Institute for Fiscal Studies

Career Progression and Formal versus On-the-Job Training

IFS Working Paper 10/13

J. Adda
C. Dustmann
C. Meghir
J.-M. Robin

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

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

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

July 27, 2010

Abstract

We evaluate the German apprenticeship system, which combines on-the-job training with classroom teaching, by modelling individual careers from the choice to join such a scheme and followed by their employment, job to job transitions and wages over the lifecycle. Our data is drawn from administrative records that report accurately job transitions and pay. We find that apprenticeships increase wages, and change wage profiles with more growth upfront, while wages in the non-apprenticeship sector grow at a lower rate but for longer. Non-apprentices face a much higher variance to the shocks of their match specific effects and a substantially larger variance in initial level of the offered wages. We find no evidence that qualified apprentices are harder to reallocate following job loss. The average life-cycle return to an apprenticeship career is about 14% and the return is mainly driven by the differences in the wage profile.

Keywords: Apprenticeship Training, Job Mobility, Labour Supply, Wages, Wage determination, Matching, Wage Growth, Dynamic Discrete Choice, In-work Benefits, EITC, Education.

*We thank the anonymous referees and the editor as well as Steve Berry, Joe Altonji, David Card, Mariacristina De Nardi, Eric French, Guy Laroque, John Pencavel, Uta Schoenberg, seminar participants at the Chicago Fed, Duke University, the European Central Bank, the European University Institute, NYU, the New York Fed, the Minneapolis Fed, the London Business School, the 2005 SITE meeting at Stanford, UC Berkeley, UC Davis, the Labor workshop at Yale, the Department of Economics at Stanford and the Econometric Society European meeting for comments. We are grateful for funding from the DfES through the Centre for Economics of Education and to the ESRC through CEMMAP and the Centre for Fiscal Policy at the IFS. Costas Meghir also thanks the ESRC for funding through a Professorial Fellowship (RES-051-27-0204)

†University College London and IFS.

‡University of Paris 1, University College London and IFS.

1 Introduction

Vocational training is an important source of skill formation for the labour market. How this should be organised and what role it should play in educational systems attracts the interests of policy makers around the world. Germany operates an apprenticeship system targeted to 16 year olds, which consists of formal vocational training courses combined with on-the-job training that lead to certification of skills. These apprenticeships train workers in both white collar and blue collar jobs and are subsidized by the state, which funds the classroom component. Approximately 60% of each cohort participates in this scheme making it a major feature of the German labour market.

This system is often credited for the low youth unemployment rates in Germany, as it allows for a smooth and structured transition from school to vocational training and then into employment.[1] Throughout the 1990s, several countries, including Australia, the U.S., the U.K., France, and Norway, have attempted to expand or implement new firm-based apprenticeship schemes.[2] For instance, the U.K. government committed to train 35% of 16 year-olds within "modern Apprenticeship" schemes by 2010 - a target that has not been achieved, and current enrolment rates are closer to 15% (see Ryan, Gospel, and Lewis (2007) and Adult Learning Inspectorate 2006 for details). In the US Career Academies are attracting attention and they bare a close resemblance to the German apprenticeship system.[3] Thus, with such policies gaining in popularity, one important 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 secondary schooling. This is the issue we address in this paper.

[1]See e.g. Ryan (2001) for evidence. Jimeno and Rodríguez-Palenzuela (2003) document substantially lower youth unemployment rates in Germany (and Austria, which operates a similar scheme) than in all other European OECD countries.

[2]See Bowers, Sonnet, and Bardone (1999), and Dustmann and Schoenberg (2009), for an extensive discussion of the German and the UK system and House of Lords (2007) for some of the debate in the UK.

[3]To quote from Kemple and Wilner (2008) "Academy students take classes together, remain with the same group of teachers over time, follow a curriculum that includes both academic and career-oriented courses, and participate in work internships and other career-related experiences outside the classroom. [...] The Career Academies produced sustained earnings gains that averaged 11 percent more per year for Academy group members than for individuals in the non-Academy group."

A number of papers have considered the impact of apprenticeship training on wages. Krueger and Pischke (1995), Winkelmann (1996) and Fersterer and Winter-Ebmer (2003) all obtain similar OLS estimates for the wage returns to apprenticeship training in Germany and Austria, of around 15%-20%. Fersterer, Pischke, and Winter-Ebmer (2008) use an instrumental variables approach, based on information about the time to failure of firms that close down during the training period, which they use as an instrument for the length of training. Their IV estimates suggest wage returns between 2.5% and 4% per year of training, which are similar to their OLS estimates.[4]

The results from approaches such as those above are hard to interpret, even if we are just interested in the impact of apprenticeship on wages, because they ignore the role of endogenous experience profiles and the effects of selection into work. More generally, if we are to understand the full impact of an apprenticeship, we need to model the entire career path, starting with the original apprenticeship choice and followed by the period by period employment transitions, job mobility and wages. Since the two career paths may differ in job attachment, in available job opportunities as well as in wage growth, a structural approach is necessary that takes into account the dynamics of the life-cycle.

In this paper, we do just that. We specify a dynamic discrete choice model of the decision to enrol in apprenticeship training, of employment decisions, of job to job mobility and of wages. In the model individuals at 16 face the choice between joining a formal apprenticeship or the standard labor market. When working, their wages grow with experience and job (firm) specific tenure and depend on a match specific component as in Wolpin (1992); 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 and allowing for a rich stochastic specification as in the literature on the dynamics of wages.[5] The wage equations are specific to the two alternative careers (qualified apprentices or not) as in a Roy type model and are subject to aggregate shocks that affect relative wages between the two groups. Underlying choices is a flow utility function that is linear in income and depends on work

[4]An apprenticeship lasts usually three years.

[5]See for example Meghir and Pistaferri (2004), Low, Meghir, and Pistaferri (2009) and Altonji, Smith, and Vidangos (2009)

status. We draw from models of education choice[6] and wage determination.[7] Our modelling approach is closest to those dynamic models of 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 choice, labour supply and wages.[8]

Our data is drawn from administrative social security records, which track the careers and wages of individuals from when they make their educational choice and enter the labour market. Our sample covers men from all states of what used to be Western Germany and for the 1960-72 birth cohorts. The high quality of the data is an important strength of our approach: all transitions between employment and work and between different jobs as well as wages are recorded accurately by the firms thus avoiding recall bias.

The results show that apprenticeships lead to different wage profiles with more growth upfront, while wages in the non-apprenticeship sector grow at a lower rate but for longer. Overall wages are higher following an apprenticeship qualification. Moreover, non-apprentices face a much higher variance to the shocks to their match specific effects and in addition they face a substantially larger variance in initial level of the wage they are offered, which leads to a much larger dispersion of their wages and greater gains from job mobility relative to that of qualified apprentices. While we do find differences in job arrival and destruction rates the key difference between the two groups is in the wage profiles. The average life-cycle return to an apprenticeship career is about 14%. Finally, we find no evidence that qualified apprentices are harder to reallocate following job loss.[9] Particularly after some years of experience their job arrival rates are very high and their job destruction rates low.

The remaining part of the paper is structured as follows. First we describe the

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

[7]Willis and Rosen (1979), 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).

[8]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.

[9]see Heckman (1993)

apprenticeship system and provide a descriptive analysis of the data in section 2. In Section 3 we describe the model. Then in Section 4 we present the estimation method followed by the results in section 5. We conclude in section 7.

2 Background and Data

In this section, we give some brief description about the firm based training system we are analysing in this paper. We then describe our data and sample, and provide some descriptive statistics.

2.1 The Apprenticeship System

The German Apprenticeship System is a vocational training programme which combines on-the-job training, provided by the firm, with school education, provided and funded by the state. No other subsidies are involved, other than the classroom component. Similar systems operate in Austria and Switzerland.

The system offers training in 541 white- and blue collar occupations[10]. However, there is a strong concentration in a fairly small number of occupations: In 1992, about 50% of all males were concentrated in 11 occupations, with slightly more than half of those being blue collar ones. Individuals typically enter apprenticeship after completion of lower or intermediate secondary school at about 16.[11] Apprenticeship for our cohorts last about 3 years. During this time, apprentices attend vocational state schools (typically one or two days a week), where they acquire general knowledge, as well as knowledge which is specific to their occupation. The remaining days, they train on the job at their firm under the supervision of qualified personnel. Having successfully completed a set of examinations, the apprentice graduates with a professional qualification.

In our analysis, we consider all individuals who enter the labour market with a lower secondary degree which is not a sufficient qualification for attending university, and is typically obtained by the age of 16. We then define two groups: those who enrol in

[10]See http://berufenet.arbeitsagentur.de/berufe/index.jsp. for details.

[11]Germany tracks children after the age of 10 in lower, intermediate and upper secondary schools. While pupils who go through lower and intermediate secondary schools would typically enrol in blue or white collar apprenticeship schemes, only pupils who attend upper secondary schools are entitle to enrol directly into University. See Dustmann (2004) for a detailed description of the German school system.

apprenticeship schemes for at least 2 years and successfully complete their training (we refer to these as "apprentices"), and those who enrol for a shorter period, but do not graduate, or do not enrol and enter the labour market directly (we refer to these as "non-apprentices" or "unskilled"). There are also one-year vocational courses (Berufsgrundbildungsjahr or Berufsvorbereitungsjahr), which do not lead to vocational degrees. Thus, among the non-apprentices, we may include some who were exposed to some apprenticeship training, or who attended a vocational preparatory classes, without following up further training.[12]

As an alternative to firm-based apprenticeship training, some youth attend vocational schools, which offer classroom training for two to three years, with unpaid work experience, and lead to a certificate equivalent to a firm-based apprenticeship. This is more important for women (who are not included in this study), and predominant in health related occupations. About 6% of our sample undertakes qualifying training in these vocational schools. We add these to the group of apprentices, but we treat them differently in some respects, as the model section makes clear.[13]

Finally, a few words on the wage setting: Germany operates a collective bargaining system at the industry- and state 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). In the late 1990's about 56% of all firms in West Germany did belong to an employer federation, employing 73% of all workers, who were thus covered by a union agreements (see Dustmann and Schoenberg (2009)). Thus, while unions play an important role over our sample period, bargained wages only set a wage floor.[14] Overall, 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. If these minimum wages bite this will be reflected in our model in increased proportions

[12]There is a delay in the start of work for non-apprentices in our data of about 8 months relative to the start of apprenticeships. This may reflects these vocational courses as well as military service (Germany has a compulsory military service) and time to locate a job as an unskilled worker.

[13]Wage profiles of those who went through firm based training and vocational schools are almost identical, with an average difference of about 0.8%. This is in line with the findings of Parey (2009).

[14]see Gathman and Schoenberg (2010) for a more detailed analysis.

out of work due to out of work utility.

2.2 Data and Descriptive Analysis

Our main data is a 2% sample of administrative social security records organized by the German IAB[15]. The data set starts in 1975 and records all work spells with exact start and end dates up to 1996. It 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. All transitions are recorded accurately with specific dates that each job started and ended.[16] 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.

