Career Progression and Formal versus On-the-Job Training[*]

J. Adda,[†] C. Dustmann,[†] C. Meghir,[†] J.-M. Robin[‡]

August 8, 2006

Abstract

We develop a dynamic discrete choice model of training choice, employment and wage growth, allowing for job mobility, in a world where wages depend on firm-worker matches, as well as experience and tenure and jobs take time to locate. We estimate this model on a large administrative panel data set which traces labour market transitions, mobility across firms and wages from the end of statutory schooling. We use the model to evaluate the life-cycle return to apprenticeship training and find that on average the costs outweigh the benefits; however for those who choose to train the returns are positive. We then use our model to consider the long-term lifecycle effects of two reforms: One is the introduction of an Earned Income Tax Credit in Germany, and the other is a reform to Unemployment Insurance. In both reforms we find very significant impacts of the policy on training choices and on the value of realised matches, demonstrating the importance of considering such longer term implications.

1 Introduction

A number of countries in Europe, with Germany being the prime example, have (or have had) apprenticeship systems which essentially are formal vocational training courses combined with on-the-job training and which lead to a certification. Such apprenticeship systems relate both to white collar and blue collar jobs. Moreover they are subsidized by the state, which offers the classroom component. In contrast, other countries, including the U.S., have no such organized formal system at least on such a massive scale. The key difference seems to be between specific and in depth training in a particular occupation, versus the possibility of more general acquisition of skills conferred directly by the labor market. The question is how the career and wages of a worker are affected by participation in apprenticeship. Potential differences relate to wage differentials, to labor market attachment and to job mobility. Understanding how people make the choice to obtain this type of vocational education is necessary to understand what the impact of policy on career paths and wage growth is likely to be, and ultimately the potential impact that this type of training institution can have on the ability of an economy to respond to reallocation shocks. To examine the impact of apprenticeship on careers and provide answers to these questions we use a very detailed German data set which includes careers starting from the moment that statutory schooling ends. We link education choices and labor market careers within a complete life cycle setting and we study the way that incentives at different parts of the life cycle affect education choices.

Careers following an apprenticeship may differ from informal acquisition of skills in a number of ways. First they may increase wages in the long run, because of educational investment, but they may also involve a substantial investment at the start of one's career. Secondly they may affect job opportunities through various channels. This includes layoff rates, job finding rates and the variability of potential matches. On the one hand those with an apprenticeship qualification may be considered more desirable because they are better trained in a particular area, which could affect both job retention and job finding. On the other hand the specificity of training could make them less flexible and thus harder to place, following job loss. In fact this lack of flexibility is a central question for understanding the pros and cons of the system in terms of allowing the economy to

absorb reallocation shocks.

To address these issues, this paper specifies and estimates a life cycle model of education choice and labor market careers for men who complete standard schooling at 16. Individuals face the choice of formal apprenticeship or the standard labor market. Once in the labor market they can search so as to improve the quality of their job match. While working they face wage growth by experience and job specific learning (tenure). Estimation of such a model requires data on complete work and earnings histories, including information on job mobility, which is available to us. We observe individuals from the moment they enter the labor market, whether as candidate apprentices or as workers. We also observe the exact date of the start of a job. Their complete history is thus available from the age of 16 onwards with all transitions and corresponding wages observed.

The model we estimate combines many features of education choice models,[1] and wage determination models.[2] The model allows for heterogeneous returns to education, experience and tenure and similarly to the Willis and Rosen (1979) model allows for comparative advantage in training choice. Thus, individuals make their apprenticeship choice following school at age 16. Whatever their choice individuals end up in work, either as apprentices or standard workers who we refer to as non-apprentices. Wages depend on experience, on firm tenure, on training and on an initial human capital endowment. They also depend on a match specific firm effect that is modelled as a random walk. Finally, utility depends linearly on income and on work status. In each period workers can change firm, subject to them obtaining offers, or they can move to unemployment either voluntarily or because of exogenous job destruction. The unemployed can choose to remain so or to move into work, subject to receiving an offer.

Our model relates to the seminal paper by Eckstein and Wolpin (1989) who model transitions between employment and unemployment jointly with wages. In some ways our specification is similar to that of Keane and Wolpin (1997) in that both models consider labor market transitions and wages jointly. However, Keane and Wolpin's model is an

---

[1] See Taber (2001), Card (2001), Cameron and Heckman (1998).

[2] See Heckman and Sedlacec (1985), Altonji and Shakotko (1987), Topel (1991), Topel and Ward (1992), Altonji and Williams (1998), Altonji and Williams (2005), Dustmann and Meghir (2005).

3

essentially static Walrasian-Roy model where the only element of non stationarity in individual trajectories is the occupation-specific experience component of productivity (numbers of years spent in blue collar, white collar or military occupations). In contrast, our model focuses on mobility between firms, rather than occupations, allows for labor market frictions and has a richer stochastic structure, which includes match specific effects, and permanent shocks.

This is a partial equilibrium model which aims at incorporating many of the distinctive features of the equilibrium models with on-the-job search, which were initiated by the seminal paper of Burdett and Mortensen (1998). They develop a wage-posting model where employee poaching forces employers to offer (ex ante) higher wages to resist competition and implies dispersed wages in equilibrium. Growth of wage with experience in their model reflects improved matches through search for better jobs. Random matching implies that job-to-job mobility should be intense in the early stage of one's career. Extensions of the Burdett and Mortensen model have been numerous: Stevens (2004) and Burdett and Coles (2003) show that tenure-contracts are another facet of the strategies that firms develop to counter the moral hazard effects of on-the-job search; and Postel-Vinay and Robin (2002) and Cahuc, Postel-Vinay, and Robin (2006) replace wage-posting by a bargaining/sequential auction mechanism and develop a more tractable theory of individual wage and employment dynamics allowing for two-sided worker/firm heterogeneity in match productivity.

Another distinctive feature of our approach is that we combine data from a large number of cohorts who enter the labor market at different points in the business cycle and in different local labor markets. This is an important advantage of our data over other sources such as the NLSY. Thus controlling for time trends and for permanent regional effects, we use the differential changes in the availability of apprenticeship positions as a source of identification within our structural model: Different regions include different concentrations of industry. As product prices fluctuate so does the local demand for labour and for apprenticeships, depending how the local industry is affected. While trade ensures local wages do not react to such shocks the number of apprenticeship positions will adjust. This argument provides us both with the required exogenous variation and

with exclusion restrictions required to identify the effect of an apprenticeship qualification on wages. Using a difference in differences approach, we demonstrate in the descriptive part of the paper that the variation we use is indeed informative as far as educational choices are concerned.

Estimation of the model provides us with a rich set of results on how career paths are determined, on the nature of wage growth and on the importance of apprenticeship training. We provide estimates of the returns to apprenticeship net of costs incurred by the individual and we are able to distinguish between opportunity costs and other costs faced by the individual. The structural model also provides some hints on differences in the determinants of careers between the U.S. and Germany. Indeed we show that there are large differences in job mobility, due to lower arrival rates of job offers in Germany. However other parameters characterizing the labor market, such as job destruction rates and match heterogeneity are similar in the two countries.

The ultimate use of structural models is for policy analysis. There has been a growing literature on programme evaluation which typically focuses on the policy impacts on targeted outcomes.[3] However interventions viewed as permanent will have longer term effects far beyond these outcomes, as individuals position themselves to best take advantage of the new environment.[4] There is little empirical work to demonstrate the importance of such considerations. A recent exception is Heckman, Lochner, and Cossa (2003) who consider the impact of Earned Income Tax Credit (EITC), a wage subsidy designed to boost employment at the lower end of the wage distribution, on human capital accumulation. The model we estimate is ideally suited for analyzing the longer term effects of interventions. We consider two reforms, designed to be revenue neutral. The first introduces a U.S. style EITC programme in Germany. The second considers changing the Unemployment Insurance (UI) system from being related to previous earnings to a flat rate equal to approximately half the minimum wage, as designed in the UK for example. We show that these interventions have substantial effects on education choices, job mobility and wages over and above the direct impacts they were designed to have.

---

[3]See Heckman, Ichimura, and Todd (1997), Heckman, LaLonde, and Smith (1999), Blundell, Costa-Dias, Meghir, and van Reenen (2004) for recent examples.

[4]See Heckman, Lochner, and Taber (1998) for an analysis of GE effects.

The remaining part of the paper is structured as follows. In Section 2 we describe the model. Then Section 3 presents the data set and descriptive statistics. In Section 4 we display the estimation results. Finally, section 5 evaluates the effect of in-work benefits and of reforming the UI system.

2 Model

2.1 An Overview of the model

The model we describe takes individuals from the first point at which they make a choice and follows them to mid career. We focus on the population that finishes secondary academic schooling at 16 years of age and at that point just has the choice of following an apprenticeship or entering the labor market as a non-apprentice. We allow for aggregate shocks to the relative wage of the two groups thus implicitly allowing for a production function where qualified apprentices and non-apprentices are not perfectly substitutable. Utility is linear in earnings making risk and the timing of consumption irrelevant for decision making. We also allow for a utility of leisure by allowing a fixed utility cost of working.

At the start individuals choose whether they will join an apprenticeship, which offers formal on the job and classroom training at a reduced wage, or no formal training. In taking this decision they trade-off current earnings of a non-apprentice with working at a lower wage at a known job and then obtaining an improved career path through the formal training. The information they possess at that point is the distribution of idiosyncratic match specific shocks as well as the distribution of aggregate shocks that affect the evolution of relative prices in the two skill categories. They also know their type/ability which affects the costs of education, the wage level as well as the returns to experience and tenure.

Once the education choice has been made the individual starts up on his career, whether as an apprentice followed by normal work once qualified or directly into a standard job without a formal training component. All individuals receive job offers at some rate, which may differ depending on whether the worker is employed or not. Associated with an offer is a draw of a match specific effect which defines the initial wage level given

6

the person's type and experience. This then evolves as a random walk while the worker remains on the job. In addition the offer carries with it "fringe benefits" for the job. During apprenticeship, individuals may move to a new employer but not to unemployment. When out of work the individual derives utility which is a function of the wage earned in the last job. Jobs can end either because of a quit or because of exogenous job destruction. Individual choices include moving between jobs when the opportunity arises and between work and unemployment as well as the initial education choice.

