Apprenticeship Training and Commitment to Training Provision

Christian Dustmann
University College London

Uta Schönberg
University of Rochester

This Draft: July 2007

Abstract

Why do apprenticeship schemes work well in some countries, like Germany and Austria, but less so in others, like the UK? This paper argues that a necessary prerequisite for apprenticeship schemes to be successful is the enforceability of the apprenticeship contract, most notably the firm's ability to commit to training provision. We hypothesize that, by linking into an existing regulatory framework, firms in Germany are able to commit, while this may not be the case in countries that run apprenticeship schemes less successfully. To test our hypothesis, we develop a model where firms have an incentive to finance training because of wage compression due to firm-specificity and asymmetric information, and analyse it under both commitment and no commitment to training provision. Drawing on the model, we provide evidence that the German apprenticeship system is indeed characterised by commitment to training provision. We then simulate our model for values of firm-specificity and asymmetric information estimated from survey and administrative data. We find that training would be substantially lower under no commitment, at most 8 % of that under commitment. This is in line with our hypothesis.

Keywords: educational policies, training, wage compression
JEL: I2, M53, J24, J31

Correspondence: Christian Dustmann, Department of Economics, Gower Street, London WC1E 6BT, England. E-mail: c.dustmann@ucl.ac.uk; Uta Schönberg, Department of Economics, University of Rochester, Harkness Hall, Rochester 14627, NY, USA. E-mail: utas@troi.cc.rochester.edu. We are grateful to V Bhaskar, David Card, and especially to Paul Ryan for insightful discussions and comments. We gratefully acknowledge financial support from the German Research Foundation (DFG) and the Anglo-German Foundation (AGF).

1 Introduction

While all countries provide a significant part of post-secondary education in the form of academic university education, they differ widely in the way in which they offer vocational post-secondary programmes for the less academically inclined. Some countries, such as Sweden, France, and Italy, offer such training opportunities in the form of full-time school-based vocational education. In contrast, other countries run large-scale apprenticeship schemes that combine firm-provided on-the-job training with state-provided school-based education. The best known example for this type of training scheme is Germany, with nearly two in three individuals in each cohort being educated within this scheme. Although less well-known, many other countries, including Australia, Austria, Denmark, the Netherlands, and Switzerland, educate at least one third of a school leaving cohort within apprenticeship schemes.

Firm-based apprenticeship schemes have a number of advantages compared to their closest substitute, school-based vocational education. Most importantly, some skills may be taught more effectively in a work environment than in the classroom. This may be most apparent in the crafts sector where experienced craftsmen are likely to be better teachers in practical applications than classroom teachers. But apprenticeship schemes may also provide useful skills in white collar professions, such as sales, banking, and, insurance.[1] First-hand experience of how an entire firm or organisation is run, how to interact with customers and other traders, and how jobs are conceptualised and implemented is a valuable knowledge that is unlikely to be taught similarly efficiently in the classroom.[2] Moreover, apprenticeship schemes have the added benefit of allowing for smooth transitions of new labour market entrants into the labour market, thus reducing youth unemployment (see for instance Ryan, 2001 for

[1] In fact, although the apprenticeship system is often considered as a provider for training in craft related professions, in Germany only one in three individuals on apprenticeship training schemes are trained in this sector, with the remaining positions in predominantly white collar professions.

[2] A recent study by Adda et al. (2006) finds that the return to apprenticeship training is comparable to that of 2-3 years of schooling in the US and UK. Using an interesting approach based on failed firms, Festerer et al. (2004) report somewhat smaller estimates for Austria, between 2 to 3 percent per year.

evidence).[3]

It may therefore not be surprising that some academic papers suggest the German apprenticeship system as a role model for similar schemes in other countries (see for instance Lehmann, 2000; Gospel, 1998; Steedman et al., 1998, Harhoff and Kane, 1997). Throughout the 1990s several countries, including the US, the UK, France, and Norway, have attempted to implement new or to expand existing apprenticeship schemes (see e.g. Bowers, Sonnet and Bardone, 1999 for an overview). The UK, a country with a long history of apprenticeship training but where participation rates declined throughout the 80s (see also Figure 1), has set a particularly ambitious target: By 2010, 35% of the 16-year-olds are to be trained within "Modern Apprenticeship" schemes. With current enrollment rates being around 13%, the UK is so far well short of this target, and the low quality of apprenticeship training is a prime concern (e.g. Ryan and Unwin, 2001; Ryan, Gospel and Lewis, 2007).

This paper asks why apprenticeship type schemes work in some countries, like Germany and Austria, and why similar schemes have been less successful in other countries, like the UK. We focus on what we believe is a key difference between *firm-based* apprenticeship training and *school-based* (vocational) education: the enforceability of the training contract, most importantly the firm's ability to commit to training provision. Unlike schools, firms may not be able to commit to training provision because training takes place within firms and may therefore not be verifiable. Since apprenticeship wages are typically about one third lower than wages for young unskilled workers, firms may increase (short-run) profits by using apprentices as cheap labour, for activities like cleaning, photocopying, and running errands, instead of providing the promised training. This is obviously less of a concern if training is offered by and takes place in (vocational) schools whose sole purpose it is to educate young workers.

Our hypothesis is that commitment to training provision is a prerequisite for apprenticeship pro-

[3] Jimeno and Rodríguez-Palenzuela, 2003 illustrate substantially lower youth unemployment rates in Germany and Austria than in all other European OECD countries.

grams to work well. We argue that ensuring commitment to training provision requires a regulation of apprenticeship training, including a legal definition of the rights and obligations of apprentices and their employers and the external testing of apprentices. These are important components of the apprenticeship system in Germany and in other countries that run successful apprenticeship programmes, but are (currently) not present in the UK system.

One may argue that market forces are sufficiently strong to ensure commitment to training provision even in the absence of any apprenticeship regulation. If apprentices are abused as cheap labour, they can go to court. Firms may also be able to build up a reputation for offering high-quality training programmes, and there are indeed examples that some firms were able to do so (e.g. Rolls-Royce in the UK). However, the lack of legal definition of what apprenticeship training entails (as it is for instance the case in the UK) may make it more difficult to enforce training provision through courts. Moreover, training may simply be too complex to specify in a contract in a way that is legally enforceable.

Figure 1 provides some first empirical evidence that is consistent with the hypothesis that regulation may help to ensure commitment to training provision. The figure plots the share of individuals in manufacturing on apprenticeship schemes in the UK between 1975 and 2005. Entries from 1975 to 1990 come from Ryan and Unwin (2001), Figure 1. We expanded the series till 2005 using the British Labour Force Survey. In the early 80s, the Thatcher government began to de-regulate the apprenticeship system. Most importantly, the *mandatory* Industrial Training Boards that had control over the content and assessment of workplace training, were abolished and replaced with a *voluntary*, *employer-led* system of Training and Enterprise Councils that have *purely advisory* capacity (see e.g. Ryan , 2000 and Gospel, 1998 for more details on the reform). The figure shows that starting in the early 80s, apprenticeship training started to fall rather sharply in the manufacturing sector. While the share of individuals on apprenticeship schemes was about 2.2% in 1980, it was only 1.1 % in 1990, and has remained roughly constant at this level ever since.[4] While there may be alternative explanations

[4]Stevens (1994b) shows a similar trend for recruitment to basic craft training in engineering. Additional evidence that the dergulation of apprenticeship training in the 1980s has led to a decline in apprenticehsip training in the UK can

for this decline, one simple explanation is that the de-regulation removed an important commitment device for firms who wanted to offer apprenticeship training.

In order to further test our hypothesis, we build on the analysis by Acemoglu and Pischke (1999a, 1999b) who point out that in a perfectly competitive labour market, the training market would break down completely if firms are unable to commit to training provision. This is because workers anticipate that firms will not offer any training in equilibrium, and are therefore not willing to accept a wage cut to finance training. This is, however, different in an imperfect labour market characterised by wage compression (see also Stevens, 1994a). In this case, firms can increase future profits through training and are therefore willing to provide and finance training.[5]

While the analysis by these authors provides several key insights, it leaves a number of important questions unanswered. Most importantly, are firms able to commit to training provision or not? If wages are compressed, how much lower will training under no commitment be compared to that under commitment? What are the levels of wage compression for which training under no commitment is close to that under commitment, and are these levels of wage compression supported by the data?

In order to address these questions, we develop a model where firms have an incentive to finance training because of wage compression, and analyse the model under the assumption of both commitment and no commitment to training provision. We allow for two reasons for wage compression and thus firm-financed training: (firm-) specific training (e.g. Becker, 1964, Stevens, 1994a, Franz and Soskice, 1995) and asymmetric information between training and outside firms (Acemoglu and Pischke 1998). We choose these reasons for wage compression because both are likely to be more important for firm- than for school-based education. Several economists have argued that apprenticeship training is more specific than school-based education, as a significant part takes place within firms and is customised for a particular profession (see e.g. Heckman et al., 1994 and Neal, 1999). Apprenticeship

be found in Gospel (1998), Finegold and Soskice (1988) and Keep and Mayhew (1988).

[5]Using data from the UK, Booth and Bryan (2005) provide some evidence that is consistent with the hypothesis that wage compression induces firms to finance training.

training–as opposed to school-based education–may also lead to informational asymmetries. Apprentices spend up to three years with the same firm so that training firms have a lot of time to get to know the worker. Since firms train workers in order to employ them later, they have no incentive to reveal information about the abilities of apprentices to outside firms. This is different for school-based vocational or post-secondary education systems: here educational institutions have an incentive to place students well in the labour market, and therefore reveal information about their achievement.

Our empirical analysis begins with a test derived from our model for commitment versus no commitment to training provision. We find that the German apprenticeship system is indeed characterised by commitment to training provision. Using direct information on skill applicability obtained from a large survey data set, we further show that apprenticeship training has only a small firm-specific component (about 5 %), but a substantial occupation-specific component (as high as 35 %).[6] Using longitudinal data on workers drawn from social security records, we also find strong evidence in favour of informational asymmetries between training and outside firms.

We then simulate our model for estimated parameters for the degree of wage compression due to firm-specificity and asymmetric information between training and outside firms. This allows us to compare the training level under commitment with the one that we would observe if firms were unable to commit to training provision. We find that for reasonable degrees of wage compression, training in the no commitment case would be substantially lower, around 8 % of that under commitment. For training to be at least half of that under commitment, we require exceedingly high levels of wage compression that are not supported by the data. These findings suggest that enrollment in apprenticeship programmes would drop dramatically if firms in Germany were no longer able to commit to training provision. This is in line with our hypothesis that apprenticeship programs are only successful if firms are able to commit to training provision.

[6] In related research, Fitzenberger and Spitz-Öner (2004) and Fitzenberger and Kunze (2005) conclude that the specificity of training does not prevent apprentices from switching occupations: They find that occupational switches are frequent, and are likely to occur in order to realize better wage and career prospects.

The remainder of the paper is organised as follows. The next section develops a model firm-specificity, asymmetric information, and firm-financed training, and analyses it under both commitment and no commitment to training provision. Section 3 describes the data. The empirical analysis begins with a test for commitment versus no commitment, and then turns to the transferability of apprenticeship skills and the importance of asymmetric information between training and outside firms (Section 4). Section 5 reports the results from the simulation of our model. We compare the regulation of the German apprenticeship system with those in other countries (in particular in the UK) in Section 6. Section 7 concludes with some policy recommendations.

2 A Model of Firm-Financed Training

This section develops a model where firms have an incentive to training provision because of wage compression due to firm-specificity and asymmetric information. We analyse the model under two scenarios, commitment and no commitment to training provision, and focus on the differences between the two.

2.1 Set-up

There are two periods, the first period is the training period. There are many workers and firms, both are risk-neutral. Firms maximise expected profits, and workers maximise expected utility.

