey of blue-collar workers in New Haven, Connecticut, in the late 1940's, Reynolds (1951) finds that 12% of workers would move to a new geographic area for a change in permanent wages of under 25%, 30% of workers would move for a permanent change in wages of between 25% and 100%, and 45% would not move to a new area under any circumstances.

In my model, workers face switching costs when they move between firms. Each employer offers a unique wage policy that I estimate. Using this framework, I empirically examine the causes of voluntary employer turnover to see if switching costs are large, and if so, how switching costs vary for switching across geographic regions, industries, and corporate boundaries. I also estimate whether turnover is responsive enough to outside offers to lead to small differences in wages causing voluntary reallocation of experienced workers across firms.

This paper examines the labor market for Swedish engineers from 1970–1990. As Section 2 discusses, Swedish institutional features, such as wage compression and restrictions on firing, justify treating the observed, across-firm employment allocation as the revealed preferences of engineers given the location decisions and wage choices of employers. Employment allocations in Sweden therefore allow the estimation of workers' underlying preferences for voluntary employer switching.

The econometric procedure estimates the importance of four categories of employer characteristics in the underlying utility function that governs each worker's decision to stay at his current employer or switch to a new one. The four employer characteristics are wages, switching costs, nonwage benefits and taste shocks. Nonwage benefits are proxied for by firm size. I also estimate heterogeneity in worker preferences by specifying three discrete worker types, each with separate preferences for employer characteristics and separate population frequencies. After estimating worker preferences, I calculate each worker type's switching rates under alternative labor market structure scenarios and the switching response to one firm's unilateral wage increase. Estimating worker preferences may also inform us about the behavior of experienced, white collar workers in less unionized labor markets, to the degree that underlying preferences are similar across countries.

Across-firm variation in wages is necessary to precisely estimate the importance of outside offers in the decision of whether or not to switch employers. In the data on Swedish engineers, across-firm variation is best captured by differences in percentage raises received by stayers at a particular firm, rather than variation in the level of wages. In reduced-form empirical work, I show that only the rate of wage increases is negatively correlated with separation at the firm level. The level of wage is not highly correlated with turnover, which suggests that this measure captures mostly unobserved heterogeneity in worker ability. By contrast, a percentage raises specification is consistent with symmetric (to the worker and firms) unfolding of new information about a worker's idiosyncratic ability. This paper abstracts from theories of personnel policies by saying that each firm offers its own stochastic process for age–wage profiles. In a first stage, I estimate firm-specific five-year percentage wage increases using the data advantage of observing multiple stayers for each firm. I must also model how a worker's wage changes when he switches between the compensation policies of his old and new firms. I find that a younger worker's wage upon a switch is more closely linked to the average wage of similar workers at the new firm than is an older worker's wage.

This paper models employer choice as a dynamic programming, discrete choice. Given his current wage, age and employer, a worker knows his individual-specific stochastic processes of future age–wage profiles at

his current and potential employers. The worker has symmetric information about his possible future wage and taste shock outcomes at all firms, and can choose to work at any Swedish employer of engineers. He picks the employer that maximizes the expected discounted sum of his future utility. The worker considers how his decision in the current period will affect his decision of whether to switch in the future. For example, a worker who moves to an employer in a large city knows that many other potential employers are close by, and his probability of switching in the future will increase. Likewise, the worker weights the stochastic process for age–wage profiles at his current period’s choice by the probability that he will remain there in the future. The worker moving to the firm in the large city may not place much emphasis on the stochastic process of wages at his destination far into the future because he feels it is unlikely that he will remain at that firm for a long time. Estimation chooses the relative importance of the four employer characteristics in the current period utility function (switching costs, wages, nonwage benefits and taste shocks) that make the observed employer choices the most likely of all outcomes within the parametric structure of the model.

The employer switching model is based upon the Rust (1987) framework for single-agent, discrete choice dynamic programming problems. Each period, the dependent variable is the employer chosen by a worker. The strategy of estimating elasticities of labor supply to individual firms using a single-agent, dynamic programming switching model is similar to methods used in the estimation of the elasticities of demand for consumer products with fluctuating prices and that are storable, durable or that have aspects of intertemporal substitutability in the preferences for consumption (Erdem, Imai and Keane, 2003; Hartmann, 2006; Hendel and Nevo, 2006). The model is also similar to the independently-developed model of geographic movement in the labor market of Kennan and Walker (2008). This paper is the first time a single-agent dynamic programming model has been used to estimate the elasticity of labor supply to individual firms. It is also one of the few papers to use techniques from even myopic discrete choice demand estimation to estimate labor supply behavior, as most researchers do not have data on the distinct compensation schemes of all firms in a labor market.

The estimates of the underlying preferences of workers have meaning for economic quantities when workers are placed in particular labor market scenarios. As an example, consider a stylized labor market with 400 employers, each in different industries, each separated by 100 km, and each offering the same stochastic process for age–wage profiles. My estimates predict that 15% of age 35–39 engineers will switch over a five-year period. When evaluated at the firm characteristics in the data, I find that a typical firm that unilaterally increases its raises by 1% for each five-year age range will reduce the five-year switching probability of age 35–39 engineers by -0.52% from a predicted base of 15%.

I compare my results to the empirical strategy of focusing on only the separation decision in two ways (Manning, 2003). First, I estimate the structural employer choice model using only the dependent variable of whether to separate. The results are qualitatively similar to those from the model that also uses dependent variable data on the chosen destination of each worker who separates. Second, I perform a series of linear separation regressions, similar to those performed in the literature, and analyze whether turnover is negatively correlated with various measures of the wage level and the rate of wage increase. Overall, I find that in many specifications turnover is positive correlated with the wage level while turnover is negatively correlated with the rate of wage increase, which theoretically better controls for unobserved heterogeneity in the level of worker ability. I use this as evidence in favor of the structural employer choice model’s focus on the rate of wage increases. I conclude the paper by discussing the implications of my estimates for wider issues.

## 2 The Swedish Labor Market for Engineers

Estimating workers' preferences of where to work and firms' preferences for whom to hire and fire is difficult, as the econometrician does not observe employment matches available to agents on either side of the market.<sup>1</sup> This paper uses an empirical model that assumes the observed employment allocation is the result of only voluntary employer choices by workers. In other words, the model severely limits the role of firms in labor markets, which means observed employment matches are the revealed preferences of workers for employer characteristics such as wages. This assumption allows empirical progress as I can tractably model voluntary employer choice by workers as the dynamic decision of whether to switch from one stochastic process of wages to one of several hundred other such processes.<sup>2</sup>

This section argues that the highly unionized nature of the Swedish labor market for engineers makes the paper's model of voluntary employer choice a better approximation of reality than in less unionized labor markets. Legal and union bargaining institutions prevent Swedish firms from making unconstrained wage setting and employment decisions. These institutions make viewing a firm's hiring and wage policies as the solutions to an unconstrained profit maximization problem difficult. However, the constraints provide an economic environment where the assumption of voluntary employment choices by workers is more appropriate.

This paper uses data on the wages and employer choices of engineers who work as white collar workers in the Swedish private sector. The data are collected by the Swedish Employers' Federation (SAF in Swedish). I use engineers who have completed a five-year undergraduate degree from one of several university equivalents. Especially during the sample period 1970–1990, these engineers were among the most highly educated workers in the Swedish private sector. The use of engineers contrasts with studies of worker behavior that focus on all workers, less-skilled workers, or on unusual workers such as CEOs and professional athletes.

As I have said, the paper's economic model of employer choice by a worker contains strong limitations on the role of employers in the labor market. The strongest assumption is that employers do not screen workers before extending offers: the model assumes that workers can pick from any firm employing Swedish engineers, although not necessarily at the same wage. A related assumption is that firms do not fire workers or involuntarily transfer them between establishments: all turnover is voluntary in the model. Together these assumptions mean the observed employment distribution reveals the preferences of workers, and the data can be used to estimate the structural parameters of workers' utility functions using a single-agent empirical model.

The model assumes that individual firms will, on the margin, hire all prospective engineers who choose to apply. While this assumption is false, it may be less so in Sweden. The number of Swedish engineers is constrained by enrollment limitations at the fixed number of government-run technical institutions allowed to offer five-year engineering degrees. Also, the socially progressive nature of centralized wage negotiations in Sweden during the sample period means that the wages of highly-skilled workers such as engineers were artificially low in order to reduce wage inequality. A firm should earn a surplus if it employs a worker at a lower wage than his marginal product. If we would ever expect to find data on a normal labor market where there is a quantity shortage at the given price, engineers in Sweden is it. One problem with the assumption of no screening is that while different industries hire the various engineering specialties (mechanical, electrical, etc.) in differing quantities, the data do not list such specialties. My empirical solution allows switching costs to vary by the industries of the origin and destination firms.

---

<sup>1</sup>In Fox (2008a), I show how to semiparametrically identify and estimate agent preferences in static two-sided matching models with endogenous wages, like the assignment models of Koopmans and Beckmann (1957) and Shapley and Shubik (1972). These endogenous wages in these models are not rich enough to reproduce the complex age-wage profiles chosen by the firms in the data.

<sup>2</sup>Equilibrium search models typically make a similar assumption that rules out screening by firms: for all workers and firms that come into contact with each other, the firm extends a wage offer to the worker (Burdett and Mortensen, 1998; Postel-Vinay and Robin, 2002).

The empirical model assumes that all employer switching is voluntary. An outcome of both negotiations between firms and unions and the 1974 Law on Employment Security is that layoffs in Sweden happen in the inverse order of seniority.<sup>3</sup> Swedish firms cannot choose whom to lay off, and they must follow a preset rule. This lack of discretion implies that firms may be reluctant to lay off workers, except in the direst circumstance. Firing for cause is also very difficult in Sweden, because firings must be justifiable to a union. Therefore, Sweden is an excellent place to study voluntary employer switching, as most switches in the data are voluntary. I clean the data by eliminating establishments that shut down during a particular five-year time interval. Workers frequently transfer between establishments owned by the same legal company, so the model includes a switching cost in the current period utility function to fit this aspect of the data. Again, the maintained assumption in the model is that all transfers between establishments owned by the same company are voluntary.<sup>4</sup>

During the sample period 1970–1990, firms and unions collectively bargained over wages for engineers at the national, industry, and firm levels. National bargaining for white collar workers ended in 1991, after the sample period. Engineers belong to the CF union, which is part of the SACO university graduate union federation, which itself bargained with the Swedish Employers' Federation (SAF) at the national level as part of the PTK white collar workers' cartel.<sup>5</sup> A large fraction of nominal wage increases is due to wage drift, or firm level wage increases. Section 5 describes the resulting across-firm wage heterogeneity. The union negotiations limit the scope of my investigation to the behavior of workers conditional on the outcomes of the negotiations. However, the negotiations do mean that age–wage profiles observed in the data do not perfectly adjust to make the total utility from all employer choices equal, after nonwage benefits unobserved in the data are included. In other words, wage differences in Sweden probably reflect local bargaining realities and not only compensating differentials for unpleasant work. The literature on discrete choice demand estimation emphasizes that the potential correlation of prices with unobserved product characteristics is a major impediment to consistent estimation of consumer preferences (Berry, 1994). If Swedish institutions affect wages but not employer choices directly, they create wage variation that aids in identifying the preferences of workers. Furthermore, the formal process behind choosing a stochastic process for age–wage profiles at a firm implies that the model's assumption of employees being promoted at different speeds within the same stochastic process for wages may be more accurate than an alternative assumption of an individual-specific market for the labor supply of each worker.

This paper estimates the relative importance of switching costs in employer switching decisions. In the model, workers have perfect information about expected outcomes at all employers. I therefore view switching costs as most likely representing psychic costs of disrupting work and living routines, rather than informational frictions. However, data on switching where workers do not arbitrage away all wage differences can also be explained by search costs and related statistical matching processes (van der Berg and Ridder, 1998; Postel-Vinay and Robin, 2002). In fact, the across-firm wage distribution can be explained as an equilibrium where the underlying firms and workers are identical (Burdett and Mortensen, 1998). On the other hand, several characteristics of the labor market for engineers suggest that pure search costs might not be the most important friction to base a model around. The employers of engineers (the firms in the data) are dominated by large manufacturing concerns, many representing firms well-known both within Sweden and internationally. Engineers and their union representatives should be familiar to a large degree with the characteristics of these employers,

---

<sup>3</sup>The legal relationship between employer seniority and layoff risk means that switching costs may be higher in Sweden. However, unemployment for engineers is close to zero over the sample period, so the cost of a layoff is more likely to be a psychic disruption cost, rather than a large salary decrease.

<sup>4</sup>For engineers, 21% of all observed moves are to an establishment owned by the same legal company.

<sup>5</sup>I do not translate Swedish acronyms for labor unions.

and at the least could conduct a targeted search for a new position. The reality is probably that both search and switching costs play a role in decreasing switching.<sup>6</sup>

While I believe most switching costs are psychic disruption costs, institutional features can also affect switching costs. In Sweden, the government provides or heavily regulates health care and pension plans, and switching firms does not reduce these welfare state benefits. On the other hand, capital gains taxes on housing may interfere with geographic mobility (Lundborg and Skedinger, 1998). Likewise, some Swedes live in rent-controlled apartments, which the renters cannot use if they switch to employers in different regions. To the extent that the switching costs reflect these institutional features of Sweden, the structural parameters will be less reflective of the preferences of non-Swedish workers.

### 3 Model and Estimation

Only 14% of the engineers in the cleaned SAF data used in the estimation switch employers during a five-year period. One explanation is Swedish workers have high switching costs that lead to employer lock-in for experienced workers. An alternative explanation is Swedish workers are not locked in, and the Swedish wage structure is not designed to attract experienced workers to new firms. I need to estimate workers' responsiveness to wages and switching costs in order to empirically distinguish between these explanations for low switching rates. This section outlines a structural model of voluntary employer switching that allows me to estimate the preferences of Swedish engineers for employer characteristics, including switching costs and wages.

#### 3.1 Current Period Utility Function

For a worker  $i$  of age  $t$  (before the retirement age  $T$ ) and with overall state vector  $s_{i,t}$ , his current period utility function if he chooses employment at firm  $j$  out of  $J$  total firms is

$$u(j_{i,t}, s_{i,t}; \beta) + \varepsilon_{i,j,t} = \beta_w E[w_{i,j,t} | s_{i,t}, j_{i,t}] + \beta_{g,t} \text{dist}(j_{i,t-1}, j_{i,t}) + \sum_{l=1}^3 \beta_{m_l,t} m_l(j_{i,t-1}, j_{i,t}) + \beta_{s,t} \text{size}(j_{i,t}) + \varepsilon_{i,j,t}. \quad (1)$$

The formal goal of the econometric procedure is to estimate the parameter vector

$$\beta = \left\{ \beta_w, \left\{ \beta_{g,t}, \beta_{m,1,t}, \beta_{m,2,t}, \beta_{m,3,t}, \beta_{s,t} \right\}_{t=1}^T \right\},$$

which describes the relative importance of the four types of employer characteristics that enter the current period utility function.

I now describe the employer characteristics. For this study, the most important characteristic is the wage  $w_{i,j,t}$  of worker  $i$  at employer  $j$ . The parameter  $\beta_w$  controls how sensitive employer switching is to wage dif-

<sup>6</sup>There are still other models that explain the persistence of employment relationships, particularly the decline in turnover with age. For example, job search may give workers a better worker-firm match as they age. To the extent that wages in the good match are higher than bad-match outside options, the good match will persist. Older engineers in my sample do not earn large wage decreases upon an (apparently) voluntary firm switch, but as this is a selected sample, it is not strong evidence against the model that emphasizes job-worker matches. To some extent, the switching costs that increase with age in my model can be seen as reflecting better psychic match values for older workers, but my model does not encompass wages that reflect the quality of the job-worker match unobservable, in either production or a worker's utility. Some previous empirical work has thought about the match between workers and careers. By looking at a relatively homogeneous worker class (engineers with five-year degrees) in one national labor market, I hope to make jobs more comparable across firms, so the notion of an unobservable match term for each pair of a worker and a firm is less relevant in explaining the switching variation in the data.