The model is set in discrete time. To be able to capture the richness of the data without making the model intractable we chose the time period to be a quarter. We restrict the arrival of the shocks to the match specific effects to occur only once a year on average.

The dynamics in the model are due to the effects of apprenticeship education on future outcomes, the effects of experience and tenure, the difference in arrival rates of job offers between the employed and the unemployed and the effects of earnings on future unemployment benefits. We now describe the model formally and then discuss estimation.

2.2   A formal presentation of the model

The aggregate economy.   We assume an economy which fluctuates in a stationary way around a deterministic trend. The model operates on a quarterly frequency. We characterize the macroeconomic fluctuations of the economy around the steady-state growth trend by detrended GDP. The macro shock is relevant because it potentially affects the relative price of the two skill groups as well as the relative attractiveness of being out of work.[5]

The macro state variable G_t is supposed to be governed by an AR(1) process:

$$G_{t+1} = \rho G_t + v_{t+1}, \tag{1}$$

where v_t is a Gaussian white noise with variance σ_v^2. In practice, we discretize this AR(1)

---

[5]An issue of concern here is the appropriate notion of a business cycle. Under full factor price equalization with the trading partners the European business cycle would perhaps be more relevant. Here we assume that the German business cycle is sufficiently correlated with the European one to capture the relevant aggregate shocks influencing relative human capital prices.

process into a Markov process of order one.

Wages and the utility of working. The central component of the model is the job contract. If a worker i and a firm match at time t, the output is split according to some unspecified rule that yields an annual wage w_{it} to the worker. In addition, a job provides a non monetary value μ_{it} (amenities) to the worker. Workers are assumed risk neutral, which also implies that liquidity constraints are not an issue of concern for this model. The instantaneous utility of (w_{it}, μ_{it}) to the worker is then defined as the sum of the wage w_{it} and the amenity μ_{it} (expressed in monetary terms):

$$R_{it}^W = w_{it} + \mu_{it}.$$

Wages are modelled as follows. Let $Ed_i \in \{A, NA\}$ denote the worker's education (A for apprentices and NA for non-apprentices). Let X_{it} be the number of years effectively spent in work since age 16.[6] Let T_{it} denote the number of years spent in the current job ($T_{it} = 0$ if the job starts in period t). Let also ε_i be a permanent individual characteristic that is unobserved by the econometrician but is known by the worker and observed by the employer. Quarterly earnings w_{it} are functions of the macroeconomic shock G_t, education (Ed_i), experience X_{it}, tenure T_{it}, the unobserved permanent heterogeneity variable ε_i, and a match-specific component κ_{it}:

$$\ln w_{it} \equiv \ln w(Ed_i, G_t, X_{it}, T_{it}, \kappa_{it}, \varepsilon_i) = \alpha_0(\varepsilon_i) + \alpha_{Ed}(\varepsilon_i)Ed_i$$
$$+\alpha_X(X_{it}, Ed_i, \varepsilon_i) + \alpha_T(T_{it}, Ed_i, \varepsilon_i) + \alpha_G(Ed_i)G_t + \kappa_{it} \tag{2}$$

where α_X and α_T are two functions of experience and tenure, which are education specific. We use a piecewise linear function, with nodes at 0, 2, 4, 6 and 30 years of experience and tenure. Unobserved heterogeneity enters these functions multiplicatively.

The match-specific components of wages, κ_{it}, and amenities, μ_{it}, evolve with tenure in the following way. When the worker and the firm first meet ($T_{it} = 0$) they draw a match specific effect (κ_{it}, μ_{it}) such that

$$\kappa_{it} \sim \mathcal{N}\left(0, \sigma_0^2\right), \quad \mu_{it} \sim \mathcal{N}\left(0, \sigma_\mu^2\right), \quad \kappa_{it} \perp\!\!\!\perp \mu_{it}.$$

---

[6] $X_{i,t+1} = X_{it} + 1$ if the worker is working in period t; otherwise, $X_{i,t+1} = X_{it}$. We do not allow for depreciation of skills while unemployed.

Then, whenever $T_{it} \geq 1$,

$$\kappa_{it} = \kappa_{i,t-1} + u_{it}, \quad \mu_{it} = \mu_{i,t-1},$$

where u_{it} is a Gaussian white noise with variance σ_u^2. This allows for the possibility that what started out as say a good job may change to a bad one, following unobserved changes.[7]

The utility of being out of work. While unemployed, the individual derives a utility from unemployment benefits calculated as a fraction of the last wage when employed (denoted $w_{i(-1)}$), as in the German UI system. We fix the replacement rate, γ_U, to 55%.[8] When UI is exhausted after about 18 months an unemployed worker moves on to the means-tested unemployment assistance. Given the length of time for eligibility and the generosity of social assistance for lower wage individuals such as ours, we have made the simplifying assumption that the replacement rate is always 55%. In addition, there is a utility of leisure $\gamma_0(Ed_i, X_{it}, \varepsilon_i)$, which varies across individuals on the basis of education, experience, unobserved heterogeneity ε_i and a Gaussian white noise η_{it} with variance σ_η^2. Thus, the instantaneous utility of unemployment is:

$$
\begin{aligned}
R_{it}^U &\equiv R^U(Ed_i, X_{it}, w_{i(-1)}, \eta_t, \varepsilon_i) \\
&= \gamma_U w_{i(-1)} + \gamma_0(Ed_i, X_{it}, \varepsilon_i) + \eta_{it},
\end{aligned}
$$

where γ_0 is parameterized as α_X.

The intertemporal value functions and the transition probabilities. Denote by $W_{it} \equiv W(Ed_i, G_t, X_{it}, T_{it}, \kappa_{it}, \mu_{it}, \varepsilon_i)$ the intertemporal utility flow of working in period t and by $U_{it} \equiv U(Ed_i, G_t, X_{it}, w_{i(-1)}, \eta_{it}, \varepsilon_i)$ the flow of utility if period t is spent out of work. These values are defined recursively (and allow for optimal actions in the future) according to the following rule. At the end of period t, unemployed individuals draw a job offer with probability $\pi_{it}^U \equiv \pi^U(G_t, Ed_i, X_{it})$; employed individuals are laid off

---

[7]Postel-Vinay and Turon (2005) show that a sequential auction model à la Postel-Vinay and Robin (2002) can generate such a random walk match-specific component.

[8]In the appendix we describe the details of the German UI system. Here we have just taken a replacement rate that seemed reasonable for our population. Modelling the entire system would imply a vastly increased state space.

with probability $\delta_{it} \equiv \delta(Ed_i, X_{it})$ and conditional on not being laid off, they draw an alternative job offer with probability $\pi_{it}^W \equiv \pi^W(G_t, Ed_i, X_{it})$.

Thus the value of unemployment satisfies the Bellman equation

$$U_{it} = R_{it}^U + \beta \pi_{it}^U \mathbb{E}_0 \max \{U_{i,t+1}, W_{i,t+1}\} + \beta \left(1 - \pi_{it}^U\right) \mathbb{E}_0 U_{i,t+1}, \tag{3}$$

where β is the discount factor,

$$\begin{cases} U_{i,t+1} \equiv U(Ed_i, \rho G_t + v_{t+1}, X_{it}, w_{i(-1)}, \eta_{i,t+1}, \varepsilon_i), \\[2mm] W_{i,t+1} \equiv W(Ed_i, \rho G_t + v_{t+1}, X_{it}, 0, \kappa_{i,t+1}, \mu_{i,t+1}, \varepsilon_i), \end{cases}$$

and with expectation $\mathbb{E}_0$ being taken with respect to the following random variables

$$v_{t+1} \sim \mathcal{N}\left(0, \sigma_v^2\right), \quad \eta_{i,t+1} \sim \mathcal{N}\left(0, \sigma_\eta^2\right), \quad \kappa_{i,t+1} \sim \mathcal{N}\left(0, \sigma_0^2\right), \quad \mu_{i,t+1} \sim \mathcal{N}\left(0, \sigma_\mu^2\right),$$

$$v_{t+1} \perp\!\!\!\perp \eta_{i,t+1} \perp\!\!\!\perp \kappa_{i,t+1} \perp\!\!\!\perp \mu_{i,t+1}.$$

We define the value of working by:

$$W_{it} = w_{it} + \mu_{it} + \beta \delta_{it} \mathbb{E}_1 U_{i,t+1} + \beta \left(1 - \delta_{it}\right) \left(1 - \pi_{it}^W\right) \mathbb{E}_1 \max \{U_{i,t+1}, W_{i,t+1}\}$$

$$+ \beta \left(1 - \delta_{it}\right) \pi_{it}^W \mathbb{E}_1 \max \left\{U_{i,t+1}, W_{i,t+1}, \widetilde{W}_{i,t+1}\right\}, \tag{4}$$

where

$$\begin{cases} U_{i,t+1} \equiv U(Ed_i, \rho G_t + v_{t+1}, X_{it} + 1, w_{i(-1)}, \eta_{i,t+1}, \varepsilon_i), \\[2mm] W_{i,t+1} \equiv W(Ed_i, \rho G_t + v_{t+1}, X_{it} + 1, T_{it} + 1, \kappa_{it} + u_{i,t+1}, \mu_{it}, \varepsilon_i), \\[2mm] \widetilde{W}_{i,t+1} \equiv W(Ed_i, \rho G_t + v_{t+1}, X_{it} + 1, 0, \widetilde{\kappa}_{i,t+1}, \widetilde{\mu}_{i,t+1}, \varepsilon_i), \end{cases} \tag{5}$$

and where the expectation operator $\mathbb{E}_1$ relates to variables

$$v_{t+1} \sim \mathcal{N}\left(0, \sigma_v^2\right), \quad \eta_{i,t+1} \sim \mathcal{N}\left(0, \sigma_\eta^2\right), \quad u_{i,t+1} \sim \mathcal{N}\left(0, \sigma_u^2\right),$$

$$\widetilde{\kappa}_{i,t+1} \sim \mathcal{N}\left(0, \sigma_0^2\right), \quad \widetilde{\mu}_{i,t+1} \sim \mathcal{N}\left(0, \sigma_\mu^2\right),$$

$$v_{t+1} \perp\!\!\!\perp \eta_{i,t+1} \perp\!\!\!\perp u_{i,t+1} \perp\!\!\!\perp \widetilde{\kappa}_{i,t+1} \perp\!\!\!\perp \widetilde{\mu}_{i,t+1}.$$

$\widetilde{W}_{i,t+1}$ is the value of working in a new job with initial draws of the match specific random variables $\widetilde{\kappa}_{i,t+1}$ and $\widetilde{\mu}_{i,t+1}$.