The worker's productivity in training and outside firms in period 2 depends on her ability η, the amount of on-the-job training received in period 1, τ, and the degree of specificity, α :

$$y = \begin{cases} h(\tau)\eta & \text{(training firm)} \\ \alpha h(\tau)\eta & \text{(outside firm)} \end{cases} \tag{1}$$

with $\frac{\partial h(\tau)}{\partial \tau} > 0$ and $\frac{\partial^2 h(\tau)}{\partial \tau^2} \leq 0$,and $0 \leq \alpha \leq 1$. If $\alpha = 0$, training is purely firm-specific; if in contrast $\alpha = 1$, training is purely general. The parameter α is therefore a measure for the transferability of training. Throughout the paper, we treat α as exogenous; hence, firms and workers cannot decide between investing in general or specific skills. In this production function, ability and training are complements, i.e. training increases the productivity of high-ability workers more than that of low-ability workers. This is necessary for asymmetric information to provide an incentive for firms to finance training.[7] To clarify: In this model, workers receive training for one period (e.g. one year), and τ measures *how much* training workers receive during this period. It is therefore best understood as a measure for the *quality* of training.

For simplicity, we assume that workers are either of low (η_L) or high ability (η_H). The share of low-ability workers in the economy is common knowledge, and denoted by p. We normalize the productivity of workers in training to zero. The productivity of untrained workers with low and high ability is given by $\widetilde{\eta}_L$and $\widetilde{\eta}_H$, respectively. This represents the opportunity or fixed cost of training, which are important in the German apprenticeship system (see Section 4.1). The direct training costs are denoted by $c(\tau)$, with $c^{'}(\tau) > 0, c^{''}(\tau) \geqslant 0$. Note that the cost of training does not depend on the transferability of training; hence, it is as costly to provide specific skills as it is to provide general skills.

In the first period, neither firms nor workers observe the worker's ability. In the second period, only training firms get to know the worker's ability whereas outside firms receive no new information about it.

We analyze the model under two assumptions. Under the first assumption, training is not *verifiable,* and firms cannot commit to training provision. Under the second assumption, training is verifiable, and firms can commit to training provision. In both cases, we assume that training is *observable* by outside firms.[8]

[7]In the German apprenticeship system, workers attend vocational schools once or twice per week. We have only modelled the on-the-job training component that takes place within firms. In order to capture skill accumulation due to the attendance of vocational schools, the production function (1) may be augmented by a term $g(\tau^{\text{school}})\eta$. This will not affect the firm's training choice.

[8]Malcomson et al. (2003) also focus on the case in which training is observable, but not verifiable.

After training is completed, workers may either stay with or leave the training firm. We endogenise mobility in a simple way and assume that during the training period workers experience a utility shock θ. This shock captures the worker's ex post evaluation of his or her work environment. For instance, it could reflect how workers like their co-workers. Only workers, but not firms, observe θ. The worker's utility in period 2 at the training firm is the sum of her wage, w, and her utility from non-pecuniary job characteristics, θ. A worker's utility at outside firms is equal to the wage offer of outside firms. The utility shock is drawn from a distribution with the cumulative distribution function G with associated pdf g and support $[\underline{\theta}, \overline{\theta}]$, $\overline{\theta} > 0$. G belongs to the family of log-concave distribution functions, i.e. $(\frac{1-G}{g})' > 0$.

We assume that wages are determined in spot-markets and rule out long-term wage contracts. In the second period, outside firms simultaneously make wage offers to workers by maximising expected profits. Training firms observe the outside offer, and make a counter offer.[9] We further impose the standard free entry condition on firms: No firm earns positive profits in the long-run.

2.2 Analysis

Wage Determination We begin with wage determination in the second period. While the amount of training offered in the first period depends on whether firms can or cannot commit to training provision, the rules for wage determination do not. Let v denote the wage offer of outside firms, and w the wage offer of training firms. A worker stays with her training firm if the wage offer of the training firm w plus the utility shock θ exceeds her outside wage offer v. Hence, the probability that the worker stays with the training firm is given by

$$\Pr(\text{stay}) = \Pr(w + \theta > v) = 1 - G(v - w).$$

[9]This wage determination process is standard in asymmetric learning models, see e.g. Greenwald (1986) and Gibbons and Katz (1991). In contrast, Li (2007) assumes that the incumbent and outside firm make their wage offers simultaneously. He shows that this assumption produces endogenous turnover even in the absence of utility shocks, and thus allleviates the lemons problem.

Training firms observe the worker's ability and therefore offer different wages to low- and high-ability workers. From now onwards, we therefore index wage offers of training firms by ability ($i = L, H$). Taking workers' outside offers and the first period training choice as given, training firms maximise

$$\max_{w_i} \qquad (1 - G(v - w_i))(h(\tau)\eta_i - w_i), \qquad i = L, H.$$

From the first-order condition, wage offers of training firms are implicitly defined as

$$w_i = h(\tau)\eta_i - \frac{1 - G(v - w_i)}{g(v - w_i)}, \qquad i = L, H. \tag{2}$$

Outside firms, in contrast, do not observe the worker's ability and therefore offer the same wage to low- and high-ability workers. Due to perfect competition, outside wages are equal to the expected productivity of those workers who leave the training firm. Low-ability workers are more likely to switch firms, and wage offers of outside firms reflect this adverse selection:[10]

$$v = E[h(\tau)\eta_i|\text{move}] = \frac{pG(v - w_L)\alpha h(\tau)\eta_L + (1-p)G(v - w_H)\alpha h(\tau)\eta_H}{pG(v - w_L) + (1-p)G(v - w_H)}. \tag{3}$$

Training Provision Next, we turn to the firm's decision to train in the first period. Suppose first that firms cannot commit to training provision. In this case, the only training level workers consider credible is the one that maximises the firm's future profit; in other words, the training level that maximises the firm's private return to training. Let Π denote the firm's expected profit in the second period. It is given by:

$$\Pi = p(1 - G(v - w_L))(h(\tau)\eta_L - w_L) + (1-p)(1 - G(v - w_H))(h(\tau)\eta_H - w_H).$$

[10]Note that because outside wages are equal to expected productivity, 'poaching externalities' as discussed in Stevens (1994) are absent in our model.

The training level chosen by the firm under no commitment, τ^{NoCom}, equates the marginal cost of training with the marginal increase of expected profits–provided that the worker receives training:

$$\frac{\partial \Pi_{\tau^{no}}}{\partial \tau} = c'(\tau^{NoCom}). \tag{4}$$

Note, however, that due to the opportunity (fixed) cost of training, not all workers will be trained. Workers will only accept training if their utility with training exceeds that without training. The lower the opportunity cost, the higher the probability that a worker will be trained.

Contrast this with the case of commitment to training provision. Under this assumption, firms choose training by maximising expected profits subject to the constraint that workers are offered a utility at least as high as that received by outside firms; otherwise, firms will not be able to attract any worker. This maximisation problem corresponds to the maximization of the worker's utility subject to the zero profit constraint:

$$\max_{\tau} \; W_\tau + U_\tau \;\; \text{s.t.} \; \Pi_\tau - c(\tau) - W_\tau = 0.$$

Here, W denotes the training wage, U denotes the worker's expected utility in the second period, and Π denotes the firm's expected profit in the second period. The training wage W is determined by the free-entry condition and thus bid up to the point where firms make zero expected profits: $W = \Pi - c(\tau)$. Substituting this into the maximisation problem, the chosen training level under commitment, τ^{Com}, satisfies (provided that the workers receives training)

$$\frac{\partial \Pi_{\tau^{Com}}}{\partial \tau} + \frac{\partial U_{\tau^{Com}}}{\partial \tau} = c'(\tau^{Com}). \tag{5}$$

Comparing the training level under no commitment (equation (4)) and commitment (equation (5)), it is apparent that firms provide more training under commitment ($\tau^{Com} > \tau^{NoCom}$). The reason is that under no commitment, firms choose training by maximising their *private* return to training, while under commitment training is chosen by maximising the firm's and worker's *joint* return to training. Equation (4) also highlights that under no commitment, training will be positive only if the firm's expected profit is increasing in training, i.e. $\frac{\partial \Pi}{\partial \tau} > 0$. This is what Acemoglu and Pischke (1999a, 1999b) refer to as wage compression. In contrast, if firms are able to commit to training provision, training will be positive even in the absence of wage compression–since workers' utility is increasing in training.

It is easy to show that in our model both firm-specific training and asymmetric information lead to wage compression. Using the first-order condition for w_L and w_H (equation (2)), $\frac{\partial \Pi}{\partial \tau}$ satisfies (under both no commitment and commitment):

$$\frac{\partial \Pi}{\partial \tau} = p(1 - G(v - w_L))(h'(\tau)\eta_L - \frac{\partial v}{\partial \tau}) + (1-p)(1 - G(v - w_H))(h'(\tau)\eta_H - \frac{\partial v}{\partial \tau}).$$

Rearranging this equation, $\frac{\partial \Pi}{\partial \tau} > 0$ if

$$\frac{\partial v}{\partial \tau} < \frac{p(1 - G(v - w_L))h'(\tau)\eta_L + (1-p)(1 - G(v - w_H))h'(\tau)\eta_H}{p(1 - G(v - w_L)) + (1-p)(1 - G(v - w_H))} = E[h'(\tau)\eta|\text{stay}].$$

This condition says that wages are compressed if training results in a larger increase in the productivity than in the outside option v for *those workers who stay with the training firm.* It holds in equilibrium for two reasons. First, due to firm-specific human capital, an increase in training by one unit rises productivity at incumbent firms by $h'(\tau)\eta$, but at outside firms only by $\alpha h'(\tau)\eta$. Second, due to asymmetric information, more able workers are more likely to stay with the training firm and training raises the productivity of high-ability workers more than that of low-ability workers.

Who bears the training costs under no commitment and commitment, firms or workers? Let $\tau^i, i = NoCom, Com$ denote the training level under no commitment and commitment, respectively. The total training cost is the sum of the opportunity and the direct training cost, $p\widetilde{\eta}_L + (1-p)\widetilde{\eta}_H + c(\tau^i)$.

The trainee receives an apprenticeship wage equal to $W_{\tau^i} = \Pi_{\tau^i} - c(\tau^i)$. If she had not been trained, she would earn a first period wage equal to $p\widetilde{\eta}_L + (1-p)\widetilde{\eta}_H + \Pi_0$–i.e. her expected productivity plus the firms' expected second period profit in the absence of training. Consequently, firms and workers share the cost of training under both commitment and no commitment: Firms pay $\Pi_{\tau^i} - \Pi_0$–i.e. the difference between their expected profits of trained and untrained workers–, while workers pay $(p\widetilde{\eta}_L + (1-p)\widetilde{\eta}_H + \Pi_0) - (\Pi_{\tau^i} - c(\tau^i))$ –i.e. the difference between the first period wage of unskilled workers and that of apprentices.

To summarise: If firms are unable to commit to training provision, wage compression is necessary to ensure a positive training level.[11] If, in contrast, firms are able to commit to training provision, training will be positive even in the absence of wage compression. Wage compression is, however, necessary for firms to bear (part of) the training costs; if wages are not compressed (i.e. $\Pi_\tau - \Pi_0$), workers alone finance training. There is a second crucial difference between commitment and no commitment. Under no commitment, apprentices receive a wage that *exceeds* their productivity, i.e. $W_{\tau^{NoCom}} > 0$. This is because firms will not choose a training level that results in negative expected profits, and workers are not willing to accept a wage cut during training. Under commitment, in contrast, workers are willing to do so, and consequently apprenticeship wages may be *lower* than the productivity of apprentices. This is a testable implication which we will use below to test between commitment and no commitment in the case of the German apprenticeship system.

It is important to stress that welfare under commitment will be at least as high as under no commitment, regardless of the degree of wage compression. To see this, first note that the appropriate welfare measure in our economy is the utility of the worker, $W + U$, because firms earn zero expected profits in equilibrium. Under commitment, the training level is chosen by maximising the worker's utility (see equation (5), while under no commitment it is chosen by maximising the firm's future profit (see equation(4)). Hence, workers are at least as well off under commitment.

[11]Note that due to the fixed cost of training, wage compression is not sufficient to ensure a positive training level.

2.3 Model Simulation

While theory unambiguously predicts that training (and welfare) will be lower under no commitment, it is an empirical question *how much lower* it will be. In an attempt to quantify the impact of commitment on training, we simulate our model under the assumption of commitment to training provision (later, we provide evidence that this is the case in the German apprenticeship system), and then compare the training level with the one that would arise if firms were unable to commit. We choose the parameter values in our model by matching key moments implied by the model to those observed in the data. This section describes in more detail how we simulate the model.

