Training and Unions *

Christian Dustmann †Uta Schönberg ‡

First version: September 2001
This version: October 2003

Abstract

We revisit the question why firms pay for general training within the German apprenticeship system. Our hypothesis is that Germany’s wage setting institutions induce firms to do so. To test this hypothesis, we develop a model of firm-financed training that incorporates the specific features of the German collective bargaining system. We derive testable implications, and compare them with those of alternative explanations for firm-financed training, such as asymmetric information. We use a new administrative data set on workers and a panel data set on firms to test these. The evidence provides strong support for the hypothesis that Germany’s wage setting institutions and consequent wage compression favours firm-financed and firm-based training schemes.

*Preliminary and incomplete. We thank David Card, Thomas Lemieux and Marco Ottaviani for comments and suggestions. We would also like to thank participants in seminars at Universities of Berkeley, British Columbia (Vancouver), Essex, Santa Barbara, St. Andrews (Scotland), Nottingham, and IZA (Bonn) for useful comments. Parts of this paper were written while Christian Dustmann was visiting The Center for Labor Economics at Berkeley. He would like to thank the Center for its hospitality. We are grateful to Arndt Koelling and Holger Alda, and in particular to Stefan Bender from the IAB for support with the data.
†Department of Economics, Gower Street, London WC1E 6BT, England. E-mail: c.dustmann@ucl.ac.uk
‡Department of Economics, University College London, Gower Street, London WC1 6BT, United Kingdom. E-mail: u.schoenberg@ucl.ac.uk.
1 Introduction

Standard human capital theory predicts that in perfectly competitive labour markets firms do not invest into general training of their workers, as the worker captures the full return to that investment (Becker 1964). Yet, there is evidence that suggests that firms provide and pay for general training. For instance, temporary help firms in the United States often provide and finance training programmes to new employees which is largely general in nature, such as computer and typing skills (Autor 2001). Another example of firm provided general training is the German apprenticeship system, which combines firm and state provided training. Although trainees (or apprentices) are mostly trained in general skills, the system is, at least partly, financed by firms.  

In a number of papers, Acemoglu and Pischke (1999a, 1999b) offer an explanation for why firms should pay for workers’ general training. The key to their argument is a compressed wage structure, due to labour market imperfections: If training increases workers’ productivity by more than workers’ outside option, then firms can increase profits by training. Firms thus have an incentive to bear the cost of training - even if training is general. In an attempt to explain training within the German apprenticeship system, Acemoglu and Pischke (1998) suggest asymmetric employer learning as a reason for wage compression. They find the empirical implications of their model confirmed by the data.

In this paper we revisit the question why firms may be willing to pay for general training. We focus on the impact of wage rigidities caused by unions on training. The evidence on whether wage rigidities are detrimental or incremental to on-the-job training schemes is so far inconclusive. For the US, Neumark and Wascher (2001) conclude that minimum wages decrease training, whereas Acemoglu and Pischke (2001) find weak evidence that points in the opposite direction.

---

2 See studies by von Bardeleben et al. (1995) and Acemoglu and Pischke (1999b) for evidence.  
3 Freeman and Medoff (1984), Lemieux (1998), Fortin and Lemieux (1997), Dinardo et al. (1996) and Card (1992) all provide indirect evidence that a minimum wage or unions lead to a compression of the wage structure.  
4 For the UK, Boeheim and Booth (2002) and Green et al. (1996) find that unionised workers as well as workers who are employed in firms that recognise unions for collective bargaining purposes receive more employer-provided training.
As Acemoglu and Pischke (1998), we test the empirical implications of our model using data from the German apprenticeship system. In Germany unions play an important role: Between 1996 and 1999, 77% of the West German work force were covered by collective bargaining agreements.\footnote{The number is based on the IAB-Betriebspanel. See table 1.} Our empirical test exploits the fact that in Germany only firms that belong to an employer association are bound to union agreements. In these firms, union agreements apply to all workers, regardless of workers’ union status. This essentially creates a unionised and non-unionised sector.

We first develop a theory of firm-sponsored training and unions. Our model extends existing existing models on unions and training, such as Acemoglu and Pischke (1999b) and Booth et al. (1999), in several ways. First, our model is set up in the same framework as Acemoglu and Pischke’s asymmetric information model. This allows us to compare the empirical implications of their model with our model. Second, in contrast to existing models on unions and training, we allow a unionised and non-unionised sector to co-exist. Furthermore, workers decide in which sector to work, and select into either sector based on worker heterogeneity. We derive wage and training determination in both sectors, and analyse the worker’s sector choice.

Our key theoretical results can be summarised as follows. First, an adverse selection of movers -the key prediction of Acemoglu and Pischke’s (1998) asymmetric learning model- is consistent with our union model. Second, union agreements compress wages for workers with a productivity around the union wage in unionised firms, inducing firms to train these workers. If firms cannot fully commit to training provision, unions move training closer to the socially optimal level. Third, the impact of unions on training is not uniform. Workers sort into the unionised sector based on the impact union agreements have on training. Furthermore, more able workers sort into the unionised sector.

The reason why unionised firms train is that union agreements force firms to pay at least the union wage in the future. In principle, this could be achieved by long-term wage contracts: At the beginning of the training period the firm could commit to paying a minimum wage in the future. Long-term wage contracts thus mitigate, though not eliminate, the problem of limited commitment to training provision. Such contracts, however, are not self-enforceable, as the firm has an incentive to deviate after the training period and to pay a lower wage than the agreed minimum wage. One interpretation of unions is therefore that unions make long-term
wage contracts enforceable.

Our empirical analysis is based on three data sets. The first data set is a one percent sample drawn from social security records, and covering the period between 1975 and 1995. It allows construct complete work histories for apprentices, and provides, besides precisely measured wages, a lot of information that is usually not available is administrative data. It also include information on firm size, which allows construction of exogenous separations. Our second data set is a matched employer-employee data set, covering the period between 1996 and 1999. This data set is based on a firm panel, to which we were able to match information about average characteristics of workers in the firm, drawn from the overall population of workers from administrative records. The third data source is the overall population of workers from administrative records, which we observe for one year (2000).

Our empirical strategy proceeds in two steps. In a first step we use matched employer-employee data to directly test whether unions increase training. Our goal is to identify the causal impact of unions on training. Our estimation strategy attempts to take into account a nonrandom selection of workers and possibly firms into the unionised sector. We consistently find that unions increase training, conditional on a wide array of worker and firm characteristics as well as fixed firm effects. We conclude that wage floors created by unions are an important reason for firm-financed training in Germany. We cannot rule out asymmetric information as an additional reason for firm-financed training.

In a second step we test for an adverse selection of movers. We find that even after ten years in the labour market, endogenous movers earn considerably lower wages than stayers and exogenous movers. We then use matched employee-employer data to compare the mover-stayer wage differential for workers trained in unionised and non-unionised firms. We find a higher wage for stayers than for movers not only for workers trained in unionised firms but also for those trained in non-unionised firms. As union agreements are not binding in non-unionised firms, this cannot be due to wage rigidities caused by unions. Furthermore, we find a higher stayer-mover wage differential in unionised than in non-unionised firms, suggesting that wage floors contribute to an adverse selection of movers.

The structure of the paper is as follows. Section 2 develops a model of employer-financed training. We begin with a base model and then incorporate union agreements into the model. Section 3 outlines the empirical implications. We then describe the empirical strategy and
results (section 4). We conclude by pointing out the policy and welfare implications of our findings (section 5).

2 A model of firm-financed training

We first analyse firms’ incentives to train when there are no union agreements (section 2.1). We then incorporate wage rigidities due to unions into the model (section 2.2).

2.1 Base model

There are many workers and firms. Firms and workers are risk-neutral. Firms maximise expected profits, workers maximise expected utility. There are two periods. The first period is the training period. There is no discounting.

Productivity and training costs

The worker’s productivity in period 2 depends on her (true) ability $\eta$ as well as on the amount of training received in period 1, $\tau$:

$$y = y(\tau, \eta).$$

We assume that $y(\tau, \eta)$ is strictly increasing, differentiable and concave in $\tau$, with $y''(\tau, \eta) > 0$, $\lim_{\tau \to -\infty} y'(\tau, \eta) = 0$, and $\lim_{\tau \to 0} y'(\tau, \eta) = \infty$. The return to training is higher for more able workers, i.e. $\frac{\partial^2 y}{\partial \tau \partial \eta} > 0$. The productivity of a worker in training is smaller than the productivity of an untrained worker by a constant $k$, which represents a fixed cost of training. There are also variable training costs which we denote by $c(\tau)$. The function $c(\tau)$ is strictly increasing, differentiable and convex, with $c(0) = c'(0) = 0$, $c''(\tau) > 0$. We further assume that the firm’s production function exhibits constant returns to scale, i.e. the total productivity of a firm is equal to the sum of each worker’s productivity.

Information structure

Workers’ ability $\eta$ is drawn from a normal distribution with mean $\bar{\eta}$ and variance $\sigma_\eta^2$. In the first period firms and workers receive a noisy signal $\tilde{\eta}$ about workers’ ability. The signal equals

$$\tilde{\eta} = \eta + \varepsilon_\eta.$$
Firms and workers use the signal to update the belief about workers’ ability. If $\epsilon_\eta$ is normally distributed with mean 0 and variance $\sigma^2_\eta$, then the updated belief about the worker’s productivity is also normally distributed with mean (DeGroot (1970))

$$E[\eta|\tilde{\eta}] = \frac{\eta_\sigma^2 + \tilde{\eta}_\sigma^2}{\sigma^2_\eta + \sigma^2_\eta} := \hat{\eta},$$

and variance

$$V[\eta|\tilde{\eta}] = \frac{1}{\sigma^2_\eta + \sigma^2_\eta} := \sigma^2_{\hat{\eta}}.$$

The updated belief about workers’ productivity is a weighted average of the prior mean, $\eta$, and the signal, $\tilde{\eta}$. Let $F_1(\eta|\tilde{\eta})$ denote the distribution of a worker with expected ability $\hat{\eta}$ after signal $\tilde{\eta}$ has been observed, i.e. $F_1 \sim N(\tilde{\eta}, \sigma^2_{\tilde{\eta}})$.

It is important to stress that in the first period information about workers’ ability is symmetric with respect to workers and firms. This assumption rules out that firms offer training in order to induce self-selection of better workers into training (see Autor 2001 and Bhaskar and Holden 2002 models of this type). In the second period both incumbent and outside firms perfectly learn about workers’ true ability. This assumption of symmetric employer learning excludes asymmetric information as a reason for firm-financed training (as in Acemoglu and Pischke 1998). The assumption that firms perfectly learn about workers’ ability is not essential for our results.

**Training decisions**

In the first period, firms - as opposed to workers - decide how much training to offer to a worker. We analyse the firm’s decision to train under the assumption that firms can only commit to training provision, but not to the amount of training. We refer to this case as *limited commitment*. Acemoglu and Pischke (1999b) refer to this as the *constrained regime*. A justification for this assumption is that training is not easily verifiable by a third party, and can thus not be contracted upon. For our particular application - apprenticeship training schemes in Germany - the assumption that firms can commit to training provision, but not to the amount of training is reasonable. Trainees take centralised exams at the end of the apprenticeship training period and receive a certificate that is widely recognised. Hence, it is clearly verifiable whether a worker has received some training. However, an important part of apprenticeship training takes place inside the firm, and is not easily observed by an outside party.
**Mobility**

At the end of the training period workers decide whether to switch firms. As in Acemgolu and Pischke (1998), we assume that during the training period workers experience a utility shock $\theta$. This shock captures the worker’s ex post evaluation of her work environment. For example, it may reflect how well the worker gets along with her co-workers and supervisors. Only the worker, but not the firm, observes $\theta$. We specify the worker’s utility at the incumbent firm, $U^i$, as a simple linear function of the incumbent firm’s wage offer, $w$, and the utility from non-pecuniary job characteristics, $\theta$:

$$U^i = w + \theta.$$  

The worker’s utility at outside firms is equal to the wage offer, $v$. The utility shock is drawn from a distribution with the cumulative distribution and probability density function $G$ and $g$ and support $[\underline{\theta}, \bar{\theta}]$. We assume that $G(.)$ belongs to the family of log-concave distribution functions, i.e. $\frac{g(\theta)}{1-G(\theta)}$ is non-decreasing in $\theta$. We also assume that the distribution of the utility shock neither depends on workers’ ability nor on training.

Notice that the assumption that the upper support of $G$ is positive, $\bar{\theta} > 0$, implies that some workers stay with the employer even if offered a lower wage than outside firms. Consequently, firms make positive profits in the second period.

**Wage determination**

Each period firms simultaneously make wage offers to the worker by maximising expected profits. Long-term wage contracts are not feasible.

**Free entry condition**

We close the model by imposing a free entry condition on firms: No firm earns positive profits in equilibrium at any point in time.

**Timing**

The exact sequence of events is as follows.

1. At the beginning of period 1 firms and workers receive a noisy signal about workers’ ability.

2. Firms offer a (first period) wage and choose whether and how much to train the worker.
3. Training/production takes place.

4. At the end of the training period all firms and workers get to know workers’ ability.
   Incumbent and outside firms make a wage offer to the worker.

5. At the beginning of period 2 workers learn their utility shock. They decide whether to
   stay with or leave the training firm.

6. Production takes place.

7. At the end of the second period workers retire.

We solve the model backwards and begin with wage determination in the second period. We then turn to firms’ incentives to train and wage determination in the first period.

**Wage determination in the second period**

Firms observe workers’ ability so that the wage offer of incumbent and outside firms depend
on ability. Wages also depend on training. Although outside firms may not be able to directly
observe the worker’s training level, in equilibrium they can infer the training level the incumbent
firm has chosen in the first period, as they know the incumbent firm’s maximisation problem.