The following remark is in order concerning the lay-off rate $\delta_{it} = \delta(Ed_i, X_{it})$. A number of young people (although not all) are called up for military service. While the reason for leaving employment is not reported in the data we capture the incidence of military service by allowing for a different job destruction rate when work experience is less than five years for those who did not follow the apprenticeship route and between 2-5 years for those who qualified (i.e. for the first three years following their qualification). Following this initial period $\delta(Ed_i, X_{it})$ can be interpreted as the standard job destruction rate, which is comparable to that estimated in other studies such as for the U.S..

The employment transition probabilities. Given this it is now possible to construct the probabilities of events we observe. These include leaving employment, moving to a new job, remaining in the same firm or, for those out of work, moving back to work. For example, consider somebody working in period t, with individual, business cycle and job characteristics $(Ed_i, G_t, X_{it}, T_{it}, \kappa_{it}, \mu_{it}, \varepsilon_i)$. At time period $t+1$ this individual may find himself in a new job, having changed employer at the end of period t. This will happen if a new offer arrives and the previous job is not destroyed and the new offer is better than staying at the previous firm given the innovation to the match specific effect and is better than quitting to unemployment. The event "change job" thus occurs if the Gaussian, independent errors v_{t+1}, $\eta_{i,t+1}$, $u_{i,t+1}$, $\widetilde{\kappa}_{i,t+1}$ and $\widetilde{\mu}_{i,t+1}$ verify the restrictions:

$$\widetilde{W}_{i,t+1} > \max\left\{U_{i,t+1}, W_{i,t+1}\right\},$$

where $\widetilde{W}_{i,t+1}$, $U_{i,t+1}$ and $W_{i,t+1}$ are defined by (5).

However, there is one serious complication to computing the probability of this event: A number of further key predetermined variables are unobserved, including the match productivity component κ_{it} and the non-pecuniary benefits of the old job, μ_{it}. These need to be integrated out and the range of integration will have to be consistent with the fact that individuals were observed making the particular choice they made. In addition the probabilities are conditional on unobserved heterogeneity ε_i. This induces dependence across the probabilities of all events for an individual, over and above the dependence due to the sequence of endogenous decisions. This term will be integrated out of the entire history for the individual. The construction of the probabilities of other events is

similar in nature and we do not describe them explicitly here.

Employment choices while training. Going back, earlier into the individual's history we consider choices available when training. During apprenticeship we assume that the training firm pays the worker only a fraction λ_A of his productivity as a non-apprentice $(w\left(Ed_i = NA, G_t, X_{it}, T_{it}, \kappa_{it}, \varepsilon_i\right))$, the rest presumably serving as payment for the general training received. Reflecting the facts in the data, we do not allow the individual to experience unemployment during apprenticeship, although he can decide to change firm if the opportunity arises. Thus, during the apprenticeship training period $(X_{it} < \tau^A)$ the value of work is:

$$W_{it}^A \equiv W^A(G_t, X_{it}, T_{it}, \kappa_{it}, \mu_{it}, \varepsilon_i) = \lambda_A w\left(NA, G_t, X_{it}, T_{it}, \kappa_{it}, \varepsilon_i\right) + \mu_{it}$$

$$+ \beta\left(1 - \pi^A\right) \mathbb{E}_A W_{i,t+1}^A + \beta\pi^A \mathbb{E}_A \max\left\{W_{i,t+1}^A, \widetilde{W}_{i,t+1}\right\},$$

(6)

where

$$\begin{cases} W_{i,t+1}^A \equiv W^A(\rho G_t + v_{t+1}, X_{it} + 1, T_{it} + 1, \kappa_{it} + u_{i,t+1}, \mu_{it}, \varepsilon_i), \\ \widetilde{W}_{i,t+1}^A \equiv W^A(\rho G_t + v_{t+1}, X_{it} + 1, 0, \widetilde{\kappa}_{i,t+1}, \widetilde{\mu}_{i,t+1}, \varepsilon_i), \end{cases}$$

(7)

and where the expectation operator $\mathbb{E}_A$ relates to variables

$$v_{t+1} \sim \mathcal{N}\left(0, \sigma_v^2\right), \quad u_{i,t+1} \sim \mathcal{N}\left(0, \sigma_u^2\right), \quad \widetilde{\kappa}_{i,t+1} \sim \mathcal{N}\left(0, \sigma_0^2\right), \quad \widetilde{\mu}_{i,t+1} \sim \mathcal{N}\left(0, \sigma_\mu^2\right),$$

$$v_{t+1} \perp\!\!\!\perp u_{i,t+1} \perp\!\!\!\perp \widetilde{\kappa}_{i,t+1} \perp\!\!\!\perp \widetilde{\mu}_{i,t+1}.$$

The first two terms of (6) represent earnings and non-pecuniary benefits of being in the firm. At the end of this period there are two possibilities: with a probability π_A, the apprentice gets an outside offer $\left(\widetilde{\kappa}_{i,t+1}, \widetilde{\mu}_{i,t+1}\right)$ and chooses optimally whether to accept it or remain in the original firm. If no offer is received, the apprentice remains in the firm and accumulates experience and tenure and the match-specific productivity component is updated.

While in the last period of apprenticeship the value function becomes as in equation 4 with all options available. However in this case if the worker qualifies and remains in the firm we observe a wage which is an average of the apprenticeship and fully qualified wage. In effect neither wage is observed and must be integrated out.

The choice to follow an apprenticeship. At 16 the individual makes his first choice, namely whether or not to follow an apprenticeship career. The choice to follow an apprenticeship is assumed to be a one off decision made by comparing the value of a career under the two training alternatives allowing for both the direct costs of training and foregone earnings. At 16, the value of starting to work is given by equation 4 evaluated at $Ed_i = NA$ (non-apprentice), and zero experience and tenure. The value of joining an apprenticeship is given by the benefits of apprenticeship expressed in equation 6 net of direct monetary and utility costs. This is expressed as

$$V_{it}^A \equiv V_A \left(G_t, \kappa_{it}, \mu_{it}, Z_{it}, \varepsilon_i, \omega_{it}\right)$$

$$= W^A(G_t, X_{it} = 0, T_{it} = 0, \kappa_{it}, \mu_{it}, \varepsilon_i) - \lambda_0(Z_{it}, G_t, \varepsilon_i) - \omega_{it},$$

where Z_{it} is a vector of exogenous regressors characterizing the local labor market of the individual at age 16.

The last two terms represent costs. The first, $\lambda_0(Z_{it}, G_t, \varepsilon_i)$, is a direct cost term, which we model as a function of region, business cycle and unobserved heterogeneity. Variability in this term provides identification information and is discussed below in section 2.2. The second term, ω_{it}, is an normally distributed i.i.d. cost shock revealed to the individual before the choice is made. The choice to become an apprentice is thus governed by

$$V^A \left(G_t, \kappa_{it}, \mu_{it}, Z_{it}, \varepsilon_i, \omega_{it}\right) > W(Ed_i = NA, G_t, X_{it} = 0, T_{it} = 0, \kappa_{it}^*, \mu_{it}^*, \varepsilon_i),$$

where κ, μ and κ^*, μ^* represent the match specific characteristics in the initial jobs in the alternative careers. The cost shock ω_{it} induces a probability for this choice. The other unobservables, including the match specific effects in both alternatives and the non-pecuniary benefits need to be integrated out, over the range that is consistent with the observed choice.

Unobserved heterogeneity. Wages and apprenticeship costs depend on unobserved heterogeneity summarized by ε_i. In general it may be far too restrictive to allow just for one factor heterogeneity (see for example Taber (2001)). We thus assume that ε_i

consists of two random variables which follow a bivariate discrete distribution, each with two points of support. One element enters the cost of apprenticeship while the other enters wages and affects the initial value as well as the returns to experience and tenure. The two elements may be positively or negatively correlated or possibly not at all. The potential correlation in the unobserved heterogeneity in costs and wages is just one source of endogeneity of education. In this structural model the other source is the dependence of the education choice on wages, which depends on an unobserved component in the vector ε_i.

We assume that the distribution of unobserved heterogeneity within the group of individuals we are considering remains constant over time. This is an important identifying assumption, because it allows us to compare the education choices and employment paths across cohorts. Our data has the great advantage that individuals are observed when they first start their labor market career, which goes a long way towards making this assumption reasonable.

The likelihood function. The likelihood contribution of an individual conditional on the unobservable characteristics ε_i is the joint probability of all observed events and of observed wages (density). The discrete events include moving in or out of work, remaining unemployed and remaining in the same firm or moving firm. Since these are conditionally independent, given ε_i this probability consists of the product of the density of wages whenever they are observed and of the probabilities for all observed events with the leading term being the probability of choosing apprenticeship or not.

To construct each probability involves solving the model conditional on permanent exogenous characteristics, including ε_i and all other state variables. To solve the model as a function of all state variables, we treat the problem as an infinite horizon one and we use value function iterations to solve it, focussing on the first part of the life-cycle up until age 35 from an average starting age of 16.7. We discretize all the state variables. The state variables include the number of periods the individual has worked (experience, tenure in the current firm, the past wage (for unemployment insurance), region, the position of the business cycle, the current value of match specific effect and unobserved heterogeneity. We fix the discount factor to 0.95 annually.

14

Once the model is solved a number of unobservables need to be integrated out of each probability, which we do by using Gaussian quadrature. Once the probabilities have been computed we need to integrate out unobserved heterogeneity from the product of all probabilities to obtain the joint unconditional probability of all observed events for one individual. Finally, the sample likelihood is assumed to be the product of these unconditional probabilities.

To maximize the likelihood function we use a combination of Simplex and Gauss-Newton optimization algorithms. Most of the computational time for estimation is used up in computing the probabilities that constitute the likelihood function. We estimate standard errors using the outer product of the scores of the log-likelihood function.