First, we need to impose functional forms on the production and cost technology, as well as on the distribution of utility shocks. We choose a particularly simple form for the production and cost technology: $h(\tau) = \tau$ and $c(\tau) = 0.5\tau^2$. We further assume that utility shocks are drawn from a logistic distribution with mean 0 and scale parameter b. We normalize the productivity of low-ability workers, η_L, to 1, and set the share of low-ability workers, p, to 0.5. We check the robustness of our results by choosing alternative specifications for the production and cost technology, and alternative values for the share of low-ability workers.

This leaves the following parameters of our model to be specified: α, the degree of specificity of training; b, the scale parameter of the logistic distribution; and η_H, the productivity of high-ability workers. Next, we describe in turn how we choose these parameter values.

Transferability of Training (α) We exploit direct information on skill applicability to estimate the degree of specificity of apprenticeship training. In our data, individuals are asked how much of the skills obtained during apprenticeship training are applicable at their current firm as well as whether they are still employed at the apprenticeship firm. We use this information to estimate α. We obtain further an estimate of the occupation specificity of apprenticeship training by comparing workers who are still employed in their apprenticeship occupation with those who are not.

The Scale Parameter of the Logistic Distribution (b) **and the productivity of high-ability workers** (η_H) We use two data moments to pin down these two parameters, the quit rate after training and the wage differential between workers who (initially) stay with or move away from the training firm.

In our model, there are three reasons why movers earn different wages from stayers. The first reason is (firm-) specific training, implying that stayers are more productive than movers. This is captured in our model by the parameter α–for which we have an estimate. The second reason is asymmetric information. This is captured in our model by the parameter η_H; the higher the productivity of high-ability workers, the larger the informational advantage of training firms. In the extreme case in which low- and high-ability workers are equally productive, training firms are no longer better informed about workers' ability than outside firms. The third reason why wages of stayers differ from those of movers is non-pecuniary job characteristics, allowing training firms to retain workers even if they pay wages below productivity. In the extreme case in which the distribution of non-pecuniary job characteristics becomes degenerate (i.e. $b \to 0$), workers will leave the firm whenever they are paid a higher outside wage. In general, the higher b, the higher the quit rate.

We compute for a fine grid of values for b and η_H the optimal training level as well as the optimal wage offers of training and outside firms, using the (first-order) conditions (2), (3), and (5), respectively. We do this under the assumption of full commitment, and for our estimates for the degree of transferability, α. We then compute the quite rate and the mover-stayer wage differential implied by our model. Let q^M and MS^M denote these two model moments; both are complicated functions of b and η_H. The corresponding data moments are denoted by q^D and MS^D. We finally pick those values for b and η_H that minimize the sum of the squared distance between the model and the data moments:[12]

$$\min_{b,\eta_H} \quad (q^M(b,\eta_H) - q^D)^2 + (MS^M(b,\eta_H) - MS^D)^2.$$

In reality, there may be reasons other than specific human capital, asymmetric information, and non-pecuniary job characteristics why wages of movers differ from wages of stayers. Alternative reasons include job search and worker sorting. Ignoring these alternative reasons may give a misleading picture of the importance of wage compression in general, and asymmetric information in particular. We discuss this concern in detail in Section 4.3.

A further parameter in our model is the opportunity or fixed cost of training, $p\widetilde{\eta}_L + (1-p)\widetilde{\eta}_H$. This parameter measures the difference between the productivity of a worker in training and an untrained worker, and determines whether a worker receives training or not. It does not affect the welfare loss due to no commitment, specific training and asymmetric information, *provided that the worker receives training.*[13] Our simulation assumes that the fixed training costs are so low that this is the case.

For these parameter values, we then compare training and welfare under commitment and no commitment to training provision.

3 Data

We use two different data sources to estimate the quit rate and the mover-stayer wage differential after apprenticeship training, and the degree of specificity of apprenticeship training. We describe each data set in turn.

[12]Figure B.1 in Appendix B shows a three-dimensional graph, with the productivity of high-ability workers η_H on the x-axis, the scale paremeter b on the y-axis, and the distance on the z-axis. The figure shows that the objective function (i.e. the sum of the squared distance between the model and the data moments) is well-behaved in the sense that there is only one minimum, suggesting that b and η_H are uniquely identified.

[13]It is possible that for some values for the opportunity costs of training, workers will get trained if firms can commit to training provision, or if training is general or information is symmetric, but not if firms are unable to commit to training provision, or if training is specific or information is asymmetric. In these cases, the opportunity cost of training affects the welfare loss due to limited commitment to training provision, specific training and asymmetric information.

Longitudinal Data on Mobility and Wages Our first data set is a one percent sample of administrative social security records in Germany from 1975 to 1995. We use this data to estimate the quit rate as well as the mover-stayer wage differential after apprenticeship training. The data are representative of all individuals covered by the social security system, roughly 80 percent of the German workforce. It excludes the self-employed, civil servants, and individuals currently doing their compulsory military service. As in many administrative data sets, our data is right-censored at the highest level of earnings that are subject to social security contributions. However, this is not a serious problem in our sample of young workers, as less than 0.5 percent of wages are top-coded.[14]

Since plant size is recorded only from 1980 onwards, we restrict our sample to men who finished apprenticeship training after 1980. We drop all workers who were employed at least once in East Germany because the level and structure of wages differs substantially between East and West Germany. We also exclude workers who completed more than one apprenticeship, as well as workers who additionally graduate from community college (*Fachhochschule*) or university. Finally, in order to observe all individuals for at least three years in the labour market, we drop individuals who finished apprenticeship training after 1992. We provide further details about the sample selection and variable definitions in Appendix A.1.

Our final sample consists of 21395 individuals; we observe 5823 individuals with more than 10 years of potential experience after apprenticeship training.[15] We refer to this data as *Sample 1.* We also extract from this data set a sub-sample where we observe more than one apprentice in the same firm. We refer to this sample as *Sample 2.*

Table 1 summarises the main characteristics of our two samples. The main difference between the two samples is the average size of the training firm: Sample 2 heavily selects on large firms (1195 vs

[14]The data used by Wachter and Bender (2006) is drawn from the same data base as ours. There are, however, important differences between their and our sample. Most importantly, Wachter and Bender (2006) discard all apprentices trained in firms with less than 50 employees or less than 5 apprentices - which amounts to more than 50 % of all apprentices.

[15]There is a total of 16012 observations for workers with at least 10 years of potential work experience.

5521). The entries in the table show that around two thirds of individuals remain initially with their training firm after graduating. Three years after apprenticeship training, about one third are still employed at the apprenticeship firm. This number reduces to 17.5 % ten years after apprenticeship training. Occupational mobility, measured at the two-digit level, is considerably lower than firm mobility, but also substantial. Firm mobility rates are smaller in sample 2 that over-samples large firms. Interestingly, this is not true for occupational mobility. This may reflect a higher occupational mobility within firms in large firms, possibly due to internal labour markets.

Cross-Sectional Data on Skill Applicability Our second data come from the repeated cross-section *German Qualification and Career Survey* (GQCS), which is conducted jointly by the Federal Institute for Vocational Education and Training (BIBB) and the Institute for Employment (IAB). We use this survey to estimate the degree of specificity of apprenticeship training. The survey, previously used for example by DiNardo and Pischke (1997) and Acemoglu and Pischke (1998), is available for four different years: 1979, 1985, 1991/92 and 1998/99. Each wave contains information from 30,000 employees between the ages of 16 and 65. In what follows, we restrict our analysis to men who completed an apprenticeship degree, but no higher degree. That is, we drop workers with a community college (*Fachhochschule*) or university degree from our sample. Since key variables are missing for wave 4, we restrict our sample to waves 1 to 3. We provide further details on the sample selection in Appendix A.2. Our final sample consists of 25380 individuals. We refer to this data as *Sample 3*.

In each wave, workers are asked about the proportion of skills acquired during apprenticeship training that are applicable at their current job. The survey also includes detailed information on the apprenticeship, such as occupation and the size and industry of the apprenticeship firm. Moreover, workers are asked about their current occupation, as well as whether they are still employed at the apprenticeship firm, allowing us to analyze whether skills are firm- or occupation-specific. Table

2 provides an overview of the main variables used in the empirical analysis. There are important differences between this sample and our previous two samples (Sample 1 and 2). While Sample 1 and 2 follow workers from apprenticeship completion onwards for up to 13 years in the labour market, the average time since apprenticeship completion in Sample 3 is 17.61 years. Moreover, in Sample 1 or 2, individuals finished the apprenticeship between 1980 and 1992. In Sample 3, in contrast, many individuals completed apprenticeship training in the 60s and 70s. This may explain why firm and occupational mobility rates based on Sample 3 are somewhat lower than those based on Sample 1; among workers who finished apprenticeship training less than a year ago, 78.26 % (84.66 %) are still employed with their training firm (occupation), compared to less than 67 % (72 %) in Sample 1.

4 Empirical Analysis

4.1 Are Firms Able to Commit to Training Provision?

Is the German apprenticeship system characterised by commitment or no commitment to training provision? Our model suggests a simple test to discriminate between the two scenarios. Under no commitment, apprentices are not willing to accept a wage cut to finance training and consequently apprenticeship wages exceed the productivity of apprentices. Under commitment, in contrast, apprenticeship wages may be below productivity.

Reliable information on the productivity of apprentices is typically difficult to find. However, a recent study by Beicht et al. (2004) quantifies the firm's gross and net cost of apprenticeship training, and assesses the productivity of apprentices. The study is based on a representative survey of 2500 firms involved in vocational training in the year 2001. In Table 3 we report estimates on productivity of apprentices, as well as direct cost of apprenticeship training from this study (columns1 and 2), as well as estimates of apprenticeship wages and wages of unskilled workers (for the year 2001), which we compute from the IAB data (columns 3, 4).

The productivity of an apprentice during training is estimated to be 7730 Euros (Column 1). For the same time period the average apprenticeship wage is 7031 Euros (Column 3). Hence, apprenticeship wages are below productivity. This is only compatible with commitment, but not with no commitment to training provision. It also implies that workers pay at least some of the training cost.

Do only workers pay for training, or do firms sponsor training too? The direct cost of training, such as personnel costs for trainers, plant and material costs, and teaching materials, amount to 2663 Euros per year (Column 2). In a competitive market in which workers bear all the training costs, workers should receive an apprenticeship wage of 5067 (7730-2663) Euros. The apprenticeship wage that workers actually receive is with 7031 Euros considerably larger. Hence, firms bear some of the training costs.

Column 4 shows that the wage of unexperienced (i.e. workers who entered the labour market at the same time as apprentices), unskilled workers is substantially larger than the apprenticeship wage, which points to a considerable *opportunity cost* of apprenticeship training. Our model implies that the worker's cost share can be computed as the ratio between the wage of unskilled workers minus that of apprentices, divided by the total (i.e. direct plus opportunity) cost of training (see Section 2.2). Since we do not observe the productivity of unskilled workers, we approximate the opportunity cost as the difference between the unskilled wage and the productivity of apprentices. According to this calculation, workers bear about 70 %, and firms 30 % of the total training cost (Column 6).[16]

In our model, firms are willing to pay for training only if wages are compressed. The next section tests for two sources of wage compression, firm-specificity and asymmetric information. This section also reports the estimates of the data moments used to pin down the parameters for our model simulation.

[16] One problem with this approach of estimating the worker's cost share is the selection into apprenticeship training. If apprentices are of higher ability than unskilled workers, then the productivity of unskilled workers is a lower bound for the hypothetical productivity of apprentices in case they had not been trained; we therefore overestimate the opportunity cost of apprenticeship training. Hence, our estimates for workers' share are best interpreted as lower bounds for the true share.

4.2 The Transferability of Apprenticeship Training

How transferable are apprenticeship skills? This section attempts to identify α, the degree of transferability. We do that by using survey responses about how much of the skills learned as an apprentice can be applied to the current job, based on the GQCS data we described above. Survey responses are categorical, distinguishing between very much, a lot, some, not too much, and very little. We first create a variable that assumes the value 1 if the worker can apply very much or a lot of the skills learned as an apprentice, and 0 otherwise. We then quantify the extent to which human capital is specific rather than general, by assigning to the five categories very much, a lot, medium, not too much, very little the values 90, 75, 50, 25, and 10.