Let \( w \) denote the wage offer of the incumbent firm and \( v \) the worker’s outside wage offer.
Due to perfect competition in the outside market, outside firms bid up workers’ wage until
it equals the worker’s (marginal) productivity, i.e. \( v = y(\tau, \eta) \). Incumbent firms set wages by
maximising expected profits, and trade off a higher chance of attracting the worker with a lower
rent per worker. A worker stays with the training firm if the utility from staying, \( w + \theta \), exceeds
the utility from moving, \( v = y \). Hence, the probability of staying is

\[
Pr(\text{stay}) = Pr(\theta > y - w) = 1 - G(y - w).
\]

If \( \theta > 0 \), this probability is positive even if \( w < y \): Some workers stay with the firm although
they receive a lower wage offer from the incumbent than from outside firms. Incumbent firms
maximise

\[
\max_w (1 - G(y - w))(y - w).
\]

\(^6\)Here we are implicitly assuming that outside firms do not only observe the worker’s revealed ability in the
second period, but also her expected ability in the first period (or her first period wage).
From the first order condition, $w$ satisfies
\[
w = y - \frac{1 - G(y - w)}{g(y - w)}.
\]
Log-concavity of $G$ guarantees that the second order condition for a maximum is satisfied. Since workers stay with a positive probability with the incumbent firm even if they receive a higher outside wage offer, firms have monopsonic power and pay wages below productivity. It can be easily verified that $\frac{dw}{dy} = 1$: A productivity increase of one unit leads to a wage increase of the same magnitude. Hence, the wage offer of the incumbent firm is equal to the worker’s productivity minus a constant, $\Delta$:
\[
w := y - \Delta. \tag{1}
\]
This is due to the assumption that the distribution of non-pecuniary job characteristics does not depend on workers’ ability and training.

**Training decision and wage determination in the first period**

We now turn to the firm’s training decision. We begin with deriving the socially optimal training level. We then show that if firms cannot commit to the amount of training, firms offer no training in equilibrium.

**Socially optimal training**

At the social optimum the marginal cost of training is equal to the marginal productivity of training:
\[
\int_{-\infty}^{\infty} \frac{\partial y(\tau, \eta)}{\partial \tau} dF_1(\eta|\hat{\eta}) = c'(\hat{\tau}).
\]
The assumptions on the cost and production function ensure that the second order condition is satisfied. As the productivity of workers during training is smaller than the productivity of untrained workers (fixed cost of training), it is not optimal to train every worker. A worker should only be trained if her productivity when trained exceeds her productivity when not trained. As the returns to training are higher for more able workers, while the fixed cost of training does not depend on workers’ ability, there exists a cut-off ability level $\hat{\eta}^*$ such that every worker with expected ability above $\hat{\eta}^*$ is trained. The socially optimal training level $\tau^*$ satisfies
\[
\tau^* = \begin{cases} 
0 & \text{if } \hat{\eta} < \hat{\eta}^*, \\
\hat{\tau} & \text{if } \hat{\eta} \geq \hat{\eta}^*.
\end{cases}
\]
Training decision

Under limited commitment firms choose the training level by maximising future profits. They do not take into account the impact training has on the utility of the worker. From (1), firms earn a rent of $\Delta$ on each retained worker. A worker stays with the incumbent firm with probability $(1 - G(\Delta))$. Hence, the firm’s profit in the second period equals $\Pi = (1 - G(\Delta)) \Delta$. Clearly, it does not depend on training. Consequently, firms offer no training in equilibrium. Although non-pecuniary job characteristics lead to firms making positive profits, they do not compress the wage structure. We summarise

**Proposition 1** Under limited commitment firms offer no training.

We close the model by analysing wage determination in the first period. Because of the free-entry condition firms make zero profits at any point of time. In the first period firms thus bid up the worker’s wage until all rents are exhausted. As firms make positive profits in the second period, first period wages are greater than workers’ expected productivity in the first period. Hence, there is front-loading. First period wages satisfy

$$ W = E[y(0, \eta)|\hat{\eta}] + (1 - G(\Delta)) \Delta. \quad (2) $$

2.2 Union agreements

We next incorporate wage rigidities due to unions into the training model. In our model, a unionised and non-unionised sector co-exist. Each sector consist of many firms. We assume that union agreements apply to all workers in the unionised, but not to workers in the non-unionised sector. This is the only difference between firms in the unionised and non-unionised sector; hence there is no systematic selection of firms into the unionised sector. In the empirical analysis, we attempt to control for non-random firm selection. We model the union wage as a wage floor. Unionised firms are obliged to pay at least the union wage, but are free to pay a higher wage.\(^7\) Non-unionised firms, on the other hand, may pay a wage below the union wage. We further assume that the union wage is exogenous. In other words, unionised firms have no

\(^7\)Our model therefore differs from union models such as Acemoglu and Pischke (1999b) and Booth et. al. (1999) who assume that firms are forced to pay the union wage.
impact on the union wage\textsuperscript{8}. To simplify the notation, we assume that the same union wage applies for workers with and without an apprenticeship. Our results also hold if the union wage for workers with training exceeds that for workers without training\textsuperscript{9}. We denote the union wage by $\bar{w}$. Union wages do not depend on ability, and do not apply for workers in training. Finally, we assume that firms can lay off workers at the end of the training period without incurring firing costs.

These assumptions were made in order to capture the crucial features of the German collective bargaining system. In Germany, only firms that belong to an employer association are bound to union agreements. Membership at an employer association is voluntary. In these firms, union agreements apply to all workers, regardless of workers’ union status. This essentially creates a unionised and non-unionised sector. Furthermore, wage negotiations in Germany take place at an industry and regional level and usually cover many firms. The assumption that firms cannot influence the union wage is therefore reasonable. In addition, in Germany union wages act as minimum wages, and many firms pay wages above the union wage (references). Finally, although in Germany firing costs are generally high, firms can lay off trainees at the end of the apprenticeship without incurring any firing costs. We believe that our model also captures many features of a bargaining system in which wage negotiations take place at the firm level, such as the US and the UK. In these countries, only union members are legally entitled to union wages. De facto, however, agreements are often extended to other workers in firm (references).

We commence by analysing wage determination in the second period. We then turn to firms’ incentives to train and wage determination in the first period. Finally, we analyse the sorting of workers into the unionised and non-unionised sector.

\subsection*{2.2.1 Wage determination in the second period}

\textit{Non-unionised firms}

First, consider non-unionised firms. Due to the free entry condition, \textit{outside} firms (i.e. firms that do not employ a particular trainee) continue to pay wages equal to workers’ productivity. This is so for both unionised and non-unionised firms. Consequently, union agreements do not

\textsuperscript{8}This assumption does not qualitively affect our results. It is easy to endogenise the union wage. Unions would want to set the union wage such that worker welfare is maximised.

\textsuperscript{9}In Germany, a different union wage applies for workers with and without an apprenticeship system.
affect wage determination of incumbent non-unionised firms: Since the worker’s outside option is unaffected by union agreements and union agreements do not apply to non-unionised firms, non-unionised incumbent firms continue to offer a wage equal to the worker’s productivity minus a constant $\Delta$ (see expression 1).

Figure 1: Wage determination in unionised firms

Unionised firms

Union agreements affect wage offers of unionised incumbent firms. Figure 1 illustrates how wages are set in these firms. In the figure, we consider untrained and trained workers. The vertical axis carries the wage and productivity of the worker, and the horizontal axis her revealed ability. Productivity and wages of untrained (trained) workers in the absence of any union agreement are indicated by the panels $y^{nt}$ ($y^t$) and $w^{nt}$ ($w^t$). From (1), they are equal to productivity minus a constant, $\Delta$. The horizontal line indicate the union wage $\bar{w}$. It is useful to distinguish between three groups of workers.

1. Workers with productivity below the union wage. In the figure these are workers with ability below $\eta^t_1$ if trained and $\eta^{nt}_1$ if untrained. These workers are worse off due to union agreements. Unionised firms do not find it profitable to employ them. As there are no firing costs at the end of the apprenticeship, these workers are laid off. They find work
in non-unionised firms and earn a wage equal to their productivity. Note that layoffs at the end of the apprenticeship occur because employers acquire new information about workers’ ability during the training period. If unionised firms had known workers’ ability in the first period, workers with an ability below $\eta_{1t}^{nt}$ ($\eta_{1t}^{t}$) would not have been hired.

2. Next, consider workers with a productivity above the union wage, but whose wage in the absence of union agreements falls below the union wage. In the figure, this refers to all workers with ability between $\eta_{1t}^{t}$ and $\eta_{2t}^{t}$ if trained, and $\eta_{1t}^{nt}$ and $\eta_{2t}^{nt}$ if untrained. These workers are better off due to unions, and earn a higher wage than they would in the absence of union agreements. Optimally, unionised incumbent firms would want to offer a wage below the union wage. As they are not allowed to do so, the best they can do is to offer just the union wage. Hence, workers with ability between $\eta_{1t}^{t}$ and $\eta_{2t}^{t}$ ($\eta_{1t}^{nt}$ and $\eta_{2t}^{nt}$) are paid the union wage.

3. Finally, consider workers whose wage in the absence of union agreements exceeds the union wage. In the figure this applies to all workers with ability above $\eta_{2t}^{t}$ if trained, and $\eta_{2t}^{nt}$ if untrained. These workers are unaffected by union agreements. The union wage is not binding for these workers. They thus earn the same wage as in the absence of union agreements.

Note that the probability that a worker turns out to be less able than $\eta_{1t}^{t}$ ($\eta_{1t}^{nt}$) depends on the worker’s expected ability, $\tilde{\eta}$.

### 2.2.2 Training decision and wage determination in the first period

We now turn to the training decision of unionised and non-unionised firms in the first period. We show that unionised firms offer more training than non-unionised firms. Hence, unions move the training level closer to the socially optimal level. The reason is that union agreements compress the wage for workers with a productivity around the union wage.

**Non-unionised firms**

Non-unionised firms offer no training in equilibrium. This is because union agreements have no impact on wage determination of non-unionised firms. The future profit of non-unionised firms is thus the same as in the absence of union agreements, and is not increasing in training (proposition 1).
Unionised firms

We next argue that training increases the future profits of unionised firms. This is best understood from figure 1. Consider a worker whose true ability is $\eta^u_1$. Without training the firm would make zero profit on this worker. With training the worker’s productivity is higher than the union wage $\bar{w}$, and the firm makes positive profits. More generally, training increases the rent on all workers with ability between $\eta^u_1$ and $\eta^u_2$. Workers with ability below $\eta^u_1$ are less productive than the union wage even after training. Workers with ability above $\eta^u_2$ are unaffected by union agreement even without training. Observe that this argument relies on firms making positive profits in the second period. Although non-pecuniary job characteristics are not sufficient to induce firms to sponsor training, they are necessary for unions to have an impact on training.

We now formalise this argument. We first derive an expression for the second period profit of unionised firms. Let $E[\Pi_u(\tau, | \eta)]$ denote the future (i.e. second period) profit on a worker with expected ability $\bar{\eta}$. Define $\eta_1$ as $y(\tau, \eta_1) = \bar{w}$, i.e. workers with ability below $\eta_1$ have a productivity below the union wage. Similarly, define $\eta_2$ as $y(\tau, \eta_2) = \bar{w} + \Delta$, i.e. workers with ability above $\eta_2$ are not affected by the union wage. Observe that $\eta_1$ and $\eta_2$ depend on the worker’s training level. Unionised firms lay off workers with ability below $\eta_1$ and hence make zero profits on these workers. For workers with ability between $\eta_1$ and $\eta_2$, unionised firms earn a rent of $y - \bar{w}$. These workers stay with the unionised firm after apprenticeship graduation with probability $1 - G(y - \bar{w})$. Finally, for workers with ability above $\eta_2$, firms make a profit of $(1 - G(\Delta))\Delta$. Hence, unionised firms maximise

$$
\max_{\tau} \quad -c(\tau) + E[\Pi_u(\tau, | \eta)] = -c(\tau) + \int_{\eta_1}^{\eta_2} (1 - G(y - \bar{w})) (y - \bar{w}) dF(\eta | \bar{\eta})
$$

$$
+ (1 - F(\eta_2 | \bar{\eta}))(1 - G(\Delta))\Delta.
$$

The training level unionised firms offer (in case they decide to train the worker), $\bar{\tau}_u$, solves

$$
c'(\bar{\tau}_u) = \frac{\partial E[\Pi_u(\bar{\tau}_u, | \eta)]}{\partial \tau} = \int_{\eta_1}^{\eta_2} (1 - G(y(\bar{\tau}_u, \eta) - \bar{w})) \frac{\partial y}{\partial \tau} dF(\eta | \bar{\eta})
$$

$$
- \int_{\eta_1}^{\eta_2} g(y(\bar{\tau}_u, \eta) - \bar{w})(y(\bar{\tau}_u, \eta) - \bar{w}) \frac{\partial y}{\partial \tau} dF(\eta | \bar{\eta}).
$$

10 Here, we have used that $y(\tau, \eta_2) = \bar{w} + \Delta$. 

14
Training affects profits in two ways. First, training increases the rent on trained workers. This effect is represented by the first term in (3). Second, training decreases the probability that the worker stays with the firm. This effect is captured by the second term in (3). In appendix B we show the first effect dominates the second effect. Hence, training increases the future profit of unionised firms, i.e. \[ \frac{\partial E[\Pi_u|\tilde{\eta}]}{\partial \tau} \geq 0 \] and \( \tilde{\tau}_u \geq 0 \).

