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

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

February 24, 2009

Abstract

We model the choice of individuals to follow or not apprenticeship training and their subsequent career. We use German administrative data, which records education, labour market transitions and wages to estimate a dynamic discrete choice model of training choice, employment and wage growth. The model allows for returns to experience and tenure, match specific effects, job mobility and search frictions. We show how apprenticeship training affects labour market careers and we quantify its benefits, relative to the overall costs. We then use our model to show how two welfare reforms change life-cycle decisions and human capital accumulation: 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 realized matches, demonstrating the importance of considering such longer term implications.

1 Introduction

Germany operates an apprenticeship system which consists of formal vocational training courses combined with on-the-job training that lead to certification of skills. Such apprenticeship systems relate both to white collar and blue collar jobs and are subsidized by the state, which funds the classroom component. In contrast, other countries, including the U.S., have no such widespread organized formal system.

Throughout the 1990s, several countries, including the U.S., the U.K., France, and Norway, have attempted to expand or implement new firm-based apprenticeship schemes.[1] Now there is renewed interest in promoting apprenticeships at least in the UK, with the government committed to spending £1.14 billion just in 2009-10 to fund training in an expanding apprenticeship sector.[2] With such policies gaining in popularity, the question is how are the career and wages of a worker affected by participation in a formal apprenticeship and how does it compare to a career with less structured training that one obtains when one starts work following the end of schooling at 16.

To address this question we have at our disposal detailed and accurate administrative data. This allows us to track the careers and wages of individuals from when they make their educational choice and enter the labour market in all German states and for many different cohorts. This is an important advantage of our data over other sources such as the NLSY, which follow one single cohort. Individuals are thus observed deciding between an apprenticeship or a job under quite different local labour market conditions, providing exogenous variation driving this choice. In the descriptive part of the paper we demonstrate that conditional on aggregate time effects and permanent regional effects, there are significant differential changes over time and across regions in educational choices. We use this information to estimate a dynamic discrete choice model of careers and wages, with labour market frictions and to understand better the relative merits of the two different paths. The high quality of the data is an important strength of our approach: all transitions and wages are recorded accurately by the firms avoiding recall

---

[1]see Bowers, Sonnet, and Bardone (1999), and Dustmann and Schoenberg (2008), for an extensive discussion of the German and the UK system and HouseofLords (2007) for some of the debate in the UK.

[2]The Daily Telegraph 19 February 2009

bias.

Careers following an apprenticeship may differ from informal acquisition of skills in a number of ways. First they may increase wages and the return to experience, because of the specialized training they offer. Second they may affect job opportunities through various channels: 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 there is a question on how the specificity of training can affect job reallocation following loss of employment and what this might imply about the ability of an economy to adjust to reallocation shocks (see Heckman (1993)). Thus job arrival rates, layoff rates and the heterogeneity of job opportunities may all differ across the two career paths.

In the model individuals at 16 face the choice of formal apprenticeship or the standard labor market. When working, their wages grow with experience and job specific tenure and depend on a match specific component; thus workers can move to new jobs so as to improve the quality of their job match, subject to receiving an offer. The match specific effect is subject to permanent shocks, which can lead to quits and job mobility. Wages are specific to sector[3] and are subject to aggregate shocks that affect relative wages between the sectors. Underlying choices is a flow utility function that is linear in income and depends on work status.

The model has the key features of an extended Roy model where individuals choose sector by comparing gains and allowing for direct costs of apprenticeship and builds on existing models of education choice[4] and wage determination;[5] it can be viewed as a dynamic extension of Willis and Rosen (1979). We build on some of the key papers on the dynamics of employment and wages. These include Eckstein and Wolpin (1989) who model transitions between employment and unemployment jointly with wages, Wolpin (1992) who estimates a search model of wages and employment and Keane and Wolpin (1997) and Eckstein and Wolpin (1999) who estimate a model of schooling, occupational

---

[3]We distinguish between qualified apprentices versus those without such a qualification, i.e. non-apprentices. These groups constitute our two sectors.

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

[5]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).

choice, labour supply and wages. We bring together many elements of this earlier work, building a model of vocational training, wages, employment and job mobility within a search and matching framework; our model has a rich stochastic structure for wages and allows both for endogenous quits and job destruction as well as search frictions.[6] Our framework, provides us with a way of assessing the value of apprenticeship and to trace the key elements that constitute the difference in the two careers.

The results show that apprenticeships lead to different wage profiles with more growth upfront and none following 10 years of experience, while wages in the non-apprenticeship sector grow at a lower rate but for longer. Overall wages are higher with apprenticeship. Non-apprentices have much higher job destruction rates but also higher job arrival rates, making them easier to reallocate, following shocks; this relates directly to the question of how the apprenticeship system affects flexibility in an economy.

In the final part of the paper we analyze how career choice and human capital accumulation interact with welfare reform. We focus on the effects of introducing tax credits, such as the US EITC, and of making the amount of unemployment insurance fixed and not related to past earnings. The first reform is motivated by the fact that it is in fact debated as a policy option in Germany[7] and is a popular programme in the US and the UK. The second is motivated by the fact that such a reform took place in the UK in the 1980s under the government of Margaret Thatcher. Setting the policy analysis in a lifecycle framework is important for understanding the effects of policy as well as from a normative perspective of designing tax and welfare systems. Similar dynamic effects of policy have been discussed and quantified in Keane and Wolpin (2000) who simulate the lifecycle effects of a wage subsidy, including on education choice and Heckman, Lochner, and Cossa (2003) who consider the impact of Earned Income Tax Credit (EITC), on human capital accumulation. We show that the reforms we consider can have substantial effects on education choices, job mobility and wages. An important limitation of our simulations is that they are partial equilibrium and take the pay structure, as reflected

---

[6]Sullivan (2006) estimates an interesting model of educational and occupational choice, labour market transitions and wages using the NLSY. The specification of his model, nature of the data and empirical focus differ substantively from ours.

[7]See Sinn, Holzner, Meister, Ochel, and Werding (2002) 2006.

in the distribution of offers as given; this is an important avenue for further research.[8]

The remaining part of the paper is structured as follows. In Section 2 we describe the model. Section 3 discusses the identification strategy. Then Section 4 presents the data set and descriptive statistics. In Section 5 we display the estimation results. Finally, Section 6 evaluates the effect of in-work benefits and of reforming the UI system.

2 Model

The model is set in discrete time and focuses on the population that chose the vocational education track at 10 and are completing this form of secondary at 16 years of age; at that point they must choose either to follow an apprenticeship or to enter the labor market as a non-apprentice. In what follows we use the term apprentices or qualified apprentices for those who followed the apprenticeship system and non-apprentices for the rest. To be able to capture the richness of the data without making the model intractable we chose the time period to be a quarter.[9]

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 working at an unskilled labor market wage with working at a lower wage as an apprentice and then obtaining an improved career path through the formal training. We assume that both an unskilled job and an apprenticeship position are available immediately. Utility is linear in earnings making risk and the timing of consumption irrelevant for decision making.

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 an apprenticeship component. All individuals receive job offers with some probability, which may differ depending on whether the worker is employed or not. 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

---

[8]Papers which have included General Equilibrium analysis within the context of dynamic models of heterogeneous agents estimated from Micro data include Heckman, Lochner, and Taber (1998), Lee (2005), Lee and Wolpin (2006).

[9]An individual is deemed to be employed in a quarter if the largest part of the three months was spent working. Otherwise they are non-workers.

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.

Aggregate shocks 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.[10] The macro state variable G_t is modelled as a discrete two state Markov process of order 1. The aggregate trend and the transition probabilities are presented in the Appendix in Table 13. We now describe the model formally and then discuss estimation.

2.1 Payoff flows

Wages and the utility of working. The central component of the model is the job contract. If a worker i and a firm f match at time t, the output is split according to some unspecified rule that yields an annual wage w_{ift} to the worker. In addition, a job provides a *one off* unobserved value μ_{if} to the worker, which can be interpreted as an amenity value of the new firm or a cost of switching to that firm (it can be positive or negative). This allows for the possibility that workers may move to a job that pays lower wages, as is observed in the data.[11] Workers are assumed risk neutral, which also implies that liquidity constraints are not an issue of concern for this model.

One simple way to think about the wage-setting mechanism is Nash bargaining. Worker i and firm f negotiate a wage given match output and job amenities. If the worker happens to meet another firm $\widetilde{f}$ while employed, she compares the two bargaining solutions and takes the best offer. Wage contracts are continuously updated following

---

[10]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.

[11]Given we cannot observe the wage left behind when moving jobs (as we do not observe the shock to the match specific effect) it is possible to rationalize moving to firms that appear to have lower wages by a large negative shock in the previous job, without having to resort to moving costs. However, this turns out to be a restrictive specification that does not fit the data well because of the normality assumption. Thus allowing for this moving cost/amenity adds flexibility to the model and effectively relaxes the distributional assumption.

shocks to match productivity, and, as in a standard Mortensen and Pissarides (1994) model, really bad productivity shocks may result in unemployment.[12]

Wages are specifically modelled as follows. Let $Ed_i \in \{A, NA\}$ denote the worker's apprenticeship qualification status (A for apprentices and NA for non-apprentices). Let X_{it} be the number of quarters spent in work (including the apprenticeship period) since age 16.[13] Let T_{ift} denote the number of years spent in the current job ($T_{ift} = 0$ if the job in firm f 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_{ift} are functions of the macroeconomic shock G_t, education (Ed_i), experience X_{it}, tenure T_{ift}, the unobserved permanent heterogeneity variable ε_i, and a match-specific component κ_{ift}:

$$
\begin{aligned}
\ln w_{ift} \equiv \ln w(Ed_i, G_t, X_{it}, T_{ift}, \kappa_{ift}, \varepsilon_i) = \alpha_0(\varepsilon_i) + \alpha_{Ed}(\varepsilon_i)Ed_i \\
+\alpha_X(X_{it}, Ed_i) + \alpha_T(T_{ift}, Ed_i) + \alpha_G(Ed_i)G_t + \kappa_{ift}
\end{aligned}
\tag{1}
$$

where α_X and α_T are two education-specific functions of experience and tenure. We use a piecewise linear function, with nodes at 0, 2, 4, 6 and 30 years of experience and tenure. Unobserved heterogeneity affects the overall level of log wages and the wage return to apprenticeship.[14] Unobserved heterogeneity allows the wage level and the return to apprenticeship to be heterogeneous in the population, as implied by numerous empirical studies.

When the worker and the firm first meet ($T_{it} = 0$) they draw a match specific effect $\kappa_{ift} = \kappa_{if}^0$ such that

$$
\kappa_{if}^0 \sim \mathcal{N}(0, \sigma_0^2(Ed_i)).
$$

which captures the heterogeneity in wages when individuals start a new job. We interpret this as match specific heterogeneity and we allow it to differ by apprenticship status

---

[12]We discuss briefly the German institutional framework below. Here it suffices to say that within it there is enough flexibility to describe wage setting in this way because Collective bargaining, for the firms where it applies, only sets minimum wages. Of course even in a context which is more regulated than this, the firms and workers can get round these regulations by redefining jobs and promotions.

[13]$X_{i,t+1} = X_{it} + 1$ 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.

[14]In earlier versions of the paper we allowed the returns to experience and tenure to also vary with the unobserved factor ε. However, this did not yield interesting results and we restricted the wage equation to the one presented in 1.

allowing us to estimate the extent to which job opportunities vary in each of the two sectors. Then, whenever $T_{it} \geq 1$,

$$
\begin{aligned}
\kappa_{ift} &= \kappa_{ift-1} + u_{ift}, \\
u_{ift} &\sim iid\mathcal{N}(0, \sigma_u^2(Ed_i)).
\end{aligned}
$$

This allows for the possibility that the value of a match and the contracted wage can change, while allowing for persistence over time. Contrary to the US and the UK, the cross sectional variance of wages does not increase over the lifecycle (figure 13), which means that a random walk of wages that continued across jobs would lead to counterfactual implications and would be inappropriate. This led us to the above specification, where the random walk component is reinitialized when changing jobs, leading to wages that are stationary over the life-cycle, because jobs have a finite expected life.

Employed workers value the current wage $w(Ed_i, G_t, X_{it}, T_{ift}, \kappa_{ift}, \varepsilon_i)$ with a linear utility function. The one off benefit/cost of taking on the job is an iid random variable μ_{if} such that

$$
\mu_{if} \sim \mathcal{N}(m_\mu(Ed_i), \sigma_\mu^2(Ed_i)).
$$

We also allow for a one off cost of quitting a job and becoming unemployed. This transition cost is denoted by $\mu_U \equiv \mu_U(Ed_i)$ and is a deterministic function of education. Its role is discussed in the empirical section.

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 unemployment insurance UI system. 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%.[15] In addition, there is a utility of leisure which varies across individuals on the basis of education, experience, unobserved

---

[15]In Appendix A we describe the details of the German UI system. Here we have taken a replacement rate that is on average correct for our population. Modelling the entire system would imply an increased state space.

heterogeneity ε_i and a Gaussian white noise η_{it} with variance σ_η^2. Thus, the instantaneous utility of unemployment is:

$$R_{it}^U \equiv R^U(Ed_i, X_{it}, w_{i(-1)}, \eta_{it}) = \gamma_U w_{i(-1)} + \gamma_X(X_{it}, Ed_i) + \eta_{it},$$

$$\eta_{it} \sim iid\mathcal{N}(0, \sigma_\eta^2(Ed_i)),$$

with $\gamma_U = 0.55$ and $\gamma_X(X_{it}, Ed_i)$ is an education-specific, piecewise constant function of experience (with nodes at 0, 2, 4, 6 and 30 years of experience).

Finally, we assume that all shocks $\{\kappa_{if}^0,\ u_{ift},\ \mu_{if},\ \eta_{it}\}$ are jointly as well as serially independent, and independent of the unobserved heterogeneity variable ε_i (see below for a complete description of unobserved heterogeneity).

2.2 The intertemporal value functions

Individual decisions to work, to move to a new job or to quit working are carried out by comparing the lifetime values of each of these states. We now describe how they are defined.

