Universität St.Gallen

Welfare Effects of Penalty Charges to Firms that Do Not Offer Apprenticeship Training Positions

Kai-Joseph Fleischhauer

July 2006 Discussion Paper no. 2006-18

Department of Economics University of St. Gallen

Editor:	Prof. Jörg Baumberger University of St. Gallen Department of Economics Bodanstr. 1 CH-9000 St. Gallen Phone +41 71 224 22 41 Fax +41 71 224 28 85 Email joerg.baumberger@unisg.ch
Publisher:	Department of Economics University of St. Gallen Bodanstrasse 8 CH-9000 St. Gallen Phone +41 71 224 23 25 Fax +41 71 224 22 98
Electronic Publication:	http://www.vwa.unisg.ch

Welfare Effects of Penalty Charges to Firms that Do Not Offer Apprenticeship Training Positions

Kai-Joseph Fleischhauer

Author's address:

Kai-Joseph Fleischhauer, lic.oec.HSG
Institute of Public Finance and Fiscal Law (IFF-HSG)
Varnbüelstrasse 19
CH-9000 St. Gallen
Tel. +41 71 224 30 86
Fax +41 71 224 26 70
Email kai-joseph.fleischhauer@unisg.ch
Website http://www.iff.unisg.ch

Abstract

In Germany, there is currently a discussion about the implementation of penalty charges if firms refuse to offer apprenticeship training positions to school graduates. This paper aims at analyzing the policy instrument of penalty charges by a theoretical model that systematically compares its costs and benefits. Building on recent training literature, a two-period partial-equilibrium model is designed that allows for worker heterogeneity in ability and covers special features of the German apprenticeship system.

With respect to overall welfare, the implementation of penalty charges solves a trade-off. On the one hand, penalty charges increase the number of apprenticeship training positions and thus the fraction of trained workers in the workforce. On the other hand, some firms will leave the market to avoid the financial burden, which generates unemployment among workers with low ability. Altogether, we demonstrate that optimal penalty charges increase the overall welfare compared to the laissez-faire equilibrium if the productivity-enhancement of apprenticeship training exceeds some lower bound.

Keywords

Human Capital Formation, Apprenticeship Training, Inefficient Training Decision of Firms, Penalty Charges

JEL Classification

I28, J24, J31

1 Introduction

In Germany, there is currently a discussion about the implementation of penalty charges if firms refuse to offer apprenticeship training positions to school graduates. The German apprenticeship system provides basic skills to a large share of the workforce and is thus considered an exemplary model for vocational education (Harhoff and Kane (1997)).[1] But in the last several years, there are more and more youths unable to find apprenticeship training positions, who therefore remain unskilled after graduation.[2] Many firms decide not to offer training places because trained workers can freely choose to change their employer upon completion of their apprenticeship.[3] These firms perceive the danger of bearing the costs of training without getting any return. In consequence, the famous German apprenticeship system is challenged by unemployment among young workers and an increasing number of unskilled workers.

In order to approach the problem of missing apprenticeship training positions, a compulsory training quota for each firm is proposed depending on the number of full-time workers. If a firm does not satisfy this quota, it has to pay a previously defined penalty for each training place lacking to meet the quota. Unfortunately, a theoretical analysis of penalty charges is still lacking. Our paper aims at closing this gap by presenting a two-period partial-equilibrium model that systematically compares the costs and benefits of penalty charges. In the literature, there are two theoretical explanations for firms providing general training at the intensive margin. First, the basic approach of Becker (1964) concentrates on the firms' current incentives during the training period.[4] If the training wage is low enough - i.e. the gap between apprentice output and training wage is big enough - to compensate for the costs generated by training activities, then firms decide to provide general training to their workers. Second, Acemoglu and Pischke (1999b) point out the compression of the wage structure in imperfect labor markets.[5] If workers stay with the training firm after the apprenticeship has been completed, then the firms' incentives are with respect to future returns because they manage to skim a rent from the increased output of trained workers.