The identification strategy. In our model there are a number of endogenous variables. These include current employment, the level of experience and tenure, the choice of firm and whether one is a qualified apprentice or not. The difficulties with identification of such models are now well understood and Altonji and Williams (1998) provide an eloquent illustration of some of the issues on a wage equation which has some similarities to ours. The difficulties are compounded by having apprenticeship qualification as endogenous. Here by modelling the entire career as a sequence of endogenous decisions which subsequently drive the events that follow we control for endogeneity using all the restrictions implied by economic theory. Beyond this, identification is achieved through exogenous variation at the time of the first decision, i.e. to train as an apprentice or not. Changes in the local demand for apprentices by firms over time provides such exogenous variation. To justify this first note that industries are not uniformly distributed across regions. Thus each region is exposed to different product market shocks. Using familiar trade arguments, wages will not be affected by such local shocks, but the local demand for labour will. As a result each year and in each region there is variability in the number of apprenticeship positions made available by firms. This differential availability across region and time affects the cost of obtaining apprenticeship training but not wages under factor price equalization. If plenty of positions are available in one's region of residence one can live at home and only commute short distances to the training workplace. However, when the available positions are few one may have to travel longer distances and

15

possibly live away from home to obtain apprenticeship, incurring greater costs. We model this by allowing the direct costs of apprenticeship to vary by region of residence at the time of apprenticeship (Z_{it}) and the business cycle as well as an unobserved component ε_i.[9] The initial region of residence is taken as exogenous. We then assume that the labor market is integrated in the country with full factor price equalization and we exclude region and region interacted with the business cycle from wages and preferences for work.[10] The availability of 20 cohorts of data for the German states provides ample differential variability in the initial exogenous conditions to be able to identify the model by in effect comparing the careers of individuals who entered the labor market at different point in time and in different regions.

3   The Data Set

We draw a sample from a data set organized by the German IAB[11] and which in its totality consists of a 1% extract from the German social security records. The data set starts in 1975 and records all work spells with exact start and end dates. The data records spells of apprenticeship training and whether a worker holds an apprenticeship qualification or not. Once an individual is in the data set they are always followed. We concentrate on those for whom we can observe the start of the labor market career so as to avoid any initial conditions problem. This means that the oldest person in our data is 35. Our observation window is 1975-1995. This is important because it offers a long time period and hence a large number of changes in the aggregate environment; this provides the required variation for identifying education choice.

The data set also reports the average daily pre-tax wage each year if the individual stays an entire year in a firm, or for the part of the year the individual works for the firm.

---

[9]More generically, we could have used an output price index by region as the factor driving costs. By using the business cycle interacted with region we are effectively splitting up industries into ones which are more pro-cyclical than others. This will only reduce effciency.

[10]Identification relies on the exclusion of time/region interactions only, not region itself. So we have imposed more restrictions than absolutely necessary. Technically, we could go further and include region effects on wages, to allow for permanent compensating differentials across regions. However this would make the model much harder to estimate because it would multiply the size of the state space by 10fold. It would also raise the further problem of regional choice.

[11]Institut für Arbeitsmarkt- und Berufsforschung (Institute for Employment Research).

Thus wages are not averaged across different firms. The data is far too detailed. We thus time-aggregate the data to obtain information on a quarterly basis. The appendix describes precisely how this is done.

Our sample consists of West-German males, who end formal education at 15/16 and who either work or join an apprenticeship after school. We drop all individuals who continue onto higher education, a rather small fraction in Germany. In total, we use 1400 individuals followed through time, quarter after quarter up until 1995. To re-iterate, our data has some key advantages for the type of work we carry out: All transitions are recorded accurately from administrative records and so are wages from the start of the labor market career, and through the period of apprenticeship training, if applicable.

3.1   Descriptive Analysis of the Data

Wage Profile and Labor Market Transitions.   Figure 1 displays the log wage profile as a function of years of labor market experience for those with an apprenticeship qualification ("skilled"), for those currently training as apprentices ("wage in apprenticeship") and for the non-apprentices ("unskilled") as well as the difference between the apprentices and non-apprentices (right hand axis). Non-apprentices have a rapid increase in their wage during the first five years on the labor market. Over the next fifteen years, the wage growth is only about twenty percent, resulting in a 1.5% real average growth rate per year. During apprenticeship training workers are paid a very low wage, possibly to cover the cost of their apprenticeship which includes classroom training but possibly other work time inputs. [12] At the end of the apprenticeship training, wages increase and overtake those of non-apprentices. From there on, the wages of those with an apprenticeship qualification increase slightly faster than those without at an average rate of 1.6% per year. After fifteen to eighteen years, the difference in wages between skilled and unskilled is about ten percent. From this graph it almost seems puzzling that anyone wishes to follow an apprenticeship career, given the large up-front investment in training that lasts about 3 years and the apparently low rate of return in terms of wages. Of

---

[12]Heckman (1993) sees the low apprenticeship wage as a means of bypassing minimum wages mandated by the unions. Given the length of apprenticeship training and the often narrow set of skills they sometimes offer this is not an unreasonable interpretation.
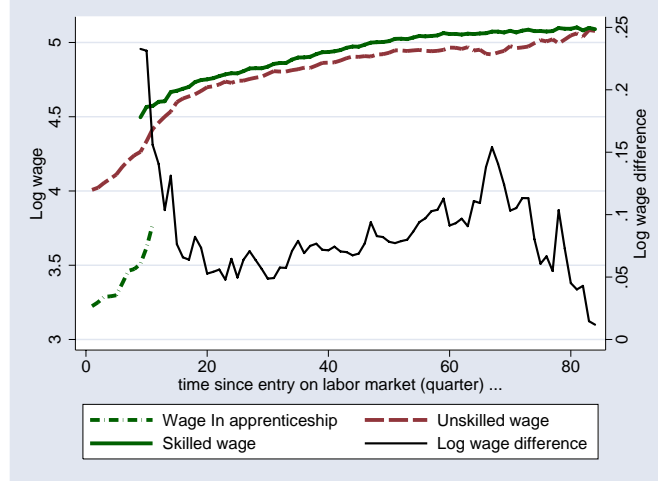
Figure 1: Log Wage by skill and the wage gain for qualified apprentices

course comparative advantage and other differences between the two career paths may well explain the large participation rates in apprenticeships and it is one of the questions we investigate.

Wages are only one dimension in which education groups may differ. Another important dimension is labor market attachment. Table 1 displays the quarterly transition probabilities by education groups and time in the labor market. Unskilled workers have a higher probability of dropping out of work. During the first five years on the labor market, each quarter, about four percent of employed skilled workers exit, while this figure is about 8% for the unskilled. The proportion decreases when we look at more senior workers, but the education difference still persists. The probability of job to job transitions are the same for both education groups, at about 2-3%. This probability decreases with time since first entry on the labor market.

Qualified apprentices have a higher probability of return to work from unemployment, by about four to five percentage points. This reinforces the effect on unemployment of the higher exit probability for the unskilled. Thus, in total, the unskilled spend less time working; over 20 years they work a total of 15 years, compared with a total of 16.5 years for skilled workers.

The education differences in exit and entry probabilities implies that non-apprentices are more mobile and have more job experiences with more firms than apprentices. Fig-

|	Work (Same Firm)	Work (New Firm)	Out of Work
Apprentices, First 5 years			
Work	92.8	2.6	4.6
Out of Labor Force	29.6		70.4
non-apprentices, First 5 Years			
Work	88.7	3.0	8.3
Out of Labor Force	25.7		74.3
Apprentices, After 5 years			
Work	96.2	1.9	1.9
Out of Labor Force	18.1		81.9
non-apprentices, After 5 Years			
Work	94.4	1.9	3.6
Out of Labor Force	13.1		86.8

Table 1: Observed Quarterly Labor Market Transitions

ure 2 displays the number of firms in which an individual has worked in as a function of time since entry on the labor market. The difference comes from the early years, where workers during their apprenticeship, are much less mobile. However they never catch up following qualification. The mobility numbers are much lower than those in the U.S. as documented in Topel and Ward (1992) amongst others.

Decomposing Wage Growth. Wage growth occurs both within firm and as a result of firm mobility. Job shopping, although sometimes vilified as engendering instability, can be a very important source of wage growth as documented in Topel and Ward (1992) and can be crucial in achieving efficient matches (see Heckman (1993)).

In Germany, despite lower mobility rates, this is no less the case. This is illustrated in figure 3 which shows within firm wage growth by potential experience and skill level and in figure 4, which displays the growth of wages following a job to job transition. The wage growth in the latter case can be substantial, at about 30% for non-apprentices and 10% to 20% for qualified apprentices (post training). The gain in wages falls over time, decreasing towards zero. If we think of wage improvements as being due to better matches, as in our model, the decline is expected because the probability of an improvement will decline

Figure 2:
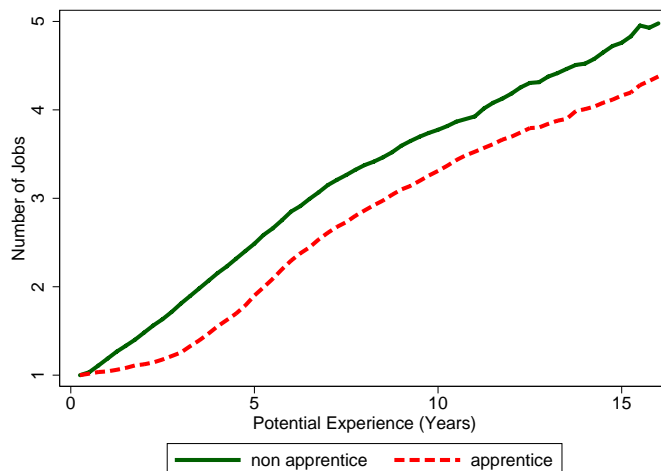
as the worker climbs up the job-quality ladder. Within firm wage growth for the non-apprentices is very high early on in the career. This probably reflects the rapid learning that takes place on the job. The equivalent training for the apprentices takes place during the official training period. Clearly job mobility is an important source of wage growth. Carrying out a simple decomposition exercise, for the unskilled 25% of growth of wages over 20 years is accounted for by job mobility. For those following an apprenticeship career the figure is 15% for wage growth that follows the training period. Whether this difference means that matching is more important for lower skill individuals or simply that qualified apprentices are less mobile and are missing out on opportunities can not be ascertained from this.

Vocational Training and Wages. Given the exogenous variation determining apprenticeship, as described earlier in section 2.2, we can follow an instrumental variables approach to estimate the effect of apprenticeship on wages, ignoring of course any selection effects due to participation. This is done mainly as a descriptive device and to illustrate what would be obtained using the more standard IV approach.