The value of unemployment. At the end of period t, unemployed individuals draw a job offer with probability $\pi_{it}^U \equiv \pi^U(G_t, Ed_i, X_{it})$ function of the aggregate shock, education and experience. They can choose to take this job, depending on how the value of working compares to the value of unemployment. The value of unemployment consists of a predetermined part and a stochastic shock η_{it} reflecting changes in the utility of being out of work. Denoting the predetermined part by $U\left(Ed_i, G_t, X_{it}, w_{i(-1)}, \varepsilon_i\right)$, we can write

$$U\left(Ed_i, G_t, X_{it}, w_{i(-1)}, \varepsilon_i\right) = \gamma_U w_{i(-1)} + \gamma_0(\varepsilon_i) + \gamma_X(X_{it}, Ed_i) \quad A$$

$$+ \beta \pi_{it}^U \, \mathbb{E} \max \left(\begin{array}{c} \underline{\mu_{if}} + W\left(Ed_i, \underline{G_{t+1}}, X_{it}, T_{ift+1} = 0, \underline{\kappa_{if}^0}, \varepsilon_i\right) \\ U\left(Ed_i, \underline{G_{t+1}}, X_{it}, w_{i(-1)}, \varepsilon_i\right) + \underline{\eta_{it+1}} \end{array} \right) \quad B \qquad (2)$$

$$+ \beta(1 - \pi_{it}^U) \, \mathbb{E} U\left(Ed_i, \underline{G_{t+1}}, X_{it}, w_{i(-1)}, \varepsilon_i\right) \quad C$$

where we underline the variables over which we are taking expectations (because they are unknown to the individual in period t) and where β is the discount factor.

In 2 the first line of the right hand side (A) represents the within period value of being out of work (up to the stochastic shock η_{it}). This consists of the unemployment insurance

income plus a value for leisure. The lines denoted by (B) represent the expected future value for the case where the worker gets a job offer, which happens with probability π_{it}^U. In that case the worker will choose the best of taking the job offer or continuing as an unemployed worker. The value of taking the job offer is equal to the sum of the present value of the future flow of earnings defined below, $W(\cdot)$, plus a (stochastic) amenity μ_{if}. The final line (C) represents the case where the individual obtains no offer and thus just has to continue out of work.

The value of employment. Employed individuals may be laid off 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 $\pi_{it}^W \equiv \pi^W(G_t, Ed_i)$. 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.

Their value of employment is then given by

$$W\left(Ed_i, G_t, X_{it}, T_{ift}, \kappa_{ift}, \varepsilon_i\right) = w_{it} \qquad A$$

$$+\beta\delta_{it} \, \mathbb{E}\left[U\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, w_{it}\right) + \underline{\eta_{it+1}}\right] \qquad B$$

$$+\beta\left(1-\delta_{it}\right)\pi_{it}^W \, \mathbb{E}\max\left(\begin{array}{c} \mu_U(Ed_i) + U\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, w_{it}, \varepsilon_i\right) + \underline{\eta_{it+1}} \\ W\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, T_{ift}+1, \kappa_{ift} + \underline{u_{ift+1}}, \varepsilon_i\right) \\ \underline{\mu_{i\widetilde{f}}} + W\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, T_{i\widetilde{f}t+1}=0, \underline{\kappa_{i\widetilde{f}}^0}, \varepsilon_i\right) \end{array}\right) \qquad C$$

$$+\beta(1-\delta_{it})(1-\pi_{it}^W) \, \mathbb{E}\max\left(\begin{array}{c} \mu_U(Ed_i) + U\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, w_{it}, \varepsilon_i\right) + \underline{\eta_{it+1}} \\ W\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, T_{ift}+1, \kappa_{ift} + \underline{u_{ift+1}}, \varepsilon_i\right) \end{array}\right) \qquad D$$

$$(3)$$

The current value of work is just the wages w_{it}. Following job destruction, which occurs with probability δ_{it} the individual will receive the value of unemployment as shown in line B. The group of lines marked C represent the events when the job is not destroyed and

the individual obtains an alternative job offer. In this case they have to choose between becoming unemployed, in which case they incur the exogenous one-off cost $\mu_U(Ed_i)$; remaining with the firm; or taking the alternative offer, which is associated with the one off random switching cost $\underline{\mu_{i\tilde{f}}}$ of joining a new firm $\tilde{f}$. The following group of lines marked by D represent the expected value of a worker not being laid off and not having access to an alternative offer. Given that a shock can occur to the match specific effect, the worker may decide it is best to quit, in which case they receive the value of unemployment. Otherwise they receive the value of working with the same firm, at the updated wage.

The value of employment while in training. Going back, earlier into the individual's history, we consider choices available when training. During apprenticeship (which lasts τ_A periods[16]) we assume that the training firm pays the worker only a fraction λ_A of his productivity as a non-apprentice ($w(Ed_i = NA, G_t, X_{it}, T_{it}, \kappa_{it}, \varepsilon_i)$), the rest presumably serving as payment for the general training received.[17] Reflecting the facts in the data, we do not allow the individual to experience unemployment during apprenticeship, although they can decide to change firm if the opportunity arises. Thus, during the apprenticeship training period ($X_{it} < \tau^A$) the value of work is:

$$W^A\left(G_t, X_{it}, T_{ift}, \kappa_{ift}, \varepsilon_i\right) = \lambda_A \cdot w\left(Ed_i = NA, G_t, X_{it}, T_{ift}, \kappa_{ift}, \varepsilon_i\right) \quad A$$

$$+\beta\pi_A(G_t)\,\mathbb{E}\max\left(\begin{array}{c} W^A\left(\underline{G_{t+1}}, X_{it}+1, T_{ift}+1, \kappa_{ift}+\underline{u_{ift+1}}, \varepsilon_i\right) \\ \underline{\mu_{i\tilde{f}}} + W^A\left(G_{t+1}, X_{it+1}, T_{i\tilde{f}t+1}=0, \underline{\kappa^0_{i\tilde{f}}}, \varepsilon_i\right) \end{array}\right) \quad B \qquad (4)$$

$$+\beta[1-\pi_A(G_t)]\,\mathbb{E}W^A\left(\underline{G_{t+1}}, X_{it}+1, T_{ift+1}, \kappa_{ift}+\underline{u_{ift+1}}, \varepsilon_i\right) \quad C$$

where and where the expectation operator $\mathbb{E}$ relates, as before, to the underlined variables, which are unknown to the individual in period t.

Similarly to the value of working described above, the first line (A) is earnings while training, (B) represents the part of the value due to the possibility of changing training firms if an offer arrives (with probability π_A). As before there is a mobility cost asso-

---

[16]Apprenticeship courses last between two and three years. We equate τ_A to whatever is the actual duration in the data.

[17]In actual fact this is only part payment towards the general training: at least the classroom component is funded by the government.

ciated with the decision to join the alternative firm $\widetilde{f}$. Finally, line (C) represents the continuation value for the case where no alternative training firm is available.

While in the last period of apprenticeship the value function becomes as in equation (3) with all options available. However in this case if the worker qualifies and remains in the firm that trained him we observe a wage which is an average of the apprenticeship and fully qualified wage: the data only records compensation over the whole calendar year and does not distinguish between pre-qualification and qualification status. Thus, in effect neither wage is observed and must be integrated out.

The ex ante value of apprenticeship. The choice to follow an apprenticeship training is assumed to be a one off decision made at age 16 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 (3) 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 (4) net of direct monetary and utility costs. This is expressed as

$$V^A\left(G_t, \kappa_{if}^0, \mu_{if}, R_i, \varepsilon_i, \omega_{it}\right) = \mu_{if} + W^A\left(G_t, X_{it} = 0, T_{ift} = 0, \kappa_{if}^0, \varepsilon_i\right) - \lambda_0(R_i, G_t, \varepsilon_i) - \omega_{it} \tag{5}$$

where R_i denotes the region the individual lives when 16.

The last two terms represent costs. The first, $\lambda_0(R_i, G_t, \varepsilon_i)$, 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 3.1. The second term, ω_{it}, is a normally distributed iid cost shock revealed to the individual before the choice is made. The choice to become an apprentice is thus governed by

$$V^A\left(G_t, \kappa_{if}^0, \mu_{if}, R_i, \varepsilon_i, \omega_{it}\right) > W(Ed_i = NA, G_t, X_{it} = 0, T_{it} = 0, \kappa_{if'}^0, \mu_{if'}, \varepsilon_i), \tag{6}$$

where κ_{if}^0, μ_{if} and $\kappa_{if'}^0$, $\mu_{if'}$ represent the match specific characteristics and one off transition costs in the initial jobs in the alternative careers. The cost shock ω_{it} induces a probability for this choice, conditional on all the other shocks, from which it is independent. These, including the match specific effects in both alternatives and the non-

12

pecuniary benefits, need to be integrated out. we allow for unobserved heterogeneity in the costs to capture the possibility that individuals may differ in their ability to train; as we will discuss below ε_i will contain two factors: one for labour market ability and one for training.

The time horizon and the terminal condition We solve the model by iterating on the Bellman equations backwards from retirement which occurs after 50 years in the labour market. At retirement the value is assigned to zero: in a linear utility framework, such as ours, this is equivalent to assuming that individuals finance retirement through their own savings out of their wages.[18] Having a terminal point beyond our observation window requires assumptions on the returns to experience and tenure. We have thus imposed that the returns to experience and tenure are constant between 10 and 30 years of experience, extrapolating from our data which stops at 20 years of experience.[19] We then assume that wage growth due to experience (and tenure) stops after 30 years of experience (and tenure respectively). The gain from this tight specification is that we avoid having to use a separately parameterized terminal value function. In general experience and age are sources of nonstationarity, the latter because of the finite life nature of the problem. The match specific random walk and tenure are not sources of nonstationarity because the jobs have finite (and relatively low) expected lifetime, although we still need to take into account of an increasing variance within jobs. However it turns out that, given our assumptions above and the fact that the oldest individual in the data is aged 35, the value functions and the simulation results were not sensitive to age, below 35. We have thus used the value function at age 35 for younger ages, simplifying the problem and reducing the state space. More on the computation of the value functions is in the appendix.

2.3 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

---

[18]Note that the model uses gross wages, before any pension contributions.

[19]This turns out to be an annual rate of zero for apprentices and 1.8% for non-apprentices. The returns to tenure turn out to be always zero.

follow a bivariate discrete distribution, each with two points of support. One element enters the cost of apprenticeship while the other enters the wage equation and affects the constant and the returns to apprenticeship. The two elements may be positively or negatively correlated or possibly not at all.[20] Education choice depends on the costs of education (observed or not) and on the expected wage gains. Hence this specification allows both for selection on unobserved returns to education and for ability bias as expressed in the labour literature.[21]

3 Identification and Estimation

We shall use maximum likelihood to estimate the parameters of the model. In this section we address some identification issues and we briefly describe how to write the likelihood, a complete derivation of the sample likelihood being provided in Appendix E.

3.1 The identification strategy

By modelling the entire sequence of choices, including the initial allocation to the vocational track and the choice to follow an apprenticeship or not, we take account of their endogeneity using the restrictions implied by economic theory and the structure of the model. However the availability of many cohorts observed in all German states gives us further identifying power. We are able to exploit the fact when each cohort in each state comes to the point of making the apprenticeship choice, it is faced with a different economic environment because each region is sensitive to different shocks depending on its industrial composition. Local demand for apprenticeships thus varies differentially in response to demand shocks affecting the cost of obtaining apprenticeship training: if plenty of positions are available in the region of residence at 16, 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 possibly live away from home to obtain apprenticeship, incurring greater costs. In addition we need to assume that the labour market is sufficiently integrated over the country that individuals not

---

[20]In practice we normalize one point of support to be zero and include a constant in the wage of each sector and in the costs of apprenticeship.

[21]See for example Griliches (1971), Card (2001), Heckman and Vytlacil (2005) and Carneiro, Heckman, and Vytlacil (2006) among many others.

14

	Birth Cohorts		
	1960	1965	1970
Academic Track	20%	21%	24%
Apprentices	64%	67%	65%
Non Apprentices	16%	12%	11%

Table 1: Proportion in different education tracks by Year of birth

involved in training search over a broad enough area so that the local shocks do not affect local wages.

To allow for these considerations we specify the direct costs of apprenticeship as functions of region of residence at the time of apprenticeship (R_i), the business cycle as well as an unobserved component ε_i.[22] The initial region of residence is taken as exogenous. We exclude region and region interacted with the business cycle from wages and preferences for work.[23] The availability of 12 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.

3.2   Initial conditions

The population whose labour market behavior we model consists of all individuals who at 10 years of age are allocated to the vocational school track, rather than the academic one. This choice is likely to depend on individual unobserved characteristics as well as the economic environment at the time and involves both parental choice and the educational authorities. As shown in Table 1, there is a steady (but small) decline in the proportion following the vocational track over time (apprentices and non-apprentices in the table).

To resolve this initial conditions problem we specify a reduced form probability of choosing the vocational versus the academic track P_i^S as a function of the region and year of birth of the individual (reflecting the economic conditions at the time) as well

---

[22]More generically, we could have used an output price index by region as the factor driving costs. We approximate this by using the business cycle indicator interacted with region.

[23]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.

as of the two factors of unobserved heterogeneity in the vector ε_i. We then specify the joint likelihood of selection into the vocational track together with all the subsequent choices and realized wages for those who followed the vocational track; the likelihood for those who did not is just the reduced form probability of this event. All probabilities and densities depend on ε_i. In practice this means that for all those who followed the vocational track and are thus in our sample, the likelihood contribution, discussed in the next subsection, is multiplied by P_i^S; for those not in the vocational track the likelihood contribution is $1 - P_i^s$. Unobserved heterogeneity, which then needs to be integrated out, accounts for the dependence of the initial education choice and the subsequent education and career path.

The key assumption in this approach is that the distribution of unobserved heterogeneity is independent of region and cohort.

3.3   The likelihood function

The likelihood function is derived in Appendix E and the computation of the value function is discussed in Appendix E.1. Here we offer a brief outline. The likelihood contribution of an individual conditional on the unobservable characteristics ε_i is the joint probability of all observed events and of observed wage growth within the firm (density) as well as the observed initial wage at the start of the job. The discrete events include the initial selection into the vocational track, the choice of apprenticeship or not, moving in or out of work, remaining unemployed and remaining in the same firm or moving firm.

To construct the probability of the events involves solving the model conditional on permanent exogenous characteristics, including ε_i and all other state variables. These 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.

Once the model is solved a number of unobservables need to be integrated out of each probability, which we do either analytically, where possible, or by using Gaussian quadra-

ture. 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.

4   The Data Set

We draw a sample from a data set organized by the German IAB[24] and which in its totality consists of a 2% extract from the German social security records. The data set starts in 1975 and records all work spells with exact start and end dates up to 1996. The data records spells of apprenticeship training and whether a worker holds an apprenticeship qualification or not as well as their overall educational qualifications. 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. Moreover, to avoid the initial conditions problem we need to model the initial education choice at age 10. But we can only infer who has made this choice once we see individuals in the labor market; then we see their educational qualifications and we can allocate them accordingly. Individuals who follow the academic track typically enter the labor market later. Hence to be sure we observe the entire cohort, whatever education choice they made, we must only use those cohorts who are old enough to be observed at age 25 years of age or older. Given our observation window this means that our population are those men born in the period 1960-1972. The length of period includes sufficient changes in the aggregate environment of each German state to provide the required variation for identifying education choice.