According to Acemoglu and Pischke, the training decision of firms bears two different kinds of inefficiencies. First, firms take into account only their own gains from higher productivity and neglect the gains for the workers through higher wages in the second period. Second, firms further underinvest in the general human capital of their workers if there is a positive probability of separation after the training period because firms do not take into account higher wages of workers and higher profits of potential employers in the future (Acemoglu (1997)). One policy measure to reduce these inefficiencies in the provision of general training at the intensive margin are firing costs. Because the probability of separation is decreased, firms invest more in the human capital of their workers.[6] However, the number of separations

[1] In Germany, two thirds of the age-group from 15 to 24 are provided with vocational training through the apprenticeship system (BMBF (2004)).

[2] From 1991 to 2003, the number of new apprenticeship contracts has decreased from 571,206 to 564,493. Hence, the number of registered apprenticeships has decreased from 1,629,312 to 1,581,629 (BMBF (2004)).

[3] In 2002, only 31.3% of all firms were providing apprenticeship training positions (BMBF (2004)).

[4] Cf. also Becker (1962).

[5] Cf. also Acemoglu (1997), Acemoglu and Pischke (1998a), Acemoglu and Pischke (1998b) and Acemoglu and Pischke (1999a).

[6] The empirical evidence shows a positive relationship between training and job tenure (Lynch (1991) and Loewenstein and Spletzer (1999)).

is also decreased and could be inefficiently low. The total effect of firing costs on unemployment is ambiguous because fewer separations lead to lower unemployment but some firms decide to leave the market and thus it becomes harder to find a job for unemployed workers (Belot, Boone, and Ours (2002)).[7]

In this paper, our formal analysis of penalty charges is based on the recent training literature with oligopsonistic labor markets, but is extended in two important ways. First, our model allows for worker heterogeneity in ability and manages to explain the extensive training decision of firms depending on the individual ability of workers. Only those workers above some critical level of individual ability are offered an apprenticeship training position while workers with low ability remain unskilled. Second, we bring together the two theoretical explanations of firm-sponsored general training in the context of the institutional setting of the German apprenticeship system which is characterized by a fixed training wage and a prescribed length of the apprenticeship. Hence, in contrast to Malcomson, Maw, and McCormick (2003), who concentrate on the regulation of apprenticeship contracts, these two parameters are not determined endogenously. We will demonstrate that the firms' incentives to provide apprenticeship training are with respect to both current and future incentives. In other words, firms may provide general training both because they are looking forward to future returns and because they currently benefit from doing so. Consequently, inefficiencies in the training decision result from the interaction of firms and workers during the training period as well as during the working period. Beyond Acemoglu and Pischke (1999b), our analysis indicates a third kind of inefficiency generated by the fixed training wage during the apprenticeship. This training wage determines the firms' costs of apprenticeship training and thus splits the workforce into apprentices and unskilled workers.

With respect to overall welfare, the implementation of penalty charges solves a trade-off. On the one hand, penalty charges increase the number of apprenticeship training positions and thus the fraction of trained workers in the workforce. On the other hand, some firms will leave the market to avoid the financial burden, which generates unemployment among workers with low ability. By also considering administration costs of implementation, our formal analysis demonstrates that the optimal policy depends on the productivity-enhancement of training. If the productivity-enhancement of apprenticeship training is low, it is optimal to reject the implementation of penalty charges and the economy will attain the laissez-faire equilibrium. However, optimal penalty charges increase the overall welfare if the enhancement in productivity exceeds some lower bound. Quite intuitively, optimal penalty charges are decreasing in the level of administration costs. Even if administration costs are very low, optimal penalty charges are bounded above by some critical value that maximizes the number of apprenticeship training positions at the cost of suppressed regular work. With welfare maximizing penalty charges, optimal overall welfare depends positively on the productivity-enhancement of apprenticeship training and negatively on the exogenous probability of separation after the apprenticeship has been completed.

The paper proceeds as follows: the next section discusses human capital theory and its two approaches to firm-sponsored human capital formation. After this, the institutional setting of the German apprenticeship system is illustrated. In sections 4 of this paper, our partial-equilibrium model is developed and

[7] The empirical evidence is mixed. Scarpetta (1996) and Elmeskov, Marint, and Scarpetta (1998) indeed find a negative correlation between firing costs and unemployment whereas Nickell (1998) does not.

the equilibrium without penalty charges is presented. In section 5, the implementation of penalty charges is analyzed and optimal penalty charges are derived. Section 6 concludes.