At age 10 children are separated into an academic track, that can eventually lead to admission to University and to a vocational track that leads up to the apprenticeship choice at 16. Although we are not modelling this choice directly we need to account for it because the composition of individuals who enter the two tracks may differ across cohorts, causing selection bias - an initial conditions problem that is.[17] We can only infer who has made this choice once we see individuals in the labor market because this is when they get included in the data set; 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 can only use those cohorts who are old enough to be observed at 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 data set reports the average daily pre-tax wage each year if the individual stays 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

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

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

[17]One could conceive of extending the structural model to that allocation as well. However, one would need to recognise that the decision process at that age will be different and involve parents to a much greater degree.

firms.

It would be intractable to model all aspects of this detailed data. We thus assume that all decisions are made on a quarterly basis. Whenever during a quarter an employment and an unemployment spell are both present we assign the spell to one of these depending on which of the two covers the largest proportion of that quarter. When the individual does not move and thus the wage we observe is an average over more than one quarter we treat this as a time aggregated wage where we do not observe the individual constituents of this average. This time aggregation problem is fully accounted for during estimation as we explain later.

Our main sample and focus of study consists of West-German male cohorts born in the period 1960-72, who end formal education at 15/16 and who either work or join an apprenticeship after school.[18] 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 data contains 57,183 apprentices and 6975 non-apprentices. These are followed through time, quarter after quarter up until 1996; we have thus a total of 2,732,394 quarterly observations. Finally, to identify the determinants of choices of school tracks at age 10, we use 69,084 individuals who follow the vocational track and 10,608 who follow the academic one. We provide more detail on the sample selection in the web Appendix.

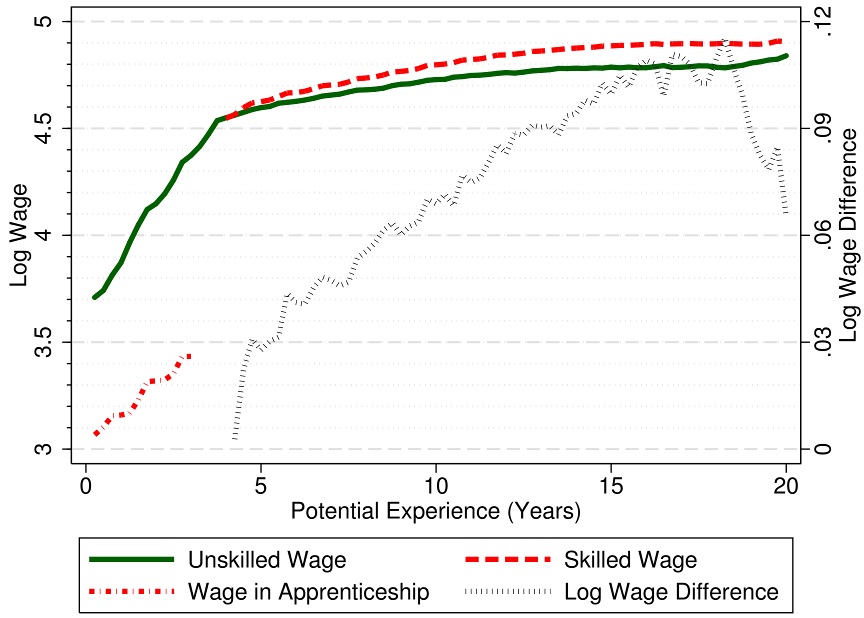
2.3 Descriptive Analysis of the Data

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%,

[18] Germany operates a school system where pupils are tracked at the age of 10 into three secondary school choices, where only the highest track allows for direct enrolment into university. Those graduating from the intermediate and lower track schools have no direct access to university, and can thus choose between vocational training and direct labour market entry. This is the population we analyse.

Figure 1: Log Wage by skill and the wage gain for qualified apprentices



resulting in a 1.2% real average growth per year. During 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. 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 by allowing for such differences in the model that will follow.

Indeed, wages are only one dimension in which education groups may differ. Another important dimension is labor market attachment. Table 1 displays the quarterly transition probabilities by education and time in the labor market. Unskilled workers have a higher probability of dropping out of work. During the first five years in 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

Table 1: Observed Quarterly Labor Market Transitions

	Potential Experience (Years)					
Labor Market Transitions	Non-Apprentices			Apprentices		
	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

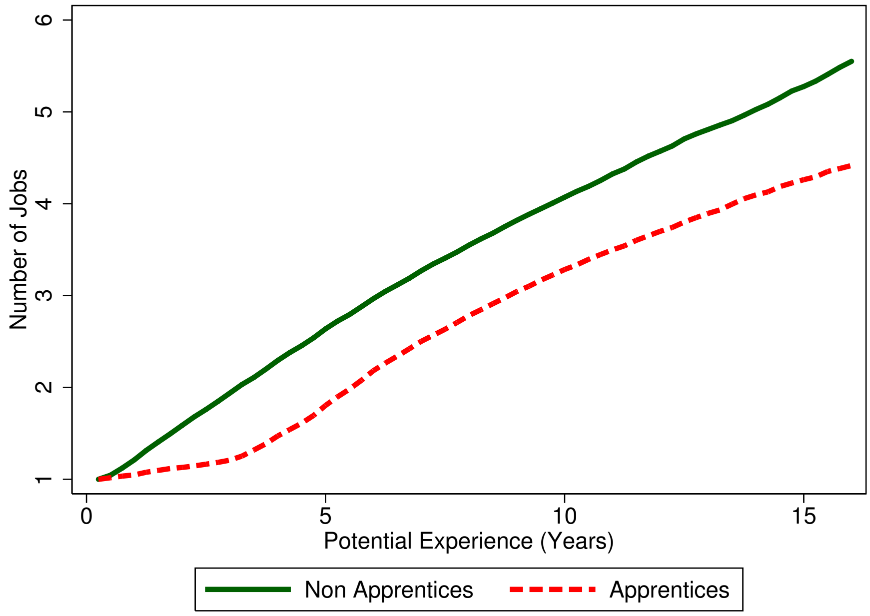
senior workers, but the education difference still persists. The probability of job to job transitions is higher at the beginning for non-apprentices and after five years declines for both groups and becomes marginally higher for the qualified apprentices.

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. The greater job attachment and the resulting higher earnings acts to "compensate" the apprentices for the lost earnings early on.

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 between the groups comes from the early years, where workers during their apprenticeship training period are less mobile. However they never catch up following qualification. Overall, the mobility numbers are much lower than those in the U.S. as documented in Topel and Ward (1992) amongst others.

Decomposing Wage Growth. Wage growth occurs both within firm and as a result of firm mobility. Job shopping, 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)).

Figure 2: Mobility: Number of Jobs, by Education



In Germany, despite lower mobility rates, this is also the case. This is illustrated in Figure 3 which shows within firm wage growth by potential experience and skill level and in Figure 4, which displays the growth of wages following a job to job transition. The wage growth in the latter case can be substantial, at 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. Job mobility is an important source of wage growth particularly for non-apprentices, which will need to be accounted for in the model. Carrying out a simple decomposition exercise, for the non-apprentices it accounts for 9.3% of growth of wages over 20 years is accounted for by job-to-job mobility. For those following an apprenticeship career the figure is smaller at 6.6% for wage growth that follows the training period. As we shall see from the model the jobs facing the non-apprentices are much more heterogeneous and hence they face greater gains from search.

Figure 3: Annual Change in Log Wage (Conditional on Staying with same Employer)

Apprenticeship Training and Wages. As a descriptive device and for comparability with more standard methods we present the results of regressing log wages on apprenticeship for different age groups. We use both OLS and IV, where the instruments we use are the same that provide the exogenous variation in our structural model, which we present later. These are the region where the individual lived when taking the apprenticeship interacted with the residual of aggregate GDP from a quadratic trend. All regressions include time dummies, and regional dummies and are estimated for individuals over 20 so that most would have completed their apprenticeship. The assumption we make, which we discuss below in section 3.2 is that the region×gdp shocks reflect varying costs of apprenticeships over time and region, and more generally local shocks to labour demand, but that the labour market is sufficiently integrated for these differential shocks not to affect wages.

To check the first stage for the IV regression, we run a probit for apprenticeship choice including region effects, time effects, and the interactions of region with the GDP residual. The latter have a p-value of zero establishing that indeed aggregate shock affect participation in apprenticeship differently in different regions.

The results for the wage regressions are presented in Table 2. The OLS results are lower than the IV ones and the pattern is different with the ones from IV declining

Figure 4: Between job wage changes

Table 2: Apprenticeship and wages by age

	(1)	(2)	(3)	(4)	(5)	(6)
Age	21-23	23-26	26-28	28-30	30-32	All
OLS	0.044	0.065	0.082	0.099	0.113	0.073
	(0.0039)	(0.0039)	(0.0042)	(0.005)	(0.005)	(0.0037)
IV	0.006	0.363	0.249	0.155	0.264	0.097
	(0.023)	(0.056)	(0.081)	(0.093)	(0.116)	(0.027)

Note: The reported coefficients are the coefficients on the apprentice dummy. Regressions include time dummies, and regional dummies. Exclusion: Interaction region-gdp in year of Labour Market Entry. Asymptotic standard errors in parenthesis. First Stage: F(10,53249)=3.69, P-value: 0.0001

after age 23. Over all ages, the OLS return is 7.3% and the IV return 9.7%. These estimates need to be interpreted with great caution because they do not control for a number of issues: these include selection into work, job mobility and growth of wages with experience. Indeed ignoring all these aspects may also invalidate the exclusion restrictions; the dynamic model that follows solves these problems and considers all aspects of the careers including costs so as to estimate the returns life-cycle returns

We then repeat the exercise for the effect of apprenticeship on the probability of being in work in any one quarter with results presented in Table 3. When all ages are pooled OLS implies a 15.6 percentage point increase in employment, while IV just an

Table 3: Apprenticeship and Probability of being employed per year by age

	(1)	(2)	(3)	(4)	(5)	(6)
Age	20-23	21-25	23-27	25-29	27-31	All
OLS	0.160	0.174	0.171	0.159	0.133	0.156
	(0.005)	(0.006)	(0.007)	(0.008)	(0.009)	(0.005)
IV	-0.012	0.05	0.236	0. 308	0.310	0.011
	(0.03)	(0.08)	(0.12)	(0.14)	(0.17)	(0.037)

The reported coefficients are the coefficients on the "apprentice" variable. Regressions include potential experience dummies, time dummies, and regional dummies. Asymptotic standard errors in parenthesis. Exclusion: Interaction region-gdp shock in year of Labour Market Entry. First Stage: F(10, 53249)=3.69, P-value: 0.0001

insignificant 1.1 pp effect. Indeed the results from the model will show that the key impact of apprenticeship lies in the wage profiles and less so in job attachment.

3 Model

The model is set in discrete time (quarters) and focuses on the population that was allocated into the vocational education track at 10 and are completing this form of secondary education 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.