The data set reports the average daily pre-tax wage each year if the individual stays

---

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

with a firm for an entire year. For individuals who move jobs we observe as many wages as firms they worked in during the year. Thus wages are not averaged across different firms. In the model we use wages that have been detrended using a quadratic trend common across the two groups as described in the appendix and in Table 12

The data is far too detailed and would be intractable to model in all fine details. We thus time-aggregate the data to obtain information on a quarterly basis. Whenever during a quarter an employment and an unemployment spell are both present we assign to one of these depending on which of the two covers the largest proportion of that quarter.

Our main sample and focus of study consists of West-German males, who end formal education at 15/16 and who either work or join an apprenticeship after school. However, individuals who are not in this group are kept so as to model the initial choice at 10 to follow or not the vocational track. The overall data set is too large, given the time it takes to compute the likelihood. We thus use a random subsample drawn from the total population of cohorts born in the period 1960-72. This subsample contains 3371 individuals in the vocational track. These are followed through time, quarter after quarter up until 1996. 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.[25]

4.1 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 just below 25%, resulting in a 1.2% real average growth per year. Dur-

---

[25]The Social security data is in principle top coded. However, this does not affect individuals in our sample, whose pay is noth high enough.

ing apprenticeship training workers are paid a very low wage, thus presumably covering the cost of their apprenticeship with the remaining output they produce during on-the-job training.[26] 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. 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 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 2 displays the quarterly transition probabilities by education 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 six percent of employed skilled workers exit, while this figure is about 14% 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 is higher at the beginning for non-apprentices and after five years declines for both groups and becomes marginally higher for the qualified apprentices.

Qualified apprentices with 5-10 years of potential experience have a higher probability of return to work from unemployment, by about 3 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 13.4 years, compared with a total of 15.3 years for skilled workers.

Figure 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

---

[26]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 offer this is not an unreasonable interpretation.

	Non-Apprentices			Apprentices		
Potential Experience (Years)	0-5	5-10	10-20	0-5	5-10	10-20
Out of work to Out of work	.84	.89	.93	.83	.86	.9
Out of Work to Work	.16	.11	.071	.17	.14	.070
Work to out of Work	.14	.073	.046	.063	.051	.023
Work to new Work	.045	.034	.022	.035	.038	.024
Work to same Work	.82	.89	.93	.91	.91	.95

Table 2: Observed Quarterly Labor Market Transitions

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.

Figure 1: Log Wage by skill and the wage gain for qualified apprentices
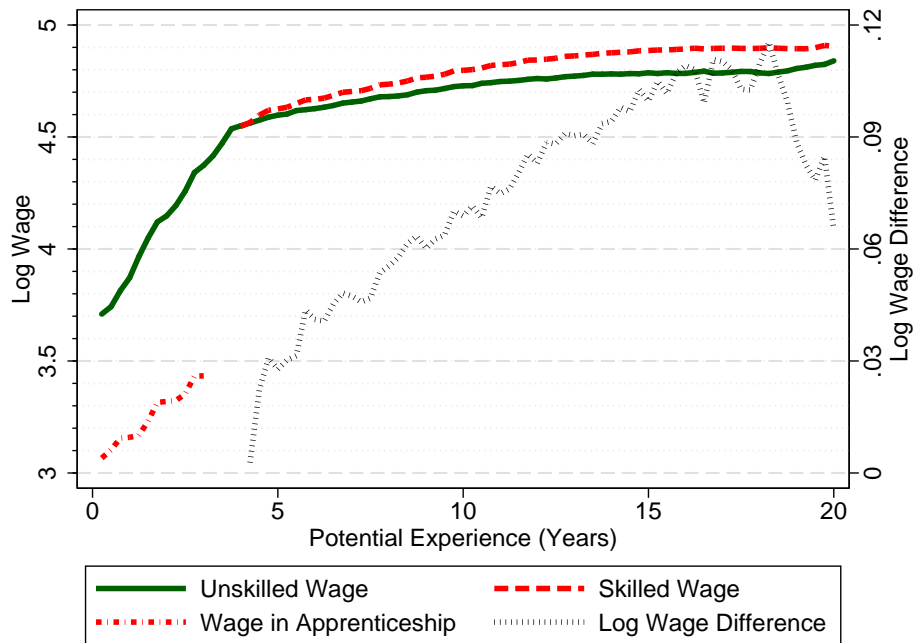
Decomposing Wage Growth. Wage growth occurs both within firm and as a result of firm mobility. Job shopping, can be a very important source of wage growth as documented in Topel and Ward (1992) and can be crucial in achieving efficient matches (see Heckman (1993)).

In Germany, despite lower mobility rates, this is also the case. This is illustrated in Figure 4 which shows within firm wage growth by potential experience and skill level and

Figure 2: Mobility: Number of Jobs, by Education
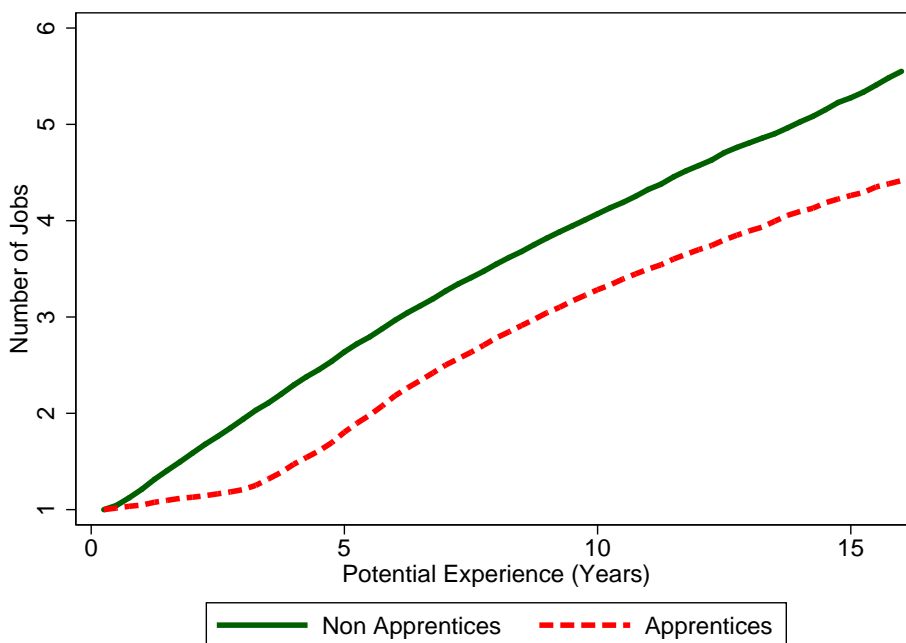
in Figure 3, which displays the growth of wages following a job to job transition. The wage growth in the latter case can be substantial, at nearly 40% for non-apprentices and for qualified apprentices (post training). The gain in wages falls over time, decreasing towards zero. If we think of wage improvements as being due to better matches, as in our model, the decline is expected because the probability of an improvement will decline 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 reflecting 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.

Finally, a few words on the institutional framework: Germany operates a collective bargaining system at the industry level. Agreed wages within this system act as minimum

wages and firms may and do pay wages above the union wage; there is no restriction on paying workers more according to merit (productivity). Union agreements are binding in firms that belong to an employer federation (Arbeitgeberverband), which constitute about 62% of employers and 83% of the workforce. Thus we can think of the German labour market as one where a negotiated minimum wage operates for many firms, with no upwards restrictions and where there is a competitive fringe with no restrictions at all. The presence of minimum wages will be reflected in our model in increased proportions out of work.[27]

Figure 3: Between job wage changes
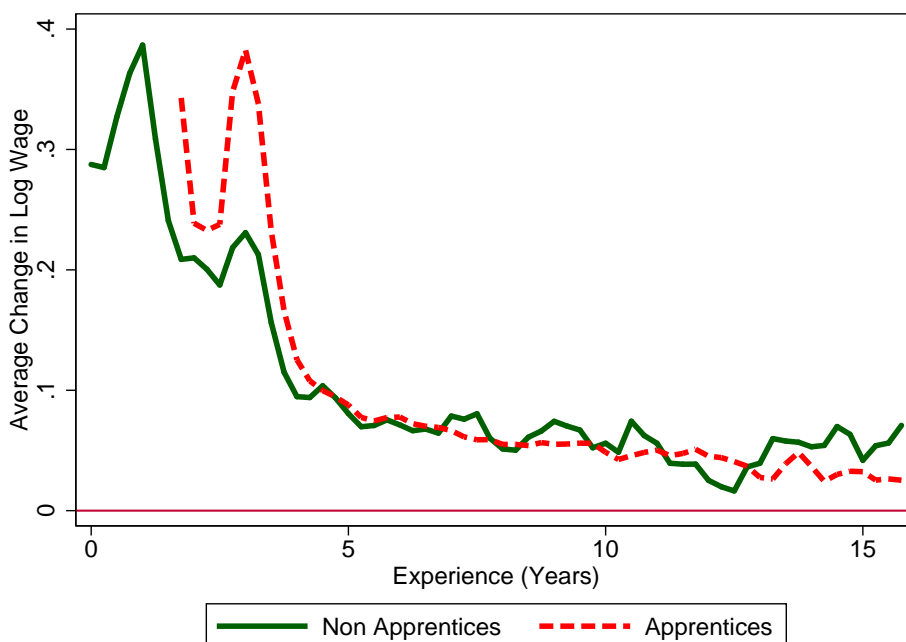
Vocational Training and Wages. Given the exogenous variation determining apprenticeship, as described earlier in section 3.1, we can follow an instrumental variables approach to estimate the effect of apprenticeship on wages, ignoring here selection effects due to participation. This is done mainly as a descriptive device and to illustrate what would be obtained using the 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

[27]For more details on the German institutional system see Dustmann and Schoenberg (2009)

Figure 4: Annual Change in Log Wage (Conditional on Staying with same Employer
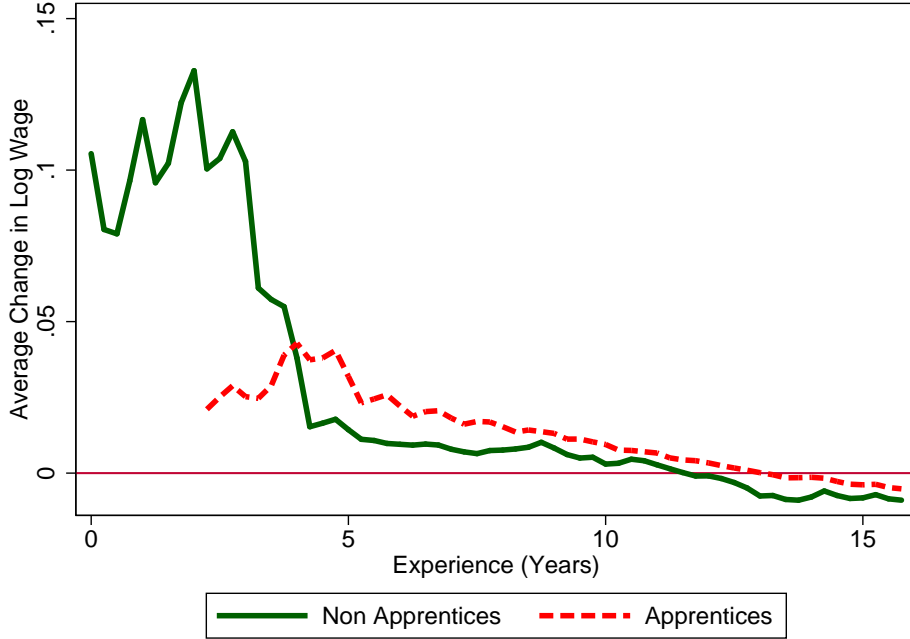
that indeed there is sufficient differential variation of apprenticeship participation, which we attribute to changing availability of positions and costs.

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.[28] 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.[29]

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

---

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

[29](see for example Blundell, Duncan, and Meghir (1998))

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. 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.

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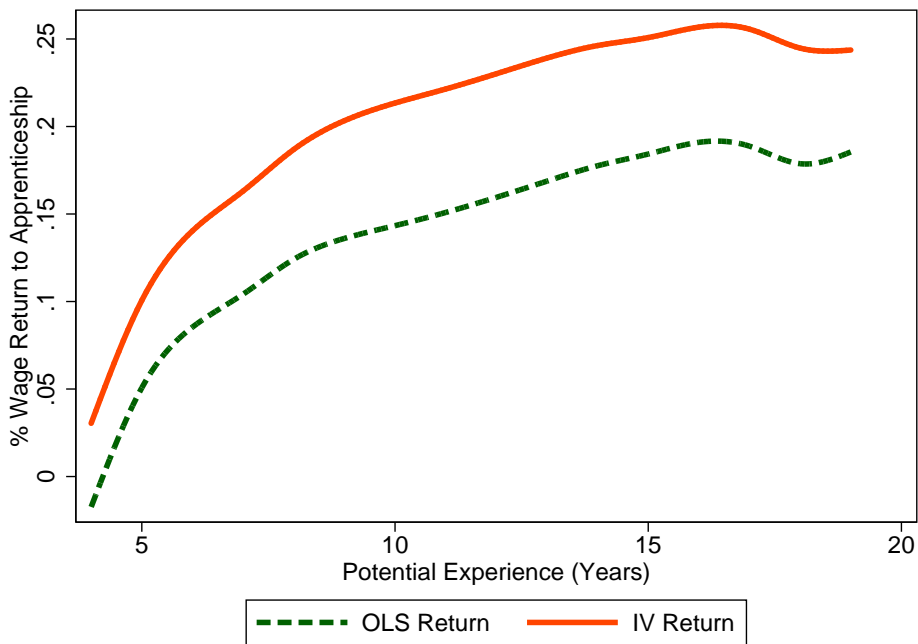
5   Estimation Results

5.1   The Fit of the Model

We evaluate the fit of the model by simulating the education decisions, the labor market transitions and the wages for a cohort of individuals over time and comparing to the actual data. The model fits remarkably well and we refer the reader to Appendix D where the results are shown in some detail in Table 14 for the labor market transitions and in Figures 10, 11 and 12 for experience and tenure profiles, the number of different jobs held and wages respectively. These graphs also serve as data description for these key aspects

of labor market careers. In Figure 13 we also show the fit of the standard deviation of wages over the life-cycle. This is interesting for the different pattern it displays to the one known for the U.S. (see Low, Meghir, and Pistaferri (2006)) where the variance is increasing over the lifecycle. In Germany this declines after a rapid increase for the young and then remains constant. This justifies our specification for the stochastic structure of wages, where the match specific shocks are not carried over to the new jobs, making them effectively transitory. Indeed, our model, while not fitting perfectly the standard deviation is very successful in capturing the broad pattern.