2 The Theory of Firm-Sponsored Human Capital Formation

"Human capital" can be defined as knowledge, skills, attitudes, aptitudes, and other acquired traits contributing to production (Goode (1959)). According to Blundell, Dearden, Meghir, and Sianesi (1999), there are two main components of human capital with strong complementarity: early ability (whether acquired or innate) and skills acquired through formal education or training on the job. An extensive review of the theory of human capital is given by Cahuc and Zylberberg (2004).

2.1 The Basic Approach with Perfect Labor Markets

In his original model, Becker (1964) distinguishes between general and specific human capital. General human capital is defined to be not only useful with the current employer but also with other potential employers. In contrast, specific human capital increases the productivity of the worker only in his current job.[8] Hence, in competitive labor markets, where workers receive wages equal to their marginal product, firms cannot recoup investments in general skills, so that they refuse to pay for general training. However, workers themselves have the right incentives to invest in general human capital because they are the sole beneficiaries of their improved productivity (either with their current or with future employers).[9] Furthermore, workers can finance such investments quite easily by accepting a wage below their productivity during the period of training (the wage may even be negative) (Becker (1962)). For example, this argument can be applied to apprenticeship systems in earlier centuries, where apprentices often paid fees or worked for very low wages until they mastered a certain grade (Hamilton (1996)).[10] If workers are not credit constrained, they efficiently invest in the accumulation of general human capital. On the other hand, Becker (1964) argues that training in specific skills is quite different because workers do not benefit from higher productivity after changing their jobs. Therefore, firms can recoup investments in specific skills and are thus willing to share some of the costs of these investments. By also sharing the returns, the accumulation of specific human capital leads to lower fluctuations because both firms and workers benefit from keeping their contractual partner.

The empirical evidence of the model by Becker (1964) is mixed. On the one hand, it is supported by the empirical analysis of Veum (1999). By using data from the National Longitudinal Survey of Youth (NLSY), he finds that firm-sponsored training is indeed negatively related to starting wages, but positively related to wage growth. On the other hand, many analyses question the validity of Becker's explanation by showing that there are investments in general human capital which are financed by the employer. For example, by further analyzing data from the NLSY, Loewenstein and Spletzer (1999) find

[8] Parsons (1974) notes that this firm-specific human capital is analytically equivalent to transfer costs for adjusting a worker to other firms.

[9] Already Eckaus (1963) criticizes that these results strictly depend on the assumption of perfect labor markets.

[10] The relevance of this argument for the German apprenticeship system is explained in section 3.2.

that the bigger part of firm-sponsored training is general. Other empirical studies also show that firms bear substantial net costs in providing general training to their apprentices. For example, Ryan (1980) examines welder apprentices in the US and Jones (1986) analyzes apprentices in British manufacturing.[11] A number of studies also investigate whether workers taking part in general training programs pay for the costs by accepting lower wages. The majority of these studies do not find evidence of lower wages, at least not in an appropriate amount to fully compensate firms for the costs. An overview about these results is given by Bishop (1997). Hence, in contradiction to the theory of Becker (1964), there is at least some empirical evidence of firm-sponsored investments in the general human capital of their employees.