Table 4 reports results. In columns 1 and 2 we report marginal effects from a simple probit model on our binary indicator, where we report estimated coefficients of a regressor whether workers are still employed at the apprenticeship firm (stayers) or not (movers). In columns 3 and 4 we report regression results where the dependent variable is the "share" of generality we assign to each category. Specifications 1 and 3 only regress on a binary variable whether the individual is a "mover", while specifications in columns 2 and 4 regress in addition on the year the apprenticeship ended, experience and its square, firms size, apprenticeship occupation, school degree, and age at end of apprenticeship. Panel A uses all observations, while Panel B distinguishes between workers who lose their job for exogenous or other reasons. Finally, Panel C distinguishes between firm- and occupation movers.

Results in columns 1 and 2 in panel A suggest that the proportion of individuals that report that they use very much or a lot of the skills they learned during apprenticeship is about 30 % higher for stayers than for movers, regardless of whether or not we condition on observable characteristics. The results in columns 3 and 4 indicate that workers who are no longer employed with their apprenticeship firm can apply 19 % less of their skills. From those estimates, we obtain an estimate for $\alpha = 1 - 0.19 = 0.81$.

A possible problem with these estimates is that workers who choose to leave the training firm are not randomly selected. On average, we would expect workers to sort into new jobs where a larger fraction of the skills acquired during apprenticeship training can be transferred. In this case, our estimates for the specificity of training would be lower bounds. In an attempt to assess the importance of selection, we follow Acemoglu and Pischke (1998) and distinguish between individuals who left the training firm because they were drafted into the compulsory military service and individuals who left the training firm for other reasons. The idea is that draft is exogenous. In our sample, 5.29 % of all apprentices, and 7.55 % of apprentices who are no longer employed at the training firm, are drafted into the military service at the end of the apprenticeship. Panel B of Table 4 reports results. Differences between exogenous and endogenous movers tend to be small in magnitude, suggesting that selection into jobs to which a larger fraction of skills can be transferred is not very important.

Our estimates above do not distinguish between workers who change firms and occupation, or workers who only change firms. We investigate the possibility of occupation-specific skills by comparing the applicability of skills for four groups of workers: workers who are still employed in the apprenticeship firm and occupation, workers who switched occupations, but are still employed in the apprenticeship firm, workers who left the apprenticeship firm but are still employed in the apprenticeship occupation, and workers who left both the apprenticeship firm and occupation. Occupations are defined at the two-digit level; we distinguish 93 different occupations.

Results can be found in Table 4, Panel C. Results in column 3 indicate that workers who left the apprenticeship firm but not the apprenticeship occupation can employ 4.53 % less of their skills, compared to 8.6 % for workers who are still employed at the training firm, but have left the training occupation, and 34.27 % for workers who left both the training firm and occupation. Results are very similar if we do condition on observable worker and training firm characteristics (column 4). Our overall conclusions are also unchanged if we estimate probability models with a dummy variable indicating that a worker can apply very much/a lot (specification 1) as the dependent variable.

We simulate our model using two different estimates for transferability of training, α. First, we only allow for firm-specific skills, $\alpha = 0.95$ (Table 4, Panel C). This is motivated by the fact that only firm-specific, but not occupation-specific training lead to wage compression, and provides an incentive for firms to finance training. In order to take into account occupation-specific skills, we repeat the simulation using $\alpha = 0.81$ (Table 4, Panel A).

4.3 The Importance of Asymmetric Information: The Mover-Stayer Wage Differential

Are training firms better informed about the worker's ability than outside firms? This section reports estimates for the quit rate and the mover-stayer wage differential. We use these data moments to pin down the scale parameter of the logistic distribution, b, and the productivity of high-ability workers, η_H–which in our model determines the importance of asymmetric information.

Baseline Results Table 5 displays our baseline results, based on sample 1. About one third of workers leaves the training firm after apprenticeship completion. Workers who move away from the training firm (initially) earn 8.6 % lower wages than workers who stay with the training firm. The wage disadvantage of movers reduces to 6.7 % if we condition on worker and firm characteristics, such as age at the beginning of the apprenticeship, apprenticeship duration, high school degree (*Abitur*), as well as the apprenticeship occupation and the size of the apprenticeship firm.

Confounding Mechanisms Our model only allows for three reasons why wages of movers differ from wages of stayers: specific training, asymmetric information, and non-pecuniary job characteristics. There may however be other reasons why wages of movers differ from wages of stayers. Ignoring these alternative reasons may bias our estimates for the degree of wage compression due to asymmetric information. Next, we discuss alternative explanations why wages of movers may differ from those of stayers, and explain how we deal with this concern.

Job Search: One reason why movers may earn lower wages than stayers that is not covered by our model is job search. Movers may be worse matched with their new firm than stayers. If this is the case, we would expect the mover-stayer wage differential to gradually disappear over time. Table 6, Panel A, analyses the evolution of the wage differential up to 10 years after apprenticeship training. We estimate a separate wage regression for each experience level. The first specification only controls for the year the apprenticeship ended, while the second specification additionally conditions on worker and firm characteristics. In both specifications, the gap roughly remains constant during the first ten years in the labour market.[17] We interpret this as evidence that movers are negatively selected in terms of ability, just as asymmetric information predicts.[18]

Worker Sorting: An adverse selection of movers is a necessary implication of asymmetric information between training and outside firms. There are, however, alternative explanations why movers are less able than stayers. One possibility is worker sorting. Suppose that high-ability workers sort into firms that offer apprenticeship programmes of higher quality, and that the separation rate after training is lower in these firms. Such a sorting model also predicts a lower ability of movers. Our results so far show that controls for observable worker and firm characteristics reduce the wage differential between movers and stayers, indicating that worker sorting is indeed important. However, this specification does not account for sorting of workers into training firms *within* observable firm characteristics. We address this concern by comparing wages of workers who have been trained in the same firm, and among whom some stay and others leave the training firm. This corresponds to conditioning on training firm fixed effects. Results can be found in Table 6, Panel B. They refer to sample 2 which heavily over-samples large firms –see Table 1. We first control for the year the apprenticeship ended only. According to this specification, movers earn 13.0 % lower wages than stayers in the year

[17]These results refer to an unbalanced panel: The wage regression right after apprenticeship training is based on more workers–who on average completed apprenticeship training earlier–than the wage regression five or ten years after apprenticeship training. This is problematic if the mover-stayer wage differential differs across cohorts. We have repeated the analysis using a balanced panel that includes only apprentices with at least ten years of potential experience (i.e. apprenticeship completion between 1980 and 1985). The results are very similar to those for the whole sample, indicating that the mover-stayer wage differential is similar across cohorts.

[18]These results are in line with Acemoglu and Pischke (1998), von Wachter and Bender (2006), and Euwals and Winkelmann (2004).

after the apprenticeship ended, and 8.4 % ten years later. This differential is considerably larger than in the whole sample (see Panel A). This is because the mover-stayer wage differential is larger for workers trained in large firms, and large firms are over-represented in our sample. We then condition on observable worker and firm characteristics. In contrast to the whole sample, this hardly reduces the mover-stayer wage differential. Finally, we additionally include fixed training firm effects. This slightly reduces the mover-stayer wage differential after apprenticeship training to 10.4 %. The gap is equally large 10 years after apprenticeship training. These results show that worker sorting alone cannot account for the lower ability of movers.

Wage Floors: A further explanation for a lower ability of movers are wage floors. According to this explanation, it is only laid-off workers who are less able than stayers, but not workers who voluntarily leave the training firm. Asymmetric information, in contrast, predicts a lower ability not only for laid-off workers, but also for voluntary movers. Unfortunately, our data does not allow us to distinguish between layoffs and quits. We approximate a layoff by a job-to-unemployment transition. In our sample, 15.21 % of workers switch to another firm without an intervening unemployment spell after apprenticeship training, compared to 19.74 % who move into unemployment first.

Table 6, Panel C, reports results. We first only control for the year the apprenticeship ended, and then additionally condition on observable worker and firm characteristics. Results refer to sample 1. In the final specification, we include fixed training firm effects, using sample 2. According to all specifications, job-to-unemployment movers earn lower wages than job-to-job movers, except in the year following the apprenticeship. Importantly, wages of job-to-job movers are lower than those of stayers even 10 years after the apprenticeship ended. Hence, wage floors do not appear to be the only reason for the adverse selection of movers.

To summarize, while job search, worker sorting, and wage floors may contribute to the disadvantage of movers, they cannot fully explain it. We interpret this as strong evidence in favor of informational asymmetries between training and outside firms.

In order to take into account these confounding mechanisms not captured by our model, we simulate the model using three different targets for the mover-stayer wage differential, and analyze how our conclusions change if different targets are used. Our first target is the mover-stayer wage differential immediately after training is completed, conditional on observable training firm characteristics (-0.067; Table 5). This target accounts for the sorting of workers into training firms based on observable firm characteristics. Our second target is the mover-stayer wage differential, conditional on observable firm characteristics, 10 years after training completion (-0.059; Table 6, Panel A). This target takes into account that movers may initially earn lower wages because they are worse matched than stayers. Our third target is the *job-to-job* mover-stayer wage differential conditional on observable training firm characteristics, 10 years after training completion (-0.012; Table 6, Panel C, Specification 2). This target additionally takes into account that movers may be of lower ability than stayers because of wage floors, rather than asymmetric information.

5 Model Simulation

Above, we provide evidence that in the German apprenticeship system, firms are able to commit to training provision. We further find evidence in favour of partially specific training and asymmetric information–which both lead to wage compression. Consequently, the training market may not break down completely if firms were unable to commit to training provision. But how much lower would training (and welfare) be under no commitment compared to under commitment? To answer this question, this section simulates our model for reasonable degrees of wage compression due to asymmetric information and specific human capital. We do this under the assumption of commitment. We then compute the training level that would arise if firms were unable to commit to training provision.

We report our results in Table 7. As discussed, we simulate the model using three different targets for the mover-stayer wage differential (Panel A to Panel C). In each panel, the first row uses our estimate for the firm-specificity of training, $\alpha = 0.95$, while the second row takes into account the loss of occupation-specific skills, $\alpha = 0.81$. Columns 1 and 2 report the scale parameter of the distribution of non-pecuniary job characteristics (b) and the productivity of high-ability workers (η_H), chosen to match the mover-stayer wage differential and the quit rate implied by our model with those observed in the data; see Section 2.3 for details. Columns 3 and 4 compare training under commitment (Com) with that under no commitment ($NoCom$), while columns 5 and 6 show the worker's life-time utility under the two scenarios. Since firms earn zero profits in our model economy, this is an appropriate welfare measure. Finally, column 7 displays welfare under the social optimum when firms are able to commit to training provision, information is symmetric, and training is purely general.

We begin with a discussion how the transferability of skills α and the alternative targets for the mover-stayer wage differential affects the estimates for scale parameter of the distribution of non-pecuniary job characteristics (b) and the productivity of high-ability workers (η_H) (Columns 1 and 2). Not surprisingly, if the wage disadvantage of movers relative to stayers used as a target is lower, the productivity of high-ability workers required to fit the data decreases (e.g. $\eta_H = 3.22$ and $\eta_H = 2.79$ when the mover-stayer wage differential is 0.067 (Panel A, row 1) and 0.012 (Panel C, row 1). Also note that compared to $\alpha = 0.95$ (row 1), the estimate for η_H decreases and the estimate for b increases when the loss of occupation-specific skills is taken into account ($\alpha = 0.81$, row 2). This is because the model now attributes a larger share of the wage disadvantage of movers to the loss of specific skills as opposed to asymmetric information–hence the lower estimate for η_H. Moreover, since a higher degree of specificity reduces the quit rate after training, the scale parameter of the logistic distribution has to increase in order to match the quit rate observed in the data.