5.2   The Parameter Estimates

Transition probabilities and costs.   Table 3 presents some key parameters that determine the careers of individuals. Exogenous quarterly destruction rates, i.e. excluding quits, are 0.02 and 0.03 a quarter for the two groups in Germany for workers who have worked over 6 years.[30] For the less experienced workers the job exogenous destruction rates are higher, but this partly reflects departures for military service, that are not explicitly observed, but occur in the early part of one's career.

The job arrival rates are allowed to vary by business cycle; we find that the arrival rate for unemployed non-apprentices is quite cyclical and in all cases it is much higher than the one for the skilled apprentices. In other words reallocating skilled apprentices seems to be much harder than for non-apprentices, indicating inflexibility possibly due to relatively narrow training they receive. When employed, the non-apprentices also receive many more alternative job offers. For the non-apprentices the arrival rate is quite pro-cyclical.

Thus among the German workers the non-apprentices loose their jobs more frequently but seem to have more job market opportunities when unemployed. All this suggests that apprentices have very stable employment, but that when this is lost they have greater difficult finding a new position: apprenticeship training may reduce mobility and may thus restrict the ability of the economy to accommodate reallocation shocks.

In the lower part of Table 3 we report the parameters driving the (stochastic) mobility

---

[30]We do not want to compare the destruction rates for the less experienced because they involve departures for military service.

Parameter	In Appren-ticeship	Qualified Apprentices	Non-Apprentices
		Job Offers and Job Destruction Rates	
Quarterly job destruction rate (δ)			
if experience $\leq$ 4 years	-	0.106	0.16
		(0.0029)	(0.0059)
if experience $\in$ [4,6] years	-	0.059	0.089
		(0.0016)	(0.0036)
if experience $>$ 6 years	-	0.022	0.031
		(0.00062)	(0.0023)
Quarterly offer arrival rate when employed (π_W)			
if business cycle low	0.0149	0.116	0.228
	(0.0012)	(0.0037)	(0.024)
if business cycle high	0.0188	0.12	0.313
	(0.0013)	(0.0037)	(0.033)
Quarterly offer arrival rate when unemployed (π_U)			
if business cycle low, experience=0	-	0.229	0.746
		(0.0084)	(0.054)
if business cycle high, experience=0	-	0.247	1
		(0.009)	(0.098)
if business cycle low, experience=10	-	0.358	1
		(0.018)	(0)
if business cycle high, experience=10	-	0.377	1
		(0.018)	(0.11)
Std dev of utility shocks to unemployment[a] (σ_η)	-	1.1%	1.1%
		(0.034)	(0.034)
Mean of mobility cost to unemployment[a] (μ_η)		-4.1	-3.81
		(0.13)	(0.26)
Mean of mobility cost[a] (m_μ)	1.1	1.1	-0.726
	(0.11)	(0.11)	(0.2)
Std dev of mobility cost[a] (σ_μ)	2%	2%	2%
	(0.043)	(0.043)	(0.043)
Utility of leisure[a] (γ_0)			
if experience $\leq$ 4 years	0.84%	0.84%	0.71%
	(0.015)	(0.015)	(0.048)
if experience $>$ 4 years	-0.37%	-0.37%	-0.25%
	(0.021)	(0.021)	(0.074)

[a]Percent of lifetime value, which is 8206.49. Likelihood: -74986.1935. Asymptotic standard Errors in parentheses

Table 3: Estimated parameters: Variance of shocks, Job destruction and job arrival rates and mobility costs

benefits or costs towards other jobs and to unemployment. For skilled individuals, the mean non-wage benefit of moving (m_μ) is 1.1% but is highly dispersed with a standard deviation (σ_μ) of 2%; Thus, on average, qualified apprentices obtain a non-wage benefit by moving equivalent to 1.1% of lifetime value and thus may be willing to move for a pay cut. In contrast, the non-apprentices on average move for a pay rise only as they face a large cost of moving of the order of 0.73% of life time value.

We have also allowed for a fixed cost of quitting work to unemployment. Without the cost of transition to unemployment the composition of those becoming unemployed was such that the average wage of employed individuals was predicted to be too high, implying that the productivity composition of the unemployed was lower than it should be. The need to allow for such a cost may reflect risk aversion, where the risk of a long unemployment spell would prevent low productivity individuals from quitting. It thus turns out that the cost of quitting is 4.1% of lifetime value for the qualified apprentices and 3.8% for the non-apprentices. Once all is accounted for the realized quit rate accounts for about 25% the flow from work into unemployment. The parameter estimates imply that over a 20 year period apprentices are expected to spend 25% of quarters out of work, while non-apprentices 35%. In Appendix Table 14 we compare the fitted transitions to those from the data.

Wage equation. Table 4 reports the parameter estimates for the wage equation. 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 heterogeneity (σ_0). For the qualified apprentices the standard deviation of the innovation is 0.035, while for the non-apprentices this is 0.038. Converting those to the variance of the growth of average annual wages we would obtain standard deviations of approximately 0.060 and 0.066 respectively. During apprenticeship the standard deviation is much larger. However, in money terms the fluctuations represent small changes because of the low salaries.

Perhaps one of the most striking result here is the standard deviation of initial match heterogeneity, which is estimated to be 0.28 for the qualified apprentices and 0.42 for the non-apprentices. This implies that the pay accompanying job offers can range ±56% and

27

±84% for the two groups respectively. Thus in Germany there is considerable heterogeneity in job matches and hence great opportunities for wage growth from job shopping.

Below the variances of the shocks we report the constant in the wage equations, the effect of the business cycle, the returns to experience and tenure. At each experience node (2,4,6,10,30 years) we report the accumulated wage growth by that level of experience. In between the nodes wages are linear in experience.[31] Similarly for tenure. For the apprentices experience (and tenure) starts counting at the start of training and the first two years (for some 3) are all spent in training: the estimated returns at two years thus refer to the growth of wages by the end of training. The returns to experience thereafter refer to the period following qualification.

Apprenticeship choice is driven partly by the opportunity cost of apprenticeship. The log difference in the wages between those starting apprenticeships and those staring regular work without such training is 0.67 as can be seen by comparing the intercepts of the non apprentices (3.55) and the apprentices during qualification (2.88) in Table 4. As we shall see the high opportunity cost of training will be a central factor driving the gains to apprenticeship. When training ends the wage equation intercept rises by 0.76 log points over and above the increased wage growth due to more rapid returns to experience.

The returns to experience for the apprenticeship sector are substantial at the start and flatten out by 10 years of experience. During the first two years wages for apprentices grow by about 32%. The incremental effects of experience after two years is smaller at 4.5% a year for the next two years, declining to 3% and then to 1.7% and finally to 0.1% between 10 and 30 years of experience. For the non-apprentices the experience profile is less concave, with returns after 10 years of experience of about 1.8% annually. This reflects the more gradual learning experience in the standard jobs. The returns to tenure are zero - specific human capital is not reflected in wages.

Finally, the German business cycle has a very small effect on relative wages for the two groups. This is of the order of 0.8% between good and bad times for both groups implying that the relative price of the human capital in the two groups is more or less

---

[31]Our data stops at 20 years of experience; beyond that we extrapolate linearly. The returns over this period are driven by wage growth between 10 and 20 years of experience.

Parameter	In Apprenticeship	Qualified Apprentices	Non-Apprentices
Std dev innovation to match specific effect (σ_u)	0.11	0.0354	0.0376
	(0.00015)	(4.2e-005)	(0.00012)
Std dev of match specific wage offers (σ_0)	0.28	0.284	0.419
	(0.0039)	(0.0016)	(0.0056)
Log Wage Constant	2.88	3.64	3.55
	(0.0082)	(0.0086)	(0.015)
Effect of high business cycle	0.008		0.00754
	(0.00027)		(0.00077)
Experience=2 yrs	0.32		0.1
	(0.0065)		(0.0078)
Experience=4 yrs	0.41		0.16
	(0.0064)		(0.013)
Experience=6 yrs	0.47		0.18
	(0.0069)		(0.015)
Experience=10 yrs	0.54		0.22
	(0.0083)		(0.022)
Experience=30 yrs	0.55		0.64
	(0.028)		(0.088)
Tenure=2 yrs	0.000		0.000
	(0.0024)		(0.0075)
Tenure=4 yrs	0.000		0.000
	(0.0045)		(0.01)
Tenure=6 yrs	0.017		0.000
	(0.0066)		(0.016)
Tenure=30 yrs	0.017		0.000
	(0.029)		(0.1)

Asymptotic standard Errors in parentheses.

Table 4: The Wage Equation and the shocks to wages

constant, which is consistent with the two inputs being perfect substitutes in production.

Job mobility and wage growth. In Figure 6 we plot the cumulative contribution of job mobility to wage growth. This is obtained by simulating wage profiles disallowing any direct job to job changes and comparing to the profiles we obtain with the full model.[32] For those in apprenticeship job mobility contributes mainly when individuals move from the training firm to a new one leading to a small wage growth of about 4%. For the non-apprentices, the "return" to mobility peaks at more than 10% with a slight decline thereafter to 9%.[33] Thus mobility is substantially more important for the non-apprentices and, perhaps as expected, more important for the younger individuals.

---

[32]In the experiment, individuals still change jobs following unemployment spells
[33]See Topel and Ward (1992) for results in the US.

Figure 6: The contribution of job mobility to wage growth
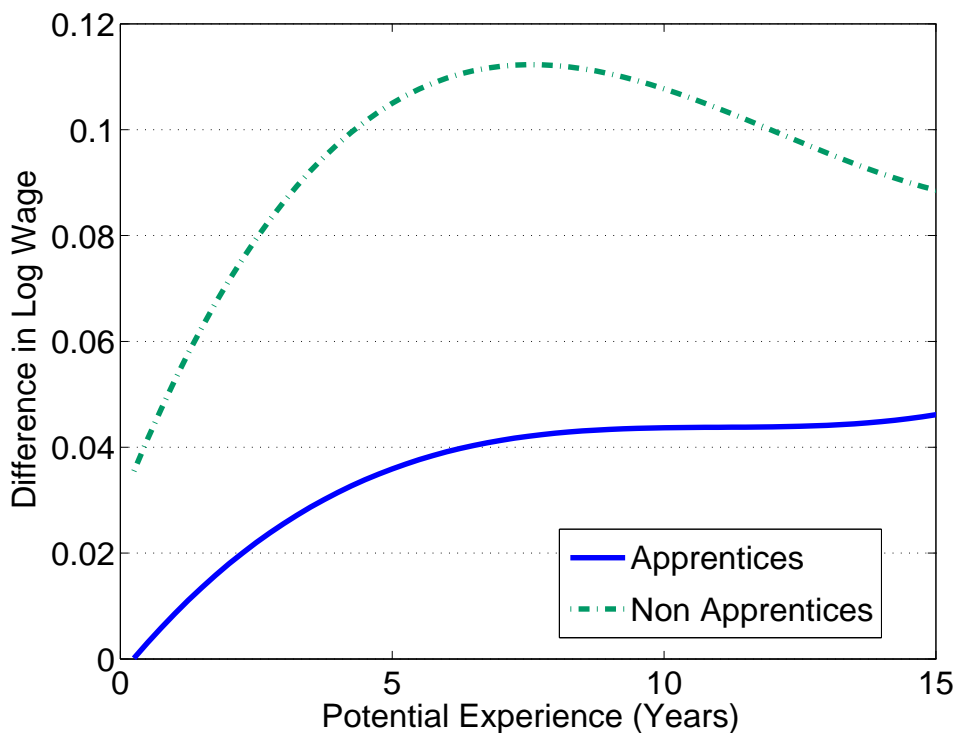
Unobserved heterogeneity and initial conditions. The model allows for two factors of unobserved heterogeneity; one factor affects the level of wages and the return to apprenticeship and another factor affects the costs of apprenticeship. Both factors enter the probability of choosing the vocational rather than the academic track at 12 years of age (the initial conditions equation). 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 18%, 14%, 64% and 3.9%. Table 5 displays summary characteristics for these groups. Individuals of Type 1 and Type 2 have a low wage, whereas Type 3 and Type 4 high wages.[34] Both Type 1 and 3 individuals have a lower cost of choosing apprenticeship equivalent to about 6% of life time value. Overall there is a positive association between having a low cost of apprenticeship and having a high wage: among those with low cost of apprenticeship (Types 1 and 3) the probability of being a high wage type is 78%; among individuals with a high cost of apprenticeship the probability is only 14%. This is reflected in the correlation between types of -0.46. Interestingly individuals with a high wage type (3 and 4) have 3% lower wage return to apprenticeship (wage constant

---

[34]The points of support are reported over and above the constant in the relevant equations, which explains why one is reported as zero.

0.31 versus 0.34). From the same table we see that there is a substantial effect of direct or utility costs of apprenticeship on the proportion taking it up. The cost effect is lower among low wage types. Moreover the effect of the wage type is positive among the low cost individuals, but negative among the high cost ones. This demonstrates that for the high cost types opportunity cost dominates their career choice.

In the final two rows of the table we report the coefficients on the two unobserved factors in the model for the selection into our sample. As a reminder for the reader, our sample consists of those who at 10 were allocated to the vocational track of the German school system. This is potentially an endogenously selected group with changing composition over time.[35] We model this initial conditions problem by introducing a reduced form probit selection equation, which depends on cohort×state effects and on the two unobserved factors. We find that participation in our sample (vocational schooling after 10) is negatively associated with the cost of training (-0.613) but positively associated with labor market ability (0.786). This is very much in line with a simple comparative advantage story. However, the factor loading on the apprenticeship cost is not significant (-0.613 with standard error 0.517), implying that the main source of sorting is the wage, with higher wage people, opting for the vocational track, which also leads to less years of education and faster entry into the labour market.

5.3   The Value of Apprenticeship

The natural approach to measuring the gains from apprenticeship is to consider the lifetime value of following that career type viewed from the point where the first choice is made; this takes into account all costs faced by the individual and all differences associated with the two paths. Thus, the overall proportional gain from apprenticeship for a given type of individual is given by

$$r = \mathbb{E}\left[\frac{V^A\left(G_t, \kappa_{if}^0, \mu_{if}, R_i, \varepsilon_i, \omega_{it}\right)}{W(Ed = NA, G, X = 0, T = 0, \kappa, \mu, \varepsilon)} - 1\right]$$

where the numerator is the discounted value of having an apprenticeship qualification as seen at the time of making the original career choice and is defined in (5), while the denominator is the equivalent value of not obtaining an apprenticeship. The gain is

---

[35]See Table 1.