Note that many search models also lack a match-specific component in either production or wages (Burdett and Mortensen, 1998; Postel-Vinay and Robin, 2002).

ferences across firms.<sup>7</sup> Because the choice of employer  $j_{i,t}$  is discrete, it is really the value of  $\beta_w$  relative to the other parameters and to the normalized distribution of the taste shocks  $\varepsilon_{i,j,t}$  that is important for understanding the role of wages in switching in the labor market.

Switching costs are an important part of utility. If I did not include switching costs, there would be little explanation for why workers do not instantaneously arbitrage wage differences across firms. In this implementation, switching costs depend on the relative locations, in employer characteristic space, of the origin and destination firms. First, the model includes a switching cost in geographic distance  $\beta_{g,t} \text{dist}(j_{i,t-1}, j_{i,t})$ , which is defined as an estimated parameter times the log distance in kilometers between the capitals of the Swedish counties in which the old,  $j_{i,t-1}$ , and potential new,  $j_{i,t}$ , firms are located.<sup>8</sup> The terms  $\{m_l(j_{i,t-1}, j_{i,t})\}_{l=1}^3$  are indicator variables that refer to employer changes that may cause a worker to incur non-geographic switching costs. One term refers to a base switching cost, which a worker incurs even if he moves to a new employer across the street. An additional switching cost is incurred if the worker switches between firms in different industries. Finally, the data show that 21% of observed switches for Swedish engineers in the estimation sample are between establishments owned by the same legal corporation.<sup>9</sup> The final switching cost is incurred if a worker transfers outside of the boundaries of his corporation. The specification allows the switching cost parameters to vary by worker age  $t$  because, empirically, employer switching decreases with age.

The term  $\text{size}(j_{i,t})$  in (1) is the log of the number of white collar workers (not only engineers) at employer  $j_{i,t}$ . I include employer size to proxy for the nonwage benefits received by workers at firm  $j_{i,t}$ .<sup>10</sup>

The final term in the current period utility function is a worker, age and employer-specific taste shock  $\varepsilon_{i,j,t}$ . Empirically, the shock allows the model to match the data, where observationally equivalent workers make different choices. Economically, the relative importance of the taste shock compared to other parameters represents the contribution of idiosyncratic life developments to employer switching. If the patterns of turnover in the data are not driven by differences in wages and other characteristics, it is likely that employer switching is caused by the idiosyncratic developments captured by  $\varepsilon_{i,j,t}$ . The empirical model assumes that  $\varepsilon_{i,j,t}$  is i.i.d.<sup>11</sup> and has the logit distribution, which greatly simplifies the estimation of the dynamic programming problem that I will now outline (Rust, 1987).<sup>12</sup>

## 3.2 Dynamic Employer Choice

The econometric method finds the parameters of the current period utility function that make the employer choices in the data the most likely of all outcomes possible in a discrete choice, dynamic programming model. Worker  $i$  of age  $t$  and overall state vector  $s_{i,t}$  picks a firm  $j_{i,t}$  to maximize his expected, present discounted value

<sup>7</sup>I discuss below the reason for using the expected value of the wage instead of the actual wage.

<sup>8</sup>I define  $\text{dist}(j_{i,t-1}, j_{i,t})$  to be 0 for firms in the same county.

<sup>9</sup>It is important to distinguish between establishments owned by the same legal corporation in order to preserve the geographic nature of an employment choice.

<sup>10</sup>In a previous investigation, I estimated a simple version of the model with employer-specific fixed effects to control for nonwage amenities, as suggested by the literature in industrial organization (Berry, 1994). I then regressed the estimated fixed effects on other employer characteristics, and found that the only statistically significant characteristic is employer size. In the model presented here, it is computationally infeasible to include employer-specific fixed effects, so the model includes employer size as a proxy for the total nonwage benefits at an employer.

<sup>11</sup>If  $\varepsilon_{i,j,t}$  is autocorrelated across a career for the same employer (an assumption that complicates computation), the model is a statistical matching model where the match quality reflects the worker's idiosyncratic evaluation of the employer's workplace environment. This contrasts to other models of statistical matching, where career-employer shocks primarily affect worker productivities and therefore wages (Jovanovic, 1979).

<sup>12</sup>The independence from irrelevant alternatives property of the static logit model does not exist in dynamic models, because the choice-specific value functions,  $v(s_{i,t}, j_{i,t})$ , are discounted and weighted (by the probabilities of future employer choices) sums of the characteristics of all employers.

of utility

$$E \left[ \sum_{\tau=t}^{60} \delta^{\tau-t} (u(s_{i,\tau}, j_{i,\tau}; \beta) + \varepsilon_{i,j,\tau}) \mid s_{i,t}, \varepsilon_{i,t} \right],$$

where the retirement age  $T$  is 60 years,  $\delta$  is a discount factor between 0 and 1, and  $\varepsilon_{i,t}$  is the vector of employer-specific taste shocks at age  $t$ .

Because this is a finite-horizon and discrete-time dynamic programming problem, I solve for the model's predicted probabilities of employer choice  $P(j_{i,t} \mid s_{i,t}; \beta, \theta_w)$  by backwards recursion. Let  $h(s_{i,t+1} \mid s_{i,t}, j_{i,t}; \theta_w)$  be the transition density of the state variable  $s_{i,t}$  conditional on the worker's choice of employer  $j_{i,t}$ . The transition density is parameterized by  $\theta_w$ . Define the continuation value to be

$$\begin{aligned} V(s_{i,t}) &= \int \max_{j_{i,t} \in \{1, \dots, J\}} E \left[ \sum_{\tau=t}^{60} \delta^{\tau-t} (u(s_{i,\tau}, j_{i,\tau}; \beta) + \varepsilon_{i,j,\tau}) \mid s_{i,t}, \varepsilon_{i,t} \right] g(\varepsilon_{i,t}) d\varepsilon_{i,t} \\ &= \int \max_{j_{i,t} \in \{1, \dots, J\}} \left[ u(s_{i,t}, j_{i,t}; \beta) + \varepsilon_{i,j,t} + \delta \int_{s_{i,t+1} \in S} V(s_{i,t+1}) h(s_{i,t+1} \mid s_{i,t}, j_{i,t}; \theta_w) ds_{i,t+1} \right] g(\varepsilon_{i,t}) d\varepsilon_{i,t} \\ &= \int \max_{j_{i,t} \in \{1, \dots, J\}} [v(s_{i,t}, j_{i,t}) + \varepsilon_{i,j,t}] g(\varepsilon_{i,t}) d\varepsilon_{i,t}, \end{aligned} \quad (2)$$

where  $g$  is the joint density of the worker's vector of taste shocks  $\varepsilon_{i,t}$ , and where  $v(s_{i,t}, j_{i,t})$  is a choice-specific value function.<sup>13</sup> If the worker-specific shock has the logit distribution, then the probability of picking firm  $j_{i,t}$  for an age  $t$  worker is

$$P(j_{i,t} \mid s_{i,t}; \beta, \theta_w) = \frac{\exp(v(s_{i,t}, j_{i,t}))}{\sum_{j=1}^J \exp(v(s_{i,t}, j_{i,t}))}.$$

This formula features the relevant choice's present discounted value of payoffs in the numerator, normalized by the sum of the payoffs of all choices in the denominator. Again, estimation proceeds by making the choice probabilities in the data the most likely of all outcomes from the model.

### 3.3 States of the Employer Choice Problem

The states of a dynamic programming model are similar to the right hand side variables that distinguish one worker's choice problem from another's. In order for the dynamic programming problem to have economic meaning, I must define the components of the vector of these state variables,  $s_{i,t}$ , and I must describe the functional forms for the transition density of the state variables,  $h(s_{i,t+1} \mid s_{i,t}, j_{i,t}; \theta_w)$ .

The state variables are the age of a worker,  $t$ , his previous employer,  $j_{i,t-1}$ , and his wage at his previous employer,  $w_{i,t-1,t-1}$ . The age of a worker captures his outside wage opportunities, his current level of switching costs, and the remaining career over which he can enjoy any wage gain. In the model, the worker chooses an employer every five years, so age,  $t$ , corresponds to a five-year interval.<sup>14</sup> The previous employer,  $j_{i,t-1}$ , of a worker captures his current position in terms of geographic location, corporate membership, and industry affiliation. If he switches firms, his switching costs are a function of how these firm characteristics change. Finally, the lagged wage,  $w_{i,t-1,t-1}$ , captures the heterogeneous positions of workers within their current employers and in the labor market as a whole. The lagged wage is a sufficient statistic for a worker's future wage outcomes,

<sup>13</sup>In practice, to calculate  $V(s_{i,t})$  I use the simplification  $V(s_{i,t}) = \sum_{j_{i,t}=1}^J P(j_{i,t} \mid s_{i,t}; \beta, \theta_w) \left\{ v(s_{i,t}, j_{i,t}) + E \left[ \varepsilon_{i,j,t} \mid j_{i,t}^* = j_{i,t} \right] \right\}$ , where  $E \left[ \varepsilon_{i,j,t} \mid j_{i,t}^* = j_{i,t} \right] = 0.577 - \log P(j_{i,t} \mid s_{i,t}; \beta, \theta_w)$  by the properties of the extreme value (logit) distribution.

<sup>14</sup>In the data, an engineer can switch multiple times over a five-year interval. Failure to account for this will tend to overstate switching costs. In the data, the number of engineers making multiple switches is small.

conditional on his path of employer choices into the future.<sup>15</sup>

In the dynamic programming model, a worker's discrete age category,  $t$ , evolves according to the laws of time, and, from the worker's point of view, his lagged employer,  $j_{i,t-1}$ , is a discrete choice he has made and will make into the future. The only non-deterministic portion of the state transition density function  $h(s_{i,t+1} | s_{i,t}, j_{i,t}; \theta_w)$  is the portion

$$f(w_{i,j,t} | w_{i,j_{i,t-1},t-1}, j_{i,t}, j_{i,t-1}, t; \theta_w)$$

corresponding to the uncertain evolution of wages. I write out the full state vector to emphasize that the density of future wages is a function of the individual-specific lagged wage. I estimate the unknown parameters  $\theta_w$  in first-stage wage regressions. While the density is identified, there are not enough engineers working in the Swedish private sector to estimate  $f(w_{i,j,t} | w_{i,j_{i,t-1},t-1}, j_{i,t}, j_{i,t-1}, t; \theta_w)$  nonparametrically, so I impose parametric functional forms for  $f$ . The parametrization of a transition density is somewhat arbitrary, and below I outline a parsimonious specification that addresses unobserved heterogeneity in worker ability while capturing across-firm differences in wage policies.

After estimating  $\theta_w$ , I compute means and variances of  $w_{i,j,t}$  conditional on  $(w_{i,j_{i,t-1},t-1}, j_{i,t}, j_{i,t-1}, t)$ , and use the means and variances as the parameters for a normal distribution of one period ahead wages,  $f$ . For a known series of future employment decisions by a worker, the combination of age-specific, one-period-ahead densities creates a flexible Markov process for both the within and across-firm evolution of wages. From now on I call the stochastic process of wages at a particular firm its *stochastic process for age-wage profiles*, to emphasize the long-term implications of employment decisions and wage changes in the current period.

The question of what parametrization to use for  $f$  remains. There are countless ways to summarize a stochastic process for wages. For example, Abowd, Kramarz and Margolis (1999) decompose the level of wages into firm-specific fixed effects, worker-specific fixed effects, and the effects of observable time-varying characteristics such as worker experience. Abowd et al. are well aware that they have performed a statistical decomposition, which does not always correspond to the primitives of an economic model, as shown by, for example, Woodcock (2003).<sup>16</sup> I now discuss the parameterizations of the wage processes for employer stayers and employer switchers separately.

### 3.3.1 Wage Process for Employer Stayers

For a parametrization for the stochastic process of age-wage profiles at an employer to be useful as an input into the employer choice model, it must capture heterogeneity in both worker abilities and in employer wage policies. In order to capture heterogeneity in worker ability, I estimate log wage regressions where the dependent variable is a percentage wage increase. In a wage increase regression, previous information on a worker's ability that has already been incorporated into wages is differenced out.<sup>17</sup> From the point of view of an agent in the employer choice model, first differencing relates to a symmetric informational assumption

<sup>15</sup>This corresponds to a model where a worker and all firms have symmetric uncertainty about his present and future ability.

<sup>16</sup>The conditional independence (i.i.d. errors) assumption of Rust (1987) allows the stochastic processes for wages to be estimated in a first stage. Intuitively, the information agents possess about the future is assumed to be equivalent to the variability in future wages observed by the econometrician. Carneiro, Hansen and Heckman (2003) and Navarro (2007) estimate the information agents possess about future outcomes, using behavioral assumptions in addition to statistical decompositions. These methods have been used for the low-dimensional choice problem of whether to attend college, and would be difficult to generalize to the dynamic employer choice problem studied here.

<sup>17</sup>In fact, specifications where the independent variable is the level (not change) in log wages produces the surprising result that firms paying higher wages have greater turnover. See Section 7.2. One explanation for a nonpositive relationship between wages and retention is that a wage level specification does not control for unobserved heterogeneity in worker ability. Firms paying a high wage level may have better workers, and those better workers may be able to get similarly-paid jobs at other firms.

about worker ability: the firm incorporates known information into wages, and the worker has no additional private information.

In order to capture heterogeneity in firm compensation policies, I estimate a firm and age-specific fixed effect in percentage raises. If workers are forward looking, their switching behavior will respond not only to the level of wages as a firm, but whether or not workers can expect to receive a large wage increase in the future. For white collar workers, descriptive empirical work in Fox (2008b) suggests that differences in the stochastic processes of wages in large and small Swedish firms are best captured by differences in slopes: the levels of percentage wage increases are higher in larger than in smaller firms.<sup>18</sup> Young workers would all go to smaller firms if they choose firms myopically. However, the rapid pace of wage increases means that the long-term payoff to working at large firms may be higher for forward-looking agents.

For those workers of age  $t$  who remain at a firm  $j$  over the period, I estimate  $\gamma_{j,t}$  and  $\sigma_{j,t}^2$  from the linear specification

$$w_{i,j,t} - w_{i,j,t-1} = \gamma_{j,t} + \eta_{i,j,t}, \text{ if } t > 30, \text{Var}(\eta_{i,j,t}) = \sigma_{j,t}^2, \quad (3)$$

where the  $\gamma_{j,t}$ 's and the  $\sigma_{j,t}^2$ 's are firm-specific parameters that capture the means and the variances of the percentage raises given to age  $t$  workers. In practice, I let the difference between ages  $t$  and  $t - 1$  represent a time period of five years, which speeds up the dynamic programming problem by reducing the state space.

Conditional on a worker staying for all future periods, the functional form choice for percentage wage increases creates a random walk with drift stochastic process for wage levels. A random walk with drift has the important property that shocks to the level of wages are permanent, in expectation.<sup>19</sup> Figure 1 shows this property for three different workers at the same firm. The solid line in the middle of the graph shows the average age–wage profile of a career stayer at the firm. The dashed line above the average profile reflects a worker who receive a higher-than-average first period wage increase, while the lower dashed line shows a worker with a negative first period wage shock. Between periods 2 and 3 the random walk with drift assumption means that all three workers are predicted to have the same percentage increase in their wage, which is reflected in the equal slopes of the second period wage change in Figure 1.

### 3.3.2 Wage Process for Employer Switchers

In the employer switching model, a worker predicts how his wage will change when he moves from his current employer's stochastic process for age-wage profiles to another firm's. The Swedish data allow the examination of the wage increases of employer switchers based upon the identities of the origin and destination employers.