To check the first stage, we run a probit for apprenticeship choice including time effects, region effects and their interactions. The latter have a p-value of zero establishing that indeed there is sufficient differential variation of apprenticeship participation, which

Figure 3:

Figure 4:
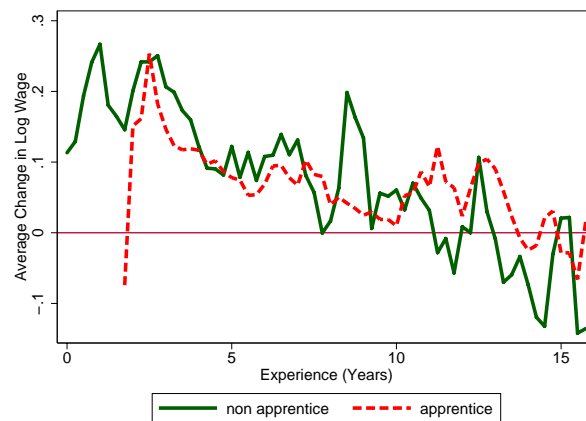
we attribute to changing availability of positions and costs.

We then use the interactions between region and cohort as the excluded instruments in a log wage equation to estimate the effect of an apprenticeship.[13] In particular we estimate the following regression

$$\ln w_{it} = (\text{region effects}) + (\text{time effects}) + \sum_{k=0}^{3} \zeta_k (PX_{it})^k + \sum_{k=0}^{3} \xi_k Ed_i \times (PX_{it})^k + \gamma \widehat{e}_{it} + v_{it}$$

where PX represents potential experience and $\widehat{e}_{it}$ is the residual from the linear reduced form regression of apprenticeship on region and time effects and their interactions. This control function approach for controlling for the endogeneity of apprenticeship choice (Ed) is identical to IV in linear models and is useful here where we have four different education terms. The regression is similar to a difference in differences approach with many time periods and regions (see Blundell, Costa-Dias, and Meghir (2003)).

This regression is estimated for all those who have at least four years of potential experience, which ensures that the trainees will have completed apprenticeship. We compare the results to those obtained by OLS (i.e. excluding the residual) in figure 5. The horizontal axis is potential experience after formal schooling ended at 16. The p-value on γ is an exogeneity test for Ed, and in this occasion it is about 3% rejecting exogeneity. The results show an IV return which is higher than OLS both of which increase with age. These facts will be replicated by our model, albeit in a richer context, where we control for selection into employment, as well as for the returns to actual experience and tenure.[14] Noting that an apprenticeship lasts between two and three years and it only involves part time schooling, the rest of the time being work, these returns are of the same order of magnitude as the returns to education.

4   Estimated Parameters

We estimate the model by maximum likelihood. We then evaluate its fit by simulating the education decisions, the labor market transitions and the wages for a cohort of individuals

---

[13]The estimates represent Local Average Treatment effects if the underlying parameters are heterogeneous. See Imbens and Angrist (1994).

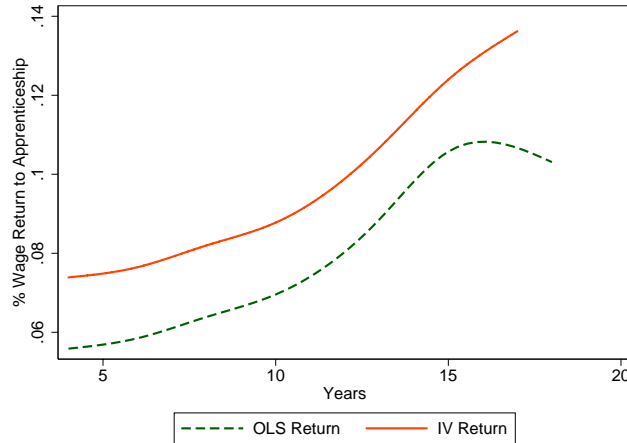[14]The differences are significant at the 5% level.

Figure 5: Wage returns to Apprenticeship (OLS and IV)

over time and comparing to the actual data. The model fits remarkably well and we refer the reader to the appendix where the results are shown in some detail.

Before we get into details, at this point we need to say a few words about specification choices. Most parameters differ by apprenticeship status. The job destruction rate is allowed to be different for the first four years of experience to allow for the fact that a number of people exit the labor force temporarily so as to complete their compulsory military service.[15] We allow the rate of arrival of job offers to differ by the business cycle, which can either be high in good times, or low. Finally, the job arrival rate on the job is allowed to vary by experience.

Common factors. Table 2 presents some key parameters that determine the careers of individuals. In presenting them we compare them to equivalent estimates for the U.S. reported in Low, Meghir, and Pistaferri (2006), allowing us to offer an explanation for the different career structures in the two countries. The comparison is appropriate because, despite their different approaches both studies allow for match specific effects, job mobility and for permanent shocks to wages.

The two parameters that characterize the stochastic structure of wages are the standard deviations of the innovation to the match specific effect (σ_u) and match hetero-

---

[15]We do not observe explicitly the reason for exit - just that they leave their job and stop working for a year or more.

geneity (σ_0). The former is about 0.09 a year for log wages and is slightly lower than the corresponding estimate for the U.S. which is about 0.114. The latter (σ_0) is 0.29 for the qualified apprentices and 0.34 for the non-apprentice group. These compare to an average across all education groups of 0.213 for the U.S.. [16] Thus in both countries there is considerable heterogeneity in job matches and hence great opportunities for wage growth from job shopping.

Exogenous job destruction rates, i.e. excluding quits, are 0.02 and 0.03 a quarter for the two groups in Germany. In the U.S., the numbers are 0.02 for the College graduates and 0.044 for those with less than College. The latter group is most comparable to our sample and it seems that jobs get destroyed at twice the rate than in Germany, pointing to the first reason why mobility rates among workers are different. When we simulate the model the overall quit rate is 0.018. Thus the total job destruction rate in Germany is 0.043.

The job arrival rates when unemployed is between 0.20 to 0.23 a quarter, while when employed these are between 0.09 and 0.11 depending on the business cycle and the skill level (see Table 2). These are considerably smaller than the equivalent U.S. numbers which are 0.5 and 0.65 for the employed and the unemployed respectively. Thus the reduced mobility of German workers compared to the U.S. ones can be attributed to lower availability of job opportunities. The lower mobility may also be consistent with the higher match heterogeneity found in Germany; high levels of mobility render the labor market more competitive and lead to more wage equalization (less frictional wage dispersion).

Apprenticeship choice is driven partly by the opportunity cost of apprenticeship. We estimate that those training for apprenticeship are paid 40% of the wage they would be paid as non-apprentices with the same tenure and experience (see λ_A in the table). As we shall see the high opportunity cost of training will be a central factor driving the returns to apprenticeship.

Finally, the German business cycle has a very small effect on relative wages for the

---

[16]In Low, Meghir, and Pistaferri (2006) the productivity shock is carried from one firm to another. This will tend to reduce the variance of the firm effect and increase the variance of the innovation to the permanent shock. Overall the stochastic properties of wages are remarkably similar.

Parameter	Qualified Apprentices		Non-Apprentices	
Standard dev. of innovation to match specific effect (σ_u)		0.086	(6e-5)	
Standard dev. of initial match specific effect (σ_0)	0.285	(0.003)	0.34	(0.005)
Quarterly job destruction rate (δ)				
if experience$\leq$4 years	0.18	(0.007)	0.07	(0.005)
if experience>4 years	0.019	(0.002)	0.029	(0.002)
Quarterly offer arrival rate when employed (π_W)				
if business cycle low,	0.106	(0.004)	0.094	(0.006)
if business cycle high	0.111	(0.004)	0.089	(0.006)
Quarterly offer arrival rate when unemployed (π_U)				
if business cycle low, experience=0	0.206	(0.009)	0.204	(0.006)
if business cycle high, experience=0	0.206	(0.009)	0.204	(0.006)
if business cycle low, experience=10	0.234	(0.009)	0.225	(0.006)
if business cycle high, experience=10	0.234	(0.009)	0.225	(0.006)
Effect of business cycles on (log) wages (α_G)	0.006	(0.001)	0.003	(0.002)
Standard dev. of utility shocks to unemployment (σ_η)		287.95	(12.2)	
Standard dev. of fringe benefits (σ_μ)		19.1	(0.32)	
Utility of unemployment (γ_0)	-68.6	(5.5)	-87.4	(7.7)
Proportion of non-apprentice wage paid to trainees λ_A	0.406	(0.005)		

Note: asymptotic standard errors in parenthesis. When only one parameter estimate and its standard error are presented in a row this parameter is restricted to be the same across the two groups

Table 2: Estimated parameters

two groups. This is of the order of 0.6% between good and bad times. This may well be due to the openness of the German economy, which would imply lower sensitivity to local aggregate shocks; however, one may expect that these are correlated with other European Union countries, the main trading partners for Germany.

Unobserved heterogeneity. The model allows for two factors of unobserved heterogeneity; one factor affects the initial level of wages and another factor affects the costs of apprenticeship. We use two points of support for each factor, which implies the existence of four types of individuals. We estimate the proportion of these types to be 4%, 3%, 72% and 21%. Table 3 displays summary characteristics for these groups. Individuals of Type 1 and Type 2 are individuals with a low initial wage, whereas Type 3 and Type 4 individuals have a high initial wage. Both Type 2 and 4 individuals have a higher cost of choosing apprenticeship equivalent to about 4% of life time value. As a result, the proportion of qualified apprentices is larger in the low cost groups. The probability of being a low education cost individual is 0.77 among the high wage people and 0.57 among the

	Type 1	Type 2	Type 3	Type 4
Proportion in Sample	4%	3%	72%	21%
Log wage constant	3.42 (0.11)	3.42 (0.11)	4.04 (0.09)	4.04 (0.09)
Utility Cost of Apprenticeship	0%	4% (0.1)	0%	4% (0.1)
(% of total lifetime value)				
Proportion Apprentices	65%	38%	75%	59%
Average Return to Experience (per year)				
Apprentices	4% (0.8)	4% (0.8)	1.2% (0.23)	1.2% (0.23)
non-apprentices	6% (0.8)	6% (0.8)	1.7% (0.23)	1.7% (0.23)
Average Return to Tenure (per year)				
Apprentices	14% (3.3)	14% (3.3)	0.01% (0.01)	0.01% (0.01)
non-apprentices	5% (1.3)	5% (1.3)	0% (0.01)	0% (0.01)

Note: asymptotic standard errors in parenthesis.