As the productivity of workers in training differs from the productivity of untrained workers, firms do not find it profitable to train every worker. The unionised firm will only offer training if the profit with training exceeds the profit without training. In appendix B we show that there exist two thresholds which we denote by \( \tilde{\eta}_1 \) and \( \tilde{\eta}_2 \). The unionised firm trains the worker if her expected ability lies in between these two thresholds. The training choice of unionised firms therefore satisfies

\[
\tau^*_u = \begin{cases} 
0 & \text{if } \tilde{\eta} < \tilde{\eta}_1 \text{ or } \tilde{\eta} > \tilde{\eta}_2, \\
\tilde{\tau}_u & \text{if } \tilde{\eta}_1 \leq \tilde{\eta} \leq \tilde{\eta}_2. 
\end{cases}
\]

The intuition for these results is as follows. Recall from figure 1 that training increases future profits only of workers with ability between \( \eta^l \) and \( \eta^u \). Consider a worker with very low expected ability. This worker is likely to turn out to be less able than \( \eta^l \). The probability that she will be laid off after training is therefore very high, and the firm is better off by not training (and not hiring) her. A worker with a very high expected productivity, on the other hand, is likely to turn out to be more able than \( \eta^u \). The probability that this worker will be affected by the union wage is therefore low even without training. Again, training has only a small impact on the firm’s future profit. In other words, union agreements compress wages only for workers with a productivity around the union wage, and firms find it most profitable to train these workers. Consequently, the impact of union agreements on training is not uniform: Unions have little or no impact on training for workers with very low and very high (expected) ability, and the strongest impact for workers with expected productivity around the union wage.

Figure 2 plots the socially optimal training level as well as the training choice of unionised firms as a function of workers’ expected ability. In the social optimum workers with expected ability above \( \tilde{\eta}^* \) get trained. Under limited commitment non-unionised firms train no worker. Unionised firms offer more training than non-unionised firms, but less training than the socially optimal level. This is because firms do not take into account the impact training has on the utility of the worker.
Next, we analyse wage determination in unionised and non-unionised firms in the first period. Details can be found in appendix B. Wages in the first period are determined by the firm’s zero profit condition. As firms make positive profits in the second period, they make negative profits in the first period. Since union agreements do not affect profits and wage determination in non-unionised firms, non-unionised firms offer the same first period wage as in the absence of union agreements, (2). The first period wage offer of unionised firms can be similarly derived. It is the firm that bears the training cost. We summarise our results in the following proposition.

**Proposition 2** Under limited commitment non-unionised firms offer no training. Unionised firms train workers with expected ability $\hat{\eta}_1 < \hat{\eta} < \hat{\eta}_2$. These workers are offered a training level of $\tilde{\tau}_u$. Training in unionised firms is less than socially optimal. Firms bear the training cost.

**Proof.** See appendix B.

### 2.2.3 Worker sorting into the unionised sector

We close the model by sketching the sorting of workers into the unionised and non-unionised sector in the first period. Details can be found in appendix C. A worker bases her decision in which sector to work on the training level as well as on the first and second period wage
unionised and non-unionised firms offer. She chooses to work in the sector in which her utility is higher.

Consider first the impact of training on worker sorting. As it is the firm which bears the training cost, workers are better off if they receive training. Hence, workers for whom union agreements increase the probability of receiving training \( \hat{\eta}_1 < \hat{\eta} < \hat{\eta}_2 \) typically prefer to work in the unionised sector. \textit{Workers sort into the unionised sector based on the impact unions have on training.}

Second, consider the impact of the worker’s expected ability on worker sorting. Suppose unionised and non-unionised firms offer the same training level. Then there exist an ability threshold such that workers with expected ability above this threshold prefer to work in unionised firms. \textit{More able workers thus self-select into the unionised sector.} The intuition for this result is simple. \textit{Ex post,} workers who will be paid the union wage are better off, while workers who turn out to be less productive than the union wage are worse off when working in the unionised sector. Workers with low expected ability are likely to be of lower productivity than the union wage, and thus choose to work in non-unionised firms.

Finally, we would like to emphasise that \textit{ex ante} all workers are better off due to unions. Since a worker can always choose to work in a non-unionised firm, she is guaranteed at least the same payoff as in the absence of union agreements.

\subsection*{2.2.4 Interpretation: Unions as a commitment device}

In this section, we give a new interpretation of why unions increase training in the economy. Recall that the reason why -in the absence of unions- the training market breaks down is that firms cannot fully commit to training provision. This problem can in principle be mitigated -though not eliminated- by a long-term wage contract that today specifies not only today’s but also future wages. In our model, unionised firms offer a special type of a long-term wage contract: They guarantee to pay at least the union wage in the future. In principle, firms could offer such a contract even without becoming unionised. The problem with this type of contract is that it is not self-enforceable. Once training is completed, the firm has an incentive to deviate and pay a lower wage than the agreed minimum wage. Hence, the role unions essentially play in our model is that they serve as a commitment device. Unionised firms credibly signal to workers that they will pay at least the agreed union wage in the future. This then provides an
incentive for firms to train workers, and improves welfare in the economy.

These arguments point to an incentive for firms to become unionised. Recall from the previous section that a contract that promises to pay a minimum wage in the future may increase workers’ utility. Clearly, the probability that a firm can attract a worker depends on the utility the worker derives from working at the firm. In perfectly competitive labour markets, a firm must offer the contract that maximises the worker’s utility (and is feasible), subject to the zero-profit constraint. This then leads to the following conclusion. If there is a single union wage in the economy which is binding only in firms that choose to be unionised, some workers are better off in unionised firms, while others are better off in non-unionised firms (see section ?). Unionised and non-unionised firms thus co-exist, and firms are indifferent between joining the unionised or non-unionised sector. Both types of firms make zero profits in the long run. Compared to non-unionised firms unionised firms make a lower profit in the second period and therefore a lower loss in the first period. There is thus less front-loading in unionised firms. If, on the other hand, firms are free to set the minimum wage, all firms would choose to be unionised and offer a union wage that does depend on workers’ expected ability.

Finally, observe that in our model, unions have relatively little power. They can only enforce that firm pays at least the union wage, but they cannot prevent layoffs after training. Furthermore, unions cannot enforce a contract that specifies the amount of training. To us, this seems a good description of the German apprenticeship and collective bargaining system. Note, however, that more powerful unions -in the sense that they can enforce other non-self-enforceable contracts- could potentially increase training and welfare in the economy.

3 Empirical implications

In this section, we derive two sets of empirical implications of our model, training in unionised and non-unionised firms as well as worker selection.

3.1 Training in unionised and non-unionised firms

According to our model, union agreements compress wages only in unionised, but not in non-unionised firms, therefore predicting training in unionised, but not in non-unionised firms. Our model addresses only one reason for wage compression - other reasons include asymmetric
information. This may lead to firm-financed training also in non-unionised firms. We therefore test whether unionised firms are more likely to train workers in apprenticeship schemes than non-unionised firms. More specifically, we test whether unionisation increases the proportion of apprentices as well as the probability of training at least one worker in apprenticeship schemes.

Testing this hypothesis is not straightforward. First, our model suggest that workers with a higher expected ability self-select into the unionised sector. If more able workers are more likely to receive training, a simple comparison of the mean training intensity in unionised and non-unionised firms will overstate the causal impact of unions on training. Furthermore, our model suggests that the impact of unions on training depends on workers’ expected ability, and that workers sort into the unionised sector based on the impact unions have on training. This may lead to an upward estimate for the average impact unions have on training (i.e. the impact of unions on the training probability of a randomly selected worker or firm). Finally, the firm’s decision to be unionised may also depend on firm characteristics. If these characteristics are correlated with the firm’s propensity to train, a simple comparison of mean training intensity is again misleading.

Our empirical test is based on a firm data panel, supplemented by information about the firm’s workforce. The panel nature of our data allows using a difference in difference estimator to identify the effects of unionisation on training. To take account of a changes in firm characteristics as well as workforce quality within the firm, we will condition on changes in observable worker and firm characteristics. We describe our empirical strategy in detail later.

3.2 Worker selection: Wages of movers and stayers in unionised and non-unionised firms

The second set of empirical implications of our model concerns worker selection. These additional implications allow us to test whether the assumptions we made about wage determination in unionised and non-unionised firms are compatible with the data. They also help us to identify other reasons for firm-financed training, and thus help to answer the question why possibly also non-unionised firms offer apprenticeship training. However, they hold even if unions have no impact on training.

Probably the most prominent explanation for firm-financed apprenticeship training in Germany is asymmetric information (Acemoglu and Pischke (1998)). Acemoglu and Pischke’s
model predicts an adverse selection of movers: Workers who leave the training firm for endogenous reasons are less able than workers who stay with the training firm or leave the training firm for exogenous reasons (i.e. because of plant closure or military service). We show that our model provides an alternative explanation for an adverse selection of movers. Furthermore, we show that information on the union status of the training firm helps to distinguish between these two alternative explanations.

In our union model, two types of worker selection are present. First, in the first period workers sort into the unionised and non-unionised sector based on their expected ability. We analysed this type of selection in section ?. Second, in the second period, workers’ decision to leave the training firm depends on their revealed ability. In this section, we derive the empirical implications of this second type of selection, ignoring for the moment the first type of selection. In the empirical analysis, we control for the first type of selection. Our union model implies the following.

Proposition 3 (i) In non-unionised firms, endogenous and exogenous movers are as productive as stayers and earn a higher wage than stayers. (ii) In unionised firms, endogenous and exogenous movers may be less productive and may earn a lower wage than stayers. (iii) The difference between the wage of endogenous movers and stayers is lower (in absolute terms) in unionised than in non-unionised firms.

Proof. See appendix D. ■

The intuition for these results is as follows. Consider first workers in non-unionised firms. Perfect competition ensures that workers who leave the training firm are paid their marginal product, while workers who remain with the incumbent firm receive a wage that is lower than their marginal product, due to a rent induced by non-pecuniary utility shocks. The difference between the incumbent and outside firm’s wage offer is Δ. Hence, conditional on productivity, endogenous as well as exogenous movers earn a higher wage than stayers. Furthermore, the difference between the incumbent and outside offer, Δ, does not depend on workers’ productivity. Consequently, the probability that a worker leaves the firm does not depend on her productivity, and stayers, endogenous as well as exogenous movers all have the same productivity on average.

Next, consider workers in unionised firms. As in non-unionised firms, exogenous and endogenous movers earn a higher wage than stayers conditional on productivity. However, in unionised
firms there is also selection. There are two opposing effects. On the one hand, unionised firms dismiss all workers with a productivity below the union wage after training. These workers find employment in non-unionised firms and are paid a wage equal to their productivity. This effect leads to a lower average productivity of movers than of stayers. On the other hand, among workers who are paid the union wage, it is the more able workers who are more likely to leave. Figure 1 illustrates this point. Recall that the probability of moving is \( G(v - w) \), where \( w \) and \( v \) denote the worker’s training and outside firm wage offer, respectively. Consider a trained worker with ability \( \eta_1 \). This worker is offered the same wage by the training and outside firms. Hence, the probability that she leaves the training firm is \( G(0) \). Workers with ability above \( \eta_1 \), on the other hand, are offered a higher wage by outside firms than by the training firm, and are thus more likely to leave the training firm. This effect leads to a higher average productivity of movers than of stayers.

Finally, the model implies that in absolute terms, the difference between the wage of endogenous movers and stayers is lower in unionised than in non-unionised firms. The reason is that for workers who are paid the union wage - and among whom it is the more able workers who are more likely to leave the training firm - , the difference between the incumbent and outside wage offer is less than \( \Delta \), i.e. what it would be in non-unionised firms.

To sum up, unconditional on the union status, our model makes no ambiguous predictions about wages of movers and stayers. A lower wage for movers must be due to a negative selection of workers from unionised firms, and should thus be permanent. Our model unambiguously predicts a higher mover-stayer wage differential in unionised than in non-unionised firms.

4 Empirical Analysis

Our empirical analysis makes use of three different data sources: a one percent sample of the German workforce, drawn from social security records between 1975 and 1995, a firm panel data set, and data set which matches information on the workforce in firms in the firm panel data set to the firm data. The workforce information is drawn from the population of social security records. We describe the three data sets and the samples we use for analysis in detail in appendix E.
4.1 Training in unionised and non-unionised firms

The first test of our union model compares the training intensity in unionised and non-unionised firms. Define \( y_{ijt} \) as an indicator variable equal to 1 if worker \( i \) in firm \( j \) receives training in period \( t \), and 0 otherwise. The probability that a worker receives training depends on the union status of the firm \( U_{jt} \), observed and unobserved worker characteristics, \( \eta_i \), as well as observed and unobserved firm characteristics, \( f_j \). Both \( \eta_i \) and \( f_j \) are defined as deviations from the population mean. Assuming linearity, this relationship can be written as

\[
y_{ijt} = \beta + \lambda_i \ U_{jt} + \theta_t + \eta_i + f_j + v_{ijt},
\]

where \( v_{jt} \) is an i.i.d. error term, and \( \lambda_i \) is the effect of unions on the training probability. This parameter is individual specific, as workers with a higher expected probability are more likely to be trained.

Averaging over all workers who work in firm \( j \) in period \( t \) results in the following expression:

\[
\bar{y}_{jt} = b + \lambda_j(t) \ U_{jt} + \theta_t + \eta_j(t) + f_j + v_{jt},
\]

(4)

where \( \bar{y}_{jt} = \frac{1}{N_{jt}} \sum_{i=1}^{N_{jt}} y_{ijt} \) is the proportion of apprentices in firm \( j \) in period \( t \) and \( \eta_j(t) = \frac{1}{N_{jt}} \sum_{i=1}^{N_{jt}} \eta_i \) is the average quality of workers in firm \( j \) in period \( t \).