The model's prediction about a worker's initial wage offer in a new firm is the convex combination of a hypothetical wage in the old firm and a hypothetical wage in the new firm. The two hypothetical wages are calculated using the within-firm age–wage profiles whose estimation I discussed in Section 3.3.1. I estimate the weights on the two hypothetical wages by regressing the actual wages received by firm switchers on the two hypothetical wages. To be more explicit, I estimate the model

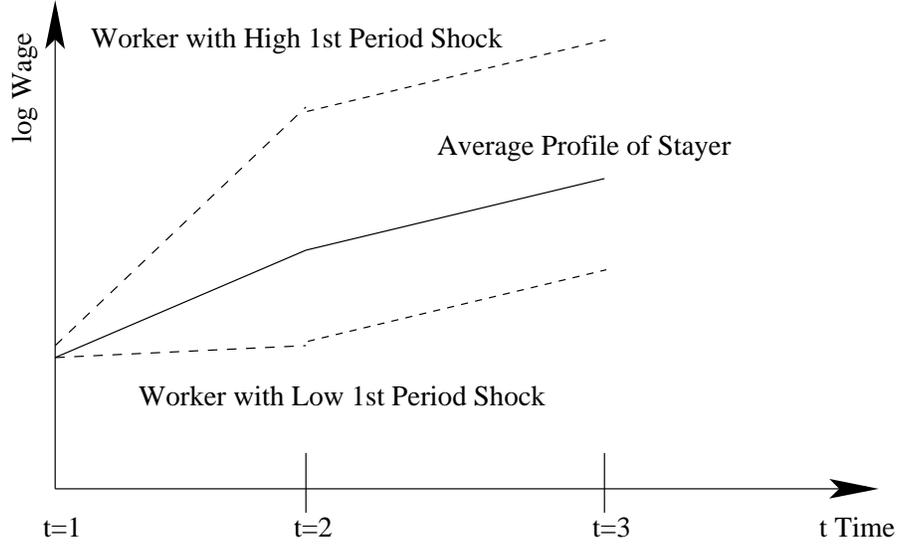
$$w_{i,j,t} = \zeta_{\text{old},t} w_{i,j_{t-1},t}^{\text{hypothetical\_old}} + \zeta_{\text{new},t} w_{j,t}^{\text{hypothetical\_new}} + v_{i,j,t}, \text{Var}(v_{i,j,t}) = \tau_{j,t}^2 \quad (4)$$

where  $w_{i,j_{t-1},t}^{\text{hypothetical\_old}}$  is the hypothetical wage at the old firm,  $j_{t-1}$ , and  $w_{j,t}^{\text{hypothetical\_new}}$  is the hypothetical wage at the new firm  $j$ . For each age category  $t$  and firm  $j$ , I estimate three parameters: the wage weights

<sup>18</sup>Fox (2008b) also shows that the cross-sectional firm-size wage gap between large and small American increases with worker age.

<sup>19</sup>Topel and Ward (1992) present evidence that a random walk with drift approximates the evolution of the wages of workers. However, Baker (1997) shows that such statistical tests have low power, and that wage processes may have individual-specific rates of wage increase. Meghir and Pistaferri (2004) estimate wages as the sum of a random walk and a transitory component.

Figure 1: Functional Form for Firm Stayers: Random Walk with Drift



$\zeta_{old,t}$  and  $\zeta_{new,t}$ , as well as the variance of the error term  $\tau_{j,t}^2$ . I do not include a constant term in the regression, which makes the weights  $\zeta_{old,t}$  and  $\zeta_{new,t}$  tend to sum to 1.0, although I do not impose this constraint in estimation. Because I only observe a few switchers at every firm and age group, I estimate the equation using all observations for firm switchers at once.

A key feature of my data is that I observe the entire workforce of individual firms, so I can create  $w_{i,j_{t-1},t}^{hypothetical\_old}$  and  $w_{j,t}^{hypothetical\_new}$  using within-firm data. My assumption that within-firm wages are a random walk with drift means that a decent prediction of the wage of a stayer is

$$w_{i,j_{t-1},t}^{hypothetical\_old} = w_{i,j_{t-1},t-1} + \hat{\gamma}_{j,t},$$

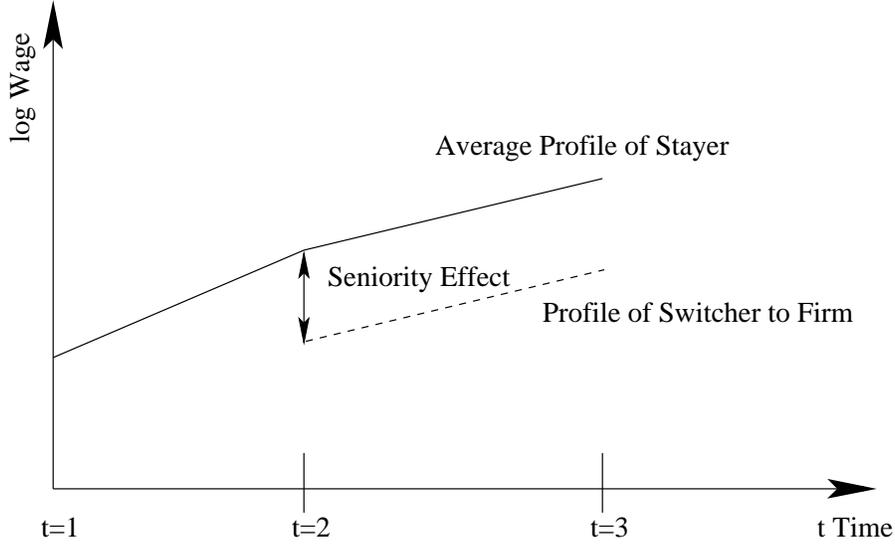
where  $w_{i,j_{t-1},t-1}$  is the observed previous wage paid to the employee, and  $\hat{\gamma}_{j,t}$  is the estimated average percentage raise paid to stayers of age  $t$  and firm  $j$ . I estimated  $\hat{\gamma}_{j,t}$  in (3) above. The wage  $w_{i,j_{t-1},t}^{hypothetical\_old}$  is an individual-specific measure, because it is a function of the observed lagged wage of a worker. I estimate  $w_{j,t}^{hypothetical\_new}$  using an even simpler method: I calculate the average log-wage (in levels) paid to firm stayers of the same age category at the potential destination firm, or

$$w_{j,t}^{hypothetical\_new} = \frac{1}{I_{j,t,stay}} \sum_{i=1}^{I_{j,t,stay}} w_{i,j,t},$$

where there are  $I_{j,t,stay}$  observed stayers in age category  $t$  at the destination firm  $j$ . Unlike the measure for the worker's current firm,  $w_{j,t}^{hypothetical\_new}$  is not an individual-specific measure, because it takes into account only the circumstances of workers currently at the potential destination firm.<sup>20</sup>

<sup>20</sup>One criticism of the functional form assumption in (4) is that a worker's heterogeneous position in the labor market is captured only by  $w_{i,j_{t-1},t}^{hypothetical\_old}$ , and not the identity of firm  $j_{t-1}$  in a more direct way. There is no reason to suspect that the same wage at all other firms should be treated the same; for example, a firm with an unpleasant work environment might pay all workers a high wage, so that an individual worker's high wage at that firm would not be a symbol of labor market success. In principal, the nonparametric density formulation for the Markov process of wages,  $f(w_{i,j,t} | w_{i,j_{t-1},t-1}, j_{i,t}, j_{i,t-1}, t; \theta_w)$ , allows for a worker's wage to be interpreted in light of the identity of his previous firm,  $j_{i,t-1}$ . In practice, I do not have enough observations to be so flexible. However, another paper could pool observations across groups in different ways to create more observations and make an origin firm-specific treatment of wages possible.

Figure 2: Age–Wage Profiles for Firm Switchers and the Port-of-Entry Seniority Effect



Some readers may be interested in how the functional form for the wage changes of employer switchers relates to the literature on distinguishing the returns to total experience in the labor market (or age for workers without career interruptions) from the returns to seniority (or tenure) at a particular employer (Topel, 1991; Altonji and Williams, 1998). My assumption is that any seniority effect happens when a worker switches firms, and afterwards that worker becomes a firm stayer and his wages increase with the random walk with drift process for stayers. Figure 2 compares the wages of a career firm stayer to a worker who moves to the firm (the dashed line) at time 2. At time 2, the wage of the switcher is placed above or below the profile of the stayer, but afterwards the switcher becomes a stayer whose percentage wage increases are the same, in expectation. The port-of-entry seniority effect is the degree that the wages of the switcher fall below the wages of the career stayer.

Note that the structural model of employer choice is internally consistent. The key assumption is the wage that enters the current period utility function (1) is the expected individual-specific wage for selecting employer  $j_{i,t}$ , or

$$E[w_{i,j,t} | s_{i,t}, j_{i,t}] = \int w_{i,j,t} f(w_{i,j,t} | w_{i,j,t-1}, j_{i,t}, j_{i,t-1}, t_i; \theta_w) dw_{i,j,t}.$$

This expectation is a function of only the state variable of the worker,  $s_{i,t}$ , which includes his lagged wage and his lagged employer choice. The force of the assumption is that workers do not receive specific wage offers before deciding between employers.<sup>21</sup> Instead, workers pick firms based upon individual-specific densities of one-period-ahead and, through the dynamic programming problem, future wages.<sup>22</sup>

<sup>21</sup>More formally, the model does not include unobserved state variables that violate the conditional independence assumption of Rust (1987) and hence directly affect the transition density of observed and unobserved state variables.

<sup>22</sup>A worker's information set is smaller than in an endogenous censoring model where workers receive specific wage offers at all firms, but more than in a search model where workers need to pay costs (or wait longer) before receiving any information on the wages at specific firms.

### 3.3.3 Selection

The data include wage offers at only firms where workers accept employment. Readers familiar with the literature on endogenous censoring may be wondering why the empirical method does not use the employer choice model to selection-correct the stochastic processes of wages at all firms. The answer is that the large number of employer choices (more than 300 per five-year interval) in this application creates serious data hurdles, even before considering the extra complexities that the model in this paper emphasizes, such as dynamic programming, the stochastic evolution of wages, and unobserved heterogeneity in the utility functions of workers.<sup>23</sup>

Parametric selection models (Heckman, 1974; Gronau, 1974) generally assume joint normality of the error terms for the wages at (in my case) all employers and the taste shocks from (again in my case) all employer choices. Because joint estimation of all equations is often computationally infeasible,<sup>24</sup> researchers typically estimate employer choice probabilities in a first stage using the state variables of workers and related nonwage employer characteristics (such as switching costs) as included covariates. In particular, next period's wage density at each of the firms is not part of the data used in the first stage employer choice estimation. This means that, in my employer choice problem, the first stage reduced form must have enough observations to accurately predict more than 300 employer choices per set of covariates without using the wages that are potentially strong drivers of employer switching.<sup>25</sup> The data on Swedish engineers do not have enough observations to accurately estimate employer choice probabilities using reduced forms. This paper reverses the two stages recommended by the selection-correction literature. Following the literature, such as Rust (1987), on estimating single-agent dynamic programming models, I estimate the firm-specific stochastic processes of wages in the first stage, and then use those processes as inputs into the second stage estimation of the underlying preferences of workers in my employer choice model.<sup>26</sup> The empirical method's focus on estimating worker preferences through the revealed preferences from employer choices also contrasts with the literature on the structural estimation of equilibrium search models, which most commonly uses theory to derive implications for the functional form and interpretation of wage equations (van der Berg and Ridder, 1998; Postel-Vinay and Robin, 2002; Woodcock, 2003).<sup>27</sup>

## 3.4 Estimating Preference Heterogeneity

In a product market, firms compete by lowering prices to attract the marginal consumers who are indifferent between keeping their current products or switching to new ones. In this application to the labor market, I wish to determine the fraction of experienced engineers who would switch firms for small wage increases. Estimating preference heterogeneity is thus crucial to understanding how labor markets for experienced workers operate. Heterogeneity could potentially enter the model both in the wage process and the current utility functions of workers. Because the wage process is a series of flexible, one-period-ahead deviations as a function

---

<sup>23</sup>Dahl (2002) estimates a selection model with many choices (the 50 American states). However, as he points out, he requires an assumption that is inconsistent with the primitives of his economic model.

<sup>24</sup>The numerical integration required by the multivariate normal distribution suffers from a curse of dimensionality.

<sup>25</sup>Heckman and Honore (1990) and Heckman (1990) investigate the nonparametric identification of endogenous censoring models. Their preferred identification strategy is identification at infinity. A covariate  $Z$  that varies greatly over the sample is included in the employer choice problem, and excluded from the wage equations.  $Z$  varies enough so that for some extreme values of  $Z$ , the employer choice is made with certainty. In my application, worker age is a variable that varies over the sample. However, age cannot be excluded from the wage equation, and variation in age affects the decision of whether to switch at all much more than the choice of the particular firm (out of 300) when switching.

<sup>26</sup>Heckman (1974), for example, uses a choice problem to get better estimates of the parameters in a wage regression. This paper uses wage regressions to get estimates of the parameters of a choice problem.

<sup>27</sup>The empirical method also differs from search models by ignoring unemployment. The market for Swedish engineers had extremely low unemployment during the sample period.

of a worker's lagged wage, there is already some heterogeneity in the stochastic process for wages. Therefore, the model adds extra heterogeneity to the unknown parameters in a worker's current period utility function. In principle, each discrete worker type has unique preferences for wages, switching costs, nonwage benefits, and taste shocks. The workers who value wages more than other characteristics are more willing to switch employers for small increases in outside wage offers. The model assumes that each firm does not condition its stochastic process for age–wage profiles on worker type.

Because I use a nested dynamic programming approach to estimate the employer choice model, the computer must solve one dynamic programming problem for each worker type, at each set of trial parameters in the nonlinear maximization routine. This computational burden forces me to use a small number  $K = 3$  of worker types. I estimate a separate vector  $\beta_k$  of parameters of the current period utility function, (1), for each type  $k$ . I also estimate a probability  $\pi_k$  of each worker being each type, which corresponds to a discrete version of random effects.<sup>28</sup>

A statistical observation for worker  $i$  of age  $t$  is a triplet  $(s_{i,t}, j_{i,t}, s_{i,t+1})$  of his state vector,  $s_{i,t}$ , his observed employer choice,  $j_{i,t}$ , and his new state vector after the choice,  $s_{i,t+1}$ . The final pseudo-likelihood for both the discrete choice model of employer choice and the transition density of wages is

$$\tilde{L} = \prod_{i=1}^I \sum_{k=1}^K \pi_k \prod_{y \in Y_i} P(j_{i,t,y} | s_{i,t,y}; \beta_k, \theta_{w,y}) h(s_{i,t+1,y} | s_{i,t,y}, j_{i,t,y}; \theta_{w,y}), \quad (5)$$

where I factor the joint density of the dependent variables  $j_{i,t,y}$  and  $s_{i,t+1,y}$  into the unconditional choice probabilities  $P(j_{i,t,y} | s_{i,t,y}; \beta_k, \theta_{w,y})$  and the conditional-on-choice transition densities of the state variable  $h(s_{i,t+1,y} | s_{i,t,y}, j_{i,t,y}; \theta_{w,y})$ . Because  $h$  is not a function of  $\beta_k$  or  $k$ , I can multiplicatively factor  $h$  out of the likelihood and estimate  $\theta_{w,y}$  consistently in a first stage procedure. The first stage is the wage regressions discussed in Section 3.3.<sup>29</sup>

The state vector,  $s_{i,t,y}$ , does not explicitly include changes in the overall labor market for Swedish engineers over time. Because computing an equilibrium to a model of a firm's entry, exit and age–wage profile choices in the labor market at every trial parameter value is difficult, the model treats the industry equilibrium as fixed from a worker's point of view during each five-year interval. The worker examines the outcomes of older workers as evidence for what will happen to him when he ages, and he is unaware that current labor market conditions will change.<sup>30</sup> I index each sample interval (a five-year period in the data) by  $y$ , and I solve a separate dynamic programming problem for each worker type  $k$  during each such  $y$ . Intuitively, identification of each  $\beta_k$  comes from the actions of workers of different ages  $t$  within the same cross section  $y$ , while identification of each  $\pi_k$  and the heterogeneity in the  $\beta_k$ 's comes from looking at repeated observations of the same worker  $i$  in different economic environments  $y$ .<sup>31</sup>

<sup>28</sup>The probability of being each type,  $\pi_k$ , is independent of worker covariates because I restrict the sample to male, full-time Swedish engineers working in the private sector. There are no extra individual-specific covariates to condition the type probabilities on.