Parameter	Type 1	Type 2	Type 3	Type 4
Proportion in sample (π_j)	0.18	0.14	0.64	0.039
	(0.27)	(0.32)	(0.27)	(0.21)
Proportion in Apprenticeship	0.91	0.78	0.88	0.71
Log wage constant Apprentices ($\alpha_0(\epsilon)$)	0	0	0.31	0.31
			(0.0067)	(0.0067)
Log wage constant Non Apprentices ($\alpha_0(\epsilon)$)	0	0	0.34	0.34
			(0.016)	(0.016)
Utility cost of apprenticeship as % of life time value ($\lambda_0(\epsilon)$)[a]	-6.1	0	-6.1	0
	(0.076)		(0.076)	
Factor loading for selection into vocational track (wage heterogeneity)	0	0	0.786	0.786
			(0.017)	(0.017)
Factor loading for selection into vocational track (cost of apprenticeship)	-0.613	0	-0.613	0
	(0.578)		(0.578)	
Correlation between types			-0.46	

[a] lifetime value used for scaling is : 8206.49. Asymptotic standard Errors in parentheses.

Table 5: Unobserved Heterogeneity and the returns to experience and tenure

computed for each individual *given* the information set at the time the decision is made and then we average over individuals. For this calculation we employ a horizon of 40 years. The results are displayed in Table 6.

Taking all individual costs into account, the average gain to apprenticeship (ATE) [36] is 11%. Netting out utility costs of apprenticeship, which turn out to be negative on average, the gain declines to -4.1%. However, netting out the opportunity cost of education the gain increases to 15%; in other words, if it was not for a preference for apprenticeship, the opportunity cost is so high that on average the gains would not justify apprenticeship training. Viewed from the point of view of age 18, after the apprentices have finished their formal training, the gains to apprenticeship are 5.5%. Here the differences between the two groups relate to wages and labor market attachment, but they exclude the costs of education as well as difference in job mobility during the training period. Finally, the gains for those who choose to qualify as apprentices (ATT gains) are a substantial 13%. The four last columns in Table 6 show the way the gains vary by type. Overall high wage individuals have lower gains to apprenticeship and among them the gains are higher for those with lower costs.

In the lower four rows of the table we consider the contribution to the gains of the var-

---

[36]ATE: Average Treatment Effect; ATT: Average treatment on the treated.

	Average	Type 1	Type 2	Type 3	Type 4
Wage		Low		High	
Cost of Education		Low	High	Low	High
Average Treatment Effect (ATE)	11%	8.4%	16%	4.3%	11%
Average Treatment on the Treated (ATT)	13%	13%	18%	8.9%	12%
ATE, at age 18	5.5%	8.2%		4.2%	
		Decomposing the gain to Apprenticeship			
ATE, net of utility of education	-4.1%	-3.2%	-3.1%	-5%	-4.5%
ATE, net of opportunity cost of education	15%	12%	20%	8%	14%
Equal distribution of firm-worker match (σ_0)	23%	19%	27%	17%	24%
Same job destruction rate	11%	8.1%	16%	3.9%	10%
No business cycle effects	13%	10%	18%	6.3%	13%
Same job offer rate	22%	20%	28%	14%	21%

Table 6: The Life-cycle Returns to Apprenticeship

ious differences in the parameters of the model between apprentices and non-apprentices. We do this by setting the relevant parameter for the apprentice group to that of the non-apprentice. It is striking that giving the higher initial match-component variance (σ_0) to the apprentice group increases the gains to 23%. The variety of jobs available to non-apprentices, is a large positive contributor to their labor market outcomes (through job shopping) and hence to the overall low gain to apprenticeship. Also giving the apprentices the same job offer rates as the non-apprentices increases the gains to 22%. Equalizing the job destruction rate has little effect, while eliminating the effects of the business cycle increases the gains slightly.

Finally, note that these gains factor in all costs faced by the individual but not the costs borne by the government or the firm.

As a further illustration of the properties of the model we carry out two experiments: in one we offer half of all individuals randomly an education subsidy of 2% of lifetime value; in the other the subsidy is increased to 20% and again randomly allocated. We then use the generated data to estimate the Local Average Treatment Effect of training using the respective instruments. We compare this to OLS and to the effects obtained when we randomly allocate training.[37] All these approaches, including the one that

---

[37]On LATE see Imbens and Angrist (1994). In our model monotonicity is satisfied, i.e following the subsidy all individuals either remain where they are or switch into training. This may not be true if one allows for General Equilibrium effects.

OLS	Instrumental Variables (LATE)[a]		Randomized Education	
	2% lifetime value	20% subsidy	No Correction	Selection correction
14.6%	31.7%	13.6%	12.6%	18.0%

[a]The instrument is an education subsidy of 2% and 20% lifetime value respectively. The effect of a 2% subsidy is 3% increase in education and the effect of 20% subsidy leads to a 10% increase

Table 7: Various estimators of the returns to training based on simulated data from the model

randomizes education will be affected by composition effects due to non-random selection into employment. To document how important this can be we also present the wage gain based on randomizing education and corrected for selection bias into work. The gains are estimated at age 24-28 and are presented in Table 7.

The OLS gains are about 15%. The IV that shifts a small marginal group into training leads to a gain of 32%, while the IV which shifts a larger number of people into apprenticeship (including many more people with lower gains) is 13.6. When we randomize education the gain is 12.6%; correcting for selection however it grows to 18%, demonstrating that it makes little sense to evaluate programme effects without correction for selection into employment, even with "perfect" instruments.

6   Labor Market Reforms and Career Decisions

In this section we consider how welfare reform can affect career decisions and outcomes. In doing this we follow Heckman and Klenow (1998) who emphasize that human capital policies should be evaluated in a life cycle setting.[38]

We focus on 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.[39] 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

---

[38]Similar considerations are discussed in Keane and Wolpin (2000), who present the effect of a wage subsidy on education and career choices.

[39]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.

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 the decision to take up an apprenticeship scheme may change because the programme compresses the returns to education for some individuals, but increases them for others. Finally, the wage subsidy will change the incentives for job 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 instead of the current German system where the young lower paid unemployed are paid about 55% of their last earnings. Such a reform was introduced in the UK in the 1980s. All policies are simulated to be revenue neutral and proportional earnings taxation is adjusted to achieve this. Both reforms are outlined in Table 8. The EITC subsidy we have introduced is described in relation to the density of observed wages in Figure 7. We also illustrate directly the effects of an unfunded tax cut to give an idea of how the requirement of revenue neutrality is likely to affect the outcome of policies.

In interpreting our results we should note some limitations. Our model assumes risk neutrality and as such ignores the insurance aspect of the policies introduced. Second, our simulation takes as fixed the overall number of people allocated to the vocational track at age 10. The change in policy may well affect the number and type of individuals joining this group, but we do not take this into account. Finally, we do not allow for general equilibrium effects, because we do not model how the pay policies of firms will change in response to the aggregate changes in supply of labour in the two sectors.

Before we present policy analysis that involves responses to complex changes we present labor supply elasticities in Table 9 so as to provide a feel for the sensitivity of labor supply to incentives.

The elasticities are the proportional change in participation resulting from a small

Name	Description
(1) EITC	A wage subsidy at a rate of 40% up to 30 euros per day, stays constant up to 73.7 euros 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 proportional tax on earnings.
(2) Flat Unemployment Benefit	40% of "minimum wage" defined as 18 euros per day. Excess revenue redistributed through proportional taxation (subsidy) on earnings.

Table 8: Simulated Policies

All workers	Apprentices	Non Apprentices
1.02	1.02	1.04

Table 9: Labour supply (participation) elasticities with respect to lifetime change in wage

proportional change in wages at all points in the lifecycle, keeping education choices constant. Since our model is linear in income and the marginal utility of wealth is constant, there is no obvious sense by which we can distinguish between Frisch and Marshalian elasticities. Note that in our model increasing wages also increases unemployment benefit.[40]

6.1 Policy analysis

To derive the implications of the two 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.

Table 10 displays the effect on education choices by type of individual. Overall in work benefits reduce take up of apprenticeship by 1.8%. This is partly because the returns to training are compressed by the subsidy:[41] Given that low wage jobs are subsidized,

---

[40]In computing the elasticity we have kept unemployment income constant. However, allowing UI to also change in line with the wage only changes the elasticities in the second significant figure.

[41]Note that the subsidy is only available to those over 19, when apprenticeship training will have finished; hence the policy is designed here not to act as a direct monetary disincentive to training.

Figure 7: Density of Wages and In-Work Benefit Scheme
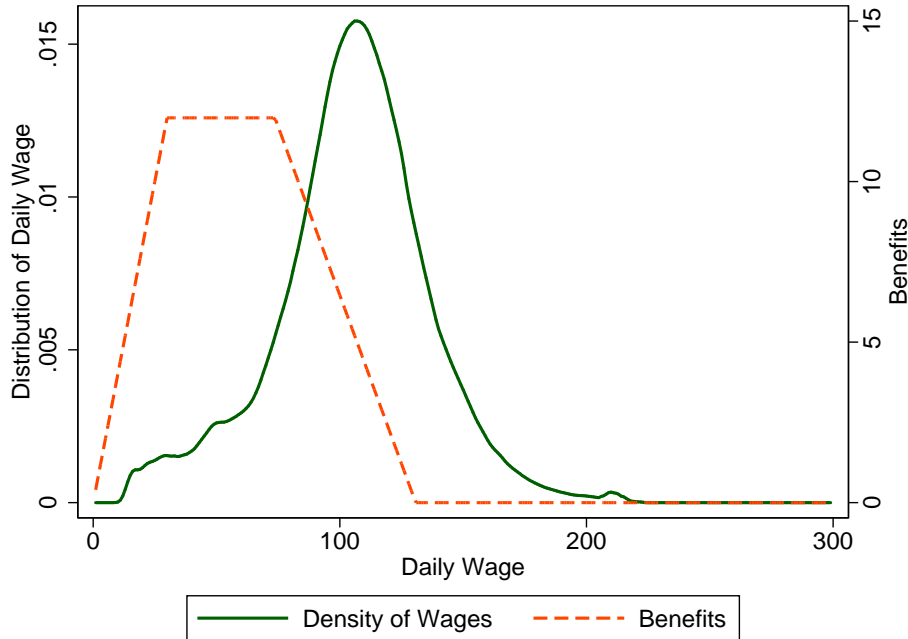
non-apprentices are clearly favored by this policy and this attracts more into the group and out of apprenticeship training. The lowest panel implies that the increase in taxation would tend to increase those obtaining skills because it reduces the opportunity cost of education.

Employment increases for all types, but particularly for the low wage types, who are much more likely to benefit from the reform. Indeed, judging from the effect of taxes, the impact would have been much higher if it was not for the fact that taxes have to increase to fund the programme. An additional channel by which policy has an impact is by changing the incentives for job mobility. A job offer consists of a new match specific effect and a mobility cost. 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 the change in EITC benefits and allowing for the switching costs, may be negative. Moreover, under EITC individuals become less choosy about the jobs they accept from unemployment. The result of this is that the quality of matches decline, which translates to an overall decline of 1.9% in wages over the lifecycle.

Replacing the earnings related UI with a flat rate, which is independent of earnings, as was done in the UK in the 1980s, has two opposite effects. On the one hand non-

|	All	Type 1	Type 2	Type 3	Type 4
Policy 1: Low Wage Subsidy (fully Funded)					
% Increase Skilled	-1.8%	-1.7%	-1.6%	-1.9%	-2.2%
% Increase in Work	4.2%	9.7%	9.7%	3%	1.6%
Difference in Match Specific Effect	-1.9%	-3.1%	-2.3%	-1.4%	-1.6%
Change in Tax	5.4%				
Policy 2: Flat Unemployment Benefits (fully Funded)					
% Increase Skilled	-1.3%	-1.3%	-1.1%	-1.4%	-1.1%
% Increase in Work	11%	18%	18%	10%	8.2%
Difference in Match Specific Effect	-5.2%	-7.5%	-6.5%	-4.9%	-4.4%
Change in Tax	-13.1%				
Policy 3: Tax Cut (unfunded)					
% Increase Skilled	-0.61%	-0.75%	0%	-0.69%	-0.73%
% Increase in Work	3.5%	7.2%	6.8%	3.2%	1.8%
Difference in Match Specific Effect	-0.47%	-1.4%	-1.2%	-1.4%	-0.0093%
Change in Tax	-5%				

Table 10: Policy Effects on Apprenticeship training, employment and wages

apprentices, that have higher job destruction rates and are more sensitive to negative wage shocks because their wage is lower, will find that unemployment is associated with lower income. However, they also have higher job arrival rates. The net effect is a decrease in those training for an apprenticeship by about 1.3% percentage points. Employment increases by 11%, particularly for the lower wage types. Finally, because individuals accept jobs much faster the average quality of the match decreases, reducing wages on average by 5.2%. Figures 8 and 9 display the overall effect of the reforms on employment and match quality respectively, over time.

7 Conclusion

In this paper we have specified and estimated by maximum likelihood a model of apprenticeship choice, employment, job mobility and wages using detailed German administrative records. This data has the rare characteristic that we can observe careers from the start when individuals make their first decisions. There is no censoring or initial conditions problem and measurement error is likely to be unimportant, because the records are reported by firms for the purpose of determining social security contributions. Moreover, we observe many cohorts of individuals in all German States; this provides useful exogenous variation in the conditions under which the early education decisions are made,

Figure 8: The Employment effects of EITC and UI reform (Apprentices)
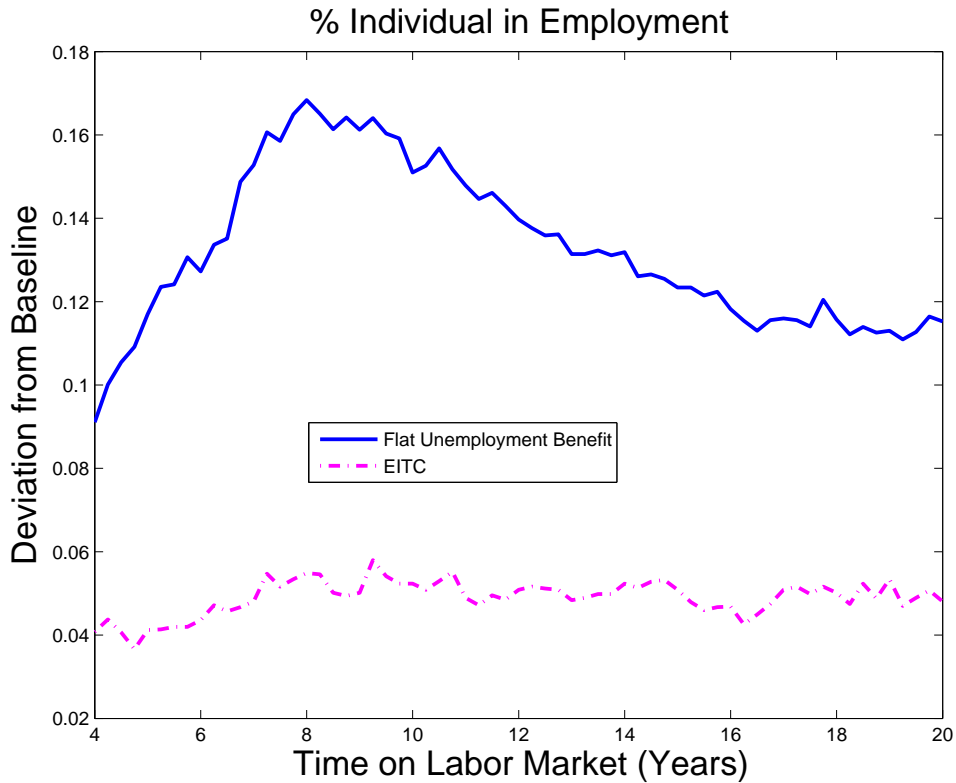
thus aiding identification.