Consider now the expectation in the change in the proportion of apprentices for firms that do not change their union status between two periods:

\[
E(\bar{y}_{jt} - \bar{y}_{jt+1}|U_{jt} = 0, U_{jt-1} = 0) = E(\Delta \theta_t + (\eta_{j(t)} - \eta_{j(t-1)}) + \Delta v_{j(t)}|U_{jt} = 0, U_{jt-1} = 0),
\]

(5)

and for firms who change their union status from being non-unionised to being unionised:

\[
E(\bar{y}_{jt} - \bar{y}_{jt+1}|U_{jt} = 1, U_{jt-1} = 0) = E(\lambda_j(t) + \Delta \theta_t + (\eta_{j(t)} - \eta_{j(t-1)}) + \Delta v_{j(t)}|U_{jt} = 1, U_{jt-1} = 0).
\]

(6)

Differencing 5 and 6, one obtains \( E(\lambda_j(t)) \). This is the difference in the training probability in a unionised and a non-unionised firm for those who choose to work in a unionised firm. This causal effect is often referred to as the effect of treatment on the treated. Our estimator is a difference in difference estimator, which identifies the change in the proportion of apprentices over time from that occurring in non-unionised firms. The identifying assumption is that,
other than through the effect of unionisation, changes in the proportion of apprentices in non-unionised firms are equal to changes in firms that change from being non-unionised to unionised. This assumption is strong. It requires that the change in the average quality $\eta_{j(t)}$ between two time periods is the same in firms that remain non-unionised in the two periods, and firms that change union status.

If better workers self-select into unionised firms, as implied by our union model, this assumption may be violated. It is not likely however that the quality of workers adjusts immediately to the new equilibrium after changes in the union status. We address this problem nevertheless by assuming that any variation in changes in the workforce quality that is correlated with changes in the union status and changes in the proportion of apprentices is absorbed by changes in observed worker characteristics as well as firm characteristics.

Assuming linearity, we define $\eta_{j(t)} = X'_{j(t)} d + e_{j(t)}$, where $X_{j(t)}$ is a vector of observed characteristics of the firm’s workforce, as well as observed firm characteristics. The key identifying assumption is now that, conditional on changes in $X_{j(t)}$, changes in the counterfactual outcome distribution of firms who changed to be unionised, had they not changed to be unionised, is the same than the observed outcome distribution of firms which are not unionised. Under this assumption, estimation identifies $E(\lambda_{j(t)})$, which is the treatment on the treated effect.

Our estimation strategy combines a linear matching estimator with a difference in difference estimator. The equation we estimate is given by

$$\Delta y_{jt} = \Delta \theta_t + \Delta X'_{j(t)} d + \Delta v_{j(t)} + \lambda_{j(t)} U_{jt}, \quad (7)$$

where the sample includes firms that are non-unionised in both periods, and firms that change from being non-unionised in period $t - 1$ to being unionised in period $t$.

This estimator uses only firms that are never unionised over the course of the panel, or firms that change from being non-unionised to being unionised. For the latter group, we could instead use firms that change from being unionised to being non-unionised; this should then identify $-E(\lambda_{j(t)})$.

An alternative reference category are firms that are unionised in both periods. In this case, identification of $E(\lambda_{j(t)})$ requires the additional assumption that $E(\lambda_{j(t)} - \lambda_{j(t-1)})$ is equal to zero for firms that are unionised in both periods. Below we present results for all these groups.
4.1.1 Empirical Findings

Table 1 displays information on union coverage as well as the proportion of trainees in unionised and non-unionised firms for the years 1996-1999. Entries are weighted so that they are representative for firms (upper panel) and workers (lower panel).

The proportion of firms that are unionised is about 56 percent, but 76 percent of all workers are employed in unionised firms. The reason is that it is predominantly large firms that recognise union agreements.

The numbers on proportion of trainees are constructed by matching information about the workforce to the firm panel data, as described in the appendix E. The information on whether or not the firms trains apprentices stems from the survey information of the firm panel. The numbers in the upper panel indicate that 29 percent of all firms train workers, and unionised firms are considerably more likely to train than non-unionised (38 percent vs. 17.5 percent). Those firms that are unionised have an average proportion of apprentices of 10 percent, while those firms that are not unionised have a proportion of only 5.42 percent.

The numbers in the lower panel indicate that about 60 percent of workers work in firms that report that they train apprentices. However, only 34 percent of workers working in non-unionised firms work at the same time in firms that train, while this is the case for 68 percent of workers employed in unionised firms.

In table (2), we compare unionised and non-unionised firms in terms of observable characteristics. The numbers in the table show that there is considerable variation in unionisation rates across industries. Also, Unionised firms are substantially larger and older than non-unionised firms. There is some evidence that the revenue per worker is higher in unionised firms, while firms’ evaluation of profits is more favourable in non-unionised firms.

In Table 3 we report conditional and unconditional wage differentials for workers employed in unionised and non-unionised firms, where we distinguish between three education levels (workers without post-secondary school education (unskilled), workers who completed an apprenticeship, and workers with a degree from a university). Estimation is based on data for the year 2000. For this year we have access to the complete linked employer-employee data set on all workers who work in one of the firms in the firm panel. Overall, wages are about 28 percent higher in unionised firms, and 8.5 percent higher when we condition on firms size, region, sector, gender, education, working part time, and being a foreign worker. This difference potentially reflects
Table 1: Training in unionised and non-unionised firms: Raw differential

<table>
<thead>
<tr>
<th>Proportion Unionised</th>
<th>All</th>
<th>Unionised</th>
<th>Non-Unionised</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>StdD</td>
<td>Mean</td>
</tr>
<tr>
<td>Proportion trainees (in percent)</td>
<td>8.09</td>
<td>17.98</td>
<td>10.18</td>
</tr>
<tr>
<td>Firm trains (in percent)</td>
<td>28.79</td>
<td>45.28</td>
<td>37.59</td>
</tr>
</tbody>
</table>

| Proportion working in unionised firms | 76.19 |
|                                      | Mean | StdD      | Mean          | StdD          | Mean | StdD          |
| Proportion trainees (in percent) | 5.97 | 10.82 | 6.29 | 10.46 | 4.93 | 11.83 |
| Firm trains (in percent)           | 59.65 | 46.06 | 67.74 | 46.74 | 33.75 | 47.29 |


both the causal impact of unions on wages and the selection of more productive workers into unionised firms. In the next columns, we report unconditional and conditional union wage differentials for different skill groups. The differential is largest for unskilled workers, decreases with the worker’s skill level, and is insignificant for highly skilled workers. This is consistent with our model, as the union wage is most binding for the unskilled and least binding for university graduates.

In table 4, we display results from our difference in difference - matching estimator, as explained in section 4.1. The first pair of columns present results of changes in unions status from being non-unionised to being unionised, where the base category are firms that are non-unionised over the entire sample period. Firms that become unionised increase their proportion of apprentices by about 2.7 percent, and this effect is significant. Results in the second row refer to changes in the response of the firm whether she trains workers, as a result of becoming unionised. The effect is again positive and significant, indicating that the probability of training workers increases by 6.8 percentage points as a result of unionisation.

The results in the next pair of columns refers to firms that change from being unionised to being non-unionised, where again non-unionised firms identify the time effects. The effect on the proportion of apprentices is practically zero, but not negative, as predicted by our model. The reason may be that downward adjustment is not immediate. Apprenticeship training schemes
Table 2: Characteristics of unionised and non-unionised firms: Firm data

<table>
<thead>
<tr>
<th></th>
<th>Unionised Mean</th>
<th>Unionised StdD</th>
<th>Non-Unionised Mean</th>
<th>Non-Unionised StdD</th>
</tr>
</thead>
<tbody>
<tr>
<td>- energy/mining/water</td>
<td>97.67 %</td>
<td>2.33 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- credit/insurance</td>
<td>91.62 %</td>
<td>8.36 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- other services</td>
<td>64.08 %</td>
<td>35.92 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- retail/wholesale</td>
<td>76.25 %</td>
<td>23.75 %</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Firm size | 962 | 3608 | 125 | 564 |
(reinvestm/worker)/1000* | 14.87 | 229.28 | 21.87 | 661.41 |
(revenue/worker)/1000* | 693.13 | 4303.73 | 313.63 | 1180.71 |
profit evaluation high | 27.56 % | 37.94 % | | |
profit evaluation low | 30.16 % | 15.15 % | | |
proportion young firms (≤ 5) | 13.03 % | 23.47 % | | |
proportion old firms (> 30) | 62.95 % | 29.94 % | | |


German Marks

last for 2 - 3 years, and reduction is therefore only possible at the margin. The effect on the self reported information of whether the firm trains is negative, but insignificant.

These results are in accordance with our model, and indicate that the proportion of apprentices is significantly increased in unionised firms, and that these firms have a significantly higher probability to train workers, as compared to non-unionised firms.

In the next two pairs of columns, we repeat the same regressions, but we now identify the time effects from firms that are unionised over the course of the panel. As we explain above, this implies the additional assumption that the selection into unionised firms over time remains constant. The results are very similar to those in columns 1 and 2.

4.2 Worker selection

We now investigate the second set of empirical implications of our model, which concerns worker selection. These additional implications allow us to test whether the assumptions we made about wage determination in unionised and non-unionised firms are compatible with the data. They also help us to identify possible other reasons for firm-financed training, such as
Table 3: Union-Non-Union Wage Differentials

<table>
<thead>
<tr>
<th>union status</th>
<th>All</th>
<th>Unskilled</th>
<th>Completed Apprenticeship</th>
<th>University/Poly.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U</td>
<td>NU</td>
<td>U</td>
<td>NU</td>
</tr>
<tr>
<td>unconditional</td>
<td>Coefficient</td>
<td>StdE</td>
<td>Coefficient</td>
<td>StdE</td>
</tr>
<tr>
<td></td>
<td>0.28 (0.006)</td>
<td>0.383 (0.014)</td>
<td>0.238 (0.005)</td>
<td>0.007 (0.012)</td>
</tr>
<tr>
<td>conditional*</td>
<td>0.085 (0.005)</td>
<td>0.145 (0.003)</td>
<td>0.107 (0.005)</td>
<td>-0.023 (0.010)</td>
</tr>
</tbody>
</table>

* The conditional coefficient is based on a log-wage regression that includes firm size (10 dummies) and dummies for industry (distinguishing between 19 sectors, region (11 states), female, part time work and foreign worker. Estimation results in columns 1 include in addition dummy variables for education.

Table 4: Effect of Unionisation on Training Intensity

<table>
<thead>
<tr>
<th></th>
<th>Reference: Nonunion-Nonunion</th>
<th>Reference: Union-Union</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nonunion-Union</td>
<td>Union-Non-Union</td>
</tr>
<tr>
<td>Proportion Apprentices</td>
<td>0.0267</td>
<td>0.0104</td>
</tr>
<tr>
<td>Firm Trains</td>
<td>0.0677</td>
<td>0.0327</td>
</tr>
</tbody>
</table>

Source: Firm panel, matched with employee information, 1996-1999. All regressions include time dummies. Firm characteristics include changes in firm size, investment per worker, revenue per worker, number of new employees, and evaluation of profit. Worker characteristics include changes in the proportion of qualified workers, the average age of workers, the average daily wage, and the ratio of females.
asymmetric information, and thus help to answer the question why also non-unionised firms offer apprenticeship training.

We commence by estimating the wage differential for workers who stay with and leave the apprenticeship firm, using longitudinal data. This data does not contain information on the worker’s union status. It has, however, several other advantages. First, it allows us to precisely distinguish between stayers, exogenous and endogenous movers. Second, and most importantly, we can study the evolution of wages, thus analyse whether differences between wages for movers and stayers are permanent.

We then use matched employee-employer data, and distinguish between workers trained in unionised and non-unionised firms. This data has the disadvantage that we observe workers’ wages only once, right after apprenticeship completion. Furthermore, we cannot distinguish between exogenous and endogenous movers, and it does not constitute a random sample of apprentices. Despite these drawbacks, it helps to discriminate between our union model and an asymmetric information model. Our union model predicts a higher wage for movers than for stayers in non-unionised firms. Hence, a higher wage for stayers among workers trained in non-unionised firms must be due to asymmetric information. A higher mover-stayer wage differential in unionised firms is evidence that union agreements contribute to an adverse selection of movers. We explain the exact empirical strategy in section 4.2.1.

4.2.1 Movers vs stayers: Longitudinal data

Our test is based on the following simple model:\(^\text{11}\):

\[
\log w_{it} = \gamma_{\mathrm{exo}} M_{i(0)}^{\mathrm{exo}} + \gamma_{\mathrm{end}} M_{i(0)}^{\mathrm{end}} + \theta j(0) + (\eta_i - \theta j(0)) + \epsilon_{ijt},
\]

where \(w_{it}\) denotes the worker \(i\)'s log wages \(t\) periods after apprenticeship completion. \(M_{i(0)}^{\mathrm{exo}}\) and \(M_{i(0)}^{\mathrm{end}}\) are indicator variables, being equal to one if the worker left the training firm right after the training period for exogenous or endogenous reasons, respectively. Separation (for exogenous or endogenous reason) of the worker from the training firm may be separation from a unionised firm \((M_{i(0)}^{\mathrm{u}})\), or a separation from a non-unionised firm \((M_{i(0)}^{\mathrm{nu}})\), with \(M_{i(0)} = M_{i(0)}^{\mathrm{u}} + M_{i(0)}^{\mathrm{nu}}\). The quality of the worker is represented by \(\eta_i\), while \(\theta j(0)\) captures the average quality of workers at firm \(j\) (at the time of apprenticeship completion) that trained the worker.