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 apprenticeship 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 apprenticeship system. We assume that both an unskilled job and an apprenticeship position are available immediately.[19] 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 stan-

[19]This is a simplification: on average individuals start the apprenticeship at 17.1 years of age and an unskilled job at 17.8. These differences between graduation age and labour market entry may be due to undertaking short vocational courses or compulsory military service before finally opting for an apprenticeship or a non-apprenticeship career. We start modelling from the point we see them joining the first job or an apprenticeship scheme.

dard 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 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 de-trended 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.[20] The macro state variable G_t is modelled as a discrete two state Markov process of order 1. The transition probabilities are presented in web Appendix B in Table 12. We now describe the model formally and then discuss estimation.

3.1 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 a rule that yields an annual wage w_{ift} to the worker; the way the split is determined is not modelled here.[21] 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 shocks to match productivity, and, as in a standard Mortensen and Pissarides (1994) model, really bad productivity shocks may result in unemployment.

Wages are modelled as follows. Let $Ed_i \in \{A, NA\}$ denote the worker's apprentice-

[20] An issue of concern here is the appropriate notion of a business cycle. Under full factor price equalisation 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.

[21] For equilibrium wage determination with shocks to firm productivity and heterogeneous workers see Lise, Meghir, and Robin (2009).

ship 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.[22] Let T_{ift} denote the number of quarters 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 ($Ed_i = 1$ for qualified apprentices and zero otherwise), 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, 10 and 30 years of experience and tenure. The specification is motivated by the fact that most of the non linearity in wages profiles is early on, so we have a denser grid between 0 and 10 years of actual experience. Unobserved heterogeneity affects the overall level of log wages and the wage return to apprenticeship.[23] 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 ($T_{ift} = 0$) they draw a match specific effect $\kappa_{ift} = \kappa^0_{if}$ such that

$$\kappa^0_{if} \sim \mathcal{N}(0, \sigma^2_0\,(Ed_i)).$$

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

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

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

[23] 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.

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 7), 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.

Workers are assumed risk neutral, which also implies that liquidity constraints are not an issue of concern for this model. Thus, employed workers value the current wage $w(Ed_i, G_t, X_{it}, T_{ift}, \kappa_{ift}, \varepsilon_i)$ with a linear utility function. In addition, we allow for a mobility cost or benefit μ_{if} when a worker moves between jobs. This allows for the possibility that workers may move to a job that pays lower wages, as is observed in the data. The one off benefit/cost of moving is an iid random variable μ_{if} such that

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

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 as $w_{i(-1)}$), 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%.[24] In addition, there is a utility of leisure which varies across individuals on the basis of education, experience, unobserved heterogeneity ε_i and a Gaussian white noise η_{it} with variance σ^2_η. Thus, the instantaneous utility of unemployment is:

$$R^U_{it} \equiv R^U(Ed_i, X_{it}, w_{i(-1)}, \eta_{it}) = \gamma_0(\varepsilon_i) + \gamma_U w_{i(-1)} + \gamma_X(X_{it}, Ed_i) + \eta_{it},$$

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

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).

[24] We have taken a replacement rate that is on average correct for our population. Modelling the entire system would imply an increased state space.

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

3.2 The 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. All value functions are indexed by a subscript a to denote their dependence on age.

The value of unemployment. At the end of period t, unemployed individuals draw a job offer with probability $\pi^U_{it} \equiv \pi^U(G_t, Ed_i, X_{it})$ 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 $U_a\left(Ed_i, G_t, X_{it}, w_{i(-1)}, \varepsilon_i\right)$, where the subscript a denotes the age of the individual, we can write

$$\begin{aligned} U_a\left(Ed_i, G_t, X_{it}, w_{i(-1)}, \varepsilon_i\right) &= \gamma_U w_{i(-1)} + \gamma_0(\varepsilon_i) + \gamma_X(X_{it}, Ed_i) & A \\ +\beta\pi^U_{it}\ \mathbb{E}\max &\begin{pmatrix} \underline{\mu_{if}} + W_{a+1}\left(Ed_i, \underline{G_{t+1}}, X_{it}, T_{ift+1}=0, \underline{\kappa^0_{if}}, \varepsilon_i\right) \\ U_{a+1}\left(Ed_i, \underline{G_{t+1}}, X_{it}, w_{i(-1)}, \varepsilon_i\right) + \underline{\eta_{it+1}} \end{pmatrix} & B \\ +\beta(1-\pi^U_{it})\ &\mathbb{E}U_{a+1}\left(Ed_i, \underline{G_{t+1}}, X_{it}, w_{i(-1)}, \varepsilon_i\right) & C \end{aligned} \tag{2}$$

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^U_{it}. 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, $W_{a+1}(\cdot)$, 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 $\delta_{it} \equiv \delta(G_t, Ed_i, X_{it})$,which depends on the state of the business cycle as well as experience and apprenticeship status. Conditional on not being laid off, they draw an alternative job offer with probability $\pi^W_{it} \equiv \pi^W(G_t, Ed_i)$. 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 3-5 years for those who qualified (i.e. for the first three years following their qualification). Following this initial period $\delta(G_t, Ed_i, X_{it})$ can be interpreted as the standard job destruction rate.

Their value of employment is then given by

$$\begin{aligned}
&W_a\left(Ed_i, G_t, X_{it}, T_{ift}, \kappa_{ift}, \varepsilon_i\right) = w_{it} && A\\
&+\beta\delta_{it}\ \mathbb{E}\left[U_{a+1}\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, w_{it}\right) + \underline{\eta_{it+1}}\right] && B\\
&+\beta\left(1-\delta_{it}\right)\pi^W_{it}\ \mathbb{E}\max\begin{pmatrix} U_{a+1}\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, w_{it}, \varepsilon_i\right) + \underline{\eta_{it+1}} \\ W_{a+1}\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, T_{ift}+1, \kappa_{ift} + \underline{u_{ift+1}}, \varepsilon_i\right) \\ \underline{\mu_{i\tilde{f}}} + W_{a+1}\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, T_{i\tilde{f}t+1}=0, \underline{\kappa^0_{i\tilde{f}}}, \varepsilon_i\right)\end{pmatrix} && C\\
&+\beta(1-\delta_{it})(1-\pi^W_{it})\ \mathbb{E}\max\begin{pmatrix} U_{a+1}\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, w_{it}, \varepsilon_i\right) + \underline{\eta_{it+1}} \\ W_{a+1}\left(Ed_i, \underline{G_{t+1}}, X_{it}+1, T_{ift}+1, \kappa_{ift} + \underline{u_{ift+1}}, \varepsilon_i\right)\end{pmatrix} && D
\end{aligned} \tag{3}$$

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; 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[25]) we assume that the training firm pays the worker only a fraction λ_A of his productivity as a non-apprentice ($w\left(Ed_i = 0, G_t, X_{it}, T_{it}, \kappa_{it}, \varepsilon_i\right)$), the rest presumably serving as payment for the general training received.[26] 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:

$$\begin{aligned}
W_a^A\left(G_t, X_{it}, T_{ift}, \kappa_{ift}, \varepsilon_i\right) &= \lambda_A \cdot w\left(Ed_i = 0, G_t, X_{it}, T_{ift}, \kappa_{ift}, \varepsilon_i\right) && A \\
+\beta\pi_A(G_t)\ \mathbb{E}\max &\begin{pmatrix} W_{a+1}^A\left(\underline{G_{t+1}}, X_{it}+1, T_{ift}+1, \kappa_{ift}+\underline{u_{ift+1}}, \varepsilon_i\right) \\ \underline{\mu_{i\widetilde{f}}} + W_{a+1}^A\left(G_{t+1}, X_{it+1}, T_{i\widetilde{f}t+1}=0, \underline{\kappa^0_{i\widetilde{f}}}, \varepsilon_i\right) \end{pmatrix} && B \qquad (4) \\
+\beta[1-\pi_A(G_t)]\ \mathbb{E}W_{a+1}^A &\left(\underline{G_{t+1}}, X_{it}+1, T_{ift+1}, \kappa_{ift}+\underline{u_{ift+1}}, \varepsilon_i\right) && C
\end{aligned}$$

where as before, the expectation operator $\mathbb{E}$ relates 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 associated 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.

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

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

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 $Ed_i = 0$ (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

$$\begin{aligned} V_a^A\left(G_t, \kappa_{if}^0, \mu_{if}, R_i, \varepsilon_i, \omega_{it}\right) = \quad & \mu_{if} + W_a^A\left(G_t, X_{it} = 0, T_{ift} = 0, \kappa_{if}^0, \varepsilon_i\right) \\ & - \left[\lambda_R(R_i, G_t) + \lambda_0(\varepsilon_i)\right] - \omega_{it} \end{aligned} \tag{5}$$

The last two terms represent direct costs. ω_{it}, is a normally distributed iid cost shock revealed to the individual before the choice is made. $\lambda_R(R_i, G_t) + \lambda_0(\varepsilon_i)$, represents the direct costs of apprenticeship, which merits some discussion. There are no fees due for apprenticeship training, but other costs, which the worker may not have to incur if instead they obtain an unskilled job, can play an important role in determining apprenticeship choice. For example if the individual needs to travel far for an apprenticeship, which may not be available close by, and even possibly move out of the parental home the direct cost of an apprenticeship will increase. On the other hand if they just obtain an unskilled job they may be able to work close to home thus economising in travel and housing costs. Thus the relative scarcity of apprenticeships will drive the cost of training. Since such scarcity is driven by the overall economic conditions we proxy these by including interactions between region (R_i) and the deviation of aggregate GDP from a quadratic trend (G_t), both measured when the choice is made at 16; these interactions reflect how aggregate shocks affect each of the eleven regions of (West) Germany. Such differential effects across regions will occur because of the differing industrial composition across regions and because some industries are more pro-cyclical than others. Thus these interactions are the source of exogenous variation driving apprenticeship choice. The availability of data for thirteen birth cohorts observed in eleven states provides plenty of variability. Indeed in section 2.3 we show that these variables have a strong and significant effect on the take up of apprenticeships. The estimates of the structural model confirm this (see Table 6).

Finally, we remove the effects of region of residence at 16 as well as aggregate trends when we compute the moments that we use to fit the model, thus controlling for cross region heterogeneity, as described in section 4.2. However, we exclude interactions of region with gdp shocks from the wage equation. This assumes that the labour market in Germany is sufficiently integrated to make fluctuations of wages across regions unimportant. While we recognise this to be just an approximation relaxing this assumption would involve accounting for mobility across regions greatly increasing the state space (see Kennan and Walker (2010)).[27]

The choice to become an apprentice is governed by

$$V_a^A\left(G_t, \kappa_{if}^0, \mu_{if}, R_i, \varepsilon_i, \omega_{it}\right) > W_a(Ed_i = 0, G_t, X_{it} = 0, T_{it} = 0, \kappa_{if'}^0, \mu_{if'}, \varepsilon_i), \qquad (6)$$

where κ_{if}^0, μ_{if}, $\kappa_{if'}^0$ and $\mu_{if'}$ and represent the match specific characteristics and one off transition costs in the alternative career options respectively. Age a is 16 at this point. 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-pecuniary benefits, need to be integrated out because they are not observed. We allow for unobserved heterogeneity in the costs so as 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.