2.2 Imperfect Labor Markets

In order to give a theoretical explanation for the empirical findings cited above, Acemoglu and Pischke (1998a) develop a model with two periods, a training period where workers have identical productivity zero and may receive an amount of general training t at costs $c(t)$, and a second period where workers have individual productivity $f(t)$ and earn a wage $w(t)$. If the labor market is competitive and workers are not credit constrained, then the results of Becker (1964) hold: firms do not invest in general training and workers invest efficiently by equating marginal returns and marginal costs of their investment: $f'(t^*) = c'(t^*)$. However, if labor market are not competitive or there are other labor market frictions which generate wage compression, the worker's wage is below his marginal product (Masters (1998)). Hence, firms manage to skim labor market rents depending on the amount of training. Due to the compression of the wage structure general skills are turned into de facto specific skills. Formally, Acemoglu and Pischke (1999b) express this by assuming $f(t) = w(t) + \triangle(t)$. Hence, the wage function is increasing in the level of training less steeply than productivity (i.e. the wage structure is compressed), so that the firm's profit, equal to the positive gap $\triangle(t)$ between productivity and wage, has a first derivative greater than zero. As a consequence, firms prefer more skilled workers to less skilled ones and invest in general training until the desired level of training satisfies $\triangle'\left(t^f\right) = c'\left(t^f\right)$ (cf. figure 1).

There are several possible sources of labor market imperfections. One possible source of wage compression is the presence of transaction costs, for example due to matching and search frictions. The costs of finding new contractual partners create a match-specific surplus that has to be shared by bargaining. The bargaining process in this oligopsonistic labor market generates profits equal to the Nash bargaining solution $\triangle(t) = (1-\beta) f(t)$ (Acemoglu (1997)). Furthermore, wage compression may arise due to the interaction of general and specific skills. If general and specific skills are complements in the production of output, the presence of specific skills increases the productivity of general human capital. On the other hand, the value of firm-specific skills increases when general skills are acquired (Acemoglu and Pischke (1999b)). Kessler and Luelfesmann (2002) as well as Balmaceda (2001) extend this idea by constructing a model with general and specific skills that constitute strategic complements although returns and costs are technologically disconnected. They find that - even in frictionless labor markets - there is firm-sponsored general training because the hold-up problem of investment in general skills is reduced.

[11] The costs of apprenticeship training in Germany are discussed in section 3.3.

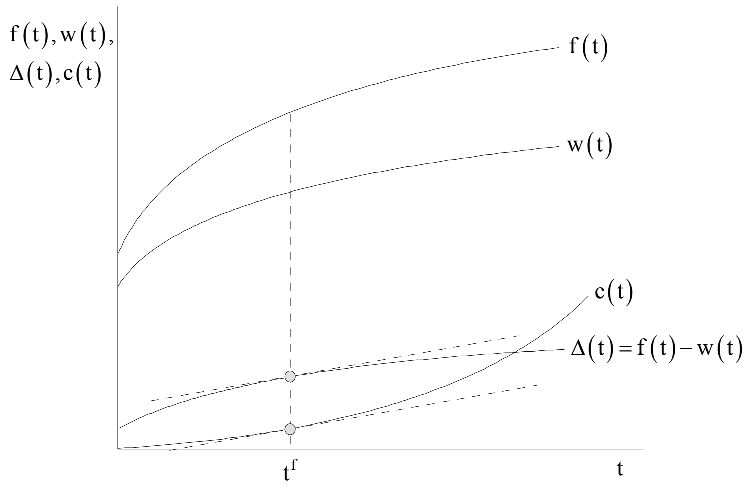

Figure 1: Training with Compressed Wage Structure

According to Bougheas and Georgellis (2004), this interaction of general and specific skills is the reason for German firms to offer apprenticeship training positions although training is largely general.

A third source of wage compression is the presence of asymmetric information between the current employer and other potential employers. There are two types of asymmetric information. The first concerns the amount of training the worker has received and is analyzed by Chang and Wang (1996). If potential employers cannot observe the correct productivity and thus pay a wage below the marginal product, the wage structure is compressed. With respect to the German apprenticeship system, this explanation for firm-sposored general training seems less important because the apprenticeship follows a prescribed curriculum.[12] A further possible asymmetry between current and potential employers is about innate ability of the worker (hidden knowledge), i.e. the employer learns about the ability of the worker by providing general training (Acemoglu and Pischke (1998b)). A fourth reason for wage compression is the presence of asymmetric information between worker and current employer concerning the worker's effort (hidden action). Hence, wages must satisfy the incentice compatibility constraints which leads to a compressed wage structure (Acemoglu and Pischke (1999b)). In a similar model, Loewenstein and Spletzer (1998) demonstrate that efficiency wages (that are paid to reduce fluctuations) can also induce firms to pay for general training. Many authors have investigated similar sources of firm-sponsored general training. For example, Bishop (1997) and Lazear (2003) point out that the firm-specific mixture of general skills makes the labor market non-competitive. Wage compression can also be generated by labor market institutions, for example minimum wages (Acemoglu and Pischke (1999a)) and worker unionization (Freeman and Medoff (1984)).