Next, consider the impact of commitment to training provision on the training level (Columns 3 and 4). For all specifications, no commitment to training provision would result in a *substantially* lower training levels, at most 8 % of that under commitment. Hence, *reasonable degrees of specificity and asymmetric information are therefore not sufficient to produce a training level that is close to that under commitment.*

To better understand why no commitment results in so much lower training levels, consider again the first order conditions for training under no commitment and commitment, equations (4) and (5). Under commitment, the chosen training level maximises the firm's and worker's *joint* return to training. Under no commitment, in contrast, firms choose training by maximising their *private* return to training. The larger wage compression, the lower is the worker's return, and the larger is the firm's return to training. For reasonable degrees of wage compression due to specificity and asymmetric information, the worker's return to training is much larger than the firm's return; training under no commitment is therefore substantially lower than under commitment.

How does the lower training level under no commitment affect welfare? Columns 5 and 6 show that for all specifications, welfare under no commitment is substantially lower (around one third) of what it would be under commitment. This computation assumes that the fixed or opportunity cost of training is so low that workers prefer low training levels over no training at all. However, in practice the opportunity cost may be high enough so that workers would actually choose no training over low training levels. We would then expect workers' enrollment in apprenticeship programmes to drop dramatically if firms in Germany were no longer able to commit to training provision.

We investigate the relationship between wage compression and training provision under no commitment further in Figure 2. Panel A plots the ratio between training under commitment and no commitment for two values of the productivity of high-ability workers, $\eta_H = 3$ and $\eta_H = 4$, and various values for the transferability of training, α. The figure assumes that the scale parameter of the logistic distribution is equal to 0.46. The lower value of $\eta_H, \eta_H = 3$, is within the range of our estimates in Table 7. When we compute the mover-stayer wage differential implied by our model (under commitment) for the higher value of $\eta_H = 4$, we obtain -0.172%, which is well outside our estimates for the wage disadvantage of movers. This value is therefore likely to overestimate the importance of asymmetric information.

As expected, the figure demonstrates that training under no commitment moves closer to that under commitment if training becomes more specific (i.e. $\alpha \rightarrow 0$), or if the productivity of high-ability workers–and therefore the importance of asymmetric information–increases. Interestingly, if training is purely firm-specific ($\alpha = 0$), training under no commitment is about 90 % of that under commitment. For such high values of wage compression, the workers' return to training is low, while the firm's return is large; training under no commitment is therefore close to that under commitment.

Note, however, for training under no commitment to be at least 50 % of that under commitment, we require, for $\eta_H = 3$, the transferability of apprenticeship skills to an outside firm to be lower than 50 %. It is important to stress the firm's profits are increasing in training only if training is firm-specific, but not if it is occupation-specific and therefore equally valuable in outside firms within the same occupation. Clearly, $\alpha = 0.5$ is outside our range of estimates for the degree of firm-specificity; see Table 4. Alternatively, if $\eta_H = 4$, training under limited commitment is about 50 % of that under commitment even if training is purely general. However, as argued above, this value of η_H implies a disadvantage of movers that is not supported by our data.

In Panel B, we compare the welfare under limited commitment and wage compression with that in the first-best situation, i.e. commitment and no wage compression (that is, purely general training and symmetric information). The figure demonstrates that welfare under no commitment moves closer to the first-best if wage compression increases. This is the well-known theory of the second-best at work: If one market failure is present, other market failures may improve welfare. The figure demonstrates that this mechanism can be very powerful. For instance, if training is purely specific, welfare under no commitment is about 80 % of the first-best situation, compared to 20 % when training is purely general.

We would like to highlight one additional finding. While some skills may be taught more effectively in a work environment than in the classroom, a school-based system of vocational education may avoid the three problems potentially inherent in a firm-based apprenticeship system: the inability of firms to commitment to training provision, the informational asymmetry between training and outside firms, and the provision of firm-specific skills. While in the German apprenticeship system appears to have solved the first problem, the two latter problems are present. In the last column (column 7), we report welfare in the social optimum when firms are able to commit to training provision, information is symmetric, and training is purely general. Comparing these numbers to those in column 5, it is apparent that specific training and asymmetric information lead to a welfare loss. While the exact numbers are admittedly somewhat difficult to interpret, the numbers suggest that welfare would be about 8-12 % higher if there were no informational asymmetries between training and outside firms and if training were purely general. We view this as a cost of firm-based apprenticeship training relative to a system of school-based vocational education.

Additional Robustness Checks We have conducted a number of robustness checks. Our welfare implications are similar if the proportion of low-ability workers is set to 0.25 and 0.75, respectively. Our overall conclusions are also unchanged if we use the cost function $c(\tau) = 1/3\tau^3$ instead of the quadratic cost function, or the production function $h(\tau) = \sqrt{\tau}$ instead of the linear production function.

Our model assumes that only training firms learn about workers' abilities, whereas outside firms receive no new information about these. A more realistic assumption would be that outside firms also learn about workers' abilities, but not as much as training firms. This may be modeled as follows (see also Schönberg 2007). At the end of the training period, training firms observe the worker's ability without error, while outside firms receive a noisy signal about the worker's ability. Suppose that the signal can take two values, good and bad. Let q denote the probability that a low-ability (high-ability) worker generates a bad (good) signal. Ideally, we would then like to identify the precision of the outside firm's signal, q, in conjunction with the ability of high-ability workers, η_H. Unfortunately, the quit rate and the mover-stayer wage differential are not sufficient to separately identify these two parameters. As an additional robustness check, we match the productivity of high-ability workers η_H

for alternative values of q. This does not change our most important finding: For those combinations of η_H and q that match the model to the data moments, no commitment would result in a substantially lower training level (at most 10 % of that under commitment).

To summarise: For reasonable values of wage compression due to specificity and asymmetric information estimated from survey and administrative data, training and welfare are *substantially* lower under no commitment than under commitment. For welfare under no commitment to be at least 50 % of that under commitment we require degrees of firm-specificity and asymmetric information that are not supported by our data. These findings suggest that participation in apprenticeship programmes would drop dramatically if firms in Germany were no longer able to commit to training provision. We interpret this as evidence that apprenticeship schemes are only successful if firms are able to commit to training provision.

6 Discussion: Regulation of Apprenticeship Schemes

Which policies can be implemented in order to ensure that firms are able to commit to training provision? This section reviews the institutional features of the German apprenticeship system that–we argue–help firms to do so. We then turn to apprenticeship schemes in other countries (most importantly in the UK), and argue that the institutional set-up in these countries makes it more difficult for firms to commit to training provision.

The German Apprenticeship System Several features of the German apprenticeship system help ensure commitment to training provision. First, the Vocational Training Act of 1969 clearly defines the legal rights and obligations of apprentices and their employers. For instance, apprentices have the right (and obligation) to attend a vocational school once or twice per week, where they are taught general subjects, such as math and English, as well as subjects specific to their occupation. This allows apprentices to regularly meet with apprentices from other firms, and thus provides an opportunity to find out about training in other firms. Vocational schools may thus also serve as an information intermediary. Apprentices are not obliged to perform tasks that are unrelated to training. Apprenticeship contracts also put restrictions on the amount of time apprentices are allowed to spend in activities with a low training content, such as cleaning or running errands (see e.g. Bundesministerium fuer Bildung und Forschung, 2005 for more details). This reduces the risk that firms exploit apprentices as cheap labour.

Second, the workplace training content is subject to several regulations, and firms that train have to follow a prescribed curriculum. The setting of nation-wide (occupation-specific) training standards may be necessary to make workplace training verifiable–which is necessary for firms to be able to commit to training provision. Chambers of crafts and industry/trade also regularly monitor training firms, and have the power to withdraw the firm's permission to train apprentices if firms do not meet the required standard (see e.g. Muench, 1992 and Franz and Soskice, 1995 for more details). To see how monitoring helps to achieve commitment, suppose that the training standard is above the training level under no commitment, τ^{NoCom} (equation (4)). Chambers may 'name and shame' firms that offer training below this standard, or even ban them from training in the future. This then provides information to young workers which firms are able to commit, and which are not. This policy–in combination with the required school attendance once or twice per week–is likely to have the added benefit of reducing the specificity of apprenticeship training. Our findings indicate that the firm-specific component of apprenticeship training in Germany is small (less than 5 %), although the occupation-specific component is substantial (up to 35 %).

Third, apprentices sit two examinations during the apprenticeship programme, one mid-term exam and one final exam (see Muench, 1992). The exams are centralised and organised by the chambers. After having successfully completed the final examination, apprentices are issued a certificate that lists in detail the subjects an apprentice has taken as well as his or her performance in each subject. Certificates are accepted nation-wide as skill qualification. To see how this policy helps to ensure commitment to training provision, suppose that the exam is designed such that a (high-ability) worker who received the training level under no commitment, τ^{NoCom} (equation (4)), fails the exam while a (low-ability) worker who received the training level under commitment, τ^{Com} (equation (5)), passes the exam. In this case, the the firm's pass rate will signal to workers whether the firm is able to commit training provision or not. The external testing of apprentices is likely to also reveal information about the apprentice's ability, thereby reducing the informational advantage of the training firm. However, our results indicate that it does not eliminate the informational asymmetry between training and outside firms. As discussed above, this may be viewed as a cost of a firm-based apprenticeship system that could potentially be avoided in a school-based system of vocational education.

In sum, the regulation of the German apprenticeship system allows firms that want to train apprentices, but have not done so in the past, to connect to an existing umbrella framework that signals commitment to new apprentices. This does not mean that regulation is perfectly effective. It is quite possible that there are some 'black sheep' (firms that do not commit to training provision) also in the German system.[19] However, the existing regulatory framework helps to identify these black sheep, and reduces the firm's incentive to shirk in the first place.

Apprenticeship Training in Other Countries How does the regulation of apprenticeship training in other countries differ from that in Germany? The legal definition of the rights and obligations of apprentices and their employers, the setting of nation-wide (occupation-specific) training standards, the external testing of apprentices, as well as the combination with vocational school-based training are also important features of apprenticeship training in other countries that run successful schemes, such as Austria, Denmark, and the Netherlands (see e.g. Ryan, 2000 for an overview). The UK, however, is different. Like Germany, it has a long history of apprenticeship training, but participation rates have decreased throughout the 80s (see also Figure 1). In a major attempt to revitalise

[19]In fact, the work by Malcomson et al. (2003) assumes that regulation can never be 100% effective at ensuring enforceability of the training contract. Among other things, they analyse how regulation may nevertheless improve welfare even if it does not solve the commitment problem.

apprenticeship training, the UK implemented the Modern Apprenticeship Act in 1993. The current goal is to increase participation rates of 16-year-olds in apprenticeship programmes to 35 % by 2010.

In contrast to Germany and other countries with successful apprenticeship programmes, the UK has avoided an external, statutory regulation of apprenticeship training (see e.g. Ryan, 2000; Ryan and Unwin, 2001; Gospel, 1998; and Steedman et al.,1998 for more details on the UK apprenticeship system). There is no legal definition of what apprenticeship training entails. While training standards are formally specified by Sector Skills Councils, there exists a variety of frameworks within particular sectors, some with low levels of learning. In contrast to the German chambers, the Training and Enterprise Councils in the UK are voluntary, operate at the local level, have purely advisory status, and lack statutory power. Perhaps most importantly, the testing of apprentices is carried out by *internal* assessors. To see why this may cause problems, firms may let apprentices pass the exam even if they received the training level under no commitment, τ^{NoCom}. Consequently, the firm's pass rate no longer signals commitment to training provision. While nowadays most apprentices receive some off-the-job training, this training often is in the form of in-house company training (see e.g. Ryan, 2000). There are several problems associated with this. Off-the-job in-company training may be subject to the same commitment problem as on-the-job workplace training. It is also likely to increase the firm-specific component of apprenticeship training. Finally, apprentices have no opportunity to learn about training opportunities in other firms.

While it is plausible that the (lack of) apprenticeship regulation in the UK makes it more difficult for firms to commit to training provision, firms in the UK may nevertheless be able to do so. After all, market forces, combined with a strong legal system, may be sufficient to ensure that training contracts can be enforced. Next, we provide three pieces of evidence which are consistent with the view that commitment to training provision may indeed be a problem of the UK apprenticeship system, and that the lack of commitment is related to the regulation of apprenticeship training.