<sup>29</sup>For dynamic programming, I discretize the wage into 20 bins based on the 5, 10, 15... quantiles of the five-year-age-interval specific wage distribution. I apply this discretization to the continuous wage processes estimated in the first stage. The wage processes then move an age 25–29 worker from a wage bin specific to age 25–29 to a wage bin specific to age 30–34.

<sup>30</sup>The assumption is workers believe that the experiences of older workers are good predictions of their own future outcomes. Even if in reality differences between older and younger workers are driven by cohort and not age effects, the assumption is correct if workers naively ignore cohort effects.

<sup>31</sup>Rust (1994) and Magnac and Thesmar (2002) prove that agent payoffs are not nonparametrically identified in most infinite-horizon settings, and in finite-horizon dynamic programming problems where the current-period payoff vector is not restricted. In this employer switching case, identification comes from the finite-horizon nature of careers, and the fact that the same linear index in parameters appears for all time periods. In the last period before retirement, a worker's employer choice is completely static, and the agent's payoffs are semiparametrically (assuming a linear index and i.i.d. errors in payoffs, but not a parametric distribution for the error terms) identified, as has been shown by Manski (1975) and Thompson (1989). The identification of the linear index in earlier time periods requires both the parametric assumption for tractable dynamic programming, and more importantly, the common set of parameters in the linear index in order to fix the intertemporal distribution of payoffs.

The number of yearly observations for a given worker is  $Y_i$ , out of a possible  $Y = 15$  (1985-1971+1) such observations. I call  $\tilde{L}$  a pseudo-likelihood because the model treats cross-sectional observations (different  $y$ 's) as independent when constructing the likelihood. As an example, the interval 1971–1976, with the state variable containing information from the 1970 data, is one  $y$ , while the interval 1972–1977, with 1971 state information, is another. Because of entry and exit of firms and other labor market changes, both intervals have unique economic environments that cannot be directly related. The number and identity of employer choices, as well as firm-specific stochastic processes for age–wage profiles, may look different from one year to another, as Section 4 discusses. I do not adjust the likelihood (or the standard errors) to enforce the fact that switching firms in 1971 changes the state variable for the same worker's choice in 1972.<sup>32</sup>

The computationally burdensome part of the estimation routine is that, for each trial parameter vector  $\{\beta_k, \pi_k\}_{k=1}^K$ , I must solve  $K \cdot Y = 3 \cdot 15 = 45$  finite horizon dynamic programming problems for the value function,  $V(s_{i,t})$ , using backwards recursion. Each dynamic programming problem features decisions that are made every time interval, for a total of five relevant five-year worker age intervals. The number of states far exceeds the number of observations in any given year,  $y$ . Both the solution of each dynamic programming problem and the outer maximization of the likelihood function suffer from separate curses of dimensionality. Estimation of  $\{\beta_k, \pi_k\}_{k=1}^K$  takes several days using a gradient-based optimization routine and a 2007-vintage computer.<sup>33</sup>

### 3.5 Worker Tenure

In Section 3.3.2, I discussed how the model addresses tenure (the length of time a worker has held a particular job) in wage regressions. Another well-known empirical fact is that the probability of a worker switching employers declines with the worker's tenure (Farber, 1999a). In the employer-choice model, a worker's age  $t$  is a state variable, but the number of years he has been employed at a particular employer is not. It would be mathematically simple to use panel data to add worker tenure as an observed state variable to the model. Unfortunately, using tenure will create many missing observations, corresponding to those workers who started a spell of employment before the beginning of the sample. Dropping these workers will drop workers with higher switching costs or lower sensitivities to outside wage offers, which will create biases in the distribution of preferences that I estimate.

Therefore, the model lacks specially-introduced features to match facts about worker tenure. This is not necessarily a major concern. The model will explain a worker with low tenure for his age as a worker with low switching costs. Likewise, workers with especially high tenures for their ages have high switching costs. Rather than any special, causal role of tenure in employer switching, heterogeneity in worker preferences can match the usual empirical regularity that switching rates decline with worker tenure. As tenure increases, the fraction of workers with high switching costs increases, as the low-switching-cost workers have already left.

---

To my knowledge, no one has addressed the more complicated issue of identifying agent-specific preferences in a finite-horizon setting, without using information on additional outcomes. Note that worker-specific preferences (fixed effects) solve the initial conditions problem (Heckman, 1981) of the current state variable being correlated with worker-specific preferences (because the preferences  $\beta_i$  comprise an estimated parameter vector), although they are computationally infeasible and data intensive.

<sup>32</sup>In addition, I do not adjust the standard errors from the employer choice problem to account for first-stage estimation error in the wage regressions. Calculation of exact standard errors is computationally burdensome, because it involves numerically computing many first and second derivatives with respect to the age and firm-specific wage process parameters. In order to minimize the distortion in my reported second-stage standard errors, I choose functional forms to maintain a decent level of precision in the first-stage wage regressions.

<sup>33</sup>The outer optimization routine is NPSOL. I first estimate versions of the model with fewer parameters, and then use these preliminary estimates to choose starting values for the full model. I also use other choices for starting values for the full model. There are multiple local maxima to the full model's likelihood; the estimates reported in the paper correspond to the highest found valid local maximum. Because I solve the finite-horizon dynamic programming problem using backwards recursion, there is no value function iteration and hence no numerical error from terminating value function iteration too early.

## 4 Data

In order to estimate my employer choice model, I need data that cover most employers that hire engineers in a national labor market, that track workers as they switch employers, and that have multiple workers per employer. In this section I detail data with these properties.

### 4.1 Overview

The data come from the Swedish Employers' Federation (SAF in Swedish), an organization that represents firms in negotiations with labor unions. The SAF is composed of industry-level employers' federations, and the data list the smaller federation an establishment belongs to. I use these sub-affiliations as my measure of an establishment's, and hence a worker's, industry. Most of the engineers in the sample work in either manufacturing or construction, although I include engineers in other industries as well.

The SAF data aim to cover all workers in member companies of the SAF from 1970–1990. Every year, each establishment owned by a member company is asked to report information on the worker ID code, age, monthly salary, education and sex of all white collar employees.<sup>34</sup> As a result, the data track workers as they move from employer to employer in the private sector, but cannot track workers if they leave the labor force or switch to an uncovered sector.<sup>35</sup> In practice, collection problems mean that not all establishments report in every year. The data contain information on roughly 60% of the Swedish private sector work force.<sup>36</sup> The coverage rate for engineers in the private sector should be higher than 60% because they are more likely to be employed by large manufacturing firms, which choose to join the SAF. Also, according to the current website of the labor union for engineers, 79% of them work in the private sector.

The most serious problem affecting data coverage is that not all establishments report the educational background of their workers. About 50% of workers in the data have reported educations. The establishments reporting on education tend to focus on the need to employ highly educated workers. Therefore, studying a group of highly educated workers such as engineers minimizes this coverage problem.<sup>37</sup>

I model workers as making employer decisions at each successive career stage. While I do not directly observe total labor market experience, age is a good proxy for experience when workers have roughly the same background and do not take time off from the labor market. Sweden had extremely low reported rates of unemployment during the period (although some workers were in government training programs), and the rate for the most highly educated workers is presumably lower. I consider only male engineers, who are less likely to take time off from work for unmodeled family reasons. A typical male engineer needs to complete his five years of university education and around a year of military service before entering the labor market. For this reason I only include engineers aged 26 and higher. Many Swedish workers choose early retirement plans, and I lose track of employees whose firms promote them to executive positions. For this reason, I do not model workers older than 61, even though the mandatory retirement age is 65. Finally, I only include engineers who work full-time, which is defined by the SAF as working more than 35 hours in a week.<sup>38</sup>

---

<sup>34</sup>The data do not contain information on the personal characteristics of workers, such as their family status and nonwage income, or on the financial performance of firms.

<sup>35</sup>There is no particular age pattern for workers who leave the data, except for the expected spike at retirement.

<sup>36</sup>I calculate this based on numbers in Calmfors and Forslund (1990). The main exceptions are the banking industry, which is represented by a different employers' federation, firms that are cooperatively owned by their workers, and firms that do not belong to any employers' federation. The data coverage is much better for workers in manufacturing.

<sup>37</sup>My attempts to accurately predict educational background from other covariates failed.

<sup>38</sup>Sweden was not a participant in World War II, so the war disrupted the careers of the relevant age cohorts less than the careers of workers in some other countries. Sweden had no baby boom, although over time the government established new technical universities, increasing the number of engineering graduates entering the labor market.

Swedish firms pay engineers a straight monthly salary, which is the wage measure I use in this paper.<sup>39</sup> While the data report contractual hours (usually 40) for white collar workers, contractual hours may be a poor proxy for actual hours of work in an office environment.<sup>40</sup> The wage that enters the current period utility function (1) is the log of the yearly salary.<sup>41</sup>

The SAF data do not follow workers before they enter the labor market, so this paper does not investigate the initial decision of where to work. In a dynamic setting, all workers must pay a switching cost for this initial employer choice. For brand new workers, the elasticity of labor supply with respect to the stochastic process for age–wage profiles at a particular firm may be especially high, and firms may strongly compete to attract new workers into potentially long careers inside a single firm.

## 4.2 Firm Estimation Sample

In my sample, 95% of the establishments employing engineers do not meet the criteria for estimating compensation schemes using only within-firm variation. The most common problem is that there are very few workers at an establishment and there is not someone who stayed at the establishment for the necessary length of time in a particular age category. For the majority of establishments that do not meet the criteria, I pool them into larger groups based upon region. Most of the establishments in the pooled groups are very small. These estimation groups will be used only to estimate age–wage profiles; in the choice model below workers will be able to choose from all firms, or at least a selection of real-life firms, as opposed to pooled groups of firms. Switching between firms in the pooled categories is important, therefore I use pooling only for the necessary steps of estimating the age–wage profiles.<sup>42</sup> I require that firms that are not pooled by region have at least three stayers (non-switchers) over the five-year cross-sectional time interval in each five-year worker-age interval. Pooled firms are grouped by Swedish county.

There are an average of more than 1000 establishments employing engineers per year. Handling this many choices in a nonlinear discrete choice model is computationally intensive. In order to increase the speed of the estimation procedure, I narrow the sample by eliminating firms with less than four engineers in either the start or end year of a two-year period. Table 1 lists sample statistics for firms employing engineers that are valid choices in the employer switching model. Establishments that shut down are excluded as a result. Table 1 labels this truncated sample the estimation sample. From now on, all results use the estimation sample. Table 1 shows that many of the firms are in Stockholm County, and presents the number of establishments in the two largest industries: “manufacturing” and “industry and chemicals.”

---

<sup>39</sup>As the dynamic programming problem uses a five-year interval, one approach might be to multiply the monthly salary by  $5 \cdot 12 = 60$  months. The constant “60” will difference out when the utility of two firms are compared in the discrete choice problem, as  $\log(60 \cdot w) = \log 60 + \log w$ .

<sup>40</sup>I equate the means and standard deviations of the engineering wages to their 1988 values. I do this instead of deflating by a price index because the deflation makes the mean of the real wage distribution change over time, because the high inflation during the sample period changed the real purchasing power of Swedish wages. The full-time, male engineers are highly committed to the labor market; I do not model any labor-leisure substitution that a decrease in the real purchasing power of wages might induce.

<sup>41</sup>I have estimated specifications where the unlogged level of wages enters payoffs. This approach subsumes multiplicative constants into  $\beta_w$ . Multiplicative constants from period to period include the the marginal tax rate paid on the incremental wage increase from an employer switch. By using the level of wages, I do not need to model saving and borrowing, as workers care only about the present discounted value of consumption. Changing the curvature of consumption in utility does not dramatically alter the final elasticities of employer switching with respect to outside wage offers.

<sup>42</sup>When I pool for the purposes of estimating age–wage profiles, I still define movers and stayers using real-life establishments, and not the pool definitions. One concern with pooling is that firms may find it optimal to play asymmetric strategies in order to create artificial product differentiation in their age–wage profiles, even if the underlying fundamentals facing each firm are the same. Pooling eliminates any ability to detect these asymmetries.

Table 1: Facts About Firms: Mean Over Five-Year Time Intervals

Variable	Estimation Sample
# of Establishments	337
Minimum Size for Inclusion	4
# of Pooled Groups	20.7
# of non-Pooled Firms	26.0
Total Estimation Groups	46.7
Mean # Engineers per Estab.	33.9
% of Estabs. in Stockholm County	32.4%
# of Estabs. in Small Counties	142
% of Estabs. in Manufacturing	36.6%
% of Estabs. in Industry and Chemicals	31.1%

For the age–wage profile estimation, I pool firms lacking the right number and types of observations. An observation in this table (for the statistics not involving pooled firms) is a non-pooled physical establishment over a five year period that appears in the estimation sample. Remember that a dynamic programming model is separately solved for each five-year cross section.

Large counties are the 1990 administrative boundaries that include the cities of Goteburg, Stockholm and Malmo. Small counties excluded those three cities.

### 4.3 Worker Estimation Sample

The estimation sample covers 16,402 unique Swedish engineers. Each engineer is observed for an average of 5.3 (overlapping) five-year intervals.

Table 2 lists descriptive statistics for the engineers in the estimation sample. Monthly salaries are in Swedish crowns normalized to their value in 1988. The percentage growth in salaries reflects the average growth in the normalized wage levels over five years; in this case it is 14%. The switching rate is the percentage of workers who switch to another firm, over a five year period with both the new and old firms remaining in business over the period. In the estimation sample, engineers have a 14% probability of switching employers.<sup>43</sup> The variable large county is an indicator variable for working in one of the three largest metropolitan areas in Sweden. A majority of Swedish engineers work in large counties. Establishment size is the number of other white collar workers at a worker’s place of employment, while the number of engineers is the number of employees with a five-year engineering degree at that establishment.<sup>44</sup> Many of the engineers are employed at quite large establishments: the mean engineer has 1690 fellow white-collar workers in his plant. Elite engineers in Sweden tend to be employed by large, multinational manufacturing firms.

## 5 Stochastic Processes for Age-Wage Profiles

Section 3.3 describes how I use first stage wage regressions to estimate the parameters  $\theta_w$  that enter the normal, one-period-ahead densities of wages,  $f(w_{i,j,t} | w_{i,j,t-1}, j_{i,t}, j_{i,t-1}, t_i; \theta_w)$ . The age-specific densities create firm-specific stochastic processes for the wages of stayers and switchers. These stochastic processes are inputs into the later employer choice model estimation. Estimation of the transition density of states in a first stage is a standard part of the approach taken in many single-agent, discrete choice, dynamic programming models, following Rust (1987).

<sup>43</sup>The number is higher when workers at firms that close down are counted as switchers. In this paper, I focus on voluntary switching.

<sup>44</sup>The estimation sample does not include all engineers as listed in Table 2, as I delete workers if they work part time, are female, or if they were not in the data in the beginning of the five-year period.