---

\(^\text{11}\)See von Wachter (2002) for a similar set-up.
When we say worker quality, we mean both her ability and training she received. Finally, $e_{it}$ is a random disturbance term. Due to non-pecuniary job characteristics, our union model implies $\gamma^{exo}$ and $\gamma^{end} > 0$, for workers trained in unionised as well as non-unionised firms. For endogenous movers, OLS estimation results in

$$
\hat{\gamma}^{end} = \frac{\text{Cov}(w_{it}, M^{end}_{i(0)})}{\text{Var}(M^{end}_{i(0)})} = \gamma^{end} + \frac{\text{Cov} (\eta_{j(0)}, M^{end}_{i(0)})}{\text{Var}(M^{end}_{i(0)})} + \frac{\text{Cov} (\eta_{i} - \eta_{j(0)}, M^{end}_{i(0)})}{\text{Var}(M^{end}_{i(0)})}.
$$

The second term, $\frac{\text{Cov} (\eta_{j(0)}, M^{end}_{i(0)})}{\text{Var}(M^{end}_{i(0)})}$, captures sorting of workers into training firms. In our union model, workers sort into the unionised sector in the first period based on their expected ability. If retention rates after apprenticeship completion are higher in unionised and non-unionised firms, then the second term is negative, otherwise it is positive$^{12}$.

The third term, $\frac{\text{Cov} (\eta_{i} - \eta_{j(0)}, M^{end}_{i(0)})}{\text{Var}(M^{end}_{i(0)})}$, captures the selection of workers after training. It tells us whether among trainees in the same firm, it is the less or more able workers who are more likely to leave. It is this covariance we are interested in. Acemoglu and Pischke’s asymmetric information model implies a negative covariance, while our union model makes no unambiguous prediction. Suppose we can control for sorting of workers into training firms. Then, as $\gamma^{end} > 0$, a negative estimate for $\hat{\gamma}^{end}$ - i.e. a lower wage for endogenous movers than for stayers - implies that it is the low quality workers who are more likely to leave the training firm. This is what we call adverse selection. Furthermore, the wage difference between movers and stayers should be permanent, and visible even many years after apprenticeship completion. We control for selection into training firms in two ways. First, we condition on observable training firm characteristics, such as firm size and sector. Recall from the previous section that these characteristics are highly correlated with unionisation. Second, there are some workers in our data who have done their apprenticeship in the same firm. This allows us to condition also on training firm fixed effects.

A similar argument holds for exogenous movers. Here, a comparison between $\hat{\gamma}^{exo}$ and $\hat{\gamma}^{end}$ is of more interest. Since both exogenous and endogenous movers are paid a wage equal to their marginal productivity, the difference between $\hat{\gamma}^{exo}$ and $\hat{\gamma}^{end}$ identifies the productivity

$^{12}$More general models may lead to worker sorting not only into unionised and non-unionised firms, but also into small and large firms, firms with high or low quality training programmes, etc.
difference between the two groups of workers.

In table 5, we report wage differentials between stayers and various groups of movers conditional on individual characteristics (columns 1) and individual as well as training firm characteristics (columns 2). The entries in the two columns are coefficients on the respective dummy variable, i.e. $\gamma^{end}$ and $\hat{\gamma}^{exo}$. Mobility after apprenticeship training is considerable. Overall, 33 percent of workers leave the training firm after training. Of those, 6.9 percent are forced to leave their training firm due to closure. If we do not condition on firm characteristics, workers who remain with the training firm earn a 10.23 percent higher wage than workers who leave the training firm for endogenous reasons. This wage differential is highly significant. It reduces to 5.68 percent when we condition on training firm characteristics in addition. Training firm characteristics are highly significant. This suggests that sorting of workers into training firms plays an important role in the German apprenticeship system.

A comparison between wages for exogenous and endogenous movers (row 2 and 3) show sensitivity to the set of conditioning variables. When conditioning only on worker characteristics, exogenous movers earn a significantly lower wage than endogenous movers. However, it is predominantly small firms which close down\textsuperscript{13}. If larger firms attract better workers, and larger firms have higher retention rates, then this could explain the difference. Indeed, when we condition on characteristics of the training firm, exogenous movers obtain wages that are 3.3% higher than endogenous movers. This difference is significant at the 1 percent level. Both groups obtain lower wages than individuals who remain with their training firm after the training period has finished.

The next panel of table 5 (rows 4-7) breaks the sample down according to whether individuals went to the service after their apprenticeship training. This is the case for 11.3 percent of our sample. Of those, 68 percent do not return to the training firm. Those who do service and return to the training firm have virtually identical wages to those who stay with the training firm (row 6). In consequence, any wage advantage of workers who do not return to the training firm after their service, compared to endogenous movers, is unlikely to be due to training received whilst in service\textsuperscript{14}. Those who were drafted into the service right after the

\textsuperscript{13}The average size (median) of the apprenticeship firm for stayers is 1560 (57), for endogenous movers 734 (35), for workers who leave due to plant closure 130 (8), and for workers who are drafted into military service 560 (34).

\textsuperscript{14}This confirms Acemoglu and Pischke’s (1998) conjecture that military service does not increase productivity.
Table 5: Wages of stayers, endogenous and exogenous movers

<table>
<thead>
<tr>
<th></th>
<th>1 worker characteristics</th>
<th>2 + training firm characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>standard errors</td>
<td>standard errors</td>
</tr>
<tr>
<td></td>
<td>N=24194</td>
<td>N=24194</td>
</tr>
<tr>
<td>1 stays</td>
<td>67 %</td>
<td></td>
</tr>
<tr>
<td>2 leaves</td>
<td>30.7 %</td>
<td>-0.1023 (0.0041)</td>
</tr>
<tr>
<td>3 leaves, plant closure</td>
<td>2.3 %</td>
<td>-0.1394 (0.0128)</td>
</tr>
<tr>
<td>Test: 2=3</td>
<td>F(1,24173)=8.14</td>
<td>P: 0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 no service, stays</td>
<td>62.8 %</td>
<td></td>
</tr>
<tr>
<td>5 service, returns</td>
<td>3.6 %</td>
<td>0.0005 (0.0100)</td>
</tr>
<tr>
<td>6 service, leaves</td>
<td>7.7 %</td>
<td>-0.0878 (0.0076)</td>
</tr>
<tr>
<td>7 no service, leaves</td>
<td>25.9 %</td>
<td>-0.1057 (0.0046)</td>
</tr>
<tr>
<td>Test: 6=7</td>
<td>F(1,22174)=4.72</td>
<td>P: 0.029</td>
</tr>
</tbody>
</table>

Standard errors in parenthesis. The first column controls for worker characteristics only. Regressors included are age at the end of the apprenticeship, duration of the apprenticeship, A-levels (Abitur), nationality and time effects. Column 2 also controls for characteristics of the apprenticeship firm, i.e. firm size (in logs) and sector. For regressions that include service quitters, ”problematic cases” are dropped. This reduces the sample size to 22196. Definition 1 is used for service quitters. Results are not sensitive to the definition used and whether problematic cases are dropped or not.

apprenticeship and did not return to their training firm have higher wages than endogenous movers (rows 7 and 8). This difference is significant at a 5 percent level for both specifications.

To sum up, these results are in line with an adverse selection of movers. However, our observable training firm characteristics may be insufficient to control for sorting of workers into training firms. Next, we present results that condition on fixed training firm effects.

Fixed training firm effects

In our sample, we observe for some larger establishments more than one worker who obtains his apprenticeship training. This sample selects on large firms - while the average (median) firm size is 1265 (58), the average (median) size of firms with multiple apprentices is 5602 (1264). In table (6), we present results. Entries in the first and second row replicate OLS estimations conditional on worker (row 1) and worker as well as training firm characteristics (row 2); thus, results correspond to those reported in the first and second column in table 5. For this selected
sample of workers, OLS estimates are 4-5 percentage points larger than those using the overall sample.

Table 6: Mover, fixed training firm effect estimates

<table>
<thead>
<tr>
<th></th>
<th>Coeff</th>
<th>StdE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 leaves, worker</td>
<td>-0.1437</td>
<td>(0.0092)</td>
</tr>
<tr>
<td>characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 leaves, worker+firm</td>
<td>-0.1079</td>
<td>(0.0096)</td>
</tr>
<tr>
<td>characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 fixed training firm</td>
<td>-0.1059</td>
<td>(0.0095)</td>
</tr>
<tr>
<td>effect</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N=3874. Same specification as table 5.

The third row reports estimates conditional on training firm fixed effects. The estimate is virtually identical to the one that conditions on observable training firm characteristics. This suggests that there is little sorting into training firms, conditional on training firm size and sector. The lower wage for endogenous movers is therefore likely to be a consequence of adverse selection. If this is the case, the wage difference between movers and stayers should be persistent, and visible many years after apprenticeship completion. We next analyse the evolution of the mover-stayer wage differential by experience.

Are differentials persistent?

In figure 3, we plot wages against experience, where we distinguish between stayers, movers, and individuals who change firm because the training firm closes down (left panel) and individuals who change firms because of service (right panel). The graphs are based on regressions which include the same set of variables as columns 2 of table 5. The graphs are constructed from estimated coefficients for experience dummies, and experience dummies interacted with an indicator variable for endogenous move, or the respective exogenous move.

Wage disadvantages of endogenous movers relative to stayers are persistent, and remain sizable even after 10 years of labour market experience. Both groups of exogenous movers have initially an intermediate position, but their wages catch up with those of stayers. The provides additional support for the hypothesis that less productive workers are more likely to leave the training firm.

We have also estimated regressions that exclude sectors commonly viewed as problematic, like agriculture, and construction\textsuperscript{15}. Results change only slightly compared to those reported

\textsuperscript{15} Acemoglu and Pischke (1998) report that in the construction sector a penalty is imposed on firms who do
above.

To sum up, we find strong evidence for an adverse selection of movers: It is the less able trainees who are more likely to leave the training firm. This is implied by Acemoglu and Pischke’s asymmetric information model. It is, however, also consistent with our union model. In the next section, we use matched data to discriminate between these two reasons for an adverse selection of movers.

4.2.2 Matched data: Unionised versus non-unionised firms

We now break down the mover-stayer wage differential by the union status of the worker’s training firm. Our union model predicts a higher wage for movers than for stayers in non-unionised firms. Hence, a higher wage of stayers among workers trained in non-unionised firms must be due to asymmetric information. A higher mover-stayer wage differential in unionised firms is evidence that union agreements contribute to an adverse selection of movers.

We run separate regressions for workers trained in unionised and non-unionised firms. Regressions include the same set of variables as columns 2 of table 5: dummy variables for A-levels, nationality, year, and sector of apprenticeship firm; furthermore, the duration of apprenticeship, age, and the firm size of the apprenticeship firm.

Table 7 reports results. We find a higher wage for stayers than movers also for workers trained in non-unionised firms. However, the wage differential is 7.1 percentage points larger for workers trained in unionised firms. This is compatible with our union model. However, not train apprentices.
Table 7: Wages of movers and stayers: Unionised versus non-unionised firms

<table>
<thead>
<tr>
<th></th>
<th>Unionised N=35067</th>
<th>Non-unionised N=692</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>proportion</td>
<td>coefficient StdE</td>
</tr>
<tr>
<td>stays</td>
<td>74.34 %</td>
<td>68.83 %</td>
</tr>
<tr>
<td>leaves</td>
<td>25.66 %</td>
<td>-0.1481 (0.0042)</td>
</tr>
</tbody>
</table>

Standard errors in parenthesis. Regressors included are age at the end of the apprenticeship, duration of apprenticeship, A-levels (Abitur), nationality, time effects, firm size (in logs) and sector of training firm.

the positive wage differential also in non-unionised firms indicates that wage floors caused by unions are only one reason for an adverse selection of movers.

5 Discussion and Conclusion

In this paper we revisit the question why firms pay for general training. The focus is on the impact of wage rigidities caused by unions on training. We address these questions within the German apprenticeship system.

We develop a simple model of firm-sponsored training and unions. Our model extends existing models on unions and training, such as Acemoglu and Pischke (1999b) and Booth et al. (1999), in several ways. First, our model is set up in the same framework as Acemoglu and Pischke’s asymmetric information model. This allows us to compare the empirical implications of their model with our model. Second, in contrast to existing models on unions and training, we allow a unionised and non-unionised sector to co-exist. We derive wage and training determination in both sectors, and explicitly analyse the worker’s decision in which sector to work. We model the union wage as a minimum wage. Non-unionised firms may pay a wage below the union wage, while unionised firms have to pay at least the union wage, but are allowed to pay a wage above the union wage.

Our key theoretical results can be summarised as follows. First, an adverse selection of movers is not unique to Acemoglu and Pischke’s asymmetric information model, but is also consistent with our union model. Second, union agreements compress wages for workers with a
productivity around the union wage in unionised firms, inducing them to train these workers. If union agreements are the only source of wage compression, non-unionised firms offer no training. Training is less than socially optimal. Third, the impact of unions on training is not uniform. Workers sort into the unionised sector based on the impact union agreements have on training. Furthermore, more able workers sort into the unionised sector. Finally, long-term wage contracts mitigate, though not eliminate, the problem of market failure due to limited commitment to training provision. Such contracts, however, are not self-enforceable. The role unions essentially play in our model is that they make long-term wage contracts enforceable.

Our empirical strategy proceeds in two steps. In a first step, we use firm panel data, matched with information about the firm’s workforce, to directly test whether unions increase training. Our goal is to identify the causal impact of unions on training. Our estimation strategy follows a difference in difference - linear matching estimator, taking account of nonrandom selection of workers and possibly firms into the unionised sector. We consistently find that unions increase training. We conclude that wage floors created by unions are an important reason for firm-financed training in Germany. However, we cannot rule out asymmetric information as an additional reason for firm-financed training.