As explained in section 2.1, about 6.8% of our apprentices instead of joining a standard apprenticeship scheme, attend vocational school for 2-3 years with unpaid work experience. In the data their outcomes are very similar in all respects to the standard apprenticeship group and the average difference in wage is less than 1%. Thus we decided to account for them in a simple way as follows: when an individual receives the apprenticeship offer, this offer is associated with some probability (which we estimate) with a zero wage (rather than the positive wage associated with the standard apprenticeship option). Second, while trainees in the standard apprenticeship scheme start post-training work with three years of experience (equal to the number of periods of training), this

[27]There is some evidence that wages respond to economy wide shocks, while labour demand is more locally determined. See for example Head and Mayer (2006). Also note that job arrival rates relative to job destruction rates turn out to be very high. So given the model, the search frictions for employment turn out to be very low, which is likely to make such an approximation more realistic.

group starts with a level of experience that we estimate as a free parameter. This allows for a shift of the wage profile of this group to account for a lower level of work experience. In all other respects we treat them as qualified apprentices.

The time horizon and the terminal condition We solve for the value functions at each age by backwards induction from retirement, which occurs at 65 years of age, to the start of the labour market career when the apprenticeship choice is made at 16. 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.[28] Having a terminal point beyond our observation window requires assumptions on the returns to experience and tenure. Noting from the data (see Figure 6 or Table 13 in the web appendix) that there is almost no wage growth beyond 11 years of potential experience we imposed that the returns to experience and tenure are constant between 10 and 30 years of actual experience.[29] We then assume that there is no wage growth beyond 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. Further computational details can be found in Appendix B.

3.3 Unobserved heterogeneity

Wages and apprenticeship costs depend on unobserved heterogeneity. In general it may be far too restrictive to allow just for one factor heterogeneity (see for example Taber (2001)). We thus assume that the vector ε_i consists of two random variables which follow a bivariate discrete distribution, each with two points of support. One element enters the cost of apprenticeship while the other enters 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.[30] 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

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

[29]Thus, extrapolating from our data which stops at 20 years of experience

[30]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.

Table 4: Proportion in different education tracks by Year of birth

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

labour literature.[31]

4 Estimation

4.1 The Selection of our Population and 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 4, there is a steady (but small) increase in the proportion following the academic track over time (apprentices and non-apprentices in the table), which could point to a change in composition of our population of interest.

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 as of the two factors of unobserved heterogeneity in the vector ε_i. The key assumption in this approach is that the distribution of unobserved heterogeneity is independent of region and cohort.

4.2 Method of Simulated Moments

Because of the nature of the administrative data we have to deal with a time aggregation problem that manifests itself in two ways: first, the wages for those who do not move jobs are recorded as an average annual pay and are observed as constant throughout the year. For those who do move the pay that is observed is the average over the period

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

of the year that they were in the respective firm.[32] The pay record gets updated in the new firm and remains constant until the end of the year or until the person changes jobs again (whichever is earlier). Second, when a worker switches from apprenticeship to a regular job, following his qualification, but does not change firm, we only observe the average between the wage during the apprenticeship and the wage in the regular job, in that year. These two features of the data lead to a difficult time aggregation problem because in our model all decisions and shocks occur at a quarterly frequency.

Given our model, constructing the likelihood for the observed data is complicated enough, without this time aggregation problem. Accounting for time aggregation would involve multidimensional integrals accounting for all the possible ways that the quarterly wages could give rise to the observed wage records (annual or part thereof). A much more practical approach is to use the method of simulated moments.[33]

Thus the parameters of the model are estimated by minimising the distance between the set of chosen moments from the data and the moments implied by the simulated careers from the model. The criterion we minimise takes the form:

$$M(\theta) = (\hat{m} - g^S(\theta))'\hat{\Omega}^{-1}(\hat{m} - g^S(\theta))$$

where $\hat{m}$ represents a vector of data moments, $g^S(\theta)$ represents the moments implied by the model and based on S simulated careers and $\hat{\Omega}$ represents a weight matrix. Here we chose $\hat{\Omega}$ to be a diagonal matrix which contains the variances of the observed moments. The standard errors are estimated as in Gourieroux, Monfort, and Renault (1993)

$$\sqrt{N}(\theta_N - \theta_0) \xrightarrow{d} \mathcal{N}(0, W(S, \Omega, \theta_0))$$

where the covariance matrix $W(S, \Omega, \theta_0)$ is given as

$$W(S, \Omega, \theta_0) = \left(1 + \frac{1}{S}\right) [H^S(\theta)'\Omega^{-1}H^S(\theta)]^{-1}H^S(\theta_0)'\Omega^{-1}\Sigma(\theta_0)\Omega^{-1}H^S(\theta_0)[H^S(\theta)'\Omega^{-1}H^S(\theta)]^{-1}, \tag{7}$$

with $H^S(\theta) = \text{plim}_{N\to\infty}\partial\overline{g}_N^S(\theta)/\partial\theta'$ being the Jacobian of the auxiliary statistics with respect to the structural parameter vector and $\Sigma(\theta_0)$ is the theoretical covariance matrix

[32] For example, if someone moved jobs once during a year, in June say, we will observe the average pay from January to June and the average pay in the new firm for the rest of the year.

[33] See Lerman and Manski (1981), McFadden (1989) and Pakes and Pollard (1989)

of the moments.

Estimation is based on the simulation of 12,000 individual careers, starting from the point when at 10 years of age they are allocated to the vocational or the academic track.

Using the simulated data we construct moments that correspond to those we construct from the observed data. Dealing with time aggregation in wages involves simply generating simulated data at the quarterly frequency by the model and then imposing the same time aggregation on the simulated data as the one that produced the real data and then constructing the moments in the same way.

We use a total of 390 moments and we have a total of 118 parameters to estimate. This decomposes into 148 moments and 72 parameters characterizing the career paths of apprentices and non apprentices, 121 moments and 13 parameters for the apprenticeship choice at age 16 and 121 moments and 33 parameters for the choice of the academic track at age eleven. We characterize the career path of individuals by a number of linear regressions. We first regress the (log) wage level on a function of experience, tenure and business cycle for skilled and unskilled individuals. This set of moments helps identify the return to experience and tenure by skill groups. Second, we regress the squared residual of this equation on a constant, a function of experience and education choice. This helps to identify the variance of wages, which are governed by the distribution of initial matches and unobserved ability. Third, we regress the change in log wage on a function of experience, tenure, business cycle and skill group, which helps to identify match specific heterogeneity, as well as the return to tenure and experience. Fourth, we regress the squared residual of this equation on skill groups, to identify the innovation to the match specific effect.

We also use linear probability models to characterize the proportion of individuals in work and linear regression to describe the number of jobs held as a function of potential experience and business cycle. For the choice of apprenticeship at age 16, we use as moments the proportion of apprentices by region and year. We proceed in a similar way for the choice of academic track at age 10 by matching the proportion of individuals who chose the lower track by region and year. A full list of moments can be found in the Tables of Web Appendix C. Finally, in constructing the moments we account for heterogeneity due to initial region of residence at 16 as well as aggregate time trends by including regional dummies and a quadratic trend.

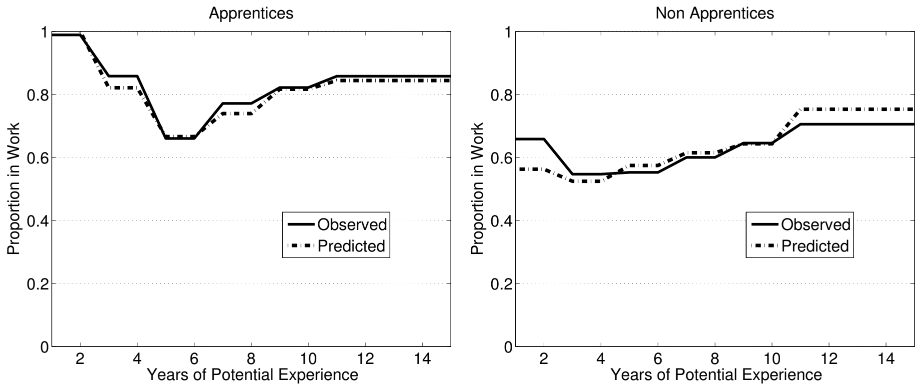
The estimation was done using a combination of Newton-Raphson methods (the e04ucf routine from the NAG library) and the simplex algorithm. To avoid local minima, we restarted the estimations from many different initial guesses.

5 Estimation Results

5.1 The Fit of the Model

We summarise the fit of the model by comparing some of its key predictions to the data. The model fits these remarkably well and we refer the reader to web Appendix C where the results are shown in detail.[34]

Figure 5: Observed and Predicted Employment Profiles



[34]We do not assess the fit of the model using chi-square tests. Given the very large amount of data we use for the moments and given the degree of overidentification, it is expected even small deviations from the data moments will be seen to be significant.

Figure 6: Observed and Predicted Wage Profiles

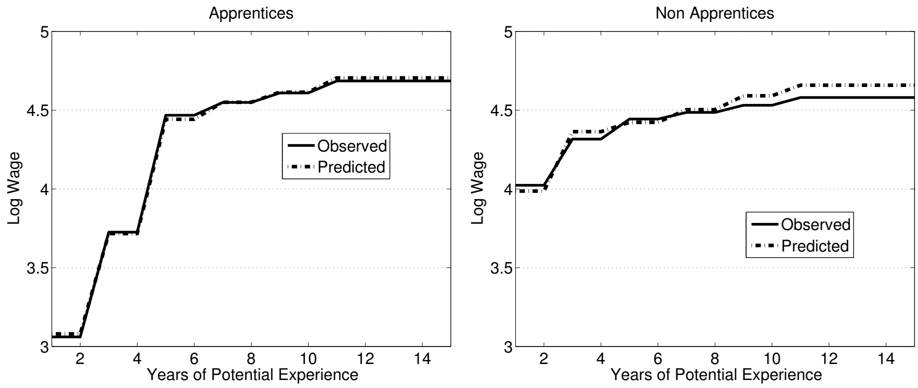
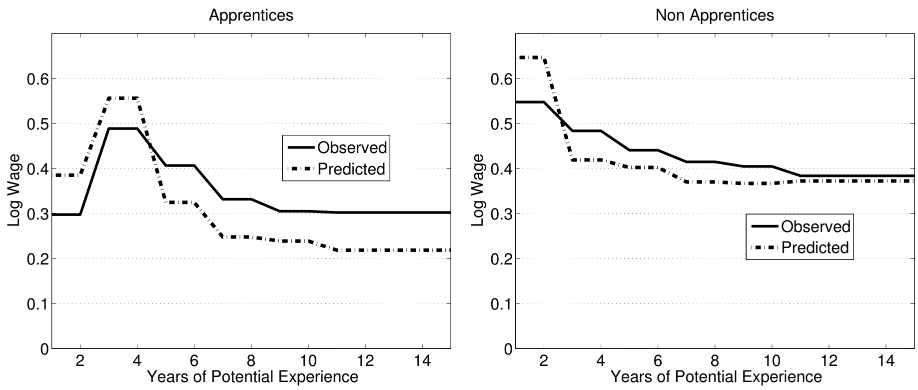


Figure 7: Observed and Predicted Standard Deviation of Wages



In the data, the proportion of non-apprentices is 11.6%, and the model prediction is 11.5%. The proportion of individuals opting for the highest academic track (and hence out of vocational training) is 13.3% in the data, and 10.4% in our model. Figures 5 and 6 present the fit of the model for the proportion of individuals working and wage profiles by skill groups. The model does a good job in matching the U shaped profile of the proportion of individuals in work. One exception is the work pattern for non-apprentices in the first two years of their career, which is slightly under-predicted.