Table 2: Descriptive Statistics about Worker Sample

Variable	Mean	Std. Dev.
Monthly Salary in 1988 Crowns	21,100	4970
Normalized Salary Percentage Increase (%)	13.3	14.0
Switching Rate (%)	14.1	34.8
Age	41.8	8.4
Large County (%)	58.5	49.2
Establishment Size (white-collar only)	1690	1980
Fellow Engineers at the Establishment	251	317
Total Worker Time Intervals	86,371	
Total Unique Workers	16,402	
Mean Time Intervals per Worker	5.27	

This table uses the estimation sample. Large counties are the 1990 administrative boundaries that include the cities of Goteburg, Stockholm and Malmo. The time interval is a five years. Time intervals overlap.

Table 3: Firm-Level Percentage Raises for Firm Stayers, by Worker Age

Measure	Mean Estimate	Std. Dev. Estimates
5 Year Percentage Increase Age 25-29	0.247	0.0650
5 Year Percentage Increase Age 30-39	0.151	0.0460
5 Year Percentage Increase Age 40-49	0.043	0.0490
5 Year Percentage Increase Age 50-59	-0.014	0.0583
Number of Estimation Groups - Cross-Sections	701	

This table reports the average means and average standard deviations across estimation groups and cross sections. The standard deviation is the dispersion of the estimates across estimation groups, not a measure of the statistical precision of individual coefficients. Generally, the coefficients are precisely estimated. I create the pooled estimation groups to eliminate imprecise results, at the risk of not accurately capturing across-firm heterogeneity.

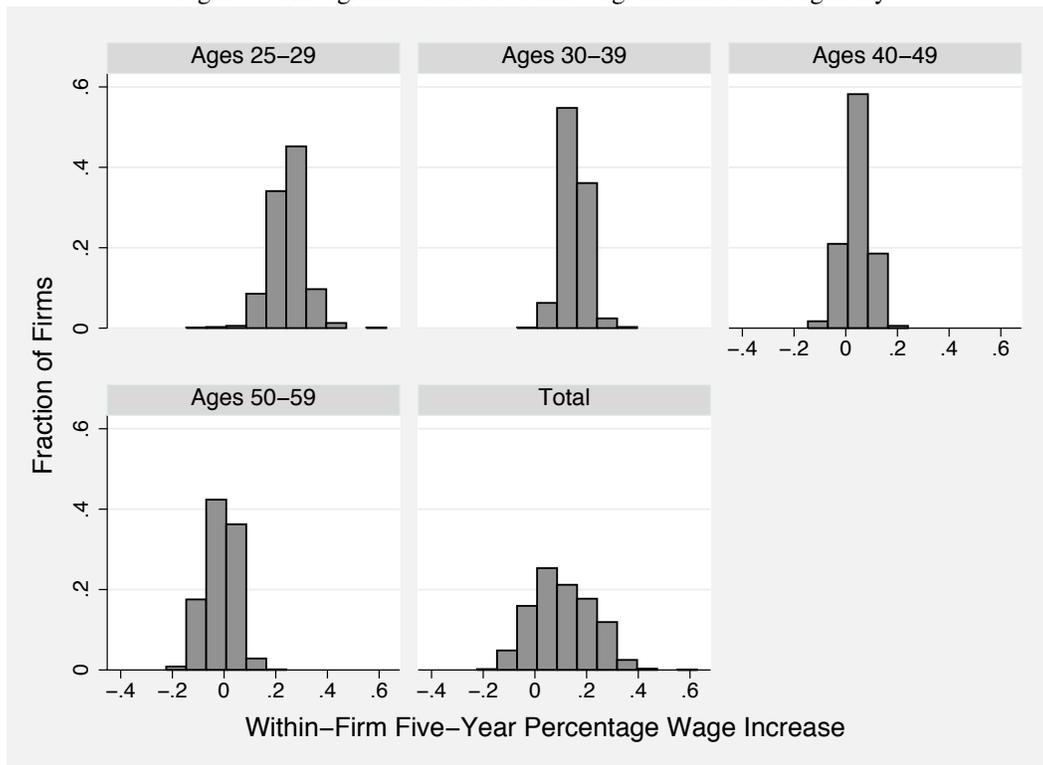
## 5.1 Estimates for Stayers

As described in Section 3.3.1, I first estimate (3) for workers who stay at the same firm. Table 3 lists estimates of average five-year raises for engineers. I first take within-firm averages, and then average those across firms to calculate the numbers in Table 3. The main result in this table is that wage increases are much higher for younger workers. Engineers who at the start of a five year cross section are between 25 and 29 receive a 25% raise over the next five years. At the other extreme, engineers between the ages of 50 and 59 receive a normalized wage reduction of 1%. These numbers, if graphed, would replicate the concave age–wage profiles familiar to labor economists.

Figure 3 shows the histograms of the combined age and firm fixed effects for an older specification. I create histograms where the vertical axis is the fraction of pooled estimation groups that are in each bin. The horizontal axis is the mean level of five-year wage increases for workers at a particular firm. I present a separate chart for each five-year age interval.<sup>45</sup> In general there appears to be important variation in the wage policies of Swedish firms, although I suspect both the within-firm and between-firm variation in wages is less than in a more diverse labor market, such as that found in the United States.

<sup>45</sup>Some of the heterogeneity reflects sampling error. As the sample size for each firm becomes large, this sampling error will be eliminated. The entire estimator is consistent, but this sampling error remains in finite samples.

Figure 3: Histograms of Across-Firm Wage Increase Heterogeneity



An observation is a five-year percentage raise at a pooled estimation group.

Table 4: Contribution of Old Firms' and New Firms' Wages for Firm Switchers, by Worker Age

Age Group in Start Year	% Hypothetical Wage in the Old Firm		% Hypothetical Wage in the New Firm	
	Mean Est.	Std. Dev. Est.	Mean Est.	Std. Dev. Est.
Age 25-29	0.640	0.0567	0.361	0.0568
Age 30-39	0.766	0.0305	0.237	0.0302
Age 40-49	0.819	0.0951	0.185	0.0922
Age 50-59	0.904	0.160	0.094	0.161
# of Switchers per Cross Section Five-Year Intervals	818 (std. dev. = 185)		15	

This table reports the average results of one regression for each five-year period. The sample is all employer switchers. The standard deviations are taken across five-year periods. In general, regression coefficients that look different from zero are estimated to be statistically different from zero. The dependent variable is the wage of a switcher. The definition of the measures of the hypothetical wages are in Section 3.3.2.

## 5.2 Estimates for Switchers

In order to fully specify the age–wage profiles of firms, I also need to estimate the policies firms use to assign wages to workers who switch firms. I discuss the functional form choices that I make in Section 3.3.2.

Table 4 lists the average slope coefficients from the regression of the wages of switchers on measures of hypothetical wages using the old and new firms' age–wage profiles for stayers. The average is taken across five-year periods. I let the coefficients vary by worker age. The main result is that the wage of an older switcher seems to be based much more on the wage policies of his old employer. For example, 64% of the wage of a younger engineer who switches is based upon his predicted wage at his old firm, while 90% of the wage of an older engineer who switches is based upon his predicted wage at his old employer. While some part of the wages of a firm switcher is due to the unique age–wage profile of his new firm, the majority of the wage is based upon the worker's past history in the labor market. A theoretical explanation is that information is revealed about a worker's ability that cannot be erased by switching firms. This information accumulates as a worker ages; by the end of a worker's life his wage is very specific to him and not subject to much variation on the market. This means that the window of higher wage opportunities from switching employers seems to be open only for younger workers. This finding is consistent with a story where the few older workers who switch move to similar positions at other firms. This is consistent with unreported empirical evidence on the job assignments of older workers before and after employer switches.

As far as I am aware, Table 4 shows the first results in the literature to analyze the wage changes of job-to-job switchers in the context of the specific wage policies of the old and new firms. Previous papers have not been able to distinguish between the part of wage change upon an employer switch specifically due to a recruitment bonus or penalty paid to switchers, and the portion due to the unique age–wage profiles of the old and new firms. Table 4 shows that for older workers, there is little or no return to switching; older switchers receive the same raises that they would have earned at their previous employers. A younger worker who switches to a high paying new firm can see a substantial wage increase, but only because the new firm pays all young workers a higher wage. Because the coefficients on the hypothetical old and new wages tend to sum to close to 1, there is no large wage gain from switching in and of itself. Young workers who switch to a new firm will be placed at about the same wage level as existing workers, subject to some adjustments based upon the wage history of

the young worker at his old firm. In other words, the return to seniority (or tenure) highlighted in the literature would appear to be about zero. The returns to seniority for older workers are also zero, as the wages of those workers change little upon employer switching. In this model, the substantial wage differences seen in older workers are due almost entirely to past labor market histories and are not accentuated by late-career employer switches.<sup>46</sup>

## 6 Employer Choice Model Estimates

### 6.1 Overview

The purpose of the dynamic model of employer choice is to estimate the utility function parameters for each worker type,  $\beta_k$ , and the population frequency of each type,  $\pi_k$ . With these parameters, I calculate the fraction of Swedish engineers who respond to unilateral outside wage offers. As the introduction discusses, having a large group of marginal workers who are sensitive to wage offers is needed for a spot market for experienced labor that responds to supply and demand shocks. Otherwise, all workers are locked-in to their current employers, and wages and employment decisions will not necessarily adjust to shocks.

Table 5 describes the parametrization of the current period utility function and the type frequencies. Because of the curse of dimensionality in nonlinear optimization, I cannot be perfectly flexible in including all possible parameters, but I try to span the parameter space by a judicious choice of parameters that vary by type and age. A type  $k$  is distinguished by separate wage parameter (the most important parameter in the model) and a base switching cost. Other switching costs vary by age in order to fit the declining pattern of turnover, but not by type in order to speed estimation. I add flexibility in age to fit the well-known pattern that employer switching decreases with worker age. The discount factor is 0.96, or a 4% real annual interest rate.<sup>47</sup>

Note that a high level of employer switching does not necessarily imply that the sensitivity of switching to outside wage offers is large. In the model, employer switching happens for nonwage reasons such as the taste shocks in the current period utility function. A labor market with lots of switching, but little sensitivity to wages, implies that firms cannot recruit experienced workers through wage increases. Also note that the ability of across-firm wage heterogeneity in Sweden to identify the effect of wages on employer choice will be seen in the standard errors of the reported statistics. The utility function parameters and type probabilities turn out to be precisely estimated.

### 6.2 Model Fit

I choose the specific parametrization of the model to match certain patterns in the data. For example, the switching costs vary by industry because switches between industries are empirically less likely than switches within industries. This section analyzes to what extent the model fits the data. Note that it is simple to fit the

<sup>46</sup>Table 4 shows that the gains from switching firms decreases with age. If this estimate is incorrect, it might seem that this wage process will lead to upward biased estimates of the sensitivity of switching to wages, as the real wage gains available to switchers are larger, so workers will appear to be over-sensitive to the smaller, predicted wage differences. On the other hand and as stated previously, these are estimated off of the sample of switchers and are not selection-corrected, because of the difficulty in jointly estimating wages and employer choice and because of the lack of natural variables included in employer choice but excluded from the wage process. This lack of selection correction might overstate wage gains to switchers, leading to an underestimate of the sensitivity of switching to wages. So, various biases can lead to under or overestimates of the benefits of switching.

<sup>47</sup>Because age–wage profiles are nonstationary, the choice of the discount factor affects the estimates of the parameters on wages,  $\{\beta_{w,k}\}_{k=1}^K$ . There is no way around this. Magnac and Thesmar (2002) discuss why identification of the current period utilities requires assuming a value for the discount factor. An econometrician can attempt to estimate the discount factor if state variables that are not payoff-relevant in the current period are observed. There are no such state variables in the employer switching application.

Table 5: Parameters of the Employer Choice Model

Variable Category	Definition	How Params. Vary	Number Included
Discount Factor	0.85		1
Wage Coefficients	$E[w_{i,j,t}   s_{i,t}, j_{i,t}]$ , wage logged	By type	3
Switching Costs:			
Geographic Distance	log of distance between regions	By age	7
Base Switching Cost	1 if switch firms	By type + By age	3 + 6
Legal Corporation Change	1 if new firm not same corp.	By age	7
Industry Change	1 if new industry	By age	7
Firm Size	# of total white collar workers	By age	7
Type Probabilities	prob. of being specific type	By type	2
Total Estimated Parameters			42

The discount factor is not included in the total parameter count.

The total base switching cost is equal to the cost for the specific type plus an extra age-specific amount common to all types.

The type probabilities must sum to 1, so one type is excluded.

data with a flexible model, but the dynamic programming model in this paper puts a premium on economic complexity to make better predictions about worker behavior in out-of-sample situations.

Perhaps the most widely known fact about employer switching is that it declines with worker age (Farber, 1999b). The model includes several features to allow for decreasing switching with age. The stochastic processes for age–wage profiles offered by firms and the switching costs of workers vary by worker age. In addition, the dynamic programming model with a mandatory retirement date assumes that younger workers have longer remaining careers to enjoy wage gains.

Table 6 presents the empirical and predicted probabilities of a worker of a given age staying at his firm for a five-year period.<sup>48</sup> The model fits the broad patterns of the switching probability decline with age quite well. However, in percentage terms, the fit of each cell is off. For example, the switching rate of 0.091 for age 40–44 workers is  $1 - 0.091/0.122 = 25\%$  off of the value. These fit errors come from the tightly parameterized structural model, which focuses on modeling a rich economic choice process over including lots of parameters in the utility function. Including more age-specific or type-specific parameters could improve the fit by age, although I do not pursue this as previous increases in the number of model parameters have not qualitatively altered the conclusions about switching costs and the sensitivity of employer switching to compensation policies.<sup>49</sup>

### 6.3 Switching Costs Relative to Taste Shocks

Rather than simply itemizing the estimated parameters, I report their implications in terms of the dependent variable, employer choice. Comparing estimates of parameters in the linear index of the current period utility function can be misleading when the linear index is only one part of a highly nonlinear, dynamic programming, discrete choice model. Because a first order cause of wage setting in Sweden is bargaining with labor unions, I do not use my parameter estimates to explain wage setting by firms or region and industry-specific turnover patterns for engineers in Sweden. Rather, I use the estimated parameters to calculate employer switching patterns under four alternative scenarios. The scenarios are intentionally stylized in order to focus on the main

<sup>48</sup>The model predictions average over the actual state variables,  $s_{i,t}$ , in the data. To compute the fit, the dynamic programming problem must be solved.

<sup>49</sup>Further, the likelihood criterion promotes matching the destination firm chosen by the worker, not just whether the worker switches at all. I discuss a specification using only the separation decision below.

Table 6: Empirical and Predicted Probabilities of Not Switching, by Worker Age Category

Ending Age	Empirical	Predicted
25-29	0.234	0.225
30-34	0.167	0.180
35-39	0.127	0.153
40-44	0.091	0.122
45-49	0.061	0.080
50-54	0.046	0.062
55-59	0.038	0.040

The empirical turnover probabilities use only the sample used in the estimation. The predicted turnover probabilities are the model probabilities summed over the data points.

implications of the parameter estimates.

Each scenario involves solving the dynamic programming problem for all three worker types and a given placement of firms in characteristic space.<sup>50</sup> I assume that there are 400 total firms.<sup>51</sup> Each firm offers the same stochastic process for age–wage profiles, which I take from the estimates from Section 5. All workers enter the labor market at age 26 with a starting wage of 16,800 crowns. When I break out the results by worker type for stayers, I assume all engineers are of the given type.

Each of the five scenarios is defined by the amount of (symmetric) differentiation in nonwage employer characteristics. In the first scenario, all firms are 1 km apart, in the same industry and owned by the same parent corporation. The next three scenarios each modify one of these choices at a time. The second scenario is the same as the first except the firms are independently owned. The third scenario has all the firms be in different industries. In the fourth scenario, all firms are 100 km apart. Finally, the fifth scenario modifies all three elements at once: the firms are in different industries, the firms are owned by different parent corporations, and all firms are 100 km apart.