Table 3: Unobserved Heterogeneity and the returns to experience and tenure

low wage individuals; thus education and labor market ability are positively correlated, although not perfectly so. Nevertheless it is quite surprising how little heterogeneity in initial wages is present. These results indicate that apart from 7% of people with really bad initial conditions, the remaining heterogeneity can be explained either by the accumulation of innovations or the endogenous factors such as the training choice, job mobility, experience and tenure. As we have shown the variances in these components are substantial. This indicates that pay rates are very homogeneous at the start in the German labor market and differences arise later.

Unobserved heterogeneity also affects the returns to experience and tenure. Individuals with low initial wages have high returns to tenure and experience and particularly so for the qualified apprentices. For them low initial labor market wage seems to be compensated by rapid learning. For the majority (93% of the population) however the returns to tenure are effectively zero. The returns to experience are higher for non-apprentices than qualified apprentices, reflecting the fact that a lot of the general learning takes place while training for an apprenticeship. Learning for the unskilled group however takes place while in a "standard" job.

4.1   Return to Apprenticeship

We use the model to estimate the life-cycle return to apprenticeship and its various components. First we compute the wage returns as a function of potential experience.

We do this by simulating the wage profile under the two education states for a set of randomly chosen individuals. The average difference between the realized profiles is the "Average Treatment effect" (ATE). We then compute the Average treatment on the treated effect (ATTE) by simulating the counterfactual wage profile for those who chose to go into apprenticeship and compare it to the one they obtain following their choice to obtain an apprenticeship qualification (also simulated). Both are compared to the profile that is obtained when endogenous selection is ignored - equivalent to OLS.

The results are shown in Figure 6 and show substantial bias in the raw (OLS) wage differences due to self-selection. The ATTE is higher than ATE; they both grow more rapidly than the OLS returns reaching about 27% by age 35, compared to the OLS results of 17%.

If we switch off permanent unobserved heterogeneity the OLS returns and the ATE returns become almost the same. The difference is driven by the fact that individuals who have better unobserved wage components are less likely to join apprenticeships, despite the fact that their utility costs of education are also lower; in other words opportunity cost considerations dominate the apprenticeship choice. Thus in the same way that IV was higher than OLS in figure 5 here ATE is higher than OLS. Note however, that our ATE estimate displays a different profile over time to the IV one shown in figure 5. Ignoring other differences between the estimators[17], this will mean that the Local Average Treatment Effect (LATE) which is applicable to the marginal worker is in fact lower than the ATE in later years but higher earlier on. An interpretation may be that the marginal entrant has a flatter profile as a qualified apprentice than the average worker.

The wage returns to apprenticeship, however, only provide a partial picture of the relative advantages of the two careers. These differ in a number of other dimensions, including, job destruction rates, income while unemployed, job arrival rates, sensitivity to business cycle fluctuations and dispersion of new job opportunities. In addition, we need to take into account the costs of apprenticeship, including direct utility costs and opportunity costs.

---

[17] The ATE effect whih is obtained form the model also controls for endogenous participation decisions, experience, tenure and mobility. The results in the figure however do not conditionon experience or tenure and in this sense are comparable to those in figure 5.
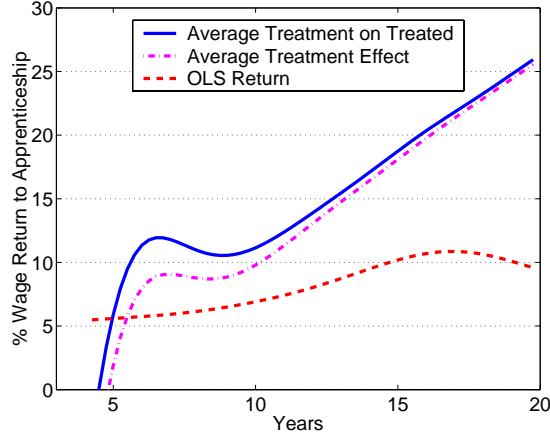
Figure 6: The wage returns to apprenticeship

Thus, the overall individual return to apprenticeship is given by

$$r\left(\varepsilon\right) = \frac{\mathbb{E}_{G,\kappa,\mu}W^A(G, X = 0, T = 0, \kappa, \mu, \varepsilon)}{\mathbb{E}_{G,\kappa,\mu}W(Ed = NA, G, X = 0, T = 0, \kappa, \mu, \varepsilon)} - 1$$

where the numerator is the discounted value of having an apprenticeship qualification as seen at the time of making the original career choice, while the denominator is the equivalent value of remaining unskilled. For this calculation we employ a horizon of 40 years. The results are displayed in Table 4.

Taking all individual costs into account, the average return to apprenticeship (ATE) is in fact negative at -1.7%. However for those who choose to qualify as apprentices the returns (ATTE) are a substantial 8.4%. The negative average return is due primarily to the opportunity cost of apprenticeship training (forgone wages) and to a lesser extent to its utility costs. This can be seen in the third and fourth rows of the Table where we in turn ignore the utility costs and the opportunity costs respectively. Thus the negative return is driven to a large extent by the pay differential between non-apprentices and those training for an apprenticeship. Finally note that the returns do not factor in any costs incurred by the government (for class room training) or by the firm.

The four last columns in Table 4 show the way the returns vary by type. Here there are some interesting patterns: the average returns to apprenticeship are higher, and indeed positive for individuals who have a low initial wage - this is driven by their lower opportunity cost of schooling. Thus the highest average returns are enjoyed by

	Average	Type 1	Type 2	Type 3	Type 4
		Low Wage		High Wage	
		Low Cost	High Cost	Low Cost	High Cost
Return to Apprenticeship at age 15					
Average Treatment Effect (ATE)	-1.7 %	5.9 %	2.2 %	-1.2 %	-5.5%
Average Treatment on the Treated (ATTE)	8.4 %	6.7 %	5.4 %	8.8 %	7.1%
ATE, net of utility of education	2.8 %	9.5 %	8.8 %	2.3%	2.3 %
ATE, net of opportunity cost of education	8.8 %	13.1 %	9.4%	9.6 %	5.3 %
Decomposing the Average return to Apprenticeship (ATE) at age 18					
Baseline	14.0 %	12.4 %		14.1 %	
Returns at age 18 when apprentices have some non-apprenticeship characteristics					
Equal distribution of firm-worker match (σ_0)	21.4 %	18.6 %		21.6 %	
Same Business Cycle effects on wages	13.8 %	12.1 %		13.9 %	
Same Job to Job offer rate	13.7 %	12.0 %		13.8 %	
Same Job Offers	13.1 %	11.4 %		13.2 %	
Same Job Destruction	13.9 %	10.8 %		14.2 %	
Same Job destruction and job offers	13.1 %	10.0 %		13.4 %	

Table 4: The Life-cycle Returns to Apprenticeship

the low cost/low initial wage individuals. Once we consider the returns enjoyed by those who choose to move into apprenticeship (ATTE in the 2nd line), the high initial wage individuals also enjoy the higher returns.

The lower part of the table effectively strips out the costs of apprenticeship by considering the returns as viewed at age 18. These are a function of the wage returns, illustrated in Figure 6, as well as of all the other differences between the groups including the different job destruction and arrival rates, leading to different levels of job attachment, the different dispersion of offered wages and the different implied levels of unemployment support. Given these factors and ignoring the apprenticeship costs, the average returns are 14%. In the lower part of the table we quantify the effect that the various differences between qualified apprentices and non-apprentice have on the returns by giving the qualified apprentices characteristics of the environment of non-apprentices. When those following the career path starting with an apprenticeship are "given" the dispersion of wage offers σ_0 that non-apprentices face the returns increase to 21.4% demonstrating the importance of that feature of the labor market in increasing the earnings of the non-apprentices through job mobility. All other features shown in the table are detrimental to the life-cycle earnings of non-apprentices but have much smaller effects overall, once

individuals are allowed to change their behavior in face of the new environment.

5 Evaluation of Labor Market Reforms

Standard evaluations, whether structural or based on experimental methods often focus on the targeted outcomes only. An intervention aimed at increasing employment, for example, is evaluated purely on this outcome. However, such interventions, particularly when viewed as permanent may well change other decisions leading to different levels of human capital accumulation over the life-cycle. This possibility has been well understood but has rarely been quantified. Our model is ideally suited for this and we demonstrate that the effect on human capital accumulation can be substantial. Such effects are almost impossible to evaluate without a structural model that considers long run career choices for individuals.

In this section we present the impact of two potential reforms. First we consider the effect of the introduction of an Earned Income Tax Credit in Germany, a type of policy currently implemented in both the U.S. and the UK and being debated for implementation in Germany. [18] The maximum subsidy can amount to 40% of the wage for those who are eligible and approximately 12% of the median wage. Heckman, Lochner, and Cossa (2003) provide an analysis of the effects of EITC on human capital accumulation, through its effect on choices for on-the-job training. They emphasize the difference in effects depending on whether human capital accumulation is rivalrous to work as in Ben-Porath (1967) and Becker (1964) or simply a by-product of work which does not require a reduction in work time and hence earnings. Our model allows for the latter form of non-rivalrous human capital accumulation when working; so from this respect an EITC type programme will lead to increased human capital accumulation because it encourages work. However, our model also allows for the possibility that individuals may refrain from joining an apprenticeship scheme because the programme compresses the returns to education. Finally, the wage subsidy will change the incentives for job

---

[18]In our simulation the rates are set to match those of the U.S. EITC policy. There is a debate in Germany to introduce programmes similar to the EITC. Perhaps the best known proposal is that of Germany's IFO institute under the name "Aktivierende Sozialhilfe" or "Kombiloehne" (Sinn, Holzner, Meister, Ochel, and Werding (2002) 2006). It proposes a permanent wage subsidy, to be paid to all low qualified workers, and is aimed at the low end of the earnings distribution.

| Name	Description
(1) EITC	A wage subsidy at a rate of 40% up to DM 36.6 per day, stays constant up to DM 54.6 per day and declines to zero at a rate of 21% thereafter (see Figure 7) EITC is available for those above 19 years of age only. It is financed by a tax on all income.
(2) Flat Unemployment Benefit	55% of "minimum wage" defined as 27 Deutchmarks. Excess revenue redistributed through taxation.