Figure 7 plots the standard deviation of wages by skill groups as a function of potential experience. This graph also serve as data description for this aspect of the German labor market. This is interesting for the different pattern it displays to the one known for the U.S.[35] 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. Note that the cross-sectional variance of wages is substantially higher for non-apprentices; this will be reflected in the results from the model discussed below.

5.2 The Parameter Estimates

Transition probabilities and costs. Table 5 presents some key parameters that determine the careers of individuals.

Exogenous quarterly destruction rates, are high at low levels of experience reflecting partly departures for military service. Amongst experienced workers they are twice the size for non-apprentices than for apprentices, although they are both quite small. Interestingly job destruction rates are not sensitive to business cycle conditions, even when we account for endogenous quits, which has been noted before in many other contexts.[36]

The job arrival rates among the unemployed increase with experience. For non-

[35]see for example Low, Meghir, and Pistaferri (2009) among many others.

[36]See for example Pissarides, Layard, and Hellwig (1986) for the UK and Petrongolo and Pissarides (2008) for more recent information on the UK, France and Spain. They interpret this as reflecting firing costs from the side of the firm due to regulations. Similar results are obtained for the US by Hall (2005) and Shimer (2007).

apprentices they display some cyclicality particularly for inexperienced workers. For apprentices they are quite low during the training period and highly cyclical in the subsequent early part of the post-training career, becoming 1 for higher levels of experience. During their early career the qualified apprentices have lower arrival rates. On the basis of the arrival and destruction rates the level of search frictions are low (π/δ is high) for both groups. However, this is just part of the story underlying the transitions between unemployment and employment because the standard deviation of job offers (σ_0) is much higher for the non-apprentices than it is for the apprentices (0.77 to 0.19). This means that many offers will be rejected by the former. On the other hand, since bad offers can always be turned down this large variance also reflects potential benefits from search for the non-apprentices. But this is not all: when in work the standard deviation of the innovations (σ_u) to the match specific effect are also four times as high for the non-apprentices.[37] Thus, relative to apprentices, they face a very heterogeneous set of jobs with quite high permanent shocks to the value of the match effect. These effects combine to imply a higher unemployment rate for non-apprentices and a larger contribution of search to wage growth, as we confirm below and as discussed in the data section.

Among the employed the arrival rate of job offers are much lower: for the qualified apprentices it is 0.2 jobs per quarter, whatever the state of the business cycle, while for the non-apprentices it fluctuates between 0.2 and 0.4. Again this contributes to the greater mobility of non-apprentices.

In the lower part of Table 5 we report the parameters driving the (stochastic) mobility costs towards other jobs measured as a percentage of the life-time value. They can reflect costs of changing location or other specific aspects of the offered job. The mobility costs between jobs are between 2.2% to 3.7% of the lifetime value for the two groups respectively. The standard deviation of these shocks turns out to be very small; this effect together with the relatively low arrival rate of job offers while on the job is part of the reason for the relatively low mobility rates for German workers. However, for non-apprentices when the business cycle is high job-to-job mobility is increased due to the increased arrival rate.

[37]The standard deviation of the innovation for apprentices is not precisely estimated but it is significantly different to that of non-apprentices.

Table 5: Estimated parameters: Variance of shocks, Job destruction and job arrival rates and mobility costs

Parameter	In Apprenticeship	Qualified Apprentices	Non-Apprentices
Std dev initial match specific effect (σ_0)	0.341	0.19	0.77
	(0.007)	(0.01)	(0.015)
Std dev innovation to match specific effect (σ_u)	0.047	0.011	0.042
	(0.011)	(0.011)	(0.009)
Job Offers and Job Destruction Rates			
Quarterly job destruction rate (δ)			
if experience $\leq$ 4 years	-	0.0958	0.0584
		(1.3e-06)	(1.4e-05)
if experience $\in$ [4,6] years	-	0.055	0.028
		(1.2e-05)	(6.3e-05)
if experience $>$ 6 years	-	0.006	0.014
		(6.1e-06)	(6.5e-06)
change in δ when business cycle is high	-	-0.00118	-0.00107
		(5.3e-06)	(7.1e-06)
Quarterly offer arrival rate when unemployed (π_U)			
if business cycle low, experience<4	-	0.30	0.543
		(9.1e-05)	(0.00016)
if business cycle high, experience<4	-	0.30	1
		(0.00014)	(0.00024)
if business cycle low, experience $\in [4, 6]$ years	-	0.356	1
		(3.1e-05)	(3.5e-14)
if business cycle high, experience $\in [4, 6]$ years	-	1	1
		(6.7e-05)	(0.00011)
if business cycle low, experience $>$ 6	-	1	0.90
		(6.0e-05)	(0.0075)
if business cycle high, experience $>$ 6	-	1	1
		(6.2e-05)	(0.0063)
Quarterly offer arrival rate when employed (π_W)			
if business cycle low	0.0279	0.205	0.204
	(1.7e-05)	(0.0012)	(0.0016)
if business cycle high	0.0333	0.21	0.397
	(0.0002)	(0.00099)	(0.002)
Std dev of utility shocks to unemployment^a (σ_η)	-	1.7%	1.7%
		(0.00021)	(0.00021)
Mean of mobility cost^a (m_μ)	-2.19%	-2.19%	-3.67%
	(0.00034)	(0.00034)	(9.1e-05)
Std dev of mobility cost^a (σ_μ)	0.6%	0.6%	0.6%
	(0.00054)	(0.00054)	(0.00054)
Utility of leisure^a (γ_0)			
if experience $\leq$ 4 years	1.1%	1.1%	0.81%
	(4.4e-05)	(4.4e-05)	(0.00014)
if experience $>$ 4 years	-0.56%	-0.56%	-1.8%
	(0.00019)	(0.00019)	(3.4e-05)

Note: ^{a}: as a percentage of lifetime value. Sample size: 54,158 individuals. Asymptotic standard errors in parenthesis. When only one parameter estimate and its standard error are presented in a row, this parameter is restricted to be the same across skill groups.

Table 6: The impact of aggregate shocks on apprenticeship choice by region

Effect of a one standard deviation of GDP at Age 15 by Region^a:		
Schleswig-Holstein	0.107%	(0.17)
Hamburg	-0.919%	(0.032)
Niedersachsen	-0.542%	(0.017)
Bremen	0.331%	(0.14)
Nordrhein-Westfalen	-1.1%	(0.0071)
Hessen	0.365%	(0.081)
Rheinland-Pfalz	0.311%	(0.083)
Baden-Wuerttemberg	1.38%	(0.0077)
Bayern	-0.0349%	(0.0068)
Saarland	-0.284%	(0.55)
Berlin	-0.103%	(0.27)
σ_ω^a	6.63% (0.000307)	

Note: ^{a}: as a percentage of lifetime value. GDP is measured per capita and it is the residual from a quadratic trend normalised by the standard deviation of the residual. This is the estimate of the function $\lambda_R(R, G)$. Asymptotic standard errors in parenthesis.

In Table 6 we present estimates of the effects of the interaction between region and the GDP shocks[38] on the apprenticeship choice, i.e. function $\lambda(R_i, G_t)$ in equation 5. This proxies for the cost of attending an apprenticeship, up to the constant which is presented in Table 8. The interactions are highly significant, as also confirmed by our reduced form regressions in section 2.3; a one standard deviation shock to GDP accounts for up to 27% (Baden-Wurttenberg) of the apprenticeship cost for high cost types which is 5.1% of lifetime value (see Table 8 below). In the last row we provide the standard deviation of the idiosyncratic shock to the cost of apprenticeship, which compared to the baseline cost is quite high, implying a high degree of unexplained variance in education choice at the individual level.

Finally, the probability of being offered the alternative vocational training (see section 2.1) is estimated to be 0.10 (0.001), which leads to a proportion of individuals in vocational school equal to 6% in the simulated data. The observed proportion is also 6%.

Wage equation. Table 7 reports the parameter estimates for the wage equation. For each experience node (2, 4, 6, 10, 30 years) we report the accumulated wage growth by

[38]Measured as the residual of real GDP from a quadratic trend.

that level of actual experience. In between the nodes wages are linear in experience.[39] The effects of tenure are modelled in the same way. For the apprentices experience (and tenure) starts counting at the start of training and the first three years are all spent in training: the estimated returns at two years thus refer to the growth of wages one year before the end of training. The returns to experience thereafter refer to the last year of training and the period following qualification.

For the small group that undergo the non-standard vocational training we estimate the amount of equivalent experience they start off with to be 1.4 years (st. error 0.02), instead of the 3 years for the usual apprentices. This allows the wage profiles to differ while otherwise treating this small group like regular apprentices.

During the first two years wages for apprentices grow by about 1.4%. The incremental effects of experience after two years is 6.1% a year for the next two years, declining to 4.4% and then to 1.5% and finally to 0.5% between 10 and 30 years of experience. For the non-apprentices the experience profile is less concave, with average returns for the first 10 years of about 1.8% annually and about 1.4% annually thereafter. This reflects the more gradual learning experience in the standard jobs. The returns to tenure are very low and for the non-apprentices insignificant. The first four years contribute about 2.3% for the qualified apprentices and 2.7% for the non-apprentices. Thereafter the incremental returns are virtually zero.

Finally, wages are procyclical and more so the wages for the non-apprentices, although the latter effect is not precisely estimated. Between lows and highs of the business cycle non-apprenticeship wages increase by 3.5%, while those of qualified apprentices by 1.9%.

Job mobility and wage growth. In Figure 8 we plot the cumulative contribution to wage growth of mobility from job to job (directly, not via unemployment). 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.[40] For those in apprenticeship job-to-job mobility contributes a maximum of 7% to wage growth, by 9 years of expe-

[39] 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.

[40] The experiment assumes that agents do not anticipate the lack of mobility. In the experiment, individuals still change jobs following unemployment spells.