First, throughout the late 60s and 70s, *mandatory* Industrial Training Boards were an integral part of apprenticeship training in the UK. The boards were endowed with statutory power over the content of apprenticeship training, similar to the German chambers. In the early 80s, the Thatcher government abolished the Boards and replaced with the *voluntary*, *employer-led* system of Training and Enterprise Councils that is still in place today and that have purely advisory capacity and lack statutory power (see e.g. Gospel, 1998 or Steedman et al.,1998 for more details). Figure 1 in the Introduction shows that apprenticeship training in manufacturing started to decline rapidly in the 80s, after apprenticeship de-regulation. Finegold and Soskice, 1988 and Keep and Mayhew, 1988 provide additional evidence that the Thatcher reforms indeed lead to a decline in apprenticeship training. One possible explanation for this trend is that the de-regulation removed an important commitment device for firms who wanted to offer apprenticeship training.

Second, the low quality of training is a prime concern of the UK apprenticeship system. We argue that this may be a consequence of firms' inability to commit to training provision, resulting in the (low quality) training level under no commitment, τ^{NoCom}, as opposed to the (high quality) training level under commitment, τ^{Com}. In an attempt to improve training quality, the Adult Learning Inspectorate (ALI) was created in 2001, and its main task is the inspection of every provider of work-based learning every four years. Inspection is centered around self-assessment, and training providers are required to submit an annual report to their local Learning and Skills Council. In the first year of inspection, 2001/02, only 42 % of the providers inspected were rated 'adequate', and only 24 % of apprentices completed their training programmes. There are signs of improvement: By 2005/06 88 % of training programmes were deemed 'adequate', and completion rates had increased to 53 %; see Lewis and Ryan, 2007 for more details.[20] A possible interpretation of this development is that the external inspection of training providers (a form of apprenticeship regulation) helped firms to signal commitment to training provision. However, the Adult Learning Inspectorate still sees room for improvement and states that 'apprenticeship success rates lower than 50 per cent persist in many areas of learning and are wholly

[20]It is important to stress that apprenticeship completion is difficult to define in a system that does not specify training standards and where the testing of apprentices is carried out internally.

unacceptable' (Adult Learning Inspectorate, 2006).

A third piece of evidence that a stricter regulation may help to foster enrollment in apprenticeship schemes comes from Ireland. Like the UK, Ireland began to reform its apprenticeship system in the 90s, but compared to the UK it put more emphasis on a *mandatory* educational content, the *external* regulation of training, and the *external* testing of apprentices. The setting of training standards were viewed as a key component of the apprenticeship scheme; and two occupations that were traditionally part of the apprenticeship system (hairdressing and upholstering) were left outside because of employers' resistance to mandatory off-the-job education and training. There is some evidence that the Irish reform has been more of a success than the UK one: In the 90s, apprenticeship enrollment has increased more in Ireland than in the UK, and completion rates are also higher (see e.g. Ryan, 2000; Ryan, 2004).

7 Conclusion

Why do apprenticeship schemes work well in some countries, such as Germany and Austria, but are less successful in others, such as the UK? This paper focuses on a key difference between firm-based apprenticeship training and its closest substitute, school-based vocational education: commitment to training provision. Our hypothesis is that the firm's ability to commit is a necessary prerequisite for apprenticeship programmes to work well. We argue that ensuring commitment to training provision requires a strict regulation of the apprenticeship system, including a legal definition of the rights and obligations of apprentices and their employers, the setting of nation-wide (occupation-specific) training standards, and the external testing of apprentices. These are important components of the apprenticeship system in Germany, but not in the UK. This may explain why young workers in the UK find apprenticeship training less attractive than young workers in Germany, despite receiving higher pay during training.

We provide several pieces of evidence that are consistent with our hypothesis. Most importantly, the simulation of a model of firm-financed training suggests that for reasonable degrees of wage compression due to asymmetric information and specific training estimated from survey and administrative data, the lack of commitment to training provision would result in *substantially* lower training levels, not more than 8 % of that under commitment. For training under no commitment to be at least one half of that under full commitment, we require exceedingly high levels of wage compression that are not supported by the data.

What policy conclusions can be drawn from our findings? Most importantly, countries that would like to expand apprenticeship training should pay careful attention that apprenticeship contracts are enforceable and that firms are able to commit to training provision. Several countries are currently providing subsidies in order to increase workers' enrollment in apprenticeship training. For instance, in the UK subsidies to apprenticeship training average nearly £7,000 per apprentice (Learning and Skill Council, 2001). This, however, does not address the commitment problem, and may therefore not be the most effective way of expanding apprenticeship training. Which policies can be implemented to ensure commitment to training provision? Recent experience from the UK suggests that the external inspection of training firms may be one way to do so. An alternative, possibly more effective way may be through the adoption of some of the key elements of the German apprenticeship system, with the setting of national or industry-wide training standards and the external testing of apprentices probably among the most important.

Another option to provide structured training opportunities for the less academically inclined may be through the expansion of *school*-based vocational education, rather than of *firm*-based apprenticeship training. While some skills may be taught more effectively in a work environment than in the classroom, commitment to training provision would be much easier to achieve.[21]

[21]This is the direction California is currently going. In 2007, governor Schwarzenegger (who completed an apprenticeship in sales in Austria) increased funds for vocational education by 18 percent to about $52 million.See e.g. http://gov.ca.gov/sots/education.html.

References

[1] Acemoglu, D. and J.-S. Pischke (1998), "Why do Firms train? Theory and Evidence", Quarterly Journal of Economics, 112, 79-119.

[2] Acemoglu, D. and J.-S. Pischke (1999a), "Beyond Becker: Training in Imperfect Labour Markets", Economic Journal Features, 109, F112-F142.

[3] Acemoglu, D. and S. Pischke (1999b), "The Structure of Wages and Investment in General Training", Journal of Political Economy, 107, 539-573.

[4] Acemoglu, D. and J.-S. Pischke (1998), "Why do Firms train? Theory and Evidence", Quarterly Journal of Economics, 112, 79-119.

[5] Adda, J., Dustmann, C., Meghir, C. & Robin, J.M. (2006). Career Progression and Formal versus On-The-Job Training. IZA Discussion Paper, No. 2260.

[6] Adult Learning Inspectorate, 2006, "The Final Annual Report of the Chief Inspector, 2005-2006",. Coventry.

[7] Becker, G. (1964), "Human Capital", Chicago: The University of Chicago Press.

[8] von Bardeleben, R., Beicht, U. and K. Feher (1995), "Betriebliche Kosten und Nutzen der Ausbildung. Repräsentative Ergebnisse aus Industrie, Handel und Handwerk", Berichte zur beruflichen Bildung, Heft 187, Bielefeld: Bertelsmann.

[9] Beicht, U., Walden, G., and H. Herget. (2004), "Cost and Benefits of In-Company Vocational Education and Training in Germany", Federal Institute for Vocational Training, Bielefeld.

[10] Bernhardt, D. (1995): "Strategic Promotion and Compensation", Review of Economic Studies, 62, 315-339.

[11] Berufsbildungsbericht (2006), Bundesministerium fuer Bildung und Forschung.

[12] BIBB (2006), "Schaubilder zur Berufsbildung - Strukturen und Entwicklungen", Bundesinstitut fuer Berufsbildung, Februar.

[13] Booth, A. and M. L. Bryan (2005), "Testing Some Predictions of Human Capital Theory: New Training Evidence from Britain", Review of Economics and Statistics, 87, 391-394.

[14] Bowers, N., A.Sonnet and L.Bardone (1999), "Giving young people a good start: the experience of OECD countries", pp. 7-84 of OECD, Preparing Youth for the Twenty-first Century: the Transition from Education to the Labour Market. (Paris: OECD).

[15] DiNardo, J. and J.-S.Pischke (1997), "The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too?", Quarterly Journal of Economics, 112, 291-303.

[16] Dustmann, C. and U. Schönberg (2005), "Training and Union Wages", IZA Discussion Paper No. 1435.

[17] Euwals, R. and R. Winkelmann (2004), "Training Intensity and First Outcomes of Apprenticeship Graduates", International Journal of Manpower, 25, 447-462.

[18] Festerer, J., Pischke, J.-S., and R. Winter-Ebmer (2004), "Returns to Apprenticeship Training in Austria: Evidence from Failed Firms", Unpublished Working Paper, London School of Economics.

[19] Finegold, D. and D. Soskice (1988), "The Failure of Training in Britain", Oxford Review of Economic Policy, 4, 21-53.

[20] Fitzenberger, B. and A. Kunze (2005), "Vocational Training and Gender: Wages and Occupational Mobility among Young Workers", Oxford Review of Economic Policy, 21, 392-415.

[21] Fitzenberger, B. and A. Spitz-Öner (2004), "Die Anatomie des Berufswechsels: Eine empirische Bestandsaufnahme auf Basis der BiBB/IAB-Daten 1998/99", in: W. Franz, H.J. Ramser und M. Stadler, Bildung, Wirtschaftswissenschaftliches Seminar in Ottobeuren, Bd. 33, Tübingen, 29-54.

[22] Franz, W. and D. Soskice (1995), "The German Apprenticeship System", in F. Buttler, W. Franz, R. Schettkat, and D. Soskice (eds), Institutional frameworks and labour market performance. London: Routledge.

[23] Gibbons, R. and L.F. Katz (1991), "Layoffs and lemons", Journal of labour Economics 9, 351-380.

[24] Gospel, H. (1998). "The Revival of Apprenticeship Training in Britain?", British Journal of Industrial Relations, 36, 435-57.

[25] Greenwald, B. (1986) "Adverse Selection in the Labour Market", Review of Economic Studies, 53, 325-347.

[26] Hamilton, S.F. (1990), "Apprenticeship for Adulthood: Preparing Youth for the Future". New York: Free Press.

[27] Harhoff, D. and T.J. Kane (1997), "Is the German apprenticeship system a panacea for the U.S. labour market?" Journal of Population Economics, 10, 171-196.

[28] Heckman, J.J., Roselius, R., and J. Smith (1994), "US Education and Training Policy: A Re-Evaluation of the Underlying Assumptions of the 'New Consensus", in: Solomon, Lewis and Alec Levenson (eds). Labour Markets, Employment Policy and Job Creation. Boulder: Westview Press.

[29] Jimeno, J. F. and D. Rodríguez-Palenzuela (2003), "Youth Unemployment In The OECD: Demographic Shifts, Labour Market Institutions And Macroeconomic Shocks", European Network of Economic Policy Research Institutes Working Paper No. 19.

[30] Keep, E. and K. Meyhew (1988), "The Assessment: Education, Training, and Economic Performance", Oxford Review of Economic Policy, 4, i-xv.

[31] Learning and Skills Council, 2001, "Learning and Skills Council Operations Guide", Version 3, 11, April.

[32] Lehmann, W. (2000), "Is Germany's Dual System Still a Model for Canadian Youth Apprenticeship Initiatives?", Canadian Public Policy 26, 225-240.

[33] Malcomson, J.M., Maw, J.W., and B. McCormick, 2003, "General Training by Firms, Apprentice Contracts, and Public Policy", European Economic Review, 47, 197-227.

[34] Muench, J. (1992), "Das Berufsbildungssystem in der Bundesrepublik Deutschland", European Centre for the Development of Vocational Training (CEDEFOP), Berlin.

[35] Li, J. (2006), "Job Mobility, Wage Dispersion, and Asymmetric Information", mimeo, MIT Department of Economics.

[36] Neal, D (1999), "The Complexity of Job Mobility Amongst Young Men", Journal of labour Economics, 17, 237-261

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

[38] Ryan, P. (2000), "The institutional requirements of apprenticeship: evidence from smaller EU countries", International Journal of Training and Development, 4, 42-65.

[39] Ryan, P. (2004), "Institutional Requirements of Apprenticeship Training", ITAC Third Conference on Apprenticeship Cork Institute of Technology.

[40] Ryan, P., H. Gospel, and P. Lewis (2007), "Large employers and apprenticeship training in Britain", British Journal of Industrial Relations, 45, 127-53.

[41] Ryan, P. and L. Unwin (2001), "Apprenticeship in the British "training market"", National Institute Economic Review, 178, 99-114.

[42] Stevens, M. (1994a), "A Theoretical Model of On-the-Job Training with Imperfect Competition", Oxford Economic Papers, 46, 537-562.