In a second step, we test for an adverse selection of movers. We find strong evidence for this: Even after ten years in the labour market, endogenous movers earn considerably lower wages than stayers and exogenous movers. We then use matched employee-employer data to compare the mover-stayer wage differential for workers trained in unionised and non-unionised firms. We find a higher wage for stayers than for movers not only for workers trained in unionised firms but also for those trained in non-unionised firms, but the differential is twice as high for workers trained in unionised firms.

Our results have implications for attempts to introduce training schemes in countries in which unions play only a minor role may fail. It is, however, important to bear in mind that the reason why unions induce firms to sponsor training is that unions make long-term wage contracts enforceable. There may be ways to do so other than through the unionisation of the economy.

Our results also have important welfare implications. Most importantly, we find no evidence that unions reduce training in the economy. Moreover, our results are compatible with unions moving training closer to the socially optimal level. Unions play a welfare-improving role.
Through binding wage floors, they shift rents from firms to workers. This welfare-increasing impact of unions as well as the impact of unions on training is the larger the greater the labour market frictions in the economy, i.e. the higher the rents firms earn. In labour markets with low mobility costs and wages close to productivity -as maybe in the US- wage floors may have only a small impact on worker welfare.

We would also like to stress that the existence of a non-unionised sector in which union agreements do not apply avoids many of the potentially harmful effects of unions. Most importantly, it allows workers with a productivity below the union wage to find employment. In this respect, the German collective bargaining system, which is based on the voluntary participation of both firms and workers, has advantages over a national-wide minimum wage and collective bargaining systems in which union agreements are extended to all firms and workers in the economy, such as in France or Austria.
6 Appendix

A Socially optimal training

We first show that in the social optimum more able workers receive more training. The socially optimal training level satisfies

$$\int_{-\infty}^{\infty} \frac{\partial y(\tau, \eta)}{\partial \tau} dF_1(\eta|\tilde{\eta}) = c'(\tilde{\tau}).$$

(10)

The increase in training due to an increase in the worker’s expected ability yields

$$\frac{d\tilde{\tau}}{d\tilde{\eta}} = \frac{\int_{-\infty}^{\infty} \frac{\partial^2 y(\tau, \eta)}{\partial \tau^2} dF_1(\eta|\tilde{\eta})}{c''(\tilde{\tau}) - \int_{-\infty}^{\infty} \frac{\partial^2 y(\tau, \eta)}{\partial \tau^2} dF_1(\eta|\tilde{\eta})} \geq 0,$$

due to complementarity between ability and training and the assumptions on the cost and production function. We next show that in the social optimum only workers with expected ability greater than $\tilde{\eta}^*$ get trained. A worker is trained if her productivity with training minus the training cost exceeds her productivity without training, i.e. if

$$\int_{-\infty}^{\infty} y(\tau, \eta)dF_1(\eta|\tilde{\eta}) - c(\tilde{\tau}) - k - \int_{-\infty}^{\infty} y(0, \eta)dF_1(\eta|\tilde{\eta}) \geq 0.$$

(11)

Totally differentiating the left hand side of (11) with respect to $\tilde{\eta}$ yields

$$\int_{-\infty}^{\infty} \frac{\partial y(\tau, \eta)}{\partial \eta} dF_1(\eta|\tilde{\eta}) - \int_{-\infty}^{\infty} \frac{\partial y(0, \eta)}{\partial \eta} dF_1(\eta|\tilde{\eta}) > 0$$

due to the complementarity between ability and training. There thus exists an ability threshold $\tilde{\eta}^*$ such that all workers with expected ability greater than $\tilde{\eta}^*$ are trained. $\tilde{\eta}^*$ is implicitly defined as

$$\int_{-\infty}^{\infty} y(\tau, \eta)dF_1(\eta|\tilde{\eta}^*) - \int_{-\infty}^{\infty} y(0, \eta)dF_1(\eta|\tilde{\eta}^*) = c(\tilde{\tau}) + k.$$

B Proof of proposition 2

Proposition 2 Under limited commitment non-unionised firms offer no training. Unionised firms train workers with expected ability $\tilde{\eta}_1 < \tilde{\eta} < \tilde{\eta}_2$. These workers are offered a training level of $\tilde{\tau}_u$. Training in unionised firms is less than socially optimal. Firms bear the training cost.

We first show that the future profit of the unionised firm, $E[\Pi_u(\tau, \eta)|\tilde{\eta}]$, is increasing in training. The increase in $E[\Pi_u(\tau, \eta)|\tilde{\eta}]$ due to training equals

$$\frac{\partial E[\Pi_u(\tau, \eta)|\tilde{\eta}]}{\partial \tau} = \int_{\eta_1}^{\eta_2} (1 - G(y(\tau, \eta) - \bar{w}) \frac{\partial y}{\partial \tau} dF_1(\eta|\tilde{\eta})$$

$$- \int_{\eta_1}^{\eta_2} g(y(\tau, \eta) - \bar{w})(y(\tau, \eta) - \bar{w}) \frac{\partial y}{\partial \tau} dF_1(\eta|\tilde{\eta}).$$
Recall that for workers with expected ability between $\eta_1$ and $\eta_2$ the union wage is greater than the wage the firm would choose optimally. Hence, from the first order condition of the second period wage, 

$$1 - G(y - \bar{w}) \geq g(y - \bar{w})(y - \bar{w}).$$

Consequently, $$\frac{\partial E[\Pi_u(\tau, \eta)]}{\partial \eta} \geq 0,$$ and $\bar{\tau}_u \geq 0$.

We next show that workers with expected ability between $\hat{\eta}_1$ and $\hat{\eta}_2$ are trained. A unionised firm trains if profits with training exceed expected profit without training, i.e.

$$E[\Pi_u(0, \eta)] < -c(\bar{\tau}_u) - k + E[\Pi_u(\bar{\tau}_u, \eta)\bar{\eta}].$$

As $\frac{\partial E[\Pi_u(\tau, \eta)]}{\partial \tau} \geq 0$, $E[\Pi_u(0, \eta)] \leq -c(\bar{\tau}_u) + E[\Pi_u(\bar{\tau}_u, \eta)\bar{\eta}]$. Hence, if the fixed cost of training is equal to 0, $(k = 0)$, every worker would get trained. Figure 4 illustrates the impact of the fixed cost of training on the probability that a worker receives training. It plots the firm’s profit with and without training as a function of the worker’s expected ability, for moderate (panel A) and high (panel B) values of the fixed cost of training, $k$. The firm’s profit is increasing in the worker’s expected ability, $\hat{\eta}$. It is first convex, then concave in $\hat{\eta}$. To prove this, we first show that $E[\Pi_u(\tau, \eta)\bar{\eta}]$ is increasing in workers’ expected ability. Differentiating $E[\Pi_u(\tau, \eta)\bar{\eta}]$ with respect to $\hat{\eta}$ yields

$$\frac{\partial E[\Pi_u(\tau, \eta)\bar{\eta}]}{\partial \eta} = \int_{\eta_1}^{\eta_2} \{(1 - G(y - \bar{w})) - g(y - \bar{w})(y - \bar{w})\} \frac{\partial y}{\partial \eta} \frac{\partial f_1(\eta)}{\partial \eta} d\eta > 0.$$

We next show that $\Pi_u$ is first convex and then concave in $\hat{\eta}$. Taking the second derivative yields

$$\frac{\partial^2 E[\Pi_u(\tau, \eta)\bar{\eta}]}{\partial \eta^2} = \int_{\eta_1}^{\eta_2} \{(1 - G(y - \bar{w})) - g(y - \bar{w})(y - \bar{w})\} \frac{\partial y}{\partial \eta} \frac{\partial f_1(\eta\bar{\eta})}{\partial \eta} d\eta.$$

Observe that $\frac{\partial f_1(\eta\bar{\eta})}{\partial \eta} > 0$ if $\hat{\eta} < \eta$, and $\frac{\partial f_1(\eta\bar{\eta})}{\partial \eta} < 0$ if $\hat{\eta} > \eta$. Hence, the second derivative is positive for low and negative for high values of $\hat{\eta}$. $E[\Pi_u(\tau, \eta)\bar{\eta}]$ is first convex, then concave in $\hat{\eta}$. Note that $\lim_{\hat{\eta} \to -\infty} E[\Pi_u(\tau, \eta)\bar{\eta}] = 0$: For a worker with very low expected ability, the probability of being more productive than the union wage is 0, independently of her training level. Furthermore, $\lim_{\hat{\eta} \to -\infty} E[\Pi_u(\tau, \eta)\bar{\eta}] = (1 - G(\Delta))\Delta$: For a worker with very high ability, the probability of being so productive that the union wage is not binding is 1, independently of her training level. Hence, when workers’ expected ability is very low, profits without training also exceed profits with training by $k$. Similarly, when workers’ expected ability is very high, profits without training also exceed profits with training by $k$. Consequently, there are two ability thresholds $\hat{\eta}_1$ and $\hat{\eta}_2$ such that expected profits with training exceed expected profits without training if $\hat{\eta}_1 < \hat{\eta} < \hat{\eta}_2$. See figure 4, panel A, for an illustration. Next, we argue that for high values of the fixed cost of training, $k$, no worker gets trained. An increase in $k$ leads to a parallel downward shift of profits with training (figure 4, panel B).
Figure 4: The impact of fixed cost of training on profits

There thus exists a $k^*$ such that profits with training are never greater than profits without training if $k > k^*$.

We next show that unionised firms offer is less training than the socially optimal level. At the socially optimal level, $c_0(\tau) = \frac{\partial E[y(\tau, \eta)]}{\partial \tau}$, while the training level unionised firms offer satisfies $c_0(\tau_u) = \frac{\partial E[\Pi_u(\tau_u, \eta)]}{\partial \tau}$. It is easy to see that $\frac{\partial E[y(\tau, \eta)]}{\partial \tau} > \frac{\partial E[\Pi_u(\tau_u, \eta)]}{\partial \tau}$. Hence, unionised firms offer a lower training level than the socially optimal level.

It remains to analyse wage determination in the first period. Wages in the first period are determined by the firm’s zero profit condition. Since union agreements do not affect profits and wage determination in non-unionised firms, non-unionised firms offer the same first period wage as in the absence of union agreements, (2). Hence,

$$W_{nu} = W = \int_{-\infty}^{\infty} y(0, \eta) dF_1(\eta | \eta) + (1 - G(\Delta)) \Delta.$$  

The wage offer of unionised firms can be similarly derived as

$$W_u = \begin{cases} 
\min \{ \bar{w}, \int_{-\infty}^{\infty} y(0, \eta) dF_1(\eta | \hat{\eta}) + E[\Pi_u(0, \eta) | \hat{\eta}] \} & \text{if } \hat{\eta} < \hat{\eta}_1 \text{ or } \hat{\eta} > \hat{\eta}_2 \text{ (no training)}. \\
\int_{-\infty}^{\infty} y(0, \eta) dF_1(\eta | \hat{\eta}) - k + E[\Pi_u(\tau_u, \eta) | \hat{\eta}] - c(\tau_u) & \text{if } \hat{\eta}_1 \leq \hat{\eta} \leq \hat{\eta}_2 \text{ (training)}. 
\end{cases}$$

(12)

It is now apparent that firms bear the training cost. As the unionised firm only trains if the profit with training, $-k + E[\Pi_u(\tau_u, \eta) | \hat{\eta}] - c(\tau_u)$, exceeds the profit without training, $E[\Pi_u(0, \eta) | \hat{\eta}]$, the worker’s training wage is higher than her first period wage would be without training.
C Worker sorting into the unionised sector

Non-unionised firms offer no training, while unionised firms offer \( \tau_u^* \). A worker chooses to work in the unionised sector if her utility from working in the unionised sector exceeds that from working in the non-unionised sector:

\[
W_u(\tau_u^*, \tilde{\eta}) + E[U_u(\tau_u^*, \eta)|\tilde{\eta}] \geq W_{nu}(0, \tilde{\eta}) + E[U_{nu}(0, \eta)|\tilde{\eta}],
\]

where \( W_j(\tau_j, \tilde{\eta}), j = u, nu \) denotes the worker’s first period wage in a unionised or non-unionised firm, and \( E[U_j(\tau_j^*, \eta)|\tilde{\eta}], j = u, nu \) denotes her second period utility when working in a unionised or non-unionised firm in the first period. We first derive the worker’s second period utility when working in a non-unionised firm in the first period, \( E[U_{nu}(0, \eta)|\tilde{\eta}] \). If the worker leaves her employer, she is paid a wage equal to her productivity, \( y(0, \eta) \). If she stays, her utility is equal to the wage the incumbent firm offers, \( y(0, \eta) - \Delta \), plus the draw of non-pecuniary job characteristics, \( \theta \). The worker stays if \( \theta > \Delta \). Hence, the worker’s expected utility in the second period equals

\[
E[U_{nu}(0, \eta)|\tilde{\eta}] = \int_{-\infty}^{\infty} y(0, \eta) dF_1(\eta|\tilde{\eta}) + \int_{\Delta}^{\infty} \theta - \Delta dG(\theta). \tag{13}
\]

Using that the worker’s first period wage equals \( W_{nu}(0, \eta) = \int_{-\infty}^{\infty} y(0, \eta) dF_1(\eta|\tilde{\eta}) + (1 - G(\Delta))\Delta \) (expression (2)), her utility from working in a non-unionised firm can be computed as

\[
W_{nu}(0, \tilde{\eta}) + E[U_{nu}(0, \eta)|\tilde{\eta}] = 2 \int_{-\infty}^{\infty} y(0, \eta) dF_1(\eta|\tilde{\eta}) + \int_{\Delta}^{\infty} \theta dG(\theta). \tag{14}
\]