Table 5: Simulated Policies
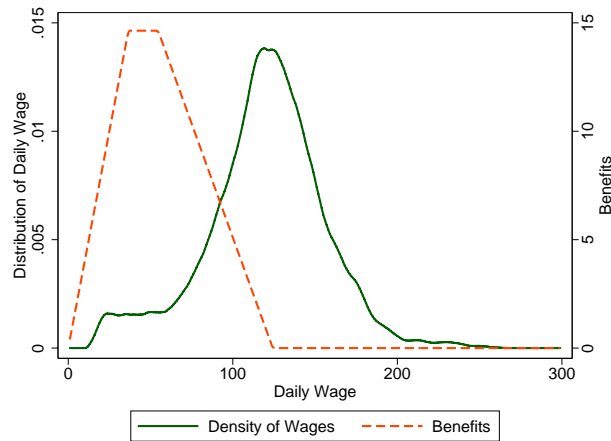
Figure 7: Density of Wages and In-Work Benefit Scheme

mobility, because it will reduce the number of jobs that arrive with improved earnings and utility, after the programme is taken into account.

The second policy we consider is the introduction of a flat unemployment benefit similar to that in the UK, instead of the current German system where the young lower paid unemployed are paid about 55% of their last earnings. Both reforms are outlined in Table 5.

All policies are simulated to be revenue neutral. We raise the funds or give back excess revenue by proportional taxation/subsidy. The effects we show include the effects of such taxation, needed to fund the policy.
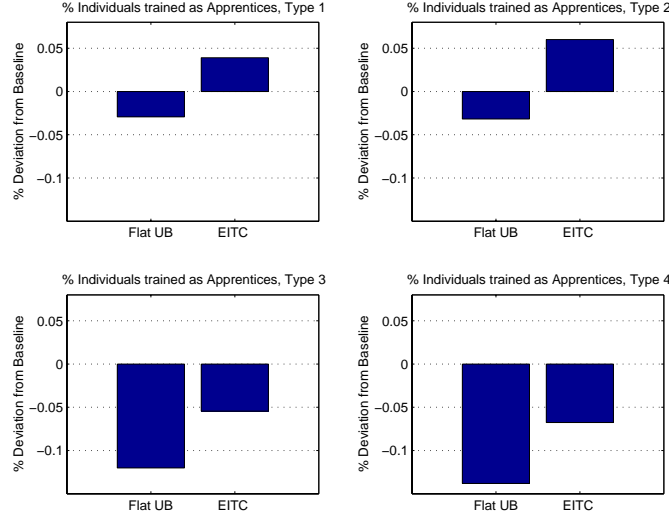
Figure 8: The Impact of Policy on the take-up of Apprenticeship by type

5.1 Results

To derive the implications of the three suggested policies we first simulate the model under baseline (no policy change) and then under each of the reforms for 10,000 individuals. We then describe the impact of the reform on three key outcomes: Education choice, employment and quality of match.

Figure 8 displays the effect on education choices by type of individual. Overall in work benefits designed on the U.S. EITC have a large effect on training, reducing it by nearly 7 percentage points. This is because the returns to training are compressed by the subsidy.[19] Given that low wage jobs are subsidized, non-apprentices are clearly favored by this policy and this attracts more into the group and out of apprenticeship training. Thus a standard policy designed to increase employment seems to have an important impact on human capital accumulation. On the other hand the employment impact of the policy is quite small, increasing employment overall by about 1%.

Considering the effect of the policy by type a more intricate picture is revealed. Among the individuals who are low initial wage types (types 1 and 2) there is an increase in takeup of apprenticeship training because their wages are low enough that training increases their EITC eligibility as well as improving their labor market attachment, leading to higher

---

[19]Note that the subsidy is only available to those over 19, when apprenticeship raining will have finished; hence the policy does not act as a direct monetary disincentive to training.
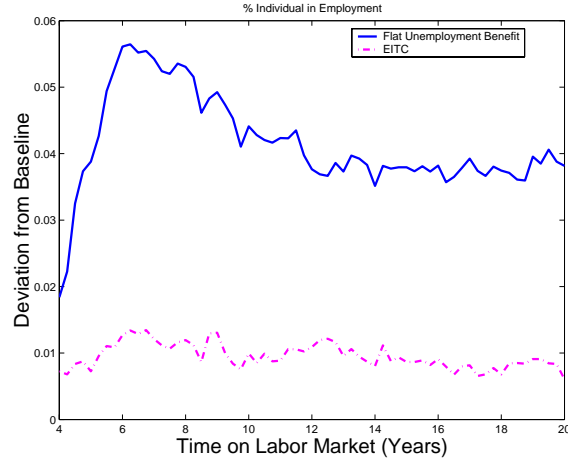
Figure 9: Proportion Individuals Working, Compared to Baseline

life-cycle benefits. The decline in training thus comes from those who are high initial wage types - the majority in our sample. For them the returns to training receive a double taxation; first it would cause them to drop out of EITC or heavily reduce the amount received potentially; second they face an increase in taxation used to fund the programme.

Replacing the earnings related UI with a flat rate, which is independent of earnings, has two opposite effects. On the one hand non-apprentices having higher job destruction rates will be worse off during unemployment and thus will have an incentive to train to improve labor market attachment. On the other hand one of the benefits of training as an apprentice, namely higher income when unemployed, is removed thus reducing the incentive to train. The net effect is a reduction in those training for an apprenticeship by about 11 percentage points. However employment rates are increased by about 5 percentage points for the young and 4 percentage points for the more experienced workers, both because unemployment is less attractive and because tax rates are now reduced because the government payouts are lower. The impact of the policy on the take up of apprenticeship is much stronger for the higher initial wage individuals because for them the actual change in the income when out of work is larger.

An additional channel by which policy has an impact is by changing the incentives for job mobility. Thus, an individual receiving EITC and being offered a job with a better match value may not move because the improvements in utility, after deducting
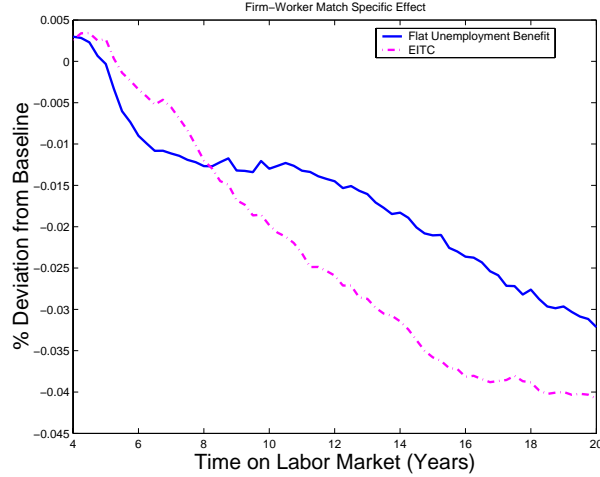
33

Figure 10: Policy Effect on Firm-Worker Match Specific Effect

the change in EITC benefits may be negative. A similar argument can be made for the UI reform; in the pre-reform regime a better job had the additional value of an increase in UI eligibility, which is no longer the case with the flat rate. In addition because unemployment is much less attractive individuals will be more likely to accept the first job offered to them when out of work. Both these effects will tend to reduce the value of the matches and consequently wages. Figure 10 shows the effect of our policy reforms on the quality of the firm-worker match. By 20 years in the labor market the quality of the match decreases under EITC by about 4% relative to the current baseline (as in the data). The flat unemployment benefit policy decreases the quality of the match between firms and workers over their life-cycle by about 3.2%.

Finally we can summarize the entire effect of the reforms by an overall effect on welfare. The EITC reform would cause a 1.4% decline in welfare while the flat UI would raise welfare by 2.5% over the lifecycle. These figures do not take into account the insurance value that EITC or an earnings related UI system might offer because utility is linear in income.

6   Conclusion

Understanding the response of individuals to incentives when planning a career is central to analyzing the longer term effects of policy. Training choices and labor supply as well as

job mobility are all interlinked and result in a path of wages and employment that define an individual's earnings and employment history. In general it is not possible to evaluate the impact of a reform without taking into account of such links. The availability of an exceptionally rich and unique administrative German data set has allowed us to model such life-cycle paths without the distractions of measurement error and recall bias.

We specify a model where individuals who stop formal schooling at 16 choose between entering directly the labor market and obtaining on-the-job training or obtaining an apprenticeship qualification. Following this initial training choice individuals make labor supply and job mobility decisions accumulating general and firm specific human capital as they go along, leading to a path of wages and employment at various firms. The model is estimated on individual career histories, which are observed from the point of first entry in the labour market.

The estimates of the model allow us to characterize the career paths and to understand the importance of the various components for wage growth and employment. We can also use the model to simulate important policy reforms and contribute to the debate on the desirability of such interventions. Thus, we complete our results by using the model to consider the longer term impacts of two policies, both of which have been implemented in either the UK or the U.S. These are the U.S. Earned Income Tax Credit and an unemployment insurance system that is not related to past earnings (as in the UK). While the policies have the desired effect on their targeted outcome, namely employment they also have substantial impacts on other aspects of behavior, namely training and job mobility. It becomes quite clear that ignoring these other effects gives a highly distorted view of the longer term impacts. Thus our results highlight the importance of combining standard evaluation approaches with structural models that are capable of addressing these longer term effects.

Appendix

The German Unemployment Insurance System

The German unemployment compensation scheme distinguishes between unemployment insurance benefit (*Arbeitslosengeld* AG) and unemployment assistance (*Arbeitslosenhilfe* AH). To be eligible for AG, the employee must have contributed for at least 12 months over the preceding 3 years to the scheme. The scheme is financed by employer and employee contributions in equal parts (amounting to 3.25 percent of the employee's salary). There is a waiting period of 12 weeks if the separation was induced by the employee, but receipt of AG starts immediately if the separation was caused by the employer. The compensation is oriented on previous net earnings, and it amounts to 67 percent of the previous net wage (or 60 percent for employees without children). There is an upper threshold (for instance, 5200 DM in 1984, and 6000 DM in 1990). AG can be received for up to 32 months, with the duration of the entitlement period depending on age and the length of contributions to the scheme. If an unemployed person fulfills the above criteria, the minimum period of eligibility is 156 days. Depending on the duration of contribution payments and the age of the applicant, this period can be extended to up to 832 days (see Kittner (1995), p. 192, for details.)