[43] Stevens, M. (1994b), "An Investment Model for the Supply of Training by Employers", Economic Journal, 104, 556-570.

[44] Steedman, H., H. Gospel, P. Ryan (1998), "Apprenticeship: A Strategy For Growth", CEP discussion paper, London.

[45] Soskice, D. (1994), "Reconciling Markets and Institutions: the German Apprenticeship System", in L.M. Lynch (ed.) Training and the private sector, International comparisons. NBER Series in Comparative labour Markets. Chicago: University of Chicago Press, 1-24.

[46] Wachter, T. and S. Bender (2006), "In the Right Place at the Wrong Time: The Role of Firms and Luck in Young Workers' Careers", American Economic Review, 96, 1679-1705.

[47] Waldmann, M. (1984): "Job Assignment, Signalling, and Efficiency", Rand Journal of Economics, 15, 255-267.

[48] Winkelmann, R. (1996), "Employment Prospects and Skill Acquisition of Apprenticeship–Trained Workers in Germany", Industrial and labour Relations Review, 49, 658-672.

A Data Appendix

A.1 Longitudinal Data on Workers: IABS

Sample Selection Our basic sample consists of all workers who finished an apprenticeship between 1980 and 1992. Apprentices are defined as workers who have been in apprenticeship training for at least 450 days. We drop all workers who were employed in East Germany at least once. We also exclude workers who have no high school degree (Abitur) and who were older than 19 at the beginning of the apprenticeship. Workers with a high school degree are dropped if they were older than 21 at the beginning of the apprenticeship. Finally, we drop workers who start more than one apprenticeship, as well as workers who are observed to be in apprenticeship training for more than 1500 days. We also drop all spells with missing information on the following variables: age at the start of the apprenticeship, A-levels, nationality, and wage (2281 observations). We include a dummy variable if the apprenticeship occupation or apprenticeship firm size is missing. These are the two variables with the largest number of missing observations. Our results are very similar if we drop these spells.

A disadvantage of our data is that we do not observe the exact end of the apprenticeship for workers who stay with their apprenticeship firm. For these workers, we observe an average of the apprenticeship wage and the post-apprenticeship wage in the year the apprenticeship ended. We deal with this problem by comparing wages of movers and stayers in the year following the year the apprenticeship ended. We only consider spells in full-time employment.

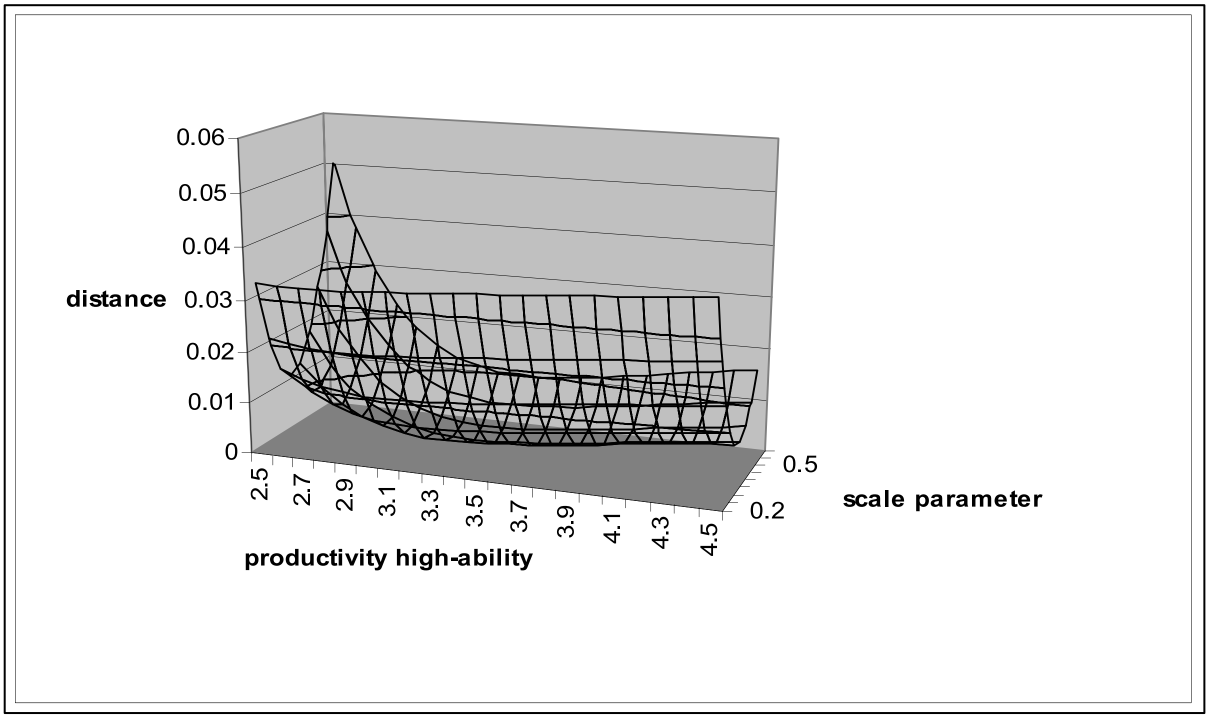
Variable Definitions Movers are workers who leave the apprenticeship firm right after apprenticeship training. As a robustness check, we have repeated the empirical analysis defining movers as workers who leave the apprenticeship firm during the year in which the apprenticeship ended. This has little impact on our results. We classify a worker as a job-to-unemployment mover if he claimed unemployment benefits and report as unemployed after apprenticeship training. In Germany, workers are entitled to unemployment benefits after apprenticeship training. It is therefore likely that workers who are searching for a job also report as unemployed. Potential experience is counted from the beginning of the year following the end of the apprenticeship. Experience $= 0$ (2, 5, $\geqslant 10$) refers to all workers who were working at least one day during the first (second, fifth, $\geqslant$ tenth) year following the apprenticeship. We deflate wages using the Consumer Price Index, with 1995 as the base year.

A.2 German Qualification and Career Survey

We drop East Germans, the self-employed, individuals who work less than 30 hours a week, as well as individuals younger than 18 and older than 55. Finally, we drops spells if information on the following variables is missing: applicability of skills, year of birth and school type. We include a dummy variable if information on the apprenticeship occupation or apprenticeship firm size is missing. These are the two variables with the largest number of missing observations. Our results are very similar if we drop these spells.

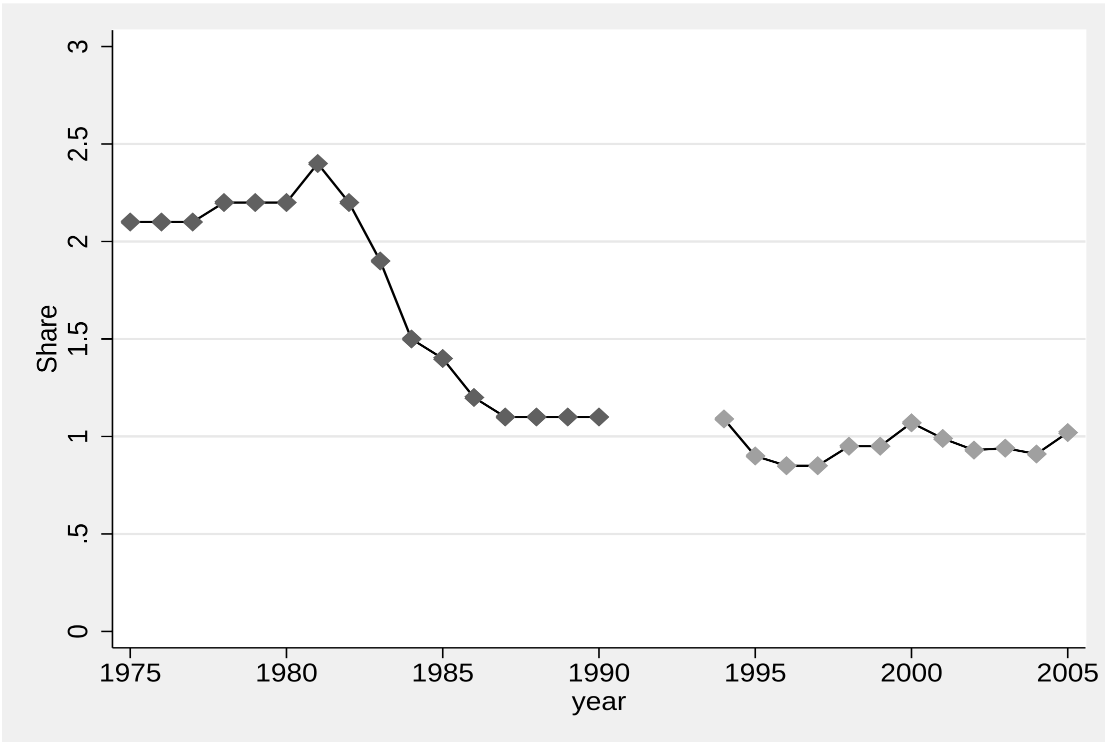
Appendix B: A Note on Identification

Figure B1: The Relationship between Distance, the Productivity of High-Abilty Workers, and the Scale Parameter of the Logistic Distribution



Note: The figure shows a three-dimensional graph, with the productivity of high-ability workers on the x-axis, the scale paremeter b on the y-axis, and the distance on the z-axis. The objective function (i.e. the sum of the squared distance between the model and the data moments) is well-behaved in the sense that there is only one minimum, suggesting that both parameters are uniquely identified.

Figure 1: The Share of Individuals in Manufacturing on Apprenticeship Schemes



Note: The figure shows the share of individuals in manufacturing on apprenticeship schemes. The entries from 1975 to 1990 are taken from Ryan and Unwin (2001). Entires from 1994 to 2005 are our own calculations based on the British Labor Force Survey. Since the industry classification is missing in the BLS for the years 1991 to 1993, the share cannot computed for these years.

Table 1: Descriptive Statistics: Longitudinal Data on Mobility and Wages

	Sample 1	*Sample 2*
Number of observations, experience = 0	17593	3020
Number of observations, experience = 2	16710	2852
Number of observations, experience = 5	13521	2220
Number of observations, experience >= 10	16012	2714
age at end of apprenticeship	19.22	19.32
proportion high school degree (Abitur)	6.12%	7.55%
duration apprenticeship training (years)	2.64	2.66
proportion non-German citizens	3.92%	4.99%
average training firm size	1195.79	5521.76
prop. still at training firm (occupation), experience = 0	67.84 % (71.00 %)	72.78 % (61.82 %)
prop. still at training firm (occupation), experience = 3	36.37 % (52.12 %)	49.65 % (48.63 %)
prop. still at training firm (occupation), experience = 5	26.68 % (42.68 %)	41.49 % (40.72 %)
prop. still at training firm (occupation), experience = 10	17.48 % (30.45 %)	34.41 % (26.01 %)

Note: Sample 1 consists of all men who finished apprenticeship training between 1980 and 1992. Sample 2 is a subsample of Sample 1 where we observe at least two apprentices in the same firm.

Table 2: Descriptive Statistics: Cross-Sectional Data on Skill Applicability

A: variables		*B: mobility rates*	
age at end of apprenticeship	18.81	still works at training firm,	29.99%
time since apprenticeship training	17.68	by time since apprenticeship: <=1	78.26%
proportion lower sec. school	71.75%	1--3	57.56%
proportion intermediate sec. school	23.42%	3--5	47.59%
proportion high school degree	4.83%	5--10	35.95%
training firm size ≤ 9	28.06%	> 10	22.89%
10--49	29.05%	still works at training occupation,	51.20%
50--500	25.88%	by time since apprenticeship: <=1	84.66%
>500	17.01%	1--3	77.33%
applicability of skills: very much	41.28%	3--5	69.49%
a lot	20.74%	5--10	60.04%
some	15.36%	> 10	44.23%
not too much	8.23%		
very little	14.39%		

Note : Results are based on Sample 3. N=25380.