In the model individuals who have followed the vocational track of education choose whether to follow an apprenticeship training or not. We then model the subsequent labour supply and job mobility decisions jointly with wages, which are allowed to grow with experience and tenure. The model allows for match specific heterogeneity and search frictions as well as permanent shocks to the match specific effects and thus allows us to understand the sources of wage growth. We are also able to estimate wage elasticities of participation as well as the stochastic properties of wages and the extent of match specific heterogeneity.

Using the model we estimate the gain relating to apprenticeship; we show that the opportunity cost of education is a major factor in the costs of training. Finally, we quantify how welfare reform can change individual decisions over the lifecycle, affecting human capital accumulation, employment and the quality of matches, demonstrating the tradeoffs that policy makers need to face when introducing policies that are designed to improve work incentives for lower productivity individuals.

39

Figure 9: The effect of the EITC and UI reforms on the value of the match over the lifecycle
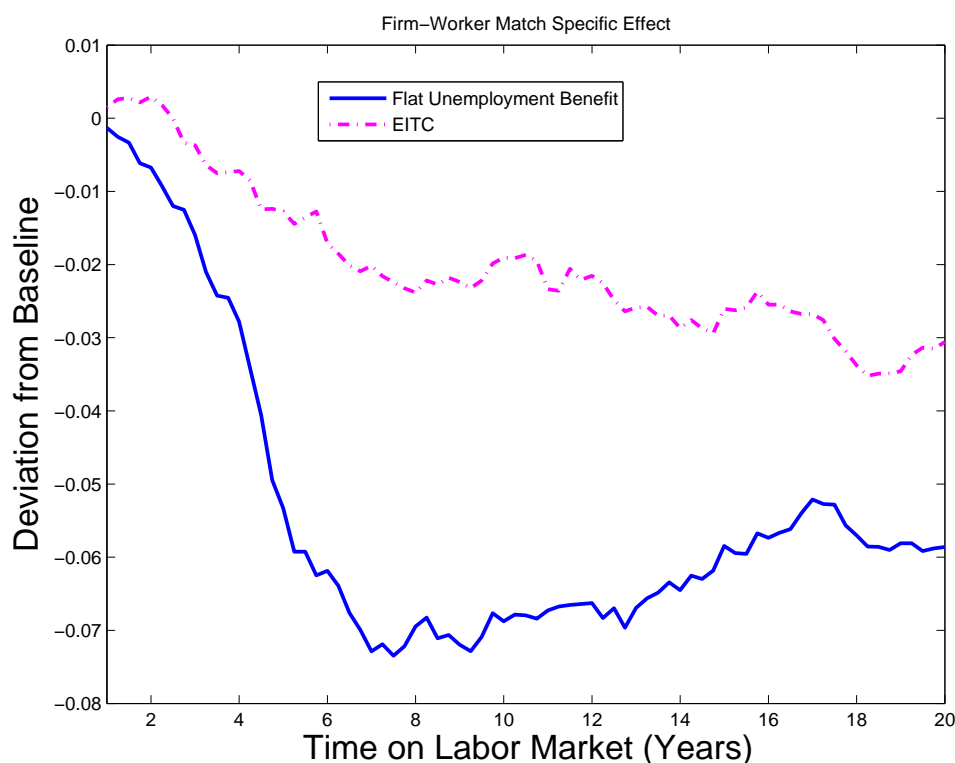
Appendix

A   The German Unemployment Insurance System

For the period we consider, 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

	Constant		Effect of GDP	
Region	Coefficient[a]	Std. err	Coefficient	Std. err
Schleswig-Holstein	-10.8	(0.038)	-0.1141	(0.0032)
Hamburg	-7.81	(0.04)	-0.0002942	(0.0036)
Niedersachsen	-10.9	(0.55)	-0.0005712	(0.002)
Bremen	-8.88	(0.0048)	0.4693	(0.0092)
Nordrhein-Westfalen	-8.21	(0.5)	-0.1786	(0.0016)
Hessen	-10.1	(0.14)	0.1509	(0.0025)
Rheinland-Pfalz	-10.2	(0.052)	0.3878	(0.0031)
Baden-Wuerttemberg	-7.11	(0.61)	0.00154	(0.002)
Bayern	-12	(0.58)	0.0342	(0.0021)
Saarland	-11	(0.021)	-0.187	(0.0061)
Berlin	-6.84	(0.0095)	0.347	(0.0056)
Parameter	Coefficient		Std. err	
Effect of GDP	0.1334		(0.11)	
σ_ω[a]	8.582		(0.19)	
[a] All coefficients are scaled as a percentage of "lifetime value": 8206.49				

Table 11: Regional and Business Cycle effects on the costs to apprenticeship

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*).

B   Some further parameter Estimates

This Table shows the effect of the business cycle and region on the cost of apprenticeship as well as the standard deviation of the cost shock to apprenticeship.

C   Wage trend, GDP growth and Markov transition matrix

Wages are deflated by the German CPI. We then detrend wages and GDP, the forme with a quadratic trend and the latter by a a linear one. Since the changes in relative wages between apprentices and non-apprentices are modelled we only remove the overall trend common to

	Coefficient	St. Error
Year	0.0022	(0.011)
Year^2	0.00032	(0.0004)

Table 12: Trend growth in wages

	Below Trend in t+1	Above Trend in t+1
Below Trend in t	0.9302 (0.039)	0.069 (0.039)
Above Trend in t	0.075 (0.042)	0.925 (0.042)
Asymptotic standard Errors in brackets		

Table 13: Quarterly transition matrix for below and above trend GDP

both. Over our modelling period (1975-1996) aggregate real wages grow very slowly as shown in Table 12. In the model we use wages after detrending by these estimates.

We also detrend real per-capita GDP (constant 1995 US$) using a linear trend. GDP grew at a rate of $479.18 (se 9.015) per year. In the model we then use transitions between above trend (good times) and below trend (bad times) GDP growth on a quarterly basis based on the estimated transition matrix in Table 13

D   The Fit of the Model

Table 14 displays the labor market transitions by education groups at a quarterly frequency. We distinguish five possible transitions, between unemployment (U), Employment E and between same job and job to job. Overall, the model matches the transition probabilities closely. It does however, underestimate the transitions from employment to unemployment particularly for non-apprentices.

A reflection of the overall good fit of the transitions above is the fit of the average experience and tenure over time for the two education groups in Figure 10 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 11).

Finally we are able to replicate almost perfectly the average profile of wages for workers since first entry on the labor market, including the apprenticeship period (see Figure 12).

In Figure 13 we show how the model fits the standard deviation of wages over the lifecycle. First note the profile of the observed standard deviation is either declining with experience or

	Apprentices		Non-Apprentices	
	Obs	Pred	Obs	Pred
U to U	0.89	0.89	0.89	0.87
U to E	0.11	0.11	0.11	0.13
E to U	0.049	0.041	0.085	0.076
E to new E	0.035	0.033	0.032	0.036
E to same E	0.92	0.93	0.88	0.89

Table 14: Model fit - Transitions

remains flat after 3 years in the labor market. This is a remarkable contrast to U.S. data where the variance is steadily increasing prompting a debate on whether wages have a unit root or not. Here clearly they do not and this justifies our modelling where the shocks are not carried from one firm to another making them effectively transitory. While our model does not fit perfectly this feature of the data it does pretty well in capturing the overall shape.

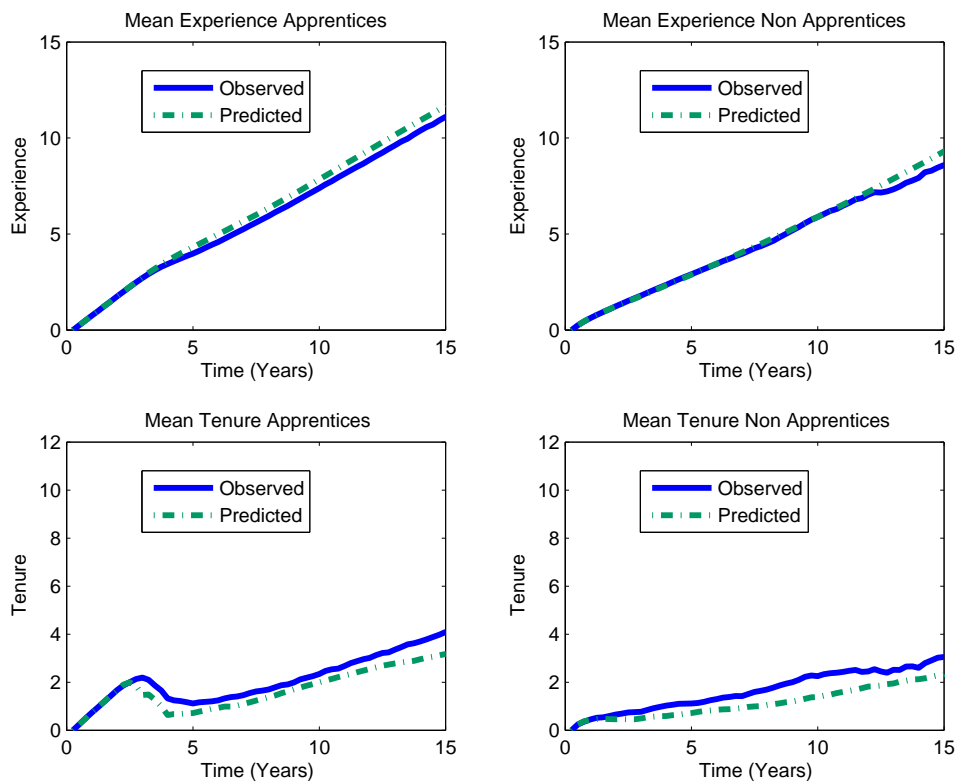Figure 10: Observed and predicted experience and tenure profiles
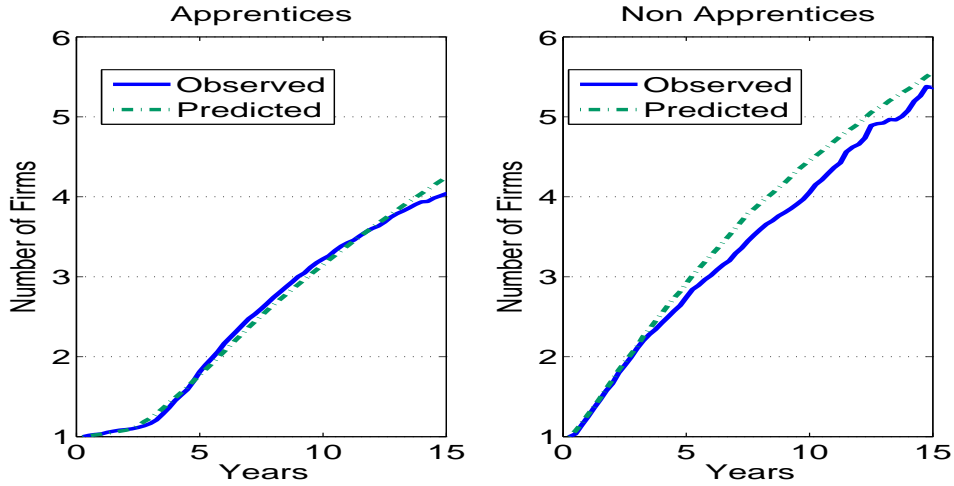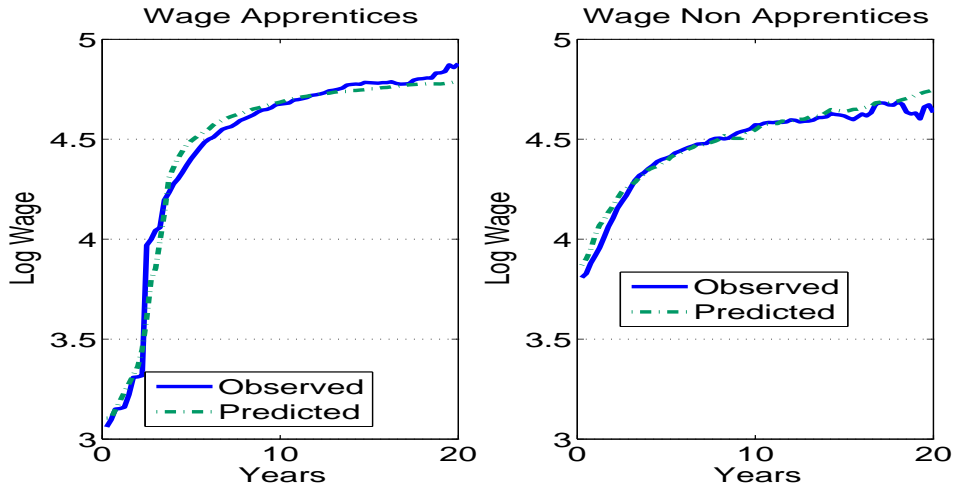
Figure 11: Observed and fitted number of jobs

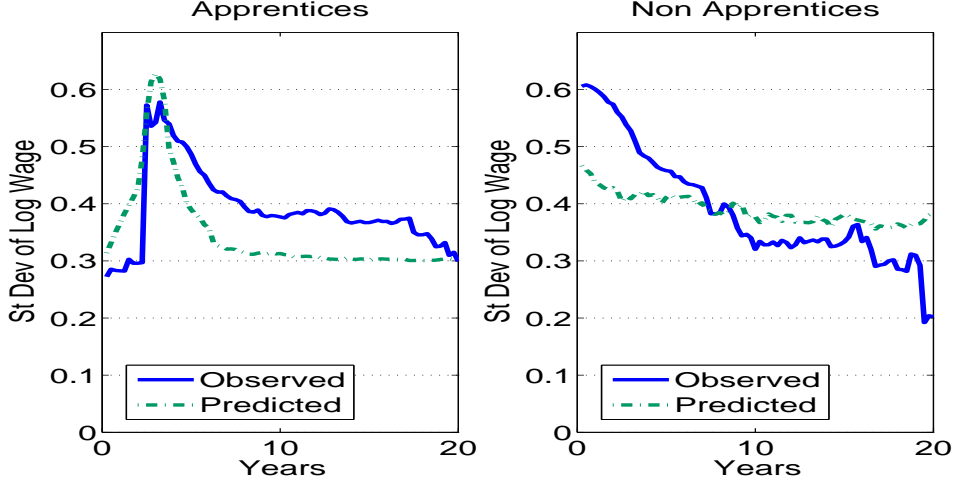Figure 12: Fitted and actual wages for apprentices and non-apprentices

E The Likelihood function

Conditional on unobserved heterogeneity, the model is Markovian. One can thus express the individual likelihood as a product of conditional densities. The likelihood is written conditional on business cycle G_t, the dynamics of which contributes multiplicatively to the sample likelihood, and conditional on unobserved heterogeneity ε_i. Unobserved heterogeneity is eventually be integrated out. The data consists in a series of wages and transitions across unemployment, and different employment spells at various employers. Here we show how to compute the likelihood of specific transitions and accepted wages given the past.