The purpose of the scenarios is to relate switching costs to turnover patterns. I include switching costs in the current period utility function (1) to stand in for many explanations for why most workers do not switch employers in a given period. These other explanations include matching and search costs. Without switching costs, workers will instantaneously switch firms in order to arbitrage away wage differences.

I report statistics from the point of view of a representative firm. Table 7 reports estimates of the probability of a worker at of a given age leaving the representative firm, under each scenario. The probabilities are evaluated at the average wage of a career stayer of the particular age.<sup>52</sup> Note that this probability compares the switching costs to the assumed i.i.d. logit distribution for the taste shocks. If switching costs are zero, the probability of staying at the representative firm is  $1/400 = 0.0025$  each period. For comparison purposes, the empirical probability of Swedish engineers staying over five years is 0.86.

Type I workers are 59% of Swedish engineers. The first scenario in Table 7 shows that if 400 alternative employers in the same industry and owned by the same parent corporation are located across the street from the representative firm, 26% of the firm’s youngest Type I engineers will leave. If the other employers are all owned by different parent corporations, only 3% will switch. If the alternative employers are all in unique industries, the youngest Type I workers will leave 6% of the time. If the 400 alternative employers are 100 km away from each other (the fourth scenario), 6% of each firm’s Type I engineers will still switch. The final

<sup>50</sup>Each scenario is a counterfactual that requires solving the dynamic programming problem for each of three worker types.

<sup>51</sup>Ackerberg and Rysman (2005) point out the importance of keeping the number of consumer choices relatively fixed in the static logit model. There are a mean 348 firms per time interval in the estimation sample.

<sup>52</sup>See Figure 1 for a picture of this notion of an average profile.

Table 7: Probabilities of Switching from a Representative Firm, Under Five Scenarios

Hypothetical Switches			Type I (prob 0.59)			Type II (prob 0.12)			Type III (prob 0.29)		
Change Company	Change Industry	Move Distance	Age 25-29	Age 35-39	Age 50-55	Age 25-29	Age 35-39	Age 50-55	Age 25-29	Age 35-39	Age 50-55
N	N	1 km	0.26	0.10	0.03	1.00	1.00	0.99	0.99	0.98	0.92
Y	N	1 km	0.03	0.01	~0	0.99	0.93	0.44	0.91	0.65	0.10
N	Y	1 km	0.06	0.05	0.03	0.99	0.99	0.99	0.96	0.95	0.93
N	N	100 km	0.07	0.03	0.02	0.99	0.99	0.98	0.96	0.93	0.87
Y	Y	100 km	~0	~0	~0	0.72	0.69	0.36	0.26	0.23	0.07

These results are derived from the solving the dynamic programming problem for each scenario and worker type. The numbers reported are the probabilities of choosing a firm other than the representative firm conditional on being previously employed at the representative firm with a wage on the average profile of career stayers, as shown in Figure 1. In general, the switching costs, and hence these reported statistics, are precisely estimated.

scenario says that no workers switch when the 400 alternative employers are 100 km away, in unique industries and independently owned. Type I workers have moderately high switching costs.

The 12% of engineers who are Type II workers and the 29% that are Type III workers have lower switching costs than Type I workers. Almost 100% of young Type II and III workers will switch firms under the first scenario. Switching rates decline with the scenarios, but not until the fifth and final scenario are the switching rates for the youngest engineers less than 90%, over five years. If all firms are 100 km away, in unique industries and independently owned, 72% of young Type I engineers and 26% of young Type II engineers will switch.

The overall level of switching under each scenario is calculated by multiplying the type-specific switching probabilities by the corresponding type frequencies. For example, the overall level of switching from the representative firm for age 35–39 workers under the fifth scenario is

$$0.59 \cdot 0.0 + 0.12 \cdot 0.69 + 0.29 \cdot 0.23 = 0.15.$$

This overall probability of staying may be more representative of worker behavior from a firm’s point of view, because the model assumes that the stochastic process of wages is not a function of worker types. Because only 15% of age 35–39 workers switch under this scenario, I estimate switching costs to be very important, compared to logit taste shocks. This paper is part of a growing literature that estimates high switching costs for various consumer choices, using individual-level panel data (Greenstein, 1993; Shum, 2004; Moshkin and Shachar, 2002; Goldfarb, 2006; Kennan and Walker, 2008).

Table 7 shows that older workers have, roughly, higher switching costs than younger workers. The estimates match the data, where older workers switch less than younger workers. The switching costs parameter estimates do not line up exactly with empirical turnover patterns, as in the model wage changes and lifetime career horizons to enjoy wage gains also cause younger workers to switch more often.

## 6.4 Unilateral Wage Changes

One goal of this paper is to see whether experienced workers respond to unilateral wage increases. This section answers this question by examining how switching changes when one particular firm unilaterally makes its stochastic process for age–wage profiles more lucrative.

While the actual final five-year raise depends upon the age of the worker, the additional amount of the raise in my experiment is the same percentage for all workers. Let the new raises be  $\gamma_{j,t} + \Delta$ , where  $\Delta$  does not

depend on age,  $t$ . Then the marginal retention effect for workers of age  $t$  and type  $k$  at firm  $j$  is

$$\frac{1}{I_{j,t}} \sum_{i=1}^{I_{j,t}} \frac{\partial (1 - P(j_{i,t} | j_{i,t-1}, t_i, w_{i,j,t-1}; \beta_k, \theta_w))}{\partial \Delta} \Big|_{\Delta=0}, \quad (6)$$

where  $1 - P(j_{i,t} | j_{i,t-1}, t_i, w_{i,j,t-1}; \beta_k, \theta_w)$  is the probability of leaving the current firm and  $I_{j,t}$  is the number of workers of age category  $t$  in firm  $j$ . I compute the marginal retention effect using a numerical derivative. To compute the marginal retention effect, I solve the dynamic programming problem for each worker type twice, once for  $\Delta = 0$  and once for a small  $\Delta > 0$ . I take the difference in  $1 - P(j_{i,t} | j_{i,t-1}, t_i, w_{i,j,t-1}; \beta_k, \theta_w)$  as the reduction in switching. As I compute the marginal retention effect for each firm separately, I must solve each type's dynamic programming problem for each of the firms in the data.<sup>53</sup>

After computing (6), I multiply it by  $\Delta = 0.01$  to judge the economic significance of the marginal retention effects. For reference purposes, it is important to understand the economic magnitudes of the units that I use to report the marginal retention effects. First it is necessary to know when a change in age–wage profiles of  $\Delta = 0.01$  represents a significant amount of money. I argue that increasing each of a firm's five-year wage increases by 0.01 is not inexpensive. Table 2 shows that the average average five year wage increase rate for engineers is 0.133. Increasing this amount by 0.01 amounts to about a  $0.01/0.133 \approx 8\%$  hike in the amount of the raise. Furthermore, young workers continue to receive larger raises over the remainder of their careers, so the compounding of this compensation scheme change means that a new 25 year old worker will receive a much higher base salary at the time he retires. For this expensive age–wage profile change to be worthwhile, it must affect employer switching significantly. Table 2 also shows that the average turnover rate for engineers is 0.141. In order for my hypothetical compensation scheme change to reduce the average turnover rate by a fourth (by 0.035), the marginal turnover effect must be -3.5. As a rough estimate, I say any marginal retention effect over -2.0 is cost-effective.<sup>54</sup>

Table 8 reports the marginal retention effects for engineers. The effects are listed by worker age. To be precise, I calculate (6) by numerical differentiation for each of the, on average, 348 firms per five-year cross section.<sup>55</sup> I then average within and then across the fifteen cross sections, and I apply the delta method to get the means and standard errors reported in the table. To recap Table 7's results on switching costs, Type I workers (59% of the population) have high switching costs, and younger Type II and III workers have low switching costs that will increase with age.

The first columns of Table 8 show that Type I workers have precisely estimated but very small marginal retention effects. Even the largest effect, 0.314 for the youngest Type I workers, is far below (in absolute magnitude) the -2.0 effect that I regard as economically meaningful. The effects are the opposite sign as the theory that workers will prefer to stay at a firm if they are offered higher wages predicts, but as the estimated effects are so small, I am not concerned. The bottom line is that Type I workers, who are the majority group (59% of all engineers) in the population, are not likely to voluntarily switch from the current employer under any situation.

<sup>53</sup>Using the delta method to compute standard errors for the marginal retention effects raises the number of dynamic programming problems by a multiple equal to the number of estimated parameters whose distribution is used to compute the standard errors.

<sup>54</sup>I can report elasticities instead of marginal effects. To compute an elasticity, I might take the weighted average of the marginal effects and multiply it by the ratio of the average log-wage increase to the turnover rate over the five-year cross-section. I find the interpretation of the elasticities confusing, because I have to consider percentages of percentages such as turnover rates and raises. For this reason, I report only the marginal effects, although readers should know that I can easily transform the results into elasticities that convey the same information.

<sup>55</sup>There are never workers of all age-groups at every single firm in a cross-section, so in practice I always take the within-cross-section average over less than the total number of firms.

Table 8: Marginal Retention Effects for Increasing the Rate of Wage Increases for all Age Groups

Starting Age	Type I (prob 0.59)		Type II (prob 0.12)		Type III (prob 0.29)	
	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
25-29	0.314	0.091	0.001	0.155	-1.42	0.486
30-34	0.080	0.017	0.002	0.268	-1.81	0.673
35-39	0.048	0.014	0.003	0.375	-1.79	0.618
40-44	0.013	0.003	0.004	0.506	-1.66	0.656
45-49	0.004	0.001	0.004	0.515	-1.02	0.419
50-54	0.003	0.001	0.003	0.406	-0.433	0.145
55-59	0.001	~0.0	0.002	0.234	-0.165	0.055

These effects are from (6). The standard errors of the effects are calculated using the delta method, where I account for estimation error in only the coefficient on wages in the utility function. This parameter has by far the largest impact on the marginal retention effects. Applying the delta method to account for variation in all of the parameter estimates is computationally prohibitive. The standard errors also do not account for the overlapping nature of the 15 cross sections and any first-stage estimation error in the age–wage profiles.

Table 8 shows that the marginal retention effects for Type II workers, 12% of the population, are estimated to be almost exactly zero with large standard errors. While from Table 7 we know that Type II workers have low switching costs, it appears that the available data on Swedish engineers are not suitable for determining whether Type II workers switch employers for wage gains. This lack of statistical precision may be related to the fact that wage dispersion across Swedish firms is presumably less than that in the United States, because of the system of national wage bargaining in place during my sample period. I can still use the estimated standard errors to construct 95% confidence intervals around the estimates. All the confidence intervals are almost within the interval -1.0 to 1.0. My benchmark for a economically significant marginal retention effect is -2.0, so I can rule out with confidence that Type II workers have large sensitivities to wages.

Finally, Table 8 shows that the 29% of all Swedish engineers who are Type III workers are the only ones who are especially sensitive to wages. Type III workers have marginal retention effects around -1.0 that are statistically different from zero at traditional levels. The largest (in absolute value) 95% confidence bound for the any of the age-groups is -3.1, so I cannot reject the hypothesis that Type III workers’ marginal retention effects are the benchmark of -2.0. I can, however, argue that the effects are less than my benchmark for a larger effect, which is -3.5. In summary, Type III workers are definitely willing to switch for wage gains, but their responses are not as large as would be predicted by traditional theories of infinitely elastic labor supply curves to individual firms.

Table 8 shows that, in general, Type III workers in the earlier part of their career are more sensitive to changes in future compensation schemes than are older workers. For the oldest engineers, the proposed increase in raises of 0.01 decreases their probability of switching by 0.0017, compared to 0.142 for age 25–29 workers. 0.0017 is a small decrease. If firms are trying to retain their workers with age–wage profiles, it appears that only workers in the earlier parts of their careers will respond. This may not be problematic from the firms’ perspectives, as young workers are more likely to switch, as Table 6 shows.

## 7 Using Data Only on Separation

A recent literature has estimated the elasticity of labor supply to individual firms by regressing the turnover rate at firm on its wage rate. The separation elasticity is an estimate of the elasticity of labor supply to individual firms under the assumptions of an equilibrium search model in (Manning, 2003). By not explicitly modeling

workers' outside options, separation regressions have much more limited data requirements than the data used in this paper. Separation regressions as implemented in the empirical literature do not include explicit detailed measures of the wage policies of alternative firms and some specifications do not even take a stand on the potential rival employers a worker could leave to join. Further, again as empirically implemented, separation regressions tend to focus on a cross-sectional notion of a firm's compensation policy: just the mean wage paid to workers at the firm. Further, there is not much of an attempt to distinguish how much of the wage reflects unobserved heterogeneity in worker ability instead of genuine differences across firms in wages for the same type of labor.

This section seeks to interact with the literature on separation elasticities in two ways. First, I estimate the paper's structural model using only the dependent variable of whether or worker switches at all, rather than also using the identity of the destination firm that a worker picks. Second, I estimate some separation regressions, where I regress a worker's decision to separate on various measures of the worker's wage.

## 7.1 Estimating the Model Using Only the Separation Decision

As argued previously, compared to other labor markets, the Swedish labor market for elite engineers corresponds more to a market where the wages of workers are below their productivities, so that the market could be better modeled as workers choosing firms given the firms' compensation policies. Still, a model of voluntary employment choice is indeed strong when the true model probably incorporates some degree of two-sided matching.

The intuition of the literature on separation regressions is that the relationship between turnover and wages tells us something about workers' responses to wages. In this subsection, I re-estimate the economic choice model in this paper, using only the dependent variable of whether or not to separate from a firm. The economic model is the same: the worker picks between all around 300 employers in the Swedish economy. The measured dependent variable is not the individual firm  $j_{i,t}$  that the worker moves to but the binary decision of whether to stay at the current firm,  $j_{i,t} = j_{i,t-1}$ , or to move to another firm,  $j_{i,t} \neq j_{i,t-1}$ .<sup>56</sup>

Solving the economic model is identical to before. Only the likelihood, previously (5), changes. Let  $d_{i,t,y} = 1$  if the worker leaves his previous employer during a five-year interval  $y$  and let  $d_{i,t,y} = 0$  if the worker stays at his job. Then the likelihood of the choice model becomes

$$\tilde{L} = \prod_{i=1}^I \sum_{k=1}^K \pi_k \prod_{y \in Y_i} P(d_{i,t,y} | s_{i,t,y}; \beta_k, \theta_{w,y}),$$

where  $P(d_{i,t,y} | s_{i,t,y}; \beta_k, \theta_{w,y}) = P(j_{i,t,y-1} | s_{i,t,y}; \beta_k, \theta_{w,y})$  if  $d_{i,t,y} = 1$  and  $P(d_{i,t,y} | s_{i,t,y}; \beta_k, \theta_{w,y}) = 1 - P(j_{i,t,y-1} | s_{i,t,y}; \beta_k, \theta_{w,y})$  if  $d_{i,t,y} = 0$ .

Table 9 lists the marginal retention effects for the specification where separation is the dependent variable.<sup>57</sup> The Type I workers now compromise 64% of the population and are even closer to 0 in their sensitivity to moving. On the other hand, the Type III workers compromise 25% of the population and now are more sensitive to outside wage offers. With the identity of the destination firm used as a dependent variable, Table 8 reported that an age 40–44 worker has a marginal retention effect of -1.66. In Table 9, the age 40–44 marginal retention effect increases to -2.52. The strangest occurrence is the 11% of workers who are Type II workers,

<sup>56</sup>The independently-developed model of Kennan and Walker (2008) first estimated a (in their case, geographic) switching model using only the separation decision.

<sup>57</sup>The parameter estimates in Table 9 come from one run of the optimization package, where the starting values for the solver come from the structural parameter values for the model with both separation and the identity of the destination firm used as dependent variables, from Section 6.