Regarding the empirical evidence, Loewenstein and Spletzer (1998) find that general training raises future wages more for workers changing their job than for workers remaining with the employer initially providing the training. This result is consistent with workers and employers sharing the returns to

[12] Cf. the description of the German apprenticeship system in section 3.2.

general training. Furthermore, Brunello (2002) suggests that wage compression and the amount of general training show a positive and significant correlation.

2.3 Regulation of Apprenticeship Contracts

In their recent approach to firm-sponsored human capital formation, Malcomson, Maw, and McCormick (2003) focus on the regulation of apprenticeship contracts. The apprenticeship contract is assumed to specify the length of the apprenticeship and the training wage. Because workers earn a wage equal to their marginal product after having completed the apprenticeship, the training wage must be sufficiently low in order to initiate firms to provide apprenticeship training. The optimal length of the apprenticeship solves a trade-off between between current returns for the training firm during the apprenticeship and future returns for all potential employers after the apprenticeship. It determines the amount of training provided for each apprentice. Depending on this optimal amount of training, the number of training firms is determined by fixed costs of training that are different for each firm. Malcomson, Maw, and McCormick (2003) show that regulation to increase the length of the apprenticeship, combined with a subsidy for each completed apprenticeship if the efficiency loss from distortionary taxation to finance the subsidy is sufficiently low, can increase the number of apprentices and the amount of training per apprentice. In our model, we will concentrate on the training decision at the extensive margin in order to analyze inefficiencies in the number of apprenticeship training positions.

3 The German Apprenticeship System

3.1 Historical Relevance of Apprenticeship Training

Historically, there have been several characteristics of apprenticeship training. First, the length of the apprenticeship was specified contractually in advance and independent of individual ability. For example, this applied to the métier in France, arte in Italy, craft guild in England, and Zunft or Innung in Germany. Second, apprenticeship training has been intensely regulated, for example by guilds in medieval times. This regulation typically implied a minimum duration of apprenticeships and the control of adequacy of training. The craft guilds of the middle ages had supervisory functions that included the right of search to insure that good materials and appropriate processes of manufacture were employed. In Germany, a range of institutions funded collectively by firms controlled the working of the apprenticeship system (Pirenne (1936)).

3.2 Institutional Setting of the German Apprenticeship System

The educational system of Germany is one of the most segregated among industrialized countries. There are four types of German secondary schools: lower (Hauptschule), middle (Realschule), upper (Gymnasium), and mixed (Gesamtschule). Upon their conclusion, all of these school tracks require the successful completion of exams which indicate whether students are qualified to enter into an apprenticeship, other vocational training, or the university (Cooke (2003)).

Apprenticeship training can be undertaken in a variety of skilled blue or white collar positions. It combines part-time schooling with a work-based component (the so-called "dual system") and is largely general. Firms offering apprenticeship training positions have to follow a prescribed curriculum and apprentices take a rigorous exam at the end of the apprenticeship. Industry or craft chambers certify whether firms fulfill the requirements to train apprentices adequately, while worker councils in the firms monitor the training. During the apprenticeship, workers receive a low training wage independent of their individual productivity. This training wage is set by negotiation of the collective bargaining parties. After having passed the exam, apprentices receive a formal skill certificate that is accepted nationwide (Bougheas and Georgellis (2004)).[13]

3.3 Costs of Apprenticeship Training in Germany

In 1991, the Bundesanstalt für Berufsbildung investigated training firms with respect to their accounting costs and apprentice productivity in order to assess the net costs of training which are defined as the difference between gross costs and apprentice output. The results of this study are described in Von-Bardeleben, Beicht, and Fehér (1995). The first step is to