Table 3: Cost Sharing in the German Apprenticeship System

	1	2	3	4	5	6
	productivity	direct cost	appr. wage	wage of unskilled worker	total cost	workers' share
all firms	7730	2663	7031	11973	6906	71.60%

Note: Column 1 lists estimates for workers' (annual) productivity during apprenticeship training. Column 2 reports estimarect cost of training, such as personnel costs for trainers, plant and material costs, as well as textbooks, teaching software, etc. The estimates in Column 1 and 2 are obtained from Beicht et al. (2004). Column 3 and 4 report the apprenticeship wage as well as the average annual wage of untrained workers with zero labor market experience. The estimates in Column 3 and 4 are obtained from social security data for the year 2000. Column 5 computes the total cost of apprenticeship training as the difference between the wage of unskilled workers and the productivity of apprentices plus the 'variable' cost of training, i.e. Column 4 - Column 1 + Column 2. Workers' share of training costs are computed as the ratio of the average wage of an unskilled worker minus the average apprenticeship wage and the total training costs, i.e. (Column 4- Column 3)/Column 5.

Table 4: Transferability of Apprenticeship Training

	Panel A: movers vs stayers			
	A: can apply a lot/very much		*B: share applicable*	
	1	2	3	4
mover	-0.303	-0.303	-20.267	-19.359
	(0.006)***	(0.006)***	(0.322)***	(0.346)***
	Panel B: exogenous vs endogenous movers			
endogenous	-0.310	-0.311	-20.380	-19.438
	(0.006)***	(0.007)***	(0.329)***	(0.353)***
exogenous (military)	-0.328	-0.338	-18.885	-18.496
	(0.014)***	(0.014)***	(0.866)***	(0.871)***
p-value	0.222	0.482	0.087	0.280
	Panel C: occupation vs firm movers			
firm stayer, occ. mover	-0.219	-0.205	-9.026	-8.603
	(0.015)***	(0.015)***	(0.619)***	(0.641)***
firm mover, occ. stayer	-0.119	-0.128	-4.473	-4.531
	(0.010)***	(0.011)***	(0.325)***	(0.349)***
firm mover, occ. mover	-0.536	-0.543	-34.707	-34.278
	(0.007)***	(0.008)***	(0.375)***	(0.404)***

Note: Sample 3, N=25380. Panel A distinguishes between movers and stayers only. Panel B additionally distinguishes between exogenous (i.e. displaced from their training firm because of military service) and endogenous movers. Panel C breaks down the analysis by occupation and firm movers. The dependent variable in specification A is equal to 1 if the worker can apply very much/a lot of the skills obtained during apprenticeship training. The dependent variable in specification B is the proportion of skills obtained during apprenticeship training that are applicable at the current job. Specification 1 includes no additional controls. Specification B additionally controls for the year the apprenticeship ended, experience, experience squared, firms size of the apprenticeship firm (7 categories), apprenticeship occupation (10), age at the end of the apprenticeship, and type of high school. Robust standard errors in parenteses.

Table 5: The Quit Rate and the Mover-Stayer Wage Differential

(1)	*Quit Rate*	34.96%
	Mover-Stayer Wage Differential	
(2)	raw	-0.086
		(0.006)***
(3)	controls	-0.067
		(0.006)***

Note: The first row shows the share of workers who leave the training firm after training completion. The second and third row shows the wage disadvantage of movers relative to stayers after apprenticeship training. Row 2 (raw) only controls for the year the apprenticeship ended. Row 3 (controls) additionally controls for citizenship, age at the beginning of apprenticeship, apprenticeship duration, high school degree (Abitur), 10 apprenticeship occupation dummies, and size of apprenticeship firm. Robust standard errors in parantheses.

Table 6: Alternative Explanations for Adverse Selection

Panel A: The Evolution of the Wage Differential						
	Sample 1, raw			*Sample 1, controls*		
0	-0.0862			-0.0668		
	(0.0060)***			(0.0062)***		
2	-0.0815			-0.0598		
	(0.0057)***			(0.0057)***		
5	-0.0914			-0.0718		
	(0.0058)***			(0.0062)***		
≥ 10	-0.081			-0.0587		
	(0.0051)***			(0.0062)***		
Panel B: Worker Sorting						
	Sample 2, raw		*Sample 2, controls*		*Sample 2, fixed effect*	
0	-0.130		-0.119		-0.104	
	(0.017)***		(0.017)***		(0.022)***	
2	-0.083		-0.074		-0.069	
	(0.014)***		(0.014)***		(0.019)***	
5	-0.100		-0.097		-0.095	
	(0.015)***		(0.015)***		(0.019)***	
≥ 10	-0.084		-0.105		-0.106	
	(0.014)***		(0.015)***		(0.017)***	
Panel C: Wage floors						
	Sample 1, raw		*Sample 1, controls*		*Sample 2, fixed effect*	
	Job-to-Job	*Job-to-Unempl.*	*Job-to-Job*	*Job-to-Unempl.*	*Job-to-Job*	*Job-to-Unempl.*
0	-0.095	-0.078	-0.071	-0.063	-0.082	-0.134
	(0.009)***	(0.007)***	(0.009)***	(0.007)***	(0.028)***	(0.032)***
2	-0.059	-0.099	-0.034	-0.081	-0.052	-0.090
	(0.009)***	(0.007)***	(0.008)***	(0.007)***	(0.023)**	(0.026)***
5	-0.065	-0.108	-0.041	-0.091	-0.068	-0.121
	(0.010)***	(0.006)***	(0.010)***	(0.007)***	(0.026)***	(0.026)***
≥ 10	-0.039	-0.114	-0.012	-0.089	-0.090	-0.121
	(0.008)***	(0.008)***	(0.008)	(0.007)***	(0.024)***	(0.021)***

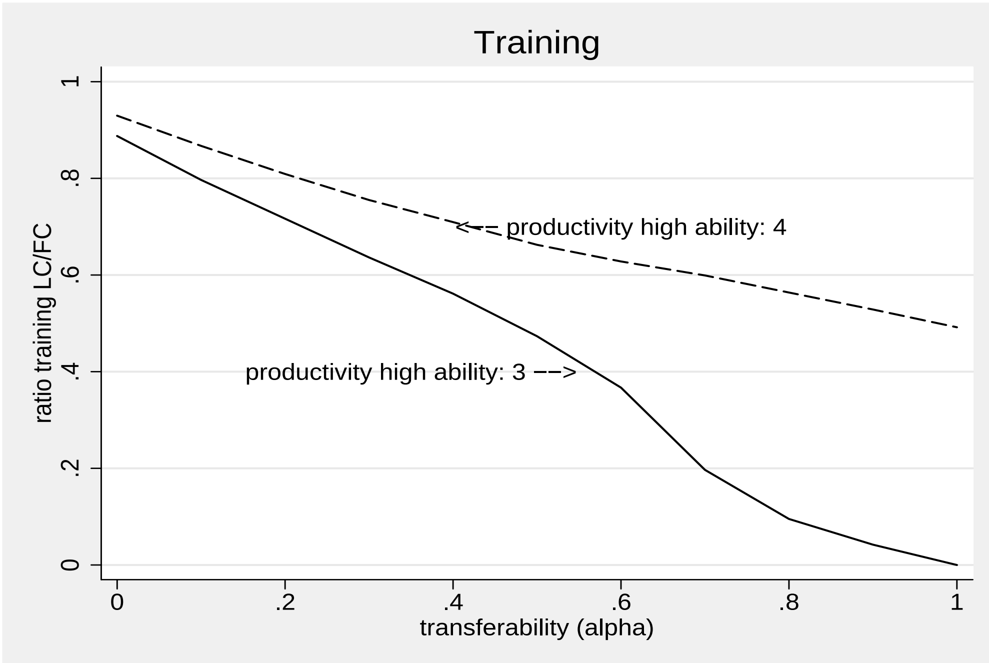
Note: Results in Panel A are based on Sample 2. Panel B distinguishes between job-to-job and job-to-unemployment movers. Specification 1 and 2 are based on Sample 1, specification 3 on Sample 2. Specification 1 controls for year the apprenticeship ended. Specification 2 additionally controls for citizenship, age at the beginning of the apprenticeship, apprenticeship duration, high school degree (Abitur), 10 apprenticeship occupation dummies and size of apprenticeship firm. Specification 3 includes fixed training firm effects. Robust standard errors in parentheses.

Table 7: Full versus Limited Commitment: Simulation Results

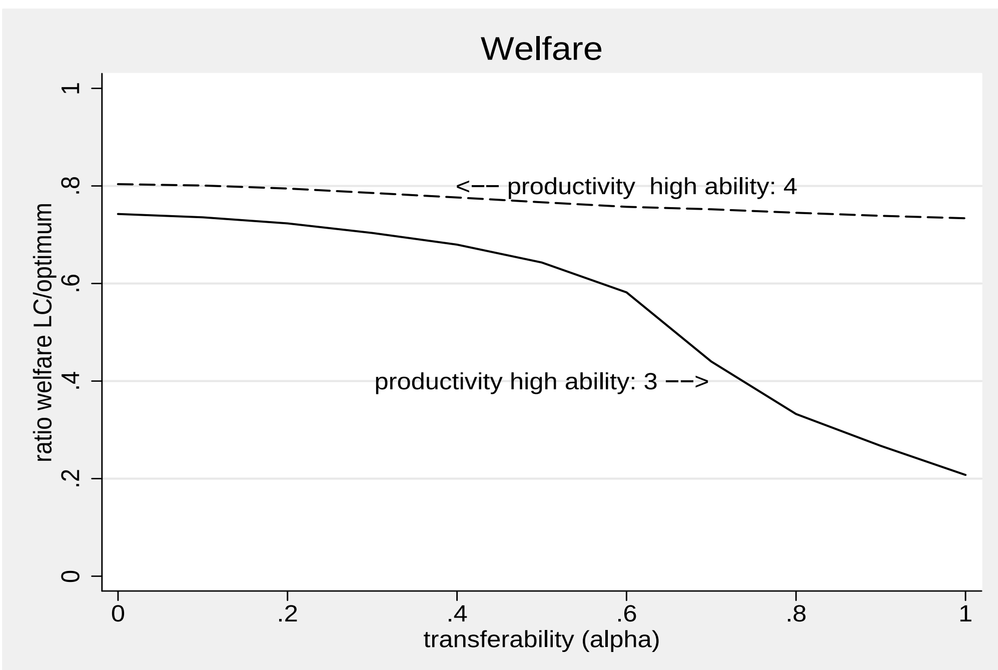
	1	2	3	4	5	6	7
	parameters		**training**		**welfare**		
	b	**η_H**	**FC**	**LC**	**FC**	**LC**	**optimum**
Panel A: Baseline Results (Mover-Stayer Wage Differential after Apprenticeship Completion							
α=0.95	0.44	3.22	2.02	**0.055**	2.51	0.64	2.75 (91.3%)
α=0.81	0.48	2.86	1.8	**0.137**	2.10	0.76	2.39 (87.8%)
Panel B: Robustness Check 1 (Mover-Stayer Wage Differential 10 years after training)							
α=0.95	0.44	3.15	1.98	**0.050**	2.45	0.62	2.68 (91.4%)
α=0.81	0.48	2.82	1.78	**0.131**	2.06	0.75	2.35 (87.7%)
Panel C: Robustness Check 2 (Job-to-Job Mover-Stayer Wage Differential 10 years after Training)							
α=0.95	0.42	2.79	1.79	**0.034**	2.14	0.59	2.32 (92.2%)
α=0.81	0.46	2.54	1.63	**0.104**	1.84	0.69	2.09 (88.0%)

Note : Column 1 and 2 report parameter values for the scale parameter of the logistic distribution, b, and the productivity of high-ability workers, ηH. They are chosen to replicate the separation rate and the mover-stayer wage differential in our data. Columns 3 and 4 compare the training level under full (FC) and limited (LC) commitment. Columns 5 to 7 report welfare under full and limited commitment as well as under the social optimum (i.e. when firms are able to training to commit to training provision, information is symmmetric, and training is fully general). We report our results for two estimates for the degree of specificity, and three alternative targets for the mover-stayer wage differential.

Figure 2: Limited versus Full Commitment: The Role of Wage Compression
Panel A



Panel B



Note: Panel A plots the ratio between training and full and limited commitment, for two values for the productivity of high-ability workers (ηH=3 and ηH=4) and various degress of transferability (alpha). Panel B plots the ratio between welfare under limited commitment and wage compression and welfare under full commitment and no wage compression, i.e. symmetric information and general training (alpha=1).