Next, we derive the worker’s second period utility when working in a unionised firm in the first period, \( E[U_u(\tau_u^*, \eta)|\tilde{\eta}] \). Workers who turn out to be less productive than \( \eta_1 \) leave the unionised firm and are paid a wage equal to their productivity \( y \). Workers whose ability is revealed to be between \( \eta_1 \) and \( \eta_2 \) get \( w + \theta \) if they stay and \( y \) if they leave. The probability that they stay is \( 1 - G(y - \bar{w}) \). Finally, the utility of workers who turn out to be more able than \( \eta_2 \) is equal to \( y \) if they leave, and \( y - \Delta + \theta \) if they stay. The probability of staying is \( 1 - G(\Delta) \). Hence, \( E[U_u(\tau_u^*, \eta)|\tilde{\eta}] \) can be computed as

\[
E[U_u(\tau_u^*, \eta)|\tilde{\eta}] = \int_{-\infty}^{\eta_1} y dF_1(\eta|\tilde{\eta}) + \int_{\eta_1}^{\eta_2} (1 - G(y - \bar{w}))\bar{w} dF_1(\eta|\tilde{\eta}) + \int_{\eta_1}^{\eta_2} G(y - \bar{w})ydF_1(\eta|\tilde{\eta}) + \int_{\eta_1}^{\eta_2} \int_{y - \bar{w}}^{\bar{w}} \theta dG(\theta) dF_1(\eta|\tilde{\eta}) + \int_{\eta_2}^{\bar{w}} y dF_1(\eta|\tilde{\eta}) + (1 - F_1(\eta_2|\tilde{\eta})) \int_{\Delta}^{\infty} \theta - \Delta dG(\theta). \tag{15}
\]
Using expression (12) for the worker’s wage in the first period, her utility from working in a unionised firm can be computed as

\[
W_u(\tau_u^*, \eta) + E[U_u(\tau_u^*, \eta)|\tilde{\eta}] = \begin{cases} 
2 \int_{-\infty}^{\infty} y(0, \eta) dF_1(\eta|\tilde{\eta}) \\
+ \int_{\eta_1}^{\eta_2} \int_0^{\tilde{\eta}} \theta dG(\theta) dF_1(\eta|\tilde{\eta}) \\
+ (1 - F_1(\eta_2|\tilde{\eta})) \int_\Delta \theta dG(m) \\
& \text{if } \tau_u^* = 0, \\
2 \int_{-\infty}^{\infty} y(0, \eta) dF_1(\eta|\tilde{\eta}) - k - c() \\
+ \int_{\eta_1}^{\eta_2} \int_0^{\tilde{\eta}} \theta dG(\theta) dF_1(\eta|\tilde{\eta}) \\
+ (1 - F_1(\eta_2|\tilde{\eta})) \int_\Delta \theta dG(m) \\
& \text{if } \tau_u^* > 0. 
\end{cases}
\]

The sorting of workers into unionised firms depends on the training level unionised firms offer as well as on their expected ability. First, consider the impact of training on worker sorting. Recall that unionised firms choose training such that the marginal cost of training is equal to the marginal profit of training: \( c'(\cdot) = \frac{\partial E[\Pi_u(\tau_u, \eta)|\tilde{\eta}]}{\partial \tau_u} \). Also note that the worker’s utility from working in a unionised firm can be alternatively be written as \( W_u(\tau_u, \tilde{\eta}) + E[U_u(\tau_u, \eta)|\tilde{\eta}] = -k - c(\tau_u) + E[\Pi_u(\tau_u, \eta)|\tilde{\eta}] + E[U_u(\tau_u, \eta)|\tilde{\eta}] \). Hence, the training level that maximises the worker’s utility.

Next, consider the impact of ability on worker sorting. Suppose unionised and non-unionised firms offer the same amount of training. Figure 5 plots the difference between the utility from working in a unionised and non-unionised firm as a function of workers’ expected ability. First observe that from (16), the utility from working in a unionised firm converges to \( 2 \int_{-\infty}^{\infty} y(0, \eta) dF_1(\eta|\tilde{\eta}) \) as the worker’s
ability becomes very low. In contrast, from (14), the utility from working in a non-unionised firm converges to \( 2 \int_{-\infty}^{\infty} y(0, \eta) dF_1(\eta|\tilde{\eta}) + \int_{\Delta}^{\infty} \theta dG(\theta) \) as expected ability becomes low. Low ability workers are therefore better off in non-unionised firms. Second, observe that \( \int_{y-w}^{\eta} \theta dG(\theta) > \int_{\Delta}^{\infty} \theta dG(\theta) \): Workers whose ability turns out to be between \( \eta_1 \) and \( \eta_2 \) are better off due to unions. Workers with expected productivity around the union wage therefore prefer to work in the unionised sector. Finally, workers who turn out to be more able than \( \eta_2 \) are unaffected by the union wage. Hence, as the worker’s expected ability becomes very high, the difference between the utility from working in a unionised or non-unionised firm converges to 0. This implies that there exists an ability threshold such that workers with an expected ability above this threshold prefer to work in the unionised sector.

D Proof of proposition 3

Proposition 3 (i) In non-unionised firms, endogenous and exogenous movers are as productive as stayers and earn a higher wage than stayers. (ii) In unionised firms, endogenous and exogenous movers may be less productive and may earn a lower wage than stayers. (iii) The difference between the wage of endogenous movers and stayers is lower (in absolute terms) in unionised than in non-unionised firms.

(i) Stayers, exogenous and endogenous movers in non-unionised firms

Outside firms offer a wage equal to workers’ productivity, i.e. \( v = y(\tau, \eta) \), while incumbent firms offer a wage equal to workers’ marginal productivity minus a constant, \( w = y(\tau, \eta) - \Delta \). Hence, conditional on workers’ productivity, wages of exogenous and endogenous movers exceed wages of stayers by \( \Delta \). Furthermore, recall that the probability of staying is \( 1 - G(v - w) \). As \( v - w = \Delta \) for all workers, the staying probability does not depend on workers’ productivity. Consequently, stayers, exogenous and endogenous movers have the same productivity on average.

(i) Stayers, exogenous and endogenous movers in unionised firms

We first compare the average productivity, and then the average wage, of stayers, exogenous and endogenous movers. For a worker with expected ability \( \tilde{\eta} \) the probability of staying with the unionised training firm is

\[
Pr_u(\text{stay}|\tilde{\eta}) = \int_{\eta_1}^{\eta_2} (1 - G(y - \bar{w})) dF_1(\eta|\tilde{\eta}) + (1 - F_1(\eta_2|\tilde{\eta}))(1 - G(\Delta)),
\]

while the probability of moving can be computed as

\[
Pr_u(\text{move}|\tilde{\eta}) = F_1(\eta_1|\tilde{\eta}) + \int_{\eta_1}^{\eta_2} G(y - \bar{w}) dF_1(\eta|\tilde{\eta}) + (1 - F_1(\eta_2|\tilde{\eta})) G(\Delta).
\]
The average productivity of stayers, conditional on expected ability, equals

\[ E_u[y|\text{stay}, \eta] = \int_{\eta_1}^{\eta_2} (1 - G(y - \bar{w}))y dF_1(\eta|\eta_1) + (1 - G(\Delta)) \int_{\eta_1}^{\infty} y dF_1(\eta|\eta_1) \frac{Pr_u(\text{stay}|\eta)}{Pr_u(\text{move}|\eta)}, \tag{18} \]

while the average productivity of exogenous movers, conditional on expected ability, can be computed as

\[ E_u[y|\text{move}, \eta] = \int_{\eta_1}^{\eta_2} y dF_1(\eta|\eta_1) + \int_{\eta_1}^{\eta_2} G(y - \bar{w})y dF_1(\eta|\eta_1) + G(\Delta) \int_{\eta_2}^{\infty} y dF_1(\eta|\eta_1) \frac{Pr_u(\text{move}|\eta)}{Pr_u(\text{stay}|\eta)} \tag{19} \]

The average productivity of exogenous movers simply is

\[ E[y|\text{exogenous}] = \int_{-\infty}^{\eta_1} y dF_1(\eta|\eta_1). \]

Comparing the average productivity of stayers and endogenous movers, we get

\[
    E_u[y|\text{stay}, \eta] - E_u[y|\text{move}, \eta] = \]

\[
    \begin{align*}
    (a) & \quad \int_{\eta_1}^{\eta_2} (1 - G(y - \bar{w}))y dF_1(\eta|\eta_1) F_1(\eta_1|\eta) - \\
     & \quad \int_{\eta_1}^{\eta_2} y dF_1(\eta|\eta_1) \int_{\eta_1}^{\eta_2} 1 - G(y - \bar{w}) dF_1(\eta|\eta_1) \\
     & \quad + (1 - G(\Delta)) \int_{\eta_1}^{\eta_2} y dF_1(\eta|\eta_1) \int_{\eta_1}^{\eta_2} G(y - \bar{w}) dF_1(\eta|\eta_1) - \\
     & \quad \int_{\eta_1}^{\eta_2} G(y - \bar{w}) y dF_1(\eta|\eta_1) (1 - F_1(\eta_2|\eta)) \\
     (b) & \quad + (1 - G(\Delta)) \int_{\eta_2}^{\infty} y dF_1(\eta|\eta_1) F_1(\eta_1|\eta) - \int_{-\infty}^{\eta_1} y dF(\eta|\eta_1) (1 - F_1(\eta_2|\eta)) \\
     & \quad - G(\Delta) \int_{\eta_2}^{\infty} y dF_1(\eta|\eta_1) \int_{\eta_1}^{\eta_2} 1 - G(y - \bar{w}) dF_1(\eta|\eta_1) - \\
     & \quad \int_{\eta_1}^{\eta_2} (1 - G(y - \bar{w})) y dF_1(\eta|\eta_1) (1 - F_1(\eta_2|\eta)) \\
     & \quad - (\int_{\eta_1}^{\eta_2} G(y - \bar{w}) y dF_1(\eta|\eta_1) \int_{\eta_1}^{\eta_2} 1 - G(y - \bar{w}) dF_1(\eta|\eta_1) - \\
     & \quad \int_{\eta_1}^{\eta_2} (1 - G(y - \bar{w}) y dF_1(\eta|\eta_1) \int_{\eta_1}^{\eta_2} G(y - \bar{w}) dF_1(\eta|\eta_1)) / \}
     \\
     & \quad \{Pr_u(\text{stay})Pr_u(\text{move})\}. 
    \end{align*}
\]

Terms (a) and (b) are clearly positive. Term (d) is negative, but adding also term (c) yields a positive expression, provided that the probability of staying exceeds the probability of leaving. This
clearly holds in the data. Term (e) is negative. To see this, observe that
\[
\frac{\int_{\eta_1}^{\eta_2} G(y-w) y d F_1(\eta|\tilde{\eta})}{\int_{\eta_1}^{\eta_2} (1-G(y-w)) y d F_1(\eta|\tilde{\eta})} > \frac{\int_{\eta_1}^{\eta_2} (1-G(y-w)) y d F_1(\eta|\tilde{\eta})}{\int_{\eta_1}^{\eta_2} 1-G(y-w) d F_1(\eta|\tilde{\eta})}.
\]
For workers with ability between \(\eta_1 < \eta < \eta_2\) it is the more able workers who are more likely to leave. Hence, conditional on workers’ expected ability, it is ambiguous whether stayers are more or less able than movers. On the one hand, the least able workers are laid off (\(\eta < \eta_1\)). On the other hand, among workers who are paid the union wage (\(\eta_1 < \eta < \eta_2\)), more able workers are more likely to quit. By the same argument, it is ambiguous whether exogenous or endogenous movers are more productive.

Finally, note that as in non-unionised firms, exogenous and endogenous movers earn a higher wage than stayers conditional on productivity. Hence, a lower productivity of movers does not necessarily translate into a lower wage of movers. It is therefore ambiguous which group of workers earns the higher wage.

(iii) We have to show that the difference between the average wage of stayers and endogenous movers is greater than \(\Delta\). The average wage of stayers, conditional on expected ability, equals
\[
E_u[w|\text{stay}, \tilde{\eta}] = \frac{\int_{\eta_1}^{\eta_2} (1 - G(y - w)) y d F_1(\eta|\tilde{\eta}) + (1 - G(\Delta)) \int_{\eta_1}^{\eta_2} y - \Delta d F_1(\eta|\tilde{\eta})}{Pr_u(\text{stay}|\tilde{\eta})},
\]
while the average wage of movers is equal to the average productivity of movers, given by (19). Taking
the difference, we get

\[ \begin{align*}
E_u[w|\text{stay}, \hat{\eta}] - E_u[v|\text{move}, \hat{\eta}] &= \\
(a) &\int_{\eta_1}^{\eta_2} (1 - G(y - \overline{w})dF_1(\eta|\hat{\eta}))F_1(\eta_1|\hat{\eta}) - \\
&\int_{\eta_1}^{\eta_2} ydF_1(\eta|\hat{\eta}) \int_{\eta_1}^{\eta_2} 1 - G(y - \overline{w})dF_1(\eta|\hat{\eta}) \\
(b) &+ (1 - G(\Delta))(\int_{\eta_2}^{\infty} y - \Delta dF_1(\eta|\hat{\eta}) \int_{\eta_1}^{\eta_2} G(y - \overline{w})dF_1(\eta|\hat{\eta}) - \\
&\int_{\eta_1}^{\eta_2} G(y - \overline{w})ydF_1(\eta|\hat{\eta})(1 - F_1(\eta_2|\hat{\eta})) \\
(c) &+ (1 - G(\Delta))(\int_{\eta_2}^{\infty} y - \Delta dF_1(\eta|\hat{\eta})F_1(\eta_1|\hat{\eta}) - \int_{\eta_1}^{\eta_2} ydF(\eta|\hat{\eta})(1 - F_1(\eta_2|\hat{\eta})) \\
(d) &- G(\Delta)(\int_{\eta_2}^{\infty} ydF_1(\eta|\hat{\eta}) \int_{\eta_1}^{\eta_2} 1 - G(y - \overline{w})dF_1(\eta|\hat{\eta}) - \\
&\int_{\eta_1}^{\eta_2} (1 - G(y - \overline{w})dF_1(\eta|\hat{\eta})(1 - F_1(\eta_2|\hat{\eta}))) \\
(e) &+ \int_{\eta_2}^{\eta_2} 1 - G(y - \overline{w})dF_1(\eta|\hat{\eta}) \int_{\eta_1}^{\eta_2} G(y - \overline{w})(y - \overline{w})dF_1(\eta|\hat{\eta}) \{/ \\
&\{Pr_u(\text{stay})Pr_u(\text{move})\}. \\
\end{align*} \]

Terms \((a)\) and \((b)\) are clearly positive. Term \((d)\) is negative, but adding also term \((c)\) yields a positive expression, provided that the probability of staying exceeds the probability of leaving. Term \((e)\) and \((f)\) are negative, but clearly smaller than \(-\Delta\).