- **Transition from unemployment to work** paid $w_{it} = w$ (observation) given $Ed_i, G_t, X_{it}, w_{i(-1)}, \varepsilon_i$:

$$L_{U-E}(w) = \Pr\{\text{offer}\} \times \underbrace{\text{pdf}\left\{\kappa_{if}^0 = \kappa\right\}}_{\text{Lik. of wage}} \times \underbrace{\Pr\left\{\mu_{if} + W(\kappa) > \eta_{it} + U\right\}}_{\text{That wage offer is better than unemployment}}$$

Figure 13: Fitted and actual standard deviation of wages for apprentices and non-apprentices

where κ_{if}^0, μ_{if} and η_{it} are random, and with

$$
\begin{aligned}
W\left(\kappa_{if}^0\right) &\equiv W\left(Ed_i, G_t, X_{it}, T_{ift} = 0, \kappa_{if}^0, \varepsilon_i\right) \\
U &\equiv U\left(Ed_i, G_t, X_{it}, w_{i(-1)}, \varepsilon_i\right)
\end{aligned}
$$

and

$$
\kappa = \ln w - \alpha_0(\varepsilon_i) - \alpha_{Ed}Ed_i - \alpha_X(X_{it}, Ed_i) - \alpha_T(T_{ift}, Ed_i) - \alpha_G(Ed_i)G_t.
$$

That is:

$$
\begin{aligned}
L_{U-E}(w) &= \pi_U \times \frac{1}{\sigma_0}\varphi\left(\frac{\kappa}{\sigma_0}\right) \times \Phi\left(\frac{m_\mu + W(\kappa) - U}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right), \\
\pi_U &\equiv \pi_U(Ed_i, X_{it}, G_t), \\
\sigma_0 &\equiv \sigma_0(Ed_i).
\end{aligned}
$$

- **Transition from unemployment to unemployment** given $Ed_i, G_t, X_{it}, w_{i(-1)}, \varepsilon_i$:

$$
L_{U-U} = 1 - \pi_U + \pi_U \times \int \Phi\left(\frac{U - m_\mu - W(\sigma_0 t)}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right) d\Phi(t),
$$

- **Transition from work to unemployment** given $Ed_i, G_t, X_{it}, T_{ift}, \kappa_{ift-1}, \varepsilon_i$:

$$
L_{E-U} = \underbrace{\delta}_{\text{fired}} + \underbrace{(1-\delta)}_{\text{not fired}}\underbrace{(1-\pi_W)}_{\text{no outside offer}}\underbrace{\Pr\{\eta_{it} + U > W(\kappa_{ift})\}}_{\text{voluntary unemployment}}
$$

$$
+ (1-\delta)\underbrace{\pi_W}_{\text{outside offer}}\underbrace{\Pr\left\{\eta_{it} + U > W(\kappa_{ift}) \,\&\, \eta_{it} + U > \mu_{i\widetilde{f}} + \widetilde{W}\left(\kappa_{i\widetilde{f}}^0\right)\right\}}_{\text{unemployment is preferred to both competing jobs}}
$$

45

where $\eta_{it}, \kappa_{ift}, \mu_{i\widetilde{f}}$ and $\kappa^0_{i\widetilde{f}}$ are random with

$$\kappa_{ift} = \kappa_{ift-1} + u_{ift}$$

$$U \equiv U\left(Ed_i, G_t, X_{it}, w_{i(-1)}, \varepsilon_i\right)$$

$$W\left(\kappa_{ift}\right) \equiv W\left(Ed_i, G_t, X_{it}, T_{ift}, \kappa_{ift}, \varepsilon_i\right)$$

$$\widetilde{W}\left(\kappa^0_{i\widetilde{f}}\right) \equiv W\left(Ed_i, G_t, X_{it}, T_{i\widetilde{f}t} = 0, \kappa^0_{i\widetilde{f}}, \varepsilon_i\right)$$

Moreover

$$\Pr\left\{\eta_{it} + U > W\left(\kappa_{ift}\right)\right\} = \int \Phi\left(\frac{U + \mu_\eta - W\left(\kappa_{ift-1} + \sigma_u t\right)}{\sigma_\eta}\right) d\Phi\left(t\right)$$

and

$$\Pr\left\{\eta_{it} + U > W\left(\kappa_{ift}\right) \& \eta_{it} + U > \mu_{i\widetilde{f}} + \widetilde{W}\left(\kappa^0_{i\widetilde{f}}\right)\right\}$$

$$= \iint \Phi_2\left(\frac{U + \mu_\eta - W\left(\kappa_{ift-1} + \sigma_u t_1\right)}{\sigma_\eta}, \frac{U + \mu_\eta - m_\mu - \widetilde{W}\left(\sigma_0 t_2\right)}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}; \frac{\sigma_\eta}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right) d\Phi\left(t_1\right) d\Phi\left(t_2\right)$$

where $\Phi_2\left(\cdot, \cdot; \rho\right)$ denotes the cdf of two standardized normal variates with correlation ρ.

- **Transition from work to work in the same firm** paying now a wage $w_{it} = w$ given $Ed_i, G_t, X_{it}, T_{ift}, \kappa_{if,t-1}, \varepsilon_i$:

$$L_{E-E}\left(w\right) = \left(1 - \delta\right) \times \text{pdf}\left\{u_{ift} = \kappa - \kappa_{ift-1}\right\}$$

$$\times \left[\pi_W \Pr\left\{W\left(\kappa\right) > \eta_{it} + U \& W\left(\kappa\right) > \mu_{i\widetilde{f}} + \widetilde{W}\left(\kappa^0_{i\widetilde{f}}\right)\right\}\right.$$

$$\left. + \left(1 - \pi_W\right) \Pr\left\{W\left(\kappa\right) > \eta_{it} + U\right\}\right]$$

where

$$\kappa = \ln w - \alpha_0(\varepsilon_i) - \alpha_{Ed} Ed_i - \alpha_X(X_{it}, Ed_i) - \alpha_T(T_{ift}, Ed_i) - \alpha_G(Ed_i) G_t$$

$$U \equiv U\left(Ed_i, G_t, X_{it}, w_{i(-1)}, \varepsilon_i\right)$$

$$W\left(\kappa_{ift}\right) \equiv W\left(Ed_i, G_t, X_{it}, T_{ift}, \kappa_{ift}, \varepsilon_i\right)$$

$$\widetilde{W}\left(\kappa^0_{i\widetilde{f}}\right) \equiv W\left(Ed_i, G_t, X_{it}, T_{i\widetilde{f}t} = 0, \kappa^0_{i\widetilde{f}}, \varepsilon_i\right)$$

Hence,

$$L_{E-E}\left(w\right) = \left(1 - \delta\right) \left[\pi_W \Pr\left\{W\left(\kappa\right) - \widetilde{W}\left(\kappa^0_{i\widetilde{f}}\right) > \mu_{i\widetilde{f}}\right\} + 1 - \pi_W\right]$$

$$\times \Phi\left(\frac{W\left(\kappa\right) - U - \mu_\eta}{\sigma_\eta}\right) \frac{1}{\sigma_u} \varphi\left(\frac{\kappa - \kappa_{if,t-1}}{\sigma_u}\right)$$

with

$$\Pr\left\{W\left(\kappa\right)-\widetilde{W}\left(\kappa_{i f}^{0}\right)>\mu_{i\widetilde{f}}\right\}=\int\Phi\left(\frac{W\left(\kappa\right)-m_{\mu}-\widetilde{W}\left(\sigma_{0}t\right)}{\sigma_{\mu}}\right)d\Phi\left(t\right).$$

- **Transition from work to work paid $w_{it}=w$ in the same firm during apprenticeship** given G_t, X_{it}, T_{ift}, κ_{ift-1}, ε_i:

$$L_{E-E}^{A}\left(w\right)=\left[\pi_A\Pr\left\{\mu_{i\widetilde{f}}+\widetilde{W}_A\left(\kappa_{i\widetilde{f}}^{0}\right)<W_A\left(\kappa\right)\right\}+1-\pi_A\right]\times\frac{1}{\sigma_u}\varphi\left(\frac{\kappa-\kappa_{ift-1}}{\sigma_u}\right)$$

with

$$
\begin{aligned}
W_A\left(\kappa_{ift}\right) &= W_A\left(G_t,X_{it},T_{ift},\kappa_{ift},\varepsilon_i\right)\\
\widetilde{W}_A\left(\kappa_{i\widetilde{f}}^{0}\right) &\equiv W_A\left(G_t,X_{it},T_{i\widetilde{f}t}=0,\kappa_{i\widetilde{f}}^{0},\varepsilon_i\right)
\end{aligned}
$$

and

$$\kappa=\ln\frac{w}{\lambda_A}-\alpha_0(\varepsilon_i)-\alpha_X\left(X_{it},NA\right)-\alpha_T\left(T_{ift},NA\right)-\alpha_G\left(NA\right)G_t.$$

Finally,

$$\Pr\left\{\mu_{i\widetilde{f}}+\widetilde{W}_A\left(\kappa_{i\widetilde{f}}^{0}\right)<W_A\left(\kappa\right)\right\}=\int\Phi\left(\frac{W_A\left(\kappa\right)-m_{\mu}-\widetilde{W}_A\left(\sigma_0\left(NA\right)t\right)}{\sigma_{\mu}}\right)d\Phi\left(t\right)$$

- **Transition from work to work paid $w_{it}=w$ in a new firm** given Ed_i, G_t, X_{it}, T_{ift}, κ_{ift-1}, ε_i:

$$
\begin{aligned}
L_{E-\widetilde{E}}\left(w\right)=(1-\delta)\,\pi_W\times\text{pdf}&\left\{\kappa_{i\widetilde{f}}^{0}=\kappa\right\}\\
&\times\Pr\left\{\mu_{i\widetilde{f}}+\widetilde{W}\left(\kappa\right)>\eta_{it}+U\&\mu_{i\widetilde{f}}+\widetilde{W}\left(\kappa\right)>W\left(\kappa_{ift}\right)\right\}.
\end{aligned}
$$

with

$$\kappa=\ln w-\alpha_0(\varepsilon_i)-\alpha_{Ed}Ed_i-\alpha_X(X_{it},Ed_i)-\alpha_T\left(T_{i\widetilde{f}t}=0,Ed_i\right)-\alpha_G(Ed_i)G_t.$$

That is:

$$
\begin{aligned}
L_{E-\widetilde{E}}\left(w\right)=(1-\delta)\,\pi_W\times\frac{1}{\sigma_0}\varphi\left(\frac{\kappa}{\sigma_0}\right)\\
\times\int\Phi_2\left(\frac{m_{\mu}+\widetilde{W}\left(\kappa\right)-U}{\sqrt{\sigma_{\eta}^{2}+\sigma_{\mu}^{2}}},\frac{m_{\mu}+\widetilde{W}\left(\kappa\right)-W\left(\kappa_{ift-1}+\sigma_u t\right)}{\sigma_{\mu}};\frac{\sigma_{\mu}}{\sqrt{\sigma_{\eta}^{2}+\sigma_{\mu}^{2}}}\right)d\Phi\left(t\right).
\end{aligned}
$$

- **Transition from work to another work paid** $w_{it} = w$ **in a new firm during apprenticeship** given $G_t, X_{it}, T_{ift}, \kappa_{ift-1}, \varepsilon_i$:

$$
\begin{aligned}
L^A_{E-\widetilde{E}}(w) &= \pi_A \times \text{pdf}\left\{\kappa^0_{if} = \kappa\right\} \times \Pr\left\{\mu_{i\widetilde{f}} + \widetilde{W}_A(\kappa) > W_A(\kappa_{ift})\right\} \\
&= \pi_A \times \frac{1}{\sigma_0(NA)}\varphi\left(\frac{\kappa}{\sigma_0(NA)}\right) \\
&\quad \times \int \Phi\left(\frac{m_\mu + \widetilde{W}_A(\kappa) - W_A(\kappa_{ift-1} + \sigma_u t)}{\sigma_\mu}\right) d\Phi(t)
\end{aligned}
$$

with

$$
\kappa = \ln\frac{w}{\lambda_A} - \alpha_0(\varepsilon_i) - \alpha_X(X_{it}, NA) - \alpha_T(T_{ift} = 0, NA) - \alpha_G(NA)G_t.
$$

- **Probability of starting apprenticeship at wage** $w_{it} = w$ **given** $G_t, REG_i, \varepsilon_i$:

$$
L_A = \text{pdf}\left\{\kappa^0_{if} = \kappa\right\} \times \Pr\left\{\mu_{if} + W^0_A(\kappa) - \lambda_0 - \omega_{it} > \mu_{i\widetilde{f}} + \widetilde{W}^0_{NA}\left(\kappa^0_{i\widetilde{f}}\right)\right\}
$$

with

$$
\begin{aligned}
W^0_A\left(\kappa^0_{if}\right) &\equiv W_A\left(G_t, X_{it} = 0, T_{ift} = 0, \kappa^0_{if}, \varepsilon_i\right), \\
\widetilde{W}^0_{NA}\left(\kappa^0_{i\widetilde{f}}\right) &\equiv W\left(Ed_i = NA, G_t, X_{it} = 0, T_{ift} = 0, \kappa^0_{i\widetilde{f}}, \varepsilon_i\right), \\
\lambda_0 &\equiv \lambda_0(REG_i, G_t),
\end{aligned}
$$

and

$$
\kappa = \ln\frac{w}{\lambda_A} - \alpha_0(\varepsilon_i) - \alpha_X(X_{it} = 0, NA) - \alpha_T(T_{ift} = 0, NA) - \alpha_G(NA)G_t.
$$

One finally obtains:

$$
L_A = \frac{1}{\sigma_0(A)}\varphi\left(\frac{\kappa}{\sigma_0(A)}\right)\int \Phi\left(\frac{W^0_A(\kappa) + m_\mu(A) - \lambda_0 - \widetilde{W}^0_{NA}(\sigma_0 t) - m_\mu(NA)}{\sqrt{2\sigma_\mu^2 + \sigma_\omega^2}}\right) d\Phi(t).
$$

- **Probability of not starting apprenticeship and being employed at wage** $w_{it} = w$ **given** $G_t, REG_i, \varepsilon_i$:

$$
L_A = \frac{1}{\sigma_0(NA)}\varphi\left(\frac{\kappa}{\sigma_0(NA)}\right)\int \Phi\left(\frac{\widetilde{W}^0_{NA}(\kappa) + m_\mu(NA) + \lambda_0 - W^0_A(\sigma_0 t) - m_\mu(A)}{\sqrt{2\sigma_\mu^2 + \sigma_\omega^2}}\right) d\Phi(t).
$$

with

$$
\kappa = \ln w - \alpha_0(\varepsilon_i) - \alpha_X(X_{it} = 0, NA) - \alpha_T(T_{ift} = 0, NA) - \alpha_G(NA)G_t.
$$

- **Sample likelihood.** Define $P^S(REG_i, YOB_i, \varepsilon_i)$ as the probability of going through the secondary or intermediate school (the vocational track, as opposed to high school which leads to university) for individual i in region REG_i and born in year YOB_i. Let $Z_i = (REG_i, YOB_i)$. We assume:

$$P^S(Z_i, \varepsilon_i) = \Phi(\theta_Z(Z_i) + \theta_\varepsilon(\varepsilon_i))$$

where $\theta_Z(Z_i)$ are region×cohort specific effects. Now assume the distribution of unobserved heterogeneity is discrete with J points of support denoted by h_j each with probability π_j, where h_j is a two dimensional vector. Given this the sample likelihood for N individuals is:

$$L = \prod_i^N \sum_{j=1}^J \pi_j \left\{ \left[P^S(Z_i, \varepsilon_i = h_j) L_{t_{0i}}(REG_i, \varepsilon_i = h_j) \right]^{S_i} \left[1 - P^S(Z_i, \varepsilon_i = h_j) \right]^{1-S_i} \right\}$$

(E1)

where $L_{t_{0i}}(REG_i, \varepsilon_i = j)$ is the likelihood of the sequence of observations for the ith individual being 16 at time t_{0i} in region i and born in year YOB_i and S_i is an indicator function, which is 1 when the individual has chosen vocational school (thus being in the population to which our sample refers to) and zero otherwise.