Table 9: Model with Separation as the Dependent Variable: Marginal Retention Effects for Increasing the Rate of Wage Increases for all Age Groups

Starting Age	Type I (prob 0.64)		Type II (prob 0.11)		Type III (prob 0.25)	
	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
25-29	0.109	0.009	0.512	0.021	-2.352	0.190
30-34	0.051	0.002	1.068	0.087	-3.021	0.418
35-39	0.038	0.001	1.446	0.153	-2.612	0.266
40-44	0.020	0.000	1.667	0.168	-2.521	0.358
45-49	0.013	0.000	2.037	0.228	-2.292	0.325
50-54	0.005	0.000	1.724	0.177	-0.778	0.023
55-59	0.002	0.000	0.981	0.067	-0.305	0.003

These effects are from (6). The standard errors of the effects are calculated using the delta method, where I account for estimation error in only the coefficient on wages in the utility function. This parameter has by far the largest impact on the marginal retention effects. Applying the delta method to account for variation in all of the parameter estimates is computationally prohibitive. The standard errors also do not account for the overlapping nature of the 15 cross sections and any first-stage estimation error in the age–wage profiles.

who now are seen to *dislike* wages. Increasing future wages for these workers will encourage them to leave the firm. This theoretically implausible result occurs when only the separation decision is used as the dependent variable.

Overall, the results for Type I and Type III workers paint a theoretically similar picture to the results with the destination firm in the dependent variable. Only the 11% of Type II workers, who in Table 8 had economically zero but statistically imprecise marginal retention effects, now behave in a substantially, and theoretically implausible, different way. Keep in mind that the two economic models are the same: the worker picks between all available firms given the how Section 5.2 predicts wages will change when a worker moves from one stochastic process for age–wage profiles to another firm’s stochastic process for age–wage profiles. This next subsection will turn to linear separation regressions, where again some unusual signs on the level of wages will appear. Together, these results show that using data on the destination firm and also using data on the slope of wages can improve on the estimates from specifications that use only separation decisions and the level of the wage.<sup>58</sup>

## 7.2 Separation Regressions for the Data on Engineers

A recent literature has estimated labor supply elasticities by combining regressions of separation (with worker-level data, the binary variable  $d_{i,t,y}$ ) on measures of a firm’s compensation policy (Hirsch, Schank and Schnabel, 2008; Manning, 2003; Naidu, 2008; Ransom and Oaxaca, 2008; Ransom and Sims, 2008). I implement separation regressions using the Swedish engineering data. To make the results comparable to the estimates of the employer choice model, I use the same estimation sample. I estimate the linear probability model

$$E [d_{i,t,y} | w_{i,t,y}, j_{i,t-1}, x_{i,t,y}] = \alpha_0 + \alpha_1 w_{i,j_{i,t-1},t,y} + x'_{i,t,y} \alpha_3, \quad (7)$$

where  $w_{i,j_{i,t-1},t,y}$  is some measure of the wage in the firm the worker is employed by at the beginning of the five-year interval  $y$ . I experiment with various wage measures. The vector  $x_{i,t,y}$  are other controls that may

<sup>58</sup>This is not to say that the separation decision is not an important aspect of the dependent variable; it is important. For example, using observations only on switchers leads to a selected sample and biased estimates. Furthermore, the base switching cost, a key part of the model, cannot be identified without data on workers who do not switch. Nor can we identify age-specific base switching costs, which are needed to fit the empirical evidence on switching rates that decline with worker age.

be correlated with wages and also separation. These always include fixed effects for worker age  $t$  and often include employer characteristics. Recall that the sample is limited to full-time, male engineers between ages 26–60.

Table 3 presents the results. The first column regresses whether a worker moves on his individual logged wage, controlling for worker-age fixed effects and year fixed effects. A worker with a 10% higher wage has a  $0.05 \cdot 0.10 = 0.5\%$  higher chance of switching firms over a five-year period. Recall from Table 2 says that the mean turnover rate is 14.1%. So 0.5% out of a base of 14.1% is about  $0.5/14.1 = 3\%$  of the base. The magnitude of the coefficient is economically small. The more surprising thing is that the coefficient on log wage has a positive sign: those with higher wages are likely to turn over more. The second column controls for firm characteristics to see if perhaps the positive sign reflects a correlation between wages and previously unmeasured firm characteristics: the coefficient drops from 0.05 to 0.024, but does not become negative. The second column says that among workers at plants with different wages that are otherwise observationally equivalent (same plant size, industry, location), the plant with higher wages will have more turnover. The third column controls for plant fixed effects: the coefficient remains about the same magnitude, at 0.03. The third column says between two workers of the same age at the same plant, the worker with the higher wage is likely to leave. The fourth column uses fixed effects for each combination of a plant and a five-year age category: the coefficient on wage remains about the same, as it does in the fifth column where the plant-and-age fixed effects are allowed to vary by the five-years intervals in the data, as well. The Swedish data contains a job assignment and a rank code (the rank applies only within job assignments) that is somewhat standardized across firms. See Fox (2008b) for background on this code. In the sixth column, I include controls for plant-worker age category-year interval-job assignment fixed effects. The coefficient remains small in magnitude, but becomes negative, suggesting that a higher paid worker of the same age and in the same job assignment will be likely to switch slightly less than a lower paid worker. Ranks apply within job assignments. The seventh column controls for plant-worker age category-year interval-job assignment-rank fixed effects. The coefficient becomes more negative, but remains small in magnitude. Not surprisingly, the standard error also increases. In all the wage regressions with individual wage data, the coefficient is small in terms of economic magnitude.

The eighth and ninth columns do away with individual-level wage variation and use as the wage measure the mean level of wages for workers in a particular age category,

$$\frac{1}{I_{j,t,y}} \sum_{i=1}^{I_{j,t,y}} w_{i,j,t,y},$$

where  $I_{j,t,y}$  is the number of workers in firm  $j$  of age category  $t$  at the beginning of the five-year time period  $y$ . The sixth column shows that plants with higher wages have more turnover, although the seventh column suggests that this effect goes to 0 in economic magnitude when firm characteristics (size, industry and county) are controlled for.

The robust finding across the first nine columns of Table 10 is that firms with higher wages have about the same levels of turnover as firms with lower wages. Also, as more controls are added for the employer and type of work being done, the coefficients on the individual wage become more negative (although less statistically precise). Perhaps the simplest explanation is that the first-order explanation for differences in wages for workers is unobserved heterogeneity in worker ability, or, almost equivalently, the higher-paying jobs that more competent workers qualify themselves for. Adding controls for job assignment and rank within a job assignment capture some but not all of this unobserved worker heterogeneity, meaning two workers of the same age, same employer, same job assignment and same rank may be more closely related in terms of ability

Table 10: Linear Probability Regressions of Separation on Various Wage Measures

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Wage Measure: (wage levels in logs)	Individual wage level	Age-specific firm mean wage level	Age-specific firm mean wage level	Career-long firm % raise	Career-long firm % raise						
Wage coefficient	0.0505 (0.0118)	0.0237 (0.0111)	0.0302 (0.0104)	0.0312 (0.0104)	0.0241 (0.0113)	-0.0143 (0.0184)	-0.0419 (0.0329)	0.300 (0.0898)	-0.00293 (0.0544)	-0.0168 (0.0162)	-0.0291 (0.0116)
Standard error	Y	Y	Y	Y	N	N	N	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	N	N	N	Y	Y	Y	Y
Worker age fixed effects	Y	Y	Y	Y	N	N	N	Y	Y	Y	Y
County fixed effects	N	Y	N	N	N	N	N	N	Y	N	Y
Industry fixed effects	N	Y	N	N	N	N	N	N	Y	N	Y
Plant size control	N	Y	N	N	N	N	N	N	Y	N	Y
Plant fixed effects	N	N	Y	N	N	N	N	N	N	N	N
Plant-age category fixed effects	N	N	N	Y	N	N	N	N	N	N	N
Plant-age category-year fixed effects	N	N	N	N	Y	N	N	N	N	N	N
Plant-age category-year- job assignment fixed effects	N	N	N	N	N	Y	N	N	N	N	N
Plant-age category-year- job assignment-rank fixed effects	N	N	N	N	N	N	Y	N	N	N	N

Each regression uses the estimation sample of 86,371 observations, from the employer choice model. Standard errors account for clustering at the plant (across all years) level. A time period is five years, so the dependent variable is whether a worker moved over a five year period. Wages are for the firm where the worker is employed at the beginning of the five-year period.

than other workers, and the difference in their wages may thus reflect idiosyncratic payments that do not reflect skills that will be equally compensated by other firms in the labor market. Having wages reflect unobserved ability may lead to close-to-zero separation elasticity, as wage differences that reflect unobserved heterogeneity should be uncorrelated with separation.

A major concern with separation regressions is that most wage variation may reflect unobserved heterogeneity in worker ability rather than different wages for workers of the same ability. This concern influenced me to estimate the firm compensation policies in Sections 3.3 and 5 in first differences: a worker's lagged wage reflects to some degree past realizations about his idiosyncratic ability and firms differ mainly in the percentage raises paid to stayers. To an extent that is estimated, new employers compensate workers for their past revelations about their individual abilities. As I stated previously, this random walk with drift specification for percentage raises is consistent with a model where (symmetric) information is revealed about worker ability as a worker progresses in his career.

This paper allows firms' compensation schemes to vary in the level of percentage raises. Do firms with greater percentage raises have lower turnover? In Section 5, I report estimates of  $\gamma_{j,t,y}$ , the five-year percentage raise given to workers in a given age category. For each firm (plant) or estimation group, these five-year raises are computed for each of four age categories. Table 4 summarizes these estimates. Here, I compute a notion of a firm  $j$ 's career-long percentage (really log) wage increase (the slope of wages with age or experience), as formed for a cross-section of workers in different ages in five-year period  $y$ , as

$$\text{wage\_increase}_{j,y} = \gamma_{j,2529,y} + 2 \cdot \gamma_{j,3039,y} + 2 \cdot \gamma_{j,4049,y} + 2 \cdot \gamma_{j,5059,y}.$$

I then use this measure of the firm's slope of wages as the wage measure in (7). The tenth column of Table 10 regresses turnover on career-long log wage increases and age and year indicators, while the eleventh column adds employer characteristics: county, industry and plant size. The coefficient on the wage in the eleventh column is -0.0291; its 95% confidence interval excludes 0. The mean of  $\text{wage\_increase}_{j,y}$  is 0.558 and its standard deviation is 0.352. So a one-standard deviation increase in an employer's wage slope is predicted to decrease worker turnover by  $-0.0291 \cdot 0.352 = 0.01$ , or 0.01 out of a mean turnover rate of 0.14, which is equivalent to a  $0.01/0.14 = 7\%$  decrease from the base turnover level. This 7% effect is, in absolute value, twice the 3% effect recorded from using the largest coefficient on the level of (not increase in) the wage, which is from the first column of Table 10, the specification without many controls. Also, the sign of the coefficient on  $\text{wage\_increase}_{j,y}$  is negative, so wage increases are predicted to decrease turnover, as theory predicts.

To conclude, in the structural model I use measures of wage increases to differentiate the compensation schemes of firms, because only measures of wage increases are negatively correlated with separation. The most likely reason is that variation in the level of wages across firms or across workers in the same firm likely reflects unmeasured worker ability, to a large extent.<sup>59</sup> Pieces of evidence for this view are the theoretically wrong signs in columns (1)–(5) of Table 10 and the fact that the coefficients on wage levels become negative (but still small in magnitude) as controls for job assignment and then also rank within a job assignment are included.

---

<sup>59</sup>This does not mean that wage increases are not correlated with worker ability as well. However, I lack plausibly valid instruments for the rate of increase in wages at a firm.

## 8 Discussion of the Results

### 8.1 Overview of the Results

The results of this paper are the estimates of the effects of switching costs and unilateral wage changes, Tables 7 and 8, respectively. Two of three estimated types (41% of workers) have relatively low switching costs, but only the Type III workers, 29% of the engineers, are responsive to wages. While my results do not confirm theories that assume experienced workers are perfectly mobile across firms, it does appear as if a minority of experienced engineers are willing to switch firms under the right circumstances, especially when young.

### 8.2 Monopsony Labor Market Equilibria

Standard models of monopsony<sup>60</sup> in the labor market (Boal and Ransom, 1997; Manning, 2003) suggest that a firm should choose a fixed wage level  $w$  so that the inverse of its elasticity of labor supply with respect to the wage equals a markdown term, or more formally

$$\frac{1}{\varepsilon_{L,w}} = \frac{\text{MPL} - \exp w}{\exp w}, \quad (8)$$

where the MPL is the marginal product of labor,  $\varepsilon_{L,w}$  is the elasticity of labor supply, and  $w$  is the logged wage to maintain notational consistency. The quantity on the left of the firm's first order condition is labeled the "exploitation index", because it describes how much firms "exploit" workers by paying less than the workers' marginal products of labor. This monopsony theory says that firms should only hire the workers with a great nonwage desire to work at the firm, and then pay these workers lower wages in order to earn higher profits. The key unknown parameter in this market equilibrium model is the elasticity of labor supply.

The static oligopsony model is probably a very poor approximation of wage setting in the unionized Swedish economy. However, the utility function parameters of workers should be relatively invariant to the particular national labor market being considered. My parameters indicate, if extrapolated to other settings, that worker labor supply is much less elastic than previously believed. Equation (8) then implies that employment and wages will both be below the levels predicted for a hypothetical perfectly competitive labor market. Alternatively, firms could be competing for the Type III, wage-sensitive workers.

I have not estimated parameters in the utility function governing the decision of where to work when a worker enters the labor market. As each brand new worker must move out of his parent's house, his college dormitory, or his military barracks anyway, it is likely that a new worker's labor supply elasticity is significantly higher than an older worker's, as the older worker has already developed comfortable workplace and living arrangements. Therefore, if workers are forward-looking dynamic agents, as in the model in Section 3, firms will compete strongly for brand new workers by offering fair terms at all age levels. It is less clear that spot markets for experienced labor will always perfectly adjust to unforeseen shocks, as workers are, at least to a small degree, locked-in to their current employers because of high switching costs.

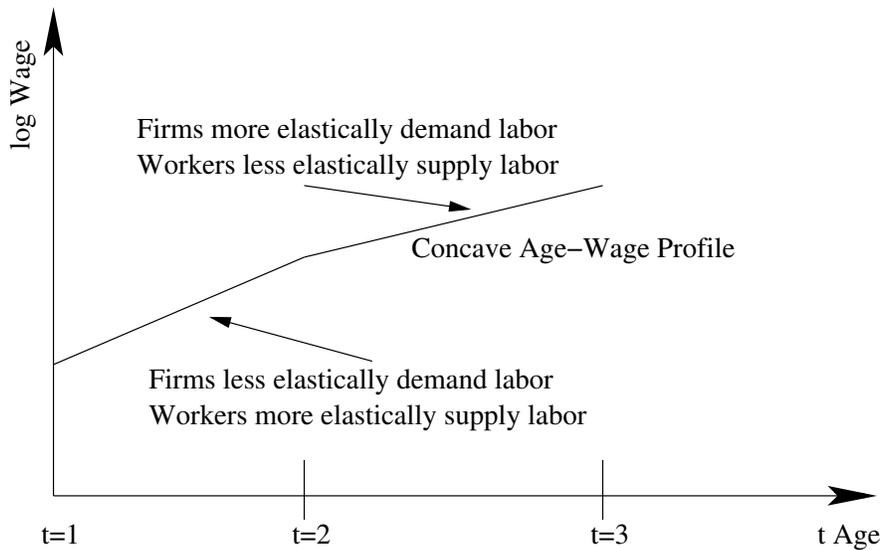
### 8.3 Wage Policies and Worker Turnover

An interesting question is whether firms have the unilateral ability to significantly reduce worker turnover by making age–wage profiles more lucrative (Parsons, 1972; Pencavel, 1972; Stiglitz, 1974; Salop and Salop, 1976). Setting aside the statistically imprecise results for Type II workers, it appears that only the 29% of

---

<sup>60</sup>Here, the term monopsony refers to imperfect competition, rather than only the special case of one employer.