**E Data**

**Worker data: Description, Sample Selection and definition of variables**

The first data set we use is of administrative nature, drawn from the German Social Security Records, the so-called IAB-Beschäftigtenstichprobe. It constitutes a one percent sample of the German labour force subject to social security contributions for the period between 1975 to 1995. The data contains an unusual set of background information for the individual, including age and education. It allows to construct the complete work history, including time spent in registered unemployment, from labour market entry onwards up to 20 years in the labour force. Furthermore, due to the administrative
nature of this data set, wages and employment spells are very accurate, and, other than in many surveys, wages are uniquely assigned to particular jobs.

Firms report to the Federal Employment Office any beginning and end of an employment relationship when these occur, as well as information on ongoing relationships at December 31st every year. The wage reported in our data is the average daily wage computed for each spell. We deflate wages using the Consumer Price Index, with 1995 as the basis year. Actual experience is measured as weeks (divided by 52) spent in full-time employment from the 1st of January following the year the apprenticeship ended. Part time employment, time as apprentice, time spent unemployed as well as time spent out of the labour force is not counted. We assume that the wage data refers to experience accumulated at the start of each employment spell.

A disadvantage of the data set is that it does not cover the entire German labour force, but only employees within the social security net. This excludes, among others, the self-employed and civil servants. As many administrative data sets, our data is right-censored at the highest level of earnings that are subject to social security contributions. In our sample top-coding is not a serious problem: Less than 0.5 % of all wage observations are top-coded.

We select from this data base a sample of male individuals who all went through apprenticeship training. We restrict analysis to former West Germany only. We define an apprentice as a worker who is reported as a trainee in one company for at least 700 days, no matter whether the education variable classifies him as a worker with or without a finished apprenticeship. In addition, we consider a worker with a work placement as an apprentice if the work placement lasted at least 1 year (position variable) and he was employed as a worker with a completed apprenticeship in at least half of his jobs (education variable).

Sample selection  An apprentice is included in the sample if

- he completed the apprenticeship after 1979;
- he has never worked in East Germany;
- he completed only one apprenticeship;
- if he was not older than 15 in 1975 and at most 19 at labour market entry and had completed no A-levels (Abitur) at labour market entry;
• if he was not older than 15 in 1975 and at most 21 at labour market entry and had completed A-levels (Abitur) at labour market entry.

We end up with a sample of 24194 individuals.

Our data set allows us to precisely determine whether a worker has left the apprenticeship firm at the end of the training period. However, for apprentices who remain with the same firm after training a separation between training spell and employment spell in the year when training is concluded is not possible. When we compare entry wages of stayers and movers, we therefore assign to both stayers and movers the first wage spell in the year following the year the training period ended.

We define two types of exogenous movers. First, we follow Acemoglu and Pischke (1998) and use draft into the military service or social service as a measure for exogenous separation from the training firm\textsuperscript{16}. We refer to this variable as service. Second, our data set contains also information on the size of the firm, as of June of the respective calendar year\textsuperscript{17}. From this information we construct a measure for firm closure by computing the year in which establishment size drops to zero.

Military and civil service last between 10 and 30 months, depending on the type of service, and the year the service started. For most of the workers in our data set, service lasted between 10 and 18 months. Our data does not contain direct information on whether workers are doing military or civil service. We approximate service by the time workers spend neither in registered employment nor in registered unemployment.\textsuperscript{18}

We use three alternative definitions. According to our first definition workers are classified as being in military or civil service if they are registered neither as employed nor as unemployed for at least 10 consecutive months and not longer than 18 months. Some workers in military service are still registered with their previous employer. These workers typically return to the previous employer after the military service. For these workers, firms report the working relationship as interrupted. We also classify these workers as in military service if the time between the interruption spell and the beginning of the new job is at least 10 months and at most 18 months. Workers are classified as not doing their military or civil service if they have no non-employment gap longer than 10 months and no interruption spell longer than 10 months. If workers have more than one non-employment gap or

\textsuperscript{16}See Acemoglu and Pischke (1998) for details on compulsory service schemes in Germany.
\textsuperscript{17}This information has been created by aggregating up the number of individuals for each firm identifier by year from the population of individuals in the 100% sample.
\textsuperscript{18}Workers who have finished apprenticeship training are entitled to unemployment benefits if they register as unemployed. Hence, it is likely that unemployed workers are also registered as unemployed.
Table 8: Military and Social Service: 3 Definitions

<table>
<thead>
<tr>
<th></th>
<th>Definition 1</th>
<th>Definition 2</th>
<th>Definition 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>total workers</td>
<td>24194</td>
<td>24194</td>
<td>24194</td>
</tr>
<tr>
<td>military service</td>
<td>6611</td>
<td>6882</td>
<td>8074</td>
</tr>
<tr>
<td>no military service</td>
<td>13516</td>
<td>13516</td>
<td>13430</td>
</tr>
<tr>
<td>problematic cases</td>
<td>4067</td>
<td>3796</td>
<td>2690</td>
</tr>
<tr>
<td>∑</td>
<td>24194</td>
<td>24194</td>
<td>24194</td>
</tr>
<tr>
<td>leaves due to service</td>
<td>2459</td>
<td>2510</td>
<td>2941</td>
</tr>
<tr>
<td>does not leave due to service</td>
<td>4152</td>
<td>4372</td>
<td>5113</td>
</tr>
<tr>
<td>∑</td>
<td>6611</td>
<td>6882</td>
<td>8074</td>
</tr>
<tr>
<td>problematic cases</td>
<td>1998</td>
<td>1923</td>
<td>1482</td>
</tr>
<tr>
<td>goes back to appr. firm</td>
<td>778</td>
<td>805</td>
<td>895</td>
</tr>
<tr>
<td>goes to new firm</td>
<td>1681</td>
<td>1705</td>
<td>2046</td>
</tr>
<tr>
<td>∑</td>
<td>2459</td>
<td>2510</td>
<td>2941</td>
</tr>
</tbody>
</table>

interruption spell longer than 10 months, they are classified as ”problematic”. We also consider a worker as problematic if both the interruption spell and the non-employment gap identifies the worker as in military service.

Our second definition follows definition 1, but workers who are classified as in service according to both the interruption spell and the non-employment gap are classified as in service, and not as problematic.

Definition 3 follows definition 2, but extends the duration of the non-employment gap and the interruption spell to 30 months.

Table 8 shows the proportion of workers in service for the three definitions. According to definition 1, 6611 out of 24194 workers do their military or civil service during the observation period. Of those, 2449 leave the apprenticeship firm due to service. Of those, 778 return to the apprenticeship firm. 4067 workers are classified as problematic. However, for our analysis it is only important whether the worker is classified as problematic in the year he finished his apprenticeship. This is the case for 1998 workers. A similar decomposition for definition 2 and 3.

To construct separations due to plant closure, we use information on the size of the firm as of June 1st each year. We classify a worker as an exogenous mover due to plant closure if the size of the training
firm drops to zero in the year the worker finishes the apprenticeship training period. Furthermore, if the worker completed the training period after June 1st, he is classified as an exogenous mover if the size of the training firm drops to zero in the next year.

**Firm data**

Our firm data is based on a yearly panel of establishments, the so-called IAB-Betriebspanel. The base population are all firms with at least one employee who pays social security contributions. In 1993 the panel started with about 4,000 establishments; in 2000 around 13,000 firms participated in the survey (see Kölling (2000)). Large firms are over-sampled. We restrict our analysis to West German firms in the private sector only.

Information on firms is collected through person-to-person interviews with the firm management. In addition to the union status, the data contains a large array of background characteristics, including firm size, industry, investment, revenue, etc. A firm is considered as "unionised" if it belongs to an employer association and is bound to the industry agreement that is relevant for their particular firm (Branchentarifvertrag), or if it engages in bilateral negotiations with the union (Haustarifvertrag).19

We are able to match for each firm and year average worker characteristics, which allows us to compute the fraction of apprentices in the firm. Average worker characteristics are computed from the social security records for each firm in our establishment sample as of June 1st each year. This data is available from 1996-1999. We therefore use only the data from 1996-1999 for the empirical analysis. We end up with an unbalanced panel of 4717 firms and 13453 observations.

Table 9 lists and defines the variables of the firm data used in the empirical analysis.

**Matched data**

It is in principle possible to match information on workers to the firm data. As our worker data, the information on workers comes from social security records. In addition to average worker characteristics for each firm for the years 1996-1999, we have access to the following two 'matched' data sets.

Mover-stayer wage differential First, for the comparison of the mover-stayer wage differential of

19 See appendix E for a brief description of the German collective bargaining system. In our sample, the proportion of workers who are employed in firms that engage in firm-level bargaining is 9% (IAB-Betriebspanel, 1996-1999, worker weights).
Table 9: Variable definitions: Firm data

<table>
<thead>
<tr>
<th>variable</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>union status</td>
<td>1 if firm is bound to industry or firm level agreement, 0 otherwise</td>
</tr>
<tr>
<td>firm size</td>
<td>total number of employees in the firm</td>
</tr>
<tr>
<td>industry</td>
<td>10 industry dummies: Energy/mining/water industry; chemical industry; metal industry/machines; electro-technical industry/automobiles/optical industry; wood/printing/paper; construction/capentry; retail/wholesale; traffic/news; credit/insurance; other services agriculture, charities/private households and public sector dropped</td>
</tr>
<tr>
<td>revenue/worker</td>
<td>total turnover in the firm in the previous year divided by number of employees</td>
</tr>
<tr>
<td>investment/worker</td>
<td>total sum of investments in the previous year divided by number of employees</td>
</tr>
<tr>
<td>age of firm</td>
<td>distinguishes between 5 years and younger, 6-15 years, 16-30 years, older than 30 years</td>
</tr>
<tr>
<td>evaluation of profit</td>
<td>firm’s current evaluation of profit; from 1 (very good) to 5 (very bad)</td>
</tr>
<tr>
<td>proportion apprentices</td>
<td>number of apprentices divided by number of employees</td>
</tr>
<tr>
<td>firm trains</td>
<td>1 if firm employs at least one apprentice, 0 otherwise</td>
</tr>
</tbody>
</table>

workers trained in unionised and non-unionised firms we select all workers who complete an apprenticeship in one of the firms in the establishment panel. A worker is included in the sample if the apprenticeship ended between 1995 and 1997 and was done in the private sector in a West German firm. For each worker we observe the first post-apprenticeship wage spell, whether he stayed with or left the training firm, as well as the same background characteristics as in the worker data. We match to this data the union status of the training firm from the establishment panel.

Comparison of workers in unionised and non-unionised firms Second, for the year 2000 we have access to the social security records of all workers who worked in one of the firms in the establishment panel on June 1st. We use this data to compare workers in unionised and non-unionised firms in terms of observable characteristics.
E Institutional background

**The German apprenticeship system**  The German apprenticeship system is a vocational training programme that combines on-the-job training, provided by firms, with off-the-job schooling, provided by the state. It takes between 2 to 3 1/2 years, depending on the apprenticeship occupation and the apprentice’s educational background. Apprentices typically attend a vocational school once or twice a week. During the rest of the time apprentices are on on-the-job training schemes in the firm. At the end of the apprenticeship, trainees take a centralised exam, and receive a certificate that is widely accepted in the economy. During the apprenticeship, apprentices receive a wage that is about 30 % of the wage of a skilled worker. Apprenticeship training is costly for firms. von Bardeleben et al. (1995) estimates that the average net costs for a firm per year and apprentice are 10.657 DM. For large firms (>500 employees), net costs increase to 15.144 DM.

**Collective bargaining in Germany**  Outcomes of wage negotiations in Germany are laid down in specific contracts. These contracts are agreed on an annual basis in negotiations between employers, usually organised in industry wide and region wide employer associations, and unions. Different wages are agreed for workers with different education- and training background. Although in principle union agreements cover only employees who are union members, in practice the union status of the employee is irrelevant.

Not all firms are bound to these agreements, however. Firms have three options. If they belong to an employer association, they are bound to the industry union agreement that is relevant for their particular firm. They can also engage in bilateral negotiations with the union. In both cases, their status is formalised, and negotiated union wages act as minimum wages for all workers in the respective firm, independent of their union status. Firms are allowed to pay wages above the wage agreement, and many firms do so. The third option for the firm is to agree individual contracts with their employees. Only in this case is the firm allowed to pay wages below those laid down in union agreements. See Lehmann (2002) for more details.
References