E.1   Computing the Value Functions

We integrated out analytically as many state variables as possible (shocks to the value of leisure (η), shocks to the cost of training ω, and shocks to cost of moving μ) as shown in the subsection below. We approximate the value functions by evaluating them at a number of discrete points in the state space and interpolating linearly in between. For experience and tenure the points where we evaluate are 0, 2, 4, 6, 10 and 30 years of experience and 0, 2, 4, 6 and 30 years of tenure; this level of detail turned out to be sufficient. The other state variable is the firm-worker match specific effect which evolves as a random walk while the worker remains in the same job. We use 6 points, on a grid which depends on education and on tenure to take into account the non-stationary nature of the process. More specifically, given the assumptions made, the match effect is a normal variable with mean zero and variance $T\sigma_U(Ed)^2 + \sigma_0(Ed)^2$ for an individual with T years of tenure. We use a quadrature-based method as in the Tauchen and Hussey (1991) procedure to generate a grid and transition matrices. We interpolate between the points. The time horizon and the way we deal with the terminal condition is discussed in the main text.

E.1.1 Emax computations

Making use of the normality of innovations allows to simplify the Bellman equations significantly. For standardized normal random variables the following identity holds true (see Tallis (1961)):

$$\mathbb{E}\left[U_1 \mathbf{1}\left\{U_1 > a, U_2 > b\right\}\right] = \varphi\left(a\right)\Phi\left(\frac{\rho a - b}{\sqrt{1-\rho^2}}\right) + \rho\varphi\left(b\right)\Phi\left(\frac{\rho b - a}{\sqrt{1-\rho^2}}\right)$$

with $\rho = \mathrm{Cov}\left(U_1, U_2\right)$, and

$$\Pr\left\{X_1 > a, X_2 > b\right\} = \Pr\left\{-X_1 < -a, -X_2 < -b\right\} = \Phi_2\left(-a, -b; \rho\right).$$

- **The deterministic value of unemployment:** Conditional on Ed_i, $\underline{G_{t+1}}$, X_{it}, $w_{i(-1)}$, ε_i, $\underline{\kappa_{if}^0}$ (where we underline the variables which will have to be integrated out), let

$$
\begin{aligned}
W &\equiv W\left(Ed_i, G_t, X_{it}, T_{ift} = 0, \kappa_{if}^0, \varepsilon_i\right), \\
U &\equiv U\left(Ed_i, G_t, X_{it}, w_{i(-1)}, \varepsilon_i\right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathbb{E}\max\left[\eta_{it+1} + U, \mu_{if} + W\right] &= \mathbb{E}\left[\left(\eta_{it+1} + U\right)\mathbf{1}\left\{\eta_{it+1} + U > \mu_{if} + W\right\}\right] \\
&\quad + \mathbb{E}\left[\left(\mu_{if} + W\right)\mathbf{1}\left\{\mu_{if} + W > \eta_{it+1} + U\right\}\right] \\
&= U\Phi\left(\frac{U - m_\mu - W}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right) + \frac{\sigma_\eta^2}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\varphi\left(\frac{m_\mu + W - U}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right) \\
&\quad + \left(m_\mu + W\right)\Phi\left(\frac{m_\mu + W - U}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right) + \frac{\sigma_\mu^2}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\varphi\left(\frac{U - m_\mu - W}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right) \\
&= U\Phi\left(\frac{U - m_\mu - W}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right) + \left(m_\mu + W\right)\Phi\left(\frac{m_\mu + W - U}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right) \\
&\quad + \sqrt{\sigma_\eta^2 + \sigma_\mu^2}\,\varphi\left(\frac{U - m_\mu - W}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right)
\end{aligned}
$$

It then remains to integrate G_{t+1} and κ_{if}^0 out of U and W.

- **The value of employment.** Conditional on Ed_i, $\underline{G_{t+1}}$, $X_{it}+1$, w_{it}, $T_{ift}+1$, $\kappa_{ift}+\underline{u_{ift+1}}$, ε_i:

$$
\begin{aligned}
\mathbb{E}\max\left(\eta_{it+1} + U, W\right) &= \mathbb{E}\left[\left(\eta_{it+1} + U\right)\mathbf{1}\left\{\eta_{it+1} + U > W\right\}\right] \\
&= U\Phi\left(\frac{U - W}{\sigma_\eta}\right) + \sigma_\eta\varphi\left(\frac{U - W}{\sigma_\eta}\right).
\end{aligned}
$$

And it remains to integrate G_{t+1} and u_{ift+1} out of U and W.

Next,

$$\mathbb{E}\max\left(\begin{array}{c} \underline{\eta_{it+1}} + U\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, w_{it}, \varepsilon_i\right) \\ W\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, T_{ift}+1, \kappa_{ift} + \underline{u_{ift+1}}, \varepsilon_i\right) \\ \underline{\mu_{i\widetilde{f}}} + W\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, T_{i\widetilde{f}t+1}=0, \underline{\kappa^0_{i\widetilde{f}}}, \varepsilon_i\right) \end{array}\right)$$

can be simplified by conditioning on $G_{t+1}, u_{ift+1}, \kappa^0_{i\widetilde{f}}$:

$$\mathbb{E}\max\left(\eta_{it+1} + U, W, \mu_{i\widetilde{f}} + \widetilde{W}\right)$$
$$= \mathbb{E}\left((\eta_{it+1} + U)\mathbf{1}\left\{\eta_{it+1} + U > W \,\&\, \eta_{it+1} + U > \mu_{i\widetilde{f}} + \widetilde{W}\right\}\right)$$
$$+ W\Pr\left\{W > \eta_{it+1} + U \,\&\, W > \mu_{i\widetilde{f}} + \widetilde{W}\right\}$$
$$+ \mathbb{E}\left(\left(\mu_{i\widetilde{f}} + \widetilde{W}\right)\mathbf{1}\left\{\mu_{i\widetilde{f}} + \widetilde{W} > W \,\&\, \mu_{i\widetilde{f}} + \widetilde{W} > \eta_{it+1} + U\right\}\right).$$

Now

$$\mathbb{E}\left((\eta_{it+1} + U)\mathbf{1}\left\{\eta_{it+1} + U > W \,\&\, \eta_{it+1} + U > \mu_{i\widetilde{f}} + \widetilde{W}\right\}\right) = Up_1$$
$$+ \sigma_\eta\mathbb{E}\left(\frac{\eta_{it+1}}{\sigma_\eta}\mathbf{1}\left\{\frac{\eta_{it+1}}{\sigma_\eta} > \frac{W-U}{\sigma_\eta} \,\&\, \frac{\eta_{it+1} - \mu_{i\widetilde{f}} + m_\mu}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}} > \frac{m_\mu + \widetilde{W} - U}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right\}\right)$$
$$= Up_1 + \sigma_\eta\varphi\left(\frac{W-U}{\sigma_\eta}\right)\Phi\left(\frac{W - m_\mu - \widetilde{W}}{\sigma_\mu}\right)$$
$$+ \frac{\sigma_\eta^2}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\varphi\left(\frac{m_\mu + \widetilde{W} - U}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right)\Phi\left(-\frac{\sigma_\eta^2\left(W - m_\mu - \widetilde{W}\right) + \sigma_\mu^2(W - U)}{\sigma_\mu\sigma_\eta\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right)$$

for

$$p_1 = \Pr\left\{\eta_{it+1} + U > W \,\&\, \eta_{it+1} + U > \mu_{i\widetilde{f}} + \widetilde{W}\right\}$$
$$= \Phi_2\left(\frac{U-W}{\sigma_\eta}, \frac{U - m_\mu - \widetilde{W}}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}; \frac{\sigma_\eta}{\sqrt{\sigma_\eta^2 + \sigma_\mu^2}}\right).$$

Moreover,

$$\Pr\left\{W > \eta_{it+1} + U \,\&\, W > \mu_{i\widetilde{f}} + \widetilde{W}\right\} = \Phi\left(\frac{W-U}{\sigma_\eta}\right)\Phi\left(\frac{W - m_\mu - \widetilde{W}}{\sigma_\mu}\right)$$

and

$$\mathbb{E}\left(\left(\mu_{i\widetilde{f}}+\widetilde{W}\right)\mathbf{1}\left\{\mu_{i\widetilde{f}}+\widetilde{W}>W \ \& \ \mu_{i\widetilde{f}}+\widetilde{W}>\eta_{it+1}+U\right\}\right) = \left(m_\mu+\widetilde{W}\right)p_2$$

$$+ \sigma_\mu\varphi\left(\frac{W-m_\mu-\widetilde{W}}{\sigma_\mu}\right)\Phi\left(\frac{W-U}{\sigma_\eta}\right)$$

$$+ \frac{\sigma_\mu^2}{\sqrt{\sigma_\eta^2+\sigma_\mu^2}}\varphi\left(\frac{U-m_\mu-\widetilde{W}}{\sqrt{\sigma_\eta^2+\sigma_\mu^2}}\right)\Phi\left(-\frac{\sigma_\eta^2\left(W-m_\mu-\widetilde{W}\right)+\sigma_\mu^2\left(W-U\right)}{\sigma_\eta\sigma_\mu\sqrt{\sigma_\eta^2+\sigma_\mu^2}}\right)$$

for

$$\begin{aligned}
p_2 &= \Pr\left\{\mu_{i\widetilde{f}}+\widetilde{W}>W \& \mu_{i\widetilde{f}}+\widetilde{W}>\eta_{it+1}+U\right\} \\
&= \Phi_2\left(\frac{m_\mu+\widetilde{W}-W}{\sigma_\mu},\frac{m_\mu+\widetilde{W}-U}{\sqrt{\sigma_\eta^2+\sigma_\mu^2}};\frac{\sigma_\mu}{\sqrt{\sigma_\eta^2+\sigma_\mu^2}}\right)
\end{aligned}$$

And it remains to integrate $G_{t+1}, u_{ift+1}, \kappa_{i\widetilde{f}}^0$ out of $U, W, \widetilde{W}$.

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., A. DUNCAN, AND C. MEGHIR (1998): "Estimating Labour Supply Responses using Tax Reforms," *Econometrica*, 66(4).

BOWERS, N., A. SONNET, AND L. BARDONE (1999): "Giving young people a good start: the experience of OECD countries," in *Preparing Youth for the Twenty-first Century: the Transition from Education to the Labour Market*, vol. 7-84. OECD.

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.

CARNEIRO, P., J. HECKMAN, AND E. VYTLACIL (2006): "Estimating Marginal and Average Returns to Education," mimeo UCL.

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

DUSTMANN, C., AND U. SCHOENBERG (2008): "Apprenticeship Training and Commitment to Training Provision," UCL, mimeo.

——— (2009): "Training and Union Wages," Forthcoming Review of Economics and Statistics.

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.

ECKSTEIN, Z., AND K. I. WOLPIN (1999): "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities," *Econometrica*, 67(6).

GRILICHES, Z. (1971): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45(1), 1–22.

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., AND E. VYTLACIL (2005): "'Structural equations, treatment effects and econometric policy evaluation', Fisher-Schultz Lecture," *Econometrica*, 73(3), 669738.

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

HECKMAN, J. J., AND P. J. KLENOW (1998): "Human capital policy," in *Policies to Promote Human Capital Formation*, ed. by M. Boskin, vol. 1. Hoover Institution.

HOUSEOFLORDS (2007): "Apprenticeship: a key route to skill," House of Lords Select Committee on Economic Affairs: Fifth Report of Session 2006-07 on Apprenticeships.

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

KEANE, M., AND K. WOLPIN (2000): "Eliminating Race Differences in School Attainment and Labor Market Success," 18(4), 614–652.

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.

LEE, D. (2005): "An Estimable Dynamic General Equilibrium Model of Work, Schooling and Occupational Choice," *International Economic Review*, 46, 1–34.

LEE, D., AND K. I. WOLPIN (2006): "Intersectoral Labor Mobility and the Growth of the Service Sector," *Econometrica*, 47.

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

MORTENSEN, D. T., AND C. A. PISSARIDES (1994): "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61(3).

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.

SULLIVAN, P. (2006): "A Dynamic Analysis of Educational Attainment, Occupational Choices, and Job Search," Bureau of Labor Statistics, http://mpra.ub.uni-muenchen.de/861/.

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.

TALLIS (1961): "The moment generating function of the truncated multi-normal distribution," *Journal of the Royal Statistical Society, Series B*.

TAUCHEN, G., AND R. HUSSEY (1991): "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models," *Econometrica*, 59, 371–396.

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.

WOLPIN, K. I. (1992): "The Determinants of Black-White Differences in Early Employment Careers: Search, Layoffs,Quits, and Endogenous Wage Growth," *Journal of Political Economy*, 100(3), 835–560.