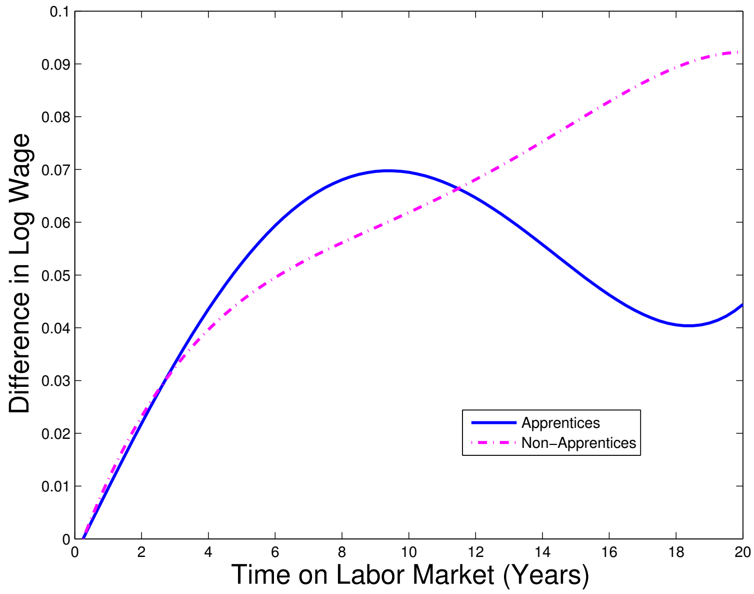
Table 7: The Wage Equations

Parameter	In Apprenticeship	Qualified Apprentices	Non-Apprentices
Log Wage Constant	2.77	3.8	2.6
	(0.0091)	(3.3e-05)	(0.00021)
Effect of high business cycle	0.0189		0.0353
	(0.006)		(0.024)
Experience=2 years	0.014		0.022
	(0.0078)		(0.03)
Experience=4 years	0.14		0.033
	(0.011)		(0.034)
Experience=6 years	0.23		0.098
	(0.013)		(0.032)
Experience=10 years	0.29		0.2
	(0.017)		(0.047)
Experience=30 years	0.3		0.52
	(0.028)		(0.1)
Tenure=2 years	0.012		0.013
	(0.0067)		(0.021)
Tenure=4 years	0.023		0.027
	(0.007)		(0.021)
Tenure=6 years	0.033		0.038
	(0.011)		(0.03)
Tenure=30 years	0.048		0.048
	(0.062)		(0.077)

Log wage is the dependent variable. The wage equation for apprentices during and following training differ only in the constant term (and the variance of the shocks). Asymptotic standard errors in parenthesis.

rience. However the effect declines with further experience. For the non-apprentices, such mobility has a continuing increasing effect on wage growth, which after 20 years of potential experience contributes about 9% to wage growth.[41] These numbers are very close to the ones we obtained in section 2.3. The much larger job heterogeneity for the non-apprentices implies greater returns to search and explains this difference. Thus job-to-job transitions are more important for the non-apprentices and, perhaps as expected, more important for the younger individuals.

Figure 8: 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. 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 9.0%, 12%, 56% and 23%. Table 8 displays summary characteristics for these groups. Individuals of Type 1 and Type 2 have a low wage, whereas Type 3 and Type 4 have high wages.[42] Both Type 1 and 3 individuals have

[41] See Topel and Ward (1992) for results in the US.

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

Table 8: Unobserved Heterogeneity

Parameter	Type 1	Type 2	Type 3	Type 4
Proportion in sample (π_j)	0.09	0.12	0.56	0.23
Proportion in Apprenticeship	0.84	0.99	0.84	0.97
Heterogeneity in the log wage				
Apprentices ($\alpha_0(\varepsilon_i) + \alpha_{ED}(\varepsilon_i)$)	0	0	0.37	0.37
			(0.013)	(0.013)
Non Apprentices ($\alpha_0(\varepsilon_i)$)	0	0	0.4	0.4
			(0.085)	(0.085)
Heterogeneity in the value of leisure ($\gamma_0(\varepsilon_i)$)^a	-	-	-8.7%	-8.7%
			(0.02)	(0.02)
Direct cost of apprenticeship ($\lambda_0(\varepsilon_i)$)^b	5.1%	-10%	5.1%	-10%
	(0.0046)	(0.0036)	(0.0046)	(0.0036)
Selection equation Probit coefficient	0	0	0.0622	0.0622
			(0.00157)	(0.00157)
Selection equation Probit coefficient	0.096	0	0.096	0
	(0.00308)		(0.00308)	

Note: ^{a}: as a percentage of the value of leisure for apprentices. ^{b}: as a percentage of lifetime value. Asymptotic standard errors in parenthesis.

a higher cost of choosing apprenticeship equivalent to about 5.1% lifetime value.[43] For types 2 and 4 the cost is negative, implying that these individuals value apprenticeship training intrinsically, over and above the monetary benefits.

There is a negative association between having a low cost of apprenticeship and being a high wage type: among those with high cost of apprenticeship (Types 1 and 3) the probability of being a high wage type is 86%; among individuals with a low cost of apprenticeship the probability is 66%. The return to ability is 3 percentage points lower among apprentices: high wage types have a 0.37 constant when non-apprentices and 0.4 otherwise. Finally, high wage types also have an 8.7% lower value of leisure relative to low wage types.

The impact of the cost of apprenticeship on its take-up is a bit higher for low wage types than it is for the high wage ones (15 percentage points versus 13). When all these combine in the model it turns out that the probability of being an apprentice is just 0.7% lower for high wage types.

In the final four rows of the table we report the coefficients (and standard errors) on

[43]So for them the intercept in the apprenticeship equation, shown in Table 6 becomes more negative.

the two unobserved factors in the model for the selection into our sample (see Section 4.1). We find that participation in our sample (vocational schooling after 10) is positively associated with higher cost of apprenticeship and with being a higher wage type in the vocational sector.

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 is given by

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

where $a = 16$ and 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 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 discount factor is 0.95 on annual basis. The results are displayed in Table 9.

The costs of an apprenticeship from the individual's perspective are the direct costs, reflected in the parameters of Tables 6 and 8, including the random shock, the heterogeneous cost $\lambda_0(\varepsilon)$ as well as the opportunity cost due to lower earnings during training. Taking all such costs into account, the average gain to apprenticeship (ATE) [44] is 14.1%. The average gain for those who took up apprenticeship is 16.5%; this reflects both a change in composition towards low cost types and shocks (ω) more favourable to taking up apprenticeship. Eliminating the opportunity cost, which is high because apprentices are paid a very low wage during training (see Figure 1) the returns increase to 23.3%. When we also eliminate the direct costs these returns slightly increase to 22.9%: on average the direct costs are slightly negative. These returns calculations include among

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

Table 9: The Life-cycle Returns to Apprenticeship

	Average	Type 1	Type 2	Type 3	Type 4
Wage		Low		High	
Cost of Education		High	Low	High	Low
Welfare Returns					
ATE	14.1%	11.7 %	32.3 %	7.39 %	22.0%
ATT	16.5%	15.2%	32.3 %	10.5%	23.3%
ATE direct cost only	23.3%	24.5 %	45.2 %	15.6 %	29.9 %
ATE no costs	22.9 %	31.5 %		20.3%	

other components the effect of income when the individuals are out of work. Given the relatively large differences in the unemployment rates (see Figure 5) between the two groups this is an important component of the returns.

The four last columns in Table 9 show the way the gains vary with unobserved heterogeneity. The returns are heavily influenced by the cost of apprenticeship ($\lambda_0(\varepsilon)$) and decline for higher wage type. However, the latter do have a higher absolute gain in lifetime value (rather than relative gain). Similar patterns emerge when we consider the returns for those who actually chose the apprenticeship career (ATT); for low cost types there is scarcely any difference between ATE and ATT because almost all join apprenticeships.

In terms of the discounted present value of income, the apprentices expect to earn (including unemployment benefits) 12% more than the non-apprentices seen from age 16. This calculation leaves out the impact of differences in preferences, but includes differences in income when unemployed. The main factors driving the return to apprenticeship are the differences in the incomes, including the influence of different variances of the shocks and of initial job offers. Equalising the job arrival rates and the job destruction rates does not have a large impact on the returns.[45]

6 Some properties of the estimated model

To illustrate the work incentives implied by the model, Table 10 presents employment elasticities. The elasticities are the proportional change in participation resulting from a small proportional change in wages at all points in the lifecycle, keeping education

[45] An interesting question for research is how risk aversion affects the choice and value of apprenticeship. At present with no asset or consumption data we could not investigate this empirically without assuming no borrowing or lending. The effect of risk be absorbed by the parameters driving apprenticeship cost

Table 10: Labour supply (participation) elasticities with respect to lifetime change in wage

All workers	**Apprentices**	**Non Apprentices**
1.14	1.07	1.41

choices constant. Since our model is linear in income and the marginal utility of wealth is constant, there is no obvious sense in which we can distinguish between Frisch and Marshallian elasticities.[46]

A dynamic life-cycle model such as ours or earlier ones such as that of Keane and Wolpin (1997) make it clear that labour market policies encouraging human capital or tax and welfare programmes, such as EITC can have an effect on the entire career, starting with education choices and continuing with employment and job mobility decisions, all of which affect earnings. Indeed Heckman and Klenow (1998) emphasize that human capital policies should be evaluated in a life cycle setting.[47]

Our ability to carry out such policy simulations here is limited by the fact that we do not model general equilibrium effects, which in this environment with search frictions is particularly complicated. This is an important problem, but well beyond the scope of the current paper.[48] Models that address such issues in the absence of search frictions include Heckman, Lochner, and Taber (1998), Lee (2005) and Lee and Wolpin (2006). Shephard (2009) using a Burdett-Mortensen type model with no match specific effects considers the extent to which the wage subsidy may be absorbed by the firm. He finds that most of the benefit is kept by the worker.

Here we just illustrate the sensitivity of lifecycle choices to changes in the policy environment by carrying out a simulation, where we introduce a low wage subsidy modelled along the lines of the US earned income tax credit. This is fully financed by a propor-

[46] In our model increasing wages also increases unemployment benefit. 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.

[47] Similar considerations are discussed in Keane and Wolpin (2000), who present the effect of a wage subsidy on education and career choices and in Heckman, Lochner, and Cossa (2003) who consider the impact of Earned Income Tax Credit (EITC), on human capital accumulation.

[48] see Postel-Vinay and Robin (2002), Cahuc, Postel-Vinay, and Robin (2006) and Lise, Meghir, and Robin (2009)

Table 11: Effects of a low wage subsidy on Apprenticeship training, employment and wages

	All	Type 1	Type 2	Type 3	Type 4
% Increase Skilled	-2.3%	-5%	-0.32%	-3.3%	-0.074%
% Increase in Work	2.4%	9.1%	11%	0.6%	-0.046%
Change in Tax	1.33%				

Note: The policy is fully funded through a proportional tax on skilled and unskilled individuals.

tional tax on earnings.[49] The results of this simulation, compared to the baseline of no change in the policy parameters are presented in Table 11.

Overall, individuals facing EITC reduce take up of apprenticeship by 2.3% and increase employment by 2.4%. The differences in employment effects between high wage and low wage types reflects the fact that among the latter more are facing increased taxes to fund the credit than are benefiting from its introduction. Thus improved work incentives through wage subsidies can have quantitatively important effects on the incentive to obtain education.