Figure 4: Potential Explanation for Concave Age–Wage Profiles



Swedish engineers that are Type III workers are responsive to outside wage pressure. Unilateral increases in wages can thus reduce worker turnover. Looking at Table 7, Type III workers are the majority of workers who do leave, so in this case the group most likely to switch is also the group most sensitive to wages. I should emphasize that this is a result of my empirical investigation, not an implication of my theoretical assumptions.

This is not an equilibrium analysis. If all firms offer higher wages in order to reduce the turnover of Type III workers, then in equilibrium no firm will gain an advantage from doing so. I conjecture one equilibrium might involve firms with more to gain from reducing worker turnover offering higher wages or more lucrative age–wage profiles.

#### 8.4 Concave Age–Wage Profiles

Economists do not lack explanations for why wages increase with age and why age–wage profiles are concave. My empirical results in this paper suggest another explanation. Figure 4 depicts a stylized rendition of a concave age–wage profile. The empirical results in Table 8 show that workers supply their labor more elastically when they are young. Switching costs for older workers are large and their sensitivities of employer switching to unilateral wage changes decrease.

Specific human capital and delayed compensation models often predict that increasing the length of a relationship will make the relationship more productive (Becker, 1962; Becker and Stigler, 1974; Lazear, 1979). If longer relationships do lead to higher worker productivity, firms may put more emphasis on recruiting younger workers, meaning that firms will demand labor less elastically for younger workers (Hu, 2003).

The results on the labor supply of older and younger workers, and my conjecture on the behavior of firms, indicate that, relative to older workers, younger workers face firms with less market power in the labor market. As traditional monopsony theory suggests that more rents should accrue to a factor that is supplied more elastically, it is not surprising that wages increase faster for younger workers than for older workers.<sup>61</sup>

My explanation is compatible with other explanations for concave age–wage profiles, such as information

<sup>61</sup>The rapid increases of wages for younger workers can also be explained if some firms choose to offer steeper wage increases in order to reduce worker turnover for younger workers (Salop and Salop, 1976).

revelation, incentives, and human capital accumulation. The results in this paper indicate that relative market power in the labor market may be part of the explanation.

## 8.5 Limitations of My Analysis

Thus far, I have analyzed my results as if the assumptions of my structural model hold. Nevertheless, my model is only an approximation of reality, and my assumptions almost certainly do not hold. I do argue in Section 2 that the assumptions are closer to reality in the labor market for Swedish engineers than elsewhere.

My estimation procedure treats the observed employment allocation as data on the revealed preferences of workers, and therefore I make the assumptions that all employer switching is voluntary and that a worker can choose any employer. Violations of these assumptions may drive my results. It would be ideal to estimate an equilibrium, two-sided matching model (as I do in Fox (2008a), for example), which would allow a firm to screen workers and hire only the most suited for a finite set of open positions. Unfortunately, I know of no matching model that can easily replicate the rich patterns of firm-specific stochastic processes for age-wage profiles that we see in the data. Thus, any tractable matching model will have to abstract with some of the realities of how wages are set in the labor market for Swedish engineers. A related equilibrium sorting model would be a search model. If it is costly to contact employers and solicit job offers, workers may move to unattractive employers, even though they would prefer higher-paying firms if they knew such firms existed. If the search frictions model is the true one, I may be interpreting the failure to move to attractive employers as reflecting insensitivity to the wage level rather than search costs. I can only argue against the search model in terms of my knowledge of the Swedish labor market for engineers during the same period. Table 2 shows that the mean engineer works at a plant with 1690 other white-collar workers and 251 fellow elite engineers. These plants correspond to large Swedish multinationals that are fairly well-known in the international business community. It seems to me that an engineer would know about the major employers of engineers and be willing to contact each major firm when deciding to switch employers

Alternative interpretations for the empirical results are always possible. For example, I find that Type I workers have high switching costs and low sensitivities to wage offers. Perhaps instead Type I workers are bad employees who cannot get outside offers of employment, and so never switch. I also find that Type II workers often switch firms, but the data cannot statistically distinguish whether this switching is in response to outside wage offers. Perhaps, contrary to supposed Swedish practice, Type II workers are bad workers whom firms repeatedly fire. Conversely, the focus on voluntary behavior by workers could actually lead to higher estimates of labor supply elasticities, if firms in reality involuntarily transfer workers to expanding establishments that offer higher wages.

Similar arguments apply to the assumption that the wage stochastic process can be estimated from the wages actually received by switchers and stayers, without correction. My functional form choices are motivated by the notion of workers progressing in the stochastic age-wage process one firm and then considering whether to switch to another firm's wage policy. I assume that taste shocks, representing idiosyncratic life developments unrelated to wages, rather than unobserved wage offers explain different choices by observationally equivalent workers. If I assume an economic framework emphasizing unobserved wage offers, such as the Roy model (Heckman and Honore, 1990), I may come to different conclusions.

## 9 Conclusions

In this paper, I argue that the institutional framework of the Swedish labor market for engineers provides an excellent situation in which to study voluntary employer switching by experienced white collar workers. Swedish labor market institutions make firing and layoffs less likely, and allow me to more comfortably interpret the observed allocation of engineers as the revealed preferences of workers. I model Swedish employers as existing in characteristic space, where each employer is distinguished by its geographic location, parent company, industry, firm size and age-wage profile. I observe data on almost all large employers of engineers and hence observe the outside option that may tempt workers to switch. I use the data to estimate the age-wage profile of each firm separately.

Using a dynamic programming, discrete choice model of employer choice, I estimate the utility function parameters for three discrete types of workers. From Table 8, my estimates of the marginal retention effects show only economically meaningful magnitudes for an estimated 29% of Swedish workers. In general, labor markets with less employer differentiation have higher switching probabilities, and younger workers supply their labor more elastically to individual firms than older workers do. Younger workers have lower switching costs and longer remaining careers with which to enjoy the wage gains after a switch.

The majority of experienced Swedish engineers are estimated to have high switching costs and are estimated to not respond to outside wage offers. My estimates show that switching costs are a leading answer for why only 14% of Swedish engineers switch employers over a five-year period. Because of these rigidities, the spot market for experienced engineers probably does not always perfectly adjust to supply and demand shocks. Still, all workers at some point are new workers who must pick a new firm. The combination of a high elasticity of labor supply for brand new but forward-looking workers, and lower but nonzero labor supply elasticities for older workers, may prevent older workers from being exploited in the monopsonistic sense. The estimates do show that a majority of workers will not voluntarily move in response to outside wage offers. Therefore, the estimates suggest that layoffs may be needed to reallocate labor to more productive firms.

## References

- Abowd, John M., Francis Kramarz, and David N. Margolis**, “High Wage Workers and High Wage Firms,” *Econometrica*, March 1999, 67 (2), 251–333.
- Akerberg, Daniel A. and Marc Rysman**, “Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects,” *RAND Journal of Economics*, Winter 2005, 36 (4), 771–788.
- Altonji, Joseph G. and Nicholas Williams**, “The Effects of Labor Market Experience, Job Seniority, and Job Mobility on Wage Growth,” *Research in Labor Economics*, 1998, 17, 233–276.
- Baker, Michael**, “Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings,” *Journal of Labor Economics*, April 1997, 15 (2), 338–375.
- Becker, Gary S.**, “Investment in Human Capital,” *The Journal of Political Economy*, 1962, 70 (5), 9–49.
- **and George J. Stigler**, “Law Enforcement, Malfeasance, and Compensation of Enforcers,” *Journal of Legal Studies*, January 1974, 3, 1–18.

- Berry, Steven T.**, “Estimating Discrete-Choice Models of Product Differentiation,” *RAND Journal of Economics*, 1994, 25 (2), 242–262.
- Boal, William M. and Michael R. Ransom**, “Monopsony in the Labor Market,” *Journal of Economic Literature*, March 1997, XXXV, 86–112.
- Burdett, Kenneth and Dale T. Mortensen**, “Wage Differentials, Employer Size, and Unemployment,” *International Economic Review*, May 1998, 2 (39), 257–273.
- Calmfors, Lars and Anders Forslund**, “Wage Formation in Sweden,” in Lars Calmfors, ed., *Wage Formation and Macroeconomic Policy in the Nordic Countries*, SNS Forlag, 1990, chapter 2.
- Carneiro, Predro, Karsten T. Hansen, and James J. Heckman**, “Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice,” *International Economic Review*, May 2003, 44 (2).
- Dahl, Gordon B.**, “Mobility and the Return to Education: Testing a Roy Model with Multiple Markets,” *Econometrica*, November 2002, 70 (6), 2367–2420.
- Erdem, Tülin, Susumu Imai, and Michael P. Keane**, “Brand and Quantity Choice Dynamics Under Price Uncertainty,” *Quantitative Marketing and Economics*, 2003, 1 (1), 5–64.
- Farber, Henry S.**, “Mobility and Stability: The Dynamics of Job Change in Labor Markets,” in Orley C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3, Elsevier Science, 1999, chapter 37, pp. 2439–2483.
- , “Mobility and Stability: The Dynamics of Job Change in Labor Markets,” in Orley C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3B, North-Holland, 1999, chapter 37, pp. 2439–2483.
- Fox, Jeremy T.**, “Estimating Matching Games with Transfers,” 2008. working paper.
- , “Firm Size Wage Gaps, Job Responsibility and Hierarchical Matching,” 2008. working paper.
- Goldfarb, Avi**, “State Dependence at Internet Portals,” *Journal of Economics & Management Strategy*, 2006, 15 (2), 317–352.
- Greenstein, Shane M.**, “Did an Installed Base Give an Incumbent any (Measurable) Advantages in Federal Computer Procurement?,” *The RAND Journal of Economics*, Spring 1993, 24 (1), 19–39.
- Gronau, Reuben**, “Wage Comparisons—A Selectivity Bias,” *The Journal of Political Economy*, Nov.–Dec. 1974, 82 (6), 1119–1143.
- Hartmann, Wesley R.**, “Intertemporal effects of consumption and their implications for demand elasticity estimates,” *Quantitative Marketing and Economics*, December 2006, 4 (4), 325–349.
- Heckman, James J.**, “Shadow Prices, Market Wages, and Labor Supply,” *Econometrica*, July 1974, 42 (4), 679–694.
- , “The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process,” in Charles F. Manski and Daniel L. McFadden, eds., *Structural Analysis of Discrete Data and Econometric Applications*, The MIT Press, 1981, chapter 4, pp. 179–195.

- , “Varieties of Selection Bias,” *The American Economic Review*, May 1990, 80 (2), 313–318.
- **and Bo E. Honore**, “The Empirical Content of the Roy Model,” *Econometrica*, September 1990, 58 (5), 1121–1149.
- Hendel, Igal and Aviv Nevo**, “Measuring the Implications of Sales and Consumer Stockpiling Behavior,” *Econometrica*, 2006, 74 (6), 1637–1673.
- Hirsch, Boris, Thorsten Schank, and Claus Schnabel**, “Differences in Labor Supply to Monopsonistic Firms and the Gender Pay Gap: An Empirical Analysis Using Linked Employer-Employee Data from Germany,” 2008. working paper.
- Hu, Luo**, “The Hiring Decisions and Compensation Structures of Large Firms,” *Industrial and Labor Relations Review*, July 2003, 56 (4), 663–681. working paper.
- Jovanovic, Boyan**, “Firm-specific Capital and Turnover,” *Journal of Political Economy*, December 1979, 87 (6), 1246–1260.
- Kennan, John and James R. Walker**, “The Effect of Expected Income on Individual Migration Decisions,” February 2008. working paper.
- Koopmans, Tjalling C. and Martin Beckmann**, “Assignment Problems and the Location of Economic Activities,” *Econometrica*, January 1957, 25 (1), 53–76.
- Lazear, Edward P.**, “Why is There Mandatory Retirement?,” *Journal of Political Economy*, 1979, 87 (61), 1261–1284.
- Lundborg, Per and Per Skedinger**, “Capital gains taxation and residential mobility in Sweden,” *Journal of Public Economics*, 1998, 67, 399–419.
- Magnac, Thierry and David Thesmar**, “Identifying Dynamic Discrete Decision Processes,” *Econometrica*, 2002, 70, 801–816.
- Manning, Alan**, *Monopsony in Motion: Imperfect Competition in Labor Markets*, Princeton, 2003.
- Manski, Charles F.**, “Maximum Score Estimation of the Stochastic Utility Model of Choice,” *Journal of Econometrics*, 1975, 3, 205–228.
- Meghir, Costas and Luigi Pistaferri**, “Income Variance Dynamics and Heterogeneity,” *Econometrica*, 2004, 72 (1), 1–32.
- Moshkin, Nickolay V. and Ron Shachar**, “The Asymmetric Information Model of State Dependence,” *Marketing Science*, Fall 2002, 21 (4), 435–454.
- Naidu, Suresh**, “Recruitment Restrictions and Labor Markets: Evidence from the Post-Bellum U.S. South,” November 2008. working paper.
- Navarro, Salvador**, “Using Observed Choices to Infer Agent’s Information: Reconsidering the Importance of Borrowing Constraints, Uncertainty and Preferences in College Attendance,” November 2007. working paper.

- Parsons, Donald O.**, “Specific Human Capital: An Application to Quit Rates and Layoff Rates,” *The Journal of Political Economy*, November 1972, 80 (6), 1120–1143.
- Pencavel, John H.**, “Wages, Specific Training, and Labor Turnover in U.S. Manufacturing Industries,” *International Economic Review*, February 1972, 13 (1), 53–64.
- Postel-Vinay, Fabien and Jean-Marc Robin**, “Equilibrium Wage Dispersion with Worker and Employer Heterogeneity,” *Econometrica*, 2002, 70 (6), 2295–2350.
- Ransom, Michael R. and David P. Sims**, “Estimating the Firm’s Labor Supply Curve in a “New Monopsony” Framework: School Teachers in Missouri,” June 2008. working paper.
- **and Ronald L. Oaxaca**, “New Market Power Models and Sex Differences in Pay,” September 2008. working paper.
- Reynolds, Lloyd G.**, *The Structure of Labor Markets*, Harper and Brothers, 1951.
- Rust, John**, “Optimal Replacement of the GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 1987.
- , “Structural Estimation of Markov Decision Processes,” in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics*, Vol. IV, Elsevier Science, 1994, chapter 51, pp. 3081–3143.
- Salop, Joanne and Steven Salop**, “Self-Selection and Turnover in the Labor Market,” *Quarterly Journal of Economics*, November 1976, 90 (4), 619–627.
- Shapley, Lloyd S. and Martin Shubik**, “The assignment game I: the core,” *International Journal of Game Theory*, 1972, 1, 111–130.
- Shum, Matthew**, “Does Advertising Overcome Brand Loyalty? Evidence from the Breakfast-Cereals Market,” *Journal of Economics & Management Strategy*, 2004, 13 (2), 241–272.
- Stiglitz, Joseph E.**, “Alternative Theories of Wage Determination and Unemployment in LDC’s: The Labor Turnover Model,” *Quarterly Journal of Economics*, May 1974, 88 (2), 194–227.
- Thompson, T. Scott**, “Identification of Semiparametric Discrete Choice Models,” 1989. University of Minnesota Discussion Paper 249.
- Topel, Robert H.**, “Specific Capital, Mobility and Wages: Wages Rise with Job Seniority,” *Journal of Political Economy*, 1991, 99 (1), 145–176.
- **and Michael P. Ward**, “Job Mobility and the Careers of Young Men,” *The Quarterly Journal of Economics*, May 1992, 107 (2), 439–479.
- van der Berg, Gerard J. and Geert Ridder**, “An Empirical Equilibrium Search Model of the Labor Market,” *Econometrica*, September 1998, 66 (5), 1183–1221.
- Woodcock, Simon D.**, “Heterogeneity and Learning in Labor Markets,” July 2003. working paper.