If AG is exhausted, or if the employee is not eligible for AG, he can claim AH. A condition for receiving AH in case of non-eligibility for AG is having been in insured employment for at least 150 days during the last year. Like AG, AH is based on previous earnings; it amounts to 57 percent of previous net earnings (50 percent for employees without children). AH is means tested, and its duration is unlimited. Both AG and AH are granted conditional on the recipient's agreement to accept a reasonable employment (*zumutbare Beschäftigung*).

The Fit of the Model

Table 6 displays the labor market transitions by education groups at a quarterly frequency. We distinguish five possible transitions, from and to unemployment, between same job and job to job. Overall, the model matches the transition probabilities closely.

A reflection of the good fit of the transitions above is the fit of the average experience and tenure over time for the two education groups in Figure 11 plots. The model does a good job in both dimension and even picks up the non linearity in the evolution of tenure for qualified apprentices. We also predict very well the average number of jobs held by both skill groups as a function of potential experience (Figure 12).

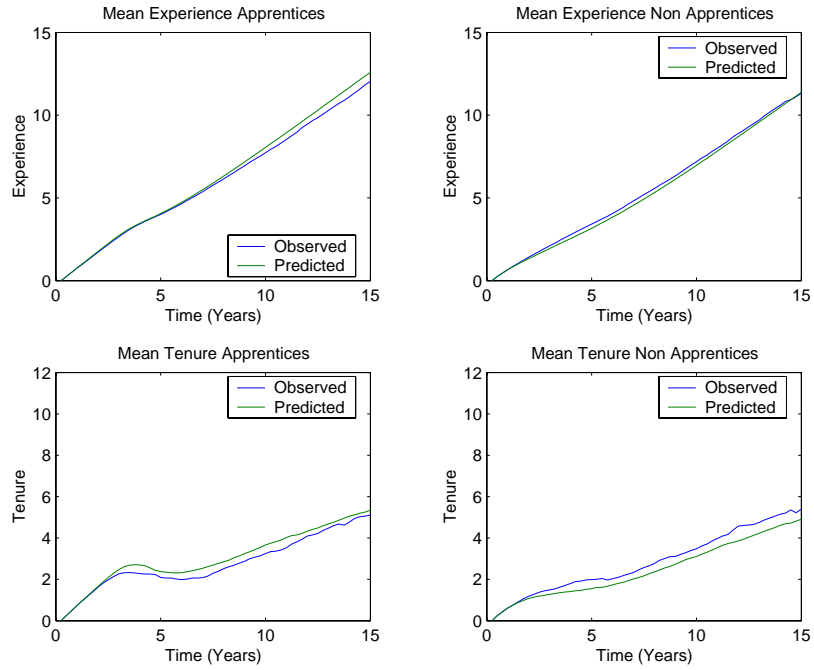|	Apprentices		Non-Apprentices	
	Obs	Pred	Obs	Pred
U to U	0.82	0.83	0.82	0.84
U to E	0.18	0.17	0.18	0.16
E to U	0.04	0.03	0.08	0.04
E to new E	0.03	0.02	0.03	0.02
E to same E	0.93	0.89	0.89	0.94

Table 6: Model fit - Transitions
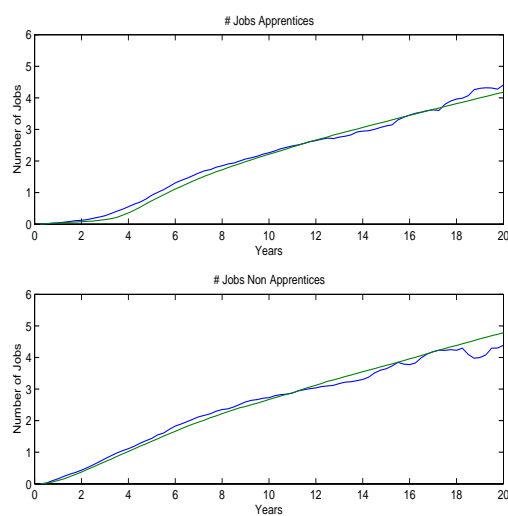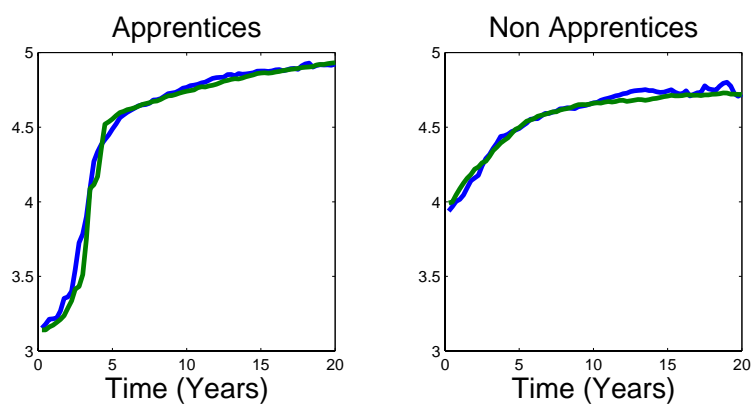
Figure 11:

Figure 12:

Figure 13:

Finally we are able to replicate almost perfectly the average profile of wages for workers as a function of time since first entry on the labor market, including the apprenticeship period (see Figure 13).

38

References

ALTONJI, J., AND R. SHAKOTKO (1987): "Do Wages Rise with Job Seniority?," in *Unemployment, trade unions, and dispute resolution*, ed. by O. Ashenfelter, and K. Hallock, vol. 47, pp. 219–241. International Library of Critical Writings in Economics.

ALTONJI, J., AND N. WILLIAMS (1998): "The Effects of Labor Market Experience, Job Seniority and Mobility on Wage Growth," *Research in Labor Economics*, 17, 233–276.

——— (2005): "Do Wages Rise with Job Seniority? A Reassessment," *Industrial and Labor Relations Review*, pp. 370–397.

BECKER, G. S. (1964): *Human Capital*. NBER.

BEN-PORATH, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75(4), 352–365.

BLUNDELL, R., M. COSTA-DIAS, AND C. MEGHIR (2003): "The Impact of Wage Subsidies on Education and Employment: A General Equilibrium Approach," mimeo, Institute for Fiscal Studies.

BLUNDELL, R., M. COSTA-DIAS, C. MEGHIR, AND J. VAN REENEN (2004): "Evaluating the Employment Impact of a Mandatory Job Search Assistance Program," *Journal of European Economic Association*, 2.

BURDETT, K., AND M. COLES (2003): "Equilibrium Wage-Tenure Contracts," *Econometrica*, 71, 1377–1404.

BURDETT, K., AND D. MORTENSEN (1998): "Wage-Differentials, Employer Size and Unemployment," *International Economic Review*, 39, 257–273.

CAHUC, P., F. POSTEL-VINAY, AND J.-M. ROBIN (2006): "Wage Bargaining with on-the-job search: Theory and Evidence," *Econometrica*, 74(2), 323–364.

CAMERON, S. V., AND J. J. HECKMAN (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, 106(2), 262–333.

CARD, D. (2001): "Estimating the Returns to Schooling: Progress on some Persistent Econometric Problems," *Econometrica*, 69, 1127–1160.

DUSTMANN, C., AND C. MEGHIR (2005): "Wages, experience and seniority," *Review of Economic Studies*, 72(1).

ECKSTEIN, Z., AND K. I. WOLPIN (1989): "Dynamic Labour Force Participation of Married Women and Endogenous Wage Growth," *Review of Economic Studies*, 56(3), 375–390.

HECKMAN, J., H. ICHIMURA, AND P. TODD (1997): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65(2), 261–294.

HECKMAN, J., L. LOCHNER, AND R. COSSA (2003): "Learning-By-Doing Versus On-the-Job Training: Using Variation Induced by the EITC to Distinguish Between Models of Skill Formation," in *Designing Inclusion: Tools to Raise Low-end Pay and Employment in Private Enterprise*, ed. by E.Phelps. Cambridge: Cambridge University Press.

HECKMAN, J., L. LOCHNER, AND C. TABER (1998): "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents," *Review of Economic Dynamics*, 1(1), 1–58.

HECKMAN, J., AND G. SEDLACEC (1985): "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy*, 93(6), 1077–1125.

HECKMAN, J. J. (1993): "Assessing Clinton's Program on Job Training, Workfare and Education in the Workplace," NBER Working Paper, 4428.

HECKMAN, J. J., R. LALONDE, AND J. SMITH (1999): "The economics and econometrics of active labor market," in *Handbook of Labor Economics, vol. 3A*, ed. by O. Ashenfelter, and D. Card. North-Holland.

IMBENS, G., AND J. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

KEANE, M. P., AND K. I. WOLPIN (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

KITTNER, M. (1995): *Arbeits- und Sozialordnung*. 20th edition, Bund Verlag, Koeln.

LOW, H., C. MEGHIR, AND L. PISTAFERRI (2006): "Wage Risk and Employment Risk over the Life Cycle," mimeo, IFS.

POSTEL-VINAY, F., AND J. M. ROBIN (2002): "Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70(6), 2295–350.

POSTEL-VINAY, F., AND H. TURON (2005): "On-the-job Search, Productivity Shocks, and the Individual Earnings Process," mimeo, Bristol University.

SINN, H., C. HOLZNER, W. MEISTER, W. OCHEL, AND M. WERDING (2002): "Aktivierende Sozialhilfe: Ein eg zu mehr Beschaeftigung und Wachstum," Ifo Schnelldienst 2/2006.

——— (2006): "Aktivierende Sozialhilfe 2006: Das Kombilohnmodell des Ifo Instituts," Ifo Schnelldienst 55(9), pp. 3-55.

STEVENS, M. (2004): "Wage-Tenure Contracts in a Frictional Labour Market: Firms' Strategies for Recruitment and Retention," *Review of Economic Studies*, 71(2), 535–551.

TABER, C. (2001): "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?," *Review of Economic Studies*, 68(3), 665–691.

TOPEL, R. (1991): "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," *Journal of Political Economy*, 99(1), 145–176.

TOPEL, R. H., AND M. P. WARD (1992): "Job Mobility and the Careers of Young Men," *Quarterly Journal of Economics*, 107(2), 439–479.

WILLIS, R., AND S. ROSEN (1979): "Education and Self-Selection," *Journal of Political Economy*, 87(5), S7–S36.