7 Conclusions

In this paper we analyse the German apprenticeship system using a life-cycle dynamic model of apprenticeship choice, employment, job mobility and wages. Our data source are detailed administrative records. This data has the rare characteristic that individuals are observed from the start, when they make their apprenticeship choice, and followed throughout their working life, recording their job spells and their pay in all firms they are employed. Measurement error is likely to be much less important than in standard survey data, because the records are reported directly by firms for the purpose of determining social security contributions.

Our population of interest are those who were allocated to the vocational track when they were 10 years old. In the model these individuals choose whether to follow an apprenticeship training or not at 16. We then model the subsequent labour supply and

[49] 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. The EITC is made available for those above 19 years of age only; hence the policy is designed here not to act as a direct monetary disincentive to training.. It is financed by a proportional tax on earnings.

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 offering a rich stochastic specification for wages and allowing us to understand the sources of wage growth.

The lifecycle pattern of wages between the two careers is quite different. The non-apprentices have profiles that are less concave, with the returns to experience remaining positive to an older age. Returns to tenure are effectively zero. Generally we find the search frictions are quite low. However, non-apprentices face much higher heterogeneity in their wages and stronger shocks to their match specific effects. These facts imply both that the non-apprentices spend more time unemployed and that they have higher returns to job search. Indeed for them the contribution of job to job movements on wages is higher than it is for apprentices. We find that exogenous job destruction rates remain constant across the business cycle, while job arrival rates are substantially pro-cyclical for low experience workers in both groups. The life-cycle returns to an apprenticeship, which lasts usually three years, are 14.1%. This does not account for any costs that the firm or the state may be paying particularly for class-room training. If all costs are eliminated the returns rise to 27.1% due entirely (on average) to the elimination of the opportunity cost

Finally, the apprenticeship system seems to offer higher earnings and greater labour market attachment and, accounting for the costs privately incurred by the individual, it is a worthwhile investment. Judging by the job arrival rates, particularly after some level of experience it does not seem to be difficult to reallocate workers who have become unemployed. On this evidence the apprenticeship system confers a number of positive effects, without the apprentices being less flexible than the ones who did not go through this system.

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.

Altonji, J. G., A. Smith, and I. Vidangos (2009): "Modeling Earnings Dynamics," NBER Working Papers 14743.

Bowers, N., A. Sonnet, and L. Bardone (1999): *Giving Young People a Good Start: the Experience of OECD Countries*. OECD Background Report.

Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006): "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 74, 323–364.

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

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

Carneiro, P., J. Heckman, and E. Vytlacil (2006): "Estimating Marginal and Average Returns to Education," mimeo UCL.

Dustmann, C. (2004): "Parental background, secondary school track choice, and wages," *Oxford Economic Papers*, pp. 209–230.

Dustmann, C., and C. Meghir (2005): "Wages, experience and seniority," *Review of Economic Studies*, 72(1).

Dustmann, C., and U. Schoenberg (2009): "Training and Union Wages," *Review of Economics and Statistics*, 91(2), 363–376.

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), 1295–1339.

Fersterer, J., J.-S. Pischke, and R. Winter-Ebmer (2008): "Returns to Apprenticeship Training in Austria: Evidence from Failed Firms," *Scandinavian Journal of Economics*, 110(4), 733–753.

Fersterer, J., and R. Winter-Ebmer (2003): "Are Austrian returns to education falling over time?," *Labour Economics*, 10(1), 73–89.

Gathman, C., and U. Schoenberg (2010): "How General is Human Capital? A Task-Based Approach," *Journal of Labor Economics*, 28, 1–49.

Gourieroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118.

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

Hall, R. E. (2005): "Job Loss, Job Finding, and Unemployment in the U.S. Economy over the Past Fifty Years," in *NBER Macroeconomics Annual 20*, ed. by M. Gertler, and K. Rogoff, p. 101 137. Cambridge, MA: MIT Press.

Head, K., and T. Mayer (2006): "Regional wage and employment responses to market potential in the EU," *Regional Sciences and Urban Economics*, 36, 573–594.

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), 669–738.

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.

House of Lords (2007): *Apprenticeship: A Key Route to Skill.* Published by the Authority of the House of Lords, London: The Stationery Office Limited.

Jimeno, and Rodríguez-Palenzuela (2003): “Youth Unemployment in the OECD: Demographic Shifts, Labour Market Institutions and Macroeconomic Shocks,” ENEPRI Working Paper No. 19.

Keane, M., and K. Wolpin (1997): “The Career Decisions of Young Men,” *Journal of Political Economy*, 105(3), 473–522.

——— (2000): “Eliminating Race Differences in School Attainment and Labor Market Success,” *Journal of Labor Economics*, 18(4), 614–652.

Kemple, and Wilner (2008): “Career Academies - Long-Term Impacts on Labor Market Outcomes, Educational Attainment, and Transitions to Adulthood,” MDRC.

Kennan, J., and J. R. Walker (2010): “The Effect of Expected Income on Individual Migration Decisions,” forthcoming Econometrica.

Krueger, A. B., and J. Pischke (1995): "A Comparative Analysis of East and West German Labor Markets: Before and After Unification," in *Differences and Changes in Wage Structures*, pp. 405–446. National Bureau of Economic Research, Inc.

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

Lee, D., and K. I. Wolpin (2006): "Intersectoral Labor Mobility and the Growth of the Service Sector," *Econometrica*, 74(1), 1–46.

Lerman, S., and C. Manski (1981): "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski, and D. McFadden, pp. 305–319. Cambridge: MIT Press.

Lise, J., C. Meghir, and J.-M. Robin (2009): "Matching, Sorting and Wages," mimeo, UCL.

Low, H., C. Meghir, and L. Pistaferri (2009): "Wage Risk and Employment Risk over the Life Cycle," Forthcoming American Economic Review.

McFadden, D. (1989): "Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 57, 995–1026.

Meghir, C., and L. Pistaferri (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72, 1–32.

Mortensen, D., and C. Pissarides (1994): "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61(3), 397–415.

Pakes, A., and R. Pollard (1989): "The Asymptotic Distribution of Simulation Experiments," *Econometrica*, 57, 1027–1057.

Parey, M. (2009): "Vocational Schooling versus Apprenticeship Training. Evidence from Vacancy Data," mimeo, University of Essex.

Petrongolo, B., and C. Pissarides (2008): "The ins and outs of European unemployment," *American Economic Review*, 98, 256–262.

Pissarides, C., R. Layard, and M. Hellwig (1986): "Unemployment and Vacancies in Britain," *Economic Policy*, 1, 500–559.

Postel-Vinay, F., and J. M. Robin (2002): "Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70(6), 2295–350.

Ryan, P. (2001): "The School-to-Work Transition: A Cross-National Perspective," *Journal of Economic Literature*, 39, 34–92.

Ryan, P., H. Gospel, and P. Lewis (2007): "Large Employers and Apprenticeship Training in Britain," *British Journal of Industrial Relations*, 45(1), 127–153.

Shephard, A. (2009): "Equilibrium search and tax credit reform," mimeo, UCL.

Shimer, R. (2007): "Reassessing the Ins and Outs of Unemployment," University of Chicago, mimeo.

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., and M. 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.

Winkelmann, R. (1996): "Training, Earnings and Mobility in Germany," *Konjunkturpolitik*, 42, 275–298.

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.

A Web Appendix

A.1 The Final Sample

We select all male individuals who are born between 1960 and 1972. Thus, we make sure that no individual is older than 15 in 1975 (the minimum age at which post-secondary labour market entry is possible), which is the first year of our data. We consider all years between 1975 and 1995. We exclude all individuals who live in East-Germany. We drop individuals who work in the agricultural industry, and individuals who work in the family businesses. We restrict our sample to those who are not older than 23 when they enter the labour market the first time, and who enter the labour market with only a lower secondary school education, who either enrol into apprenticeship training directly, or who enter the labour market without further training.[50][51] We further exclude individuals with multiple apprenticeships (which is about 6% of the sample), and workers who are still in training at the end of the observation window, or who have no valid wage spells after apprenticeship training. We also exclude individuals who had a work spell before starting apprenticeship training, and we drop individuals with unreasonably long apprenticeship training periods (which we set to 1600 days). We restrict our analysis to individuals with German citizenship, as individuals with non-German citizenship may have acquired (part of) their education abroad.

The wage information in the data is the average daily wage for the length of the working spell. A spell is at most 365 days long if the individual does not change firm, as firms have to report yearly on their employees. If individuals change firm during the calendar year, or exit into non-employment, we observe the average daily wage for the period for which the individual has been in employment. Thus, every wage we observe belongs to one particular worker-firm spell. We compute real wages in 1995 prices.

The precise distinction between individuals who enrol in a traditional apprenticeship

[50]In Germany, children enter primary school at the age of about 6. Primary school takes 4 years. After primary school, and at the age of 10, individuals decide whether to enter one of three secondary school branches: lower secondary school (which takes another 5-6 years), intermediate secondary school (which takes another 6 years), and higher secondary school (which takes another 9 years). For our analysis, we concentrate on individuals who choose lower or intermediate secondary school. These two options do not allow for direct access to university, and individuals typically enrol into apprenticeship training, or enter the labour market directly.

[51]As the comparison group of individuals who choose upper track secondary school, which we use to implement our selection correction, we define all those individuals who enter the labour market either with an upper secondary degree (with or without further training), and before the age of 23, or with college- or university education, and before the age of 32.

scheme, and individuals who enter the labour market without that training, is as follows. We define as "apprentices" all those individuals who entered the labour market with a lower or intermediate secondary school degree, and who can be observed after entry on an apprenticeship training scheme for at least 24 months, and who transit to a "skilled" status afterwards.[52] We define as "non-apprentices" all those individuals who enter the labour market without further training, or who have been on an apprenticeship training schemes for less than 7 months, without obtaining a degree (i.e. dropouts). This group may include individuals who enrolled in one-year vocational courses before entering the labour market – preparatory courses that do not lead to vocational degrees. Thus, among our non-apprentices may be individuals who did receive some preparatory training.

Another mode of entry, as discussed in Parey (2009), is attendance of 2-3 vocational schools, which provide vocational training with unpaid work experience in specialised schools for a limited number of occupations.[53] These occupations are mainly in female-dominated occupation groups, like caring and health-related occupations. In our sample, these individuals constitute about 6% of individuals.[54] In line with Parey (2009), we find that the wage paths of this group are very similar to those of individuals undergoing firm-based training, and higher than those of individuals entering the labour market without further training. We also find that they are experiencing lower employment probabilities than apprentices. The way we deal with these individuals is to include them among our apprentices, assuming that the choice to undergo training at a full time school rather than within the firm is equivalent to choosing apprenticeship training in a firm.

B Computational Details

This section presents the computational details for solving and estimating our model.

[52]For apprentices who finish their training within a calendar without changing firms, we do not observe the date of graduation, neither can we distinguish the apprenticeship wage during that year from the skilled worker wage. To compute the number of apprenticeship training months, we assign to these individual 6 months of training. Further, when we compute wages after the apprenticeship period, we discard these observations.

[53]According to the Central Labour Office (Bundesagentur fuer Arbeit), firm based apprenticeship schemes train for 541 occupations, while full-time colleges train for only 133 occupations.

[54]The size of this group is smaller than in Parey (2009). One reason for this is that we consider only the years up to 1996, where these school based vocational schemes were less frequent than in later years.

Table 12: Quarterly transition matrix for below and above trend GDP

	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

B.1 GDP growth and Markov transition matrix

To compute business cycles, we use the per capita West-German GDP expressed in constant 1995 US$. We detrend it with a linear trend for the period 1975 to 1996. GDP grew at a rate of $479.18 (9.0) per year. In the model we then use transitions between above trend (good times) and below trend (bad times). Table 12 presents the transition matrix for this first order Markov process, estimated over our sample period.

B.2 Construction of the Moments

As the model does not include regional variation in wages or employment, nor aggregate time trends, we remove those variations from the moments, by including regional indicator variables and a quadratic trend in all our regressions.

B.3 Computing the Value Functions

The model is solved recursively backward, starting at age 65 and until age 16. We allow the value function to depend on age as well as the other state variables.

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 10 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 code was solved using parallel processing to increase speed. Solving, simulating and computing the moments for a particular set of parameters takes about 20 seconds.

B.4 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,

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

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

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

Now

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

for

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

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

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

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 $G_{t+1}, u_{ift+1}, \kappa^0_{i\widetilde{f}}$ out of U, W, $\widetilde{W}$.

C The Fit of the Model

In this section, we present the fit of the model in detail in Tables 13 to 22. The tables list all the moments used in the estimation, apart from the ones used to identify the educational choices at age 10 and 16, as they involve more than 100 entries each and are too long to display.

Table 13: Goodness of Fit: Wage Level and Potential Experience

	Apprentices			Non Apprentices		
	Observed	Std Error	Simulated	Observed	Std Error	Simulated
Potential Exp ∈ [0,2]	3.05	(0.0006)	3.08	4.01	(0.003)	3.97
Potential Exp ∈]2,4]	3.71	(0.001)	3.72	4.3	(0.003)	4.28
Potential Exp ∈]4,6]	4.45	(0.0006)	4.43	4.43	(0.002)	4.36
Potential Exp ∈]6,8]	4.54	(0.0005)	4.54	4.47	(0.002)	4.45
Potential Exp ∈]8,10]	4.6	(0.0006)	4.61	4.52	(0.002)	4.5
Potential Exp ∈]10,15]	4.67	(0.0005)	4.7	4.57	(0.002)	4.58
Potential Exp ∈]15,30]	4.73	(0.0008)	4.75	4.61	(0.003)	4.64
Business Cycle Good	0.0269	(0.0005)	0.0176	0.0284	(0.002)	0.147

Table 14: Goodness of Fit: Proportion Working and Potential Experience

	Apprentices			Non Apprentices		
	Observed	Std Error	Simulated	Observed	Std Error	Simulated
Potential Exp ∈ [0,2]	0.976	(0.0003)	0.998	0.635	(0.002)	0.574
Potential Exp ∈]2,4]	0.845	(0.0007)	0.821	0.524	(0.002)	0.516
Potential Exp ∈]4,6]	0.647	(0.0009)	0.665	0.53	(0.002)	0.536
Potential Exp ∈]6,8]	0.758	(0.0008)	0.749	0.577	(0.003)	0.551
Potential Exp ∈]8,10]	0.809	(0.0008)	0.813	0.623	(0.003)	0.606
Potential Exp ∈]10,30]	0.845	(0.0006)	0.836	0.683	(0.002)	0.703
Business Cycle Good	0.0262	(0.0005)	0.0052	0.046	(0.002)	0.0301

Table 15: Goodness of Fit: Experience Levels and Potential Experience

	Apprentices			Non Apprentices		
	Observed	Std Error	Simulated	Observed	Std Error	Simulated
Potential Exp ∈ [0,2]	0.861	(0.001)	0.875	0.481	(0.002)	0.581
Potential Exp ∈]2,4]	2.76	(0.001)	2.78	1.51	(0.005)	1.63
Potential Exp ∈]4,6]	4.17	(0.001)	4.22	2.5	(0.009)	2.73
Potential Exp ∈]6,8]	5.59	(0.002)	5.63	3.52	(0.01)	3.84
Potential Exp ∈]8,10]	7.13	(0.003)	7.2	4.58	(0.02)	5.04
Potential Exp ∈]10,30]	10.3	(0.004)	11.1	6.42	(0.02)	8.34

Table 16: Goodness of Fit: Firm Seniority and Potential Experience

	Apprentices			Non Apprentices		
	Observed	Std Error	Simulated	Observed	Std Error	Simulated
Potential Exp $\in$ [0,2]	0.933	(0.002)	0.82	0.817	(0.009)	0.554
Potential Exp $\in$]2,4]	2.66	(0.003)	2.4	2.18	(0.01)	1.61
Potential Exp $\in$]4,6]	2.99	(0.005)	2.3	3.06	(0.01)	2.41
Potential Exp $\in$]6,8]	3.19	(0.006)	2.37	3.9	(0.02)	3.06
Potential Exp $\in$]8,10]	4.11	(0.007)	3.03	4.89	(0.02)	3.81
Potential Exp $\in$]10,30]	6.54	(0.006)	4.84	7.09	(0.02)	5.74
Business Cycle Good	-0.139	(0.004)	0.117	-0.298	(0.01)	-0.0511

Table 17: Goodness of Fit: Number of Firms and Potential Experience

	Apprentices			Non Apprentices		
	Observed	Std Error	Simulated	Observed	Std Error	Simulated
Potential Exp $\in$ [0,2]	1.01	(0.001)	0.982	1.16	(0.005)	1.1
Potential Exp $\in$]2,4]	1.12	(0.001)	1.12	1.67	(0.006)	1.48
Potential Exp $\in$]4,6]	1.63	(0.002)	1.64	2.17	(0.007)	1.77
Potential Exp $\in$]6,8]	2.27	(0.002)	2.21	2.66	(0.01)	2.01
Potential Exp $\in$]8,10]	2.78	(0.003)	2.62	3.1	(0.01)	2.23
Potential Exp $\in$]10,30]	3.56	(0.003)	3.35	3.96	(0.01)	2.68
Business Cycle Good	-0.0133	(0.002)	0.0397	0.0172	(0.007)	0.135

Table 18: Goodness of Fit: Standard Deviations of Wages and Potential Experience

	Apprentices			Non Apprentices		
	Observed	Std Error	Simulated	Observed	Std Error	Simulated
Potential Exp $\in$ [0,2]	0.297	(0.01)	0.382	0.547	(0.005)	0.605
Potential Exp $\in$]2,4]	0.489	(0.01)	0.559	0.483	(0.005)	0.408
Potential Exp $\in$]4,6]	0.406	(0.01)	0.333	0.44	(0.005)	0.38
Potential Exp $\in$]6,8]	0.332	(0.01)	0.249	0.415	(0.005)	0.376
Potential Exp $\in$]8,10]	0.305	(0.01)	0.238	0.404	(0.005)	0.365
Potential Exp $\in$]10,30]	0.302	(0.005)	0.216	0.383	(0.003)	0.372

Table 19: Goodness of Fit: Wages, Experience and Tenure

	Apprentices			Non Apprentices		
	Observed	Std Error	Simulated	Observed	Std Error	Simulated
Exp $\in$]2,4]	0.231	(0.001)	0.138	0.27	(0.003)	0.229
Exp $\in$]4,6]	0.448	(0.002)	0.332	0.365	(0.003)	0.313
Exp $\in$]6,8]	0.538	(0.002)	0.441	0.401	(0.003)	0.395
Exp $\in$]8,10]	0.59	(0.002)	0.517	0.433	(0.004)	0.431
Exp $\in$]10,15]	0.646	(0.002)	0.59	0.463	(0.004)	0.487
Exp $\in$]15,30]	0.693	(0.002)	0.62	0.5	(0.005)	0.581
Tenure $\in$]2,4]	-0.0338	(0.0007)	0.0146	0.0375	(0.002)	0.0406
Tenure $\in$]4,6]	0.00367	(0.0007)	0.0613	0.071	(0.003)	0.0804
Tenure $\in$]6,8]	0.01	(0.0008)	0.0855	0.0812	(0.003)	0.0951
Tenure $\in$]8,10]	0.0271	(0.0009)	0.103	0.0892	(0.003)	0.104
Tenure $\in$]10,30]	0.0403	(0.001)	0.151	0.0943	(0.003)	0.132
Business Cycle Good	0.0265	(0.0004)	0.00765	0.0297	(0.002)	0.183
In Apprenticeship Training	-0.992	(0.001)	-1.02	-	-	-
Constant	4.05	(0.001)	4.13	4.06	(0.003)	4.07

Table 20: Goodness of Fit: Standard Deviation of Wages, Experience and Tenure

	Apprentices			Non Apprentices		
	Observed	Std Error	Simulated	Observed	Std Error	Simulated
Exp	-0.0175	(0.0002)	-0.00809	-0.0527	(0.0007)	-0.0662
Exp squared	0.000748	(1e-05)	2.69e-05	0.00272	(5e-05)	0.00381
Tenure	0.00547	(0.0001)	0.0103	0.00566	(0.0005)	0.00487
Tenure squared	-0.000463	(1e-05)	-0.00042	-0.00046	(4e-05)	-0.00058
Business Cycle Good	-0.00124	(0.0002)	-0.00356	-0.00395	(0.001)	-0.126
In Apprenticeship Training	-0.023	(0.0008)	0.0904	-	-	-
Constant	0.137	(0.0009)	0.0967	0.275	(0.003)	0.476

Table 21: Goodness of Fit: Wages Changes, Experience and Tenure

	Apprentices			Non Apprentices		
	Observed	Std Error	Simulated	Observed	Std Error	Simulated
Exp	-0.0534	(0.0003)	-0.0233	-0.00512	(0.0003)	-0.0013
Exp squared	0.00249	(1e-05)	0.00089	0.000226	(2e-05)	4.27e-05
Tenure	0.00892	(0.0001)	0.0117	-0.00352	(0.0003)	0.000226
Tenure squared	-0.000712	(1e-05)	-0.000772	0.000237	(2e-05)	-1.51e-05
In Apprenticeship Training	-0.158	(0.0009)	-0.068	-	-	-
Constant	0.251	(0.001)	0.121	0.0366	(0.001)	0.0161

Table 22: Goodness of Fit: Standard Deviation of Wages Changes, Experience and Tenure

	Apprentices			Non Apprentices		
	Observed	Std Error	Simulated	Observed	Std Error	Simulated
Exp	0.00312	(0.0001)	0.00983	-0.00571	(0.0003)	0.000216
Exp squared	-0.000267	(1e-05)	-0.000656	0.00038	(2e-05)	-2.16e-05
Tenure	-0.0364	(0.0002)	-0.0231	-0.00351	(0.0003)	-0.00393
Tenure squared	0.00168	(1e-05)	0.000968	0.000163	(2e-05)	0.000209
In Apprenticeship Training	-0.125	(0.0008)	-0.067	-	-	-
Constant	0.181	(0.001)	0.108	0.034	(0.001)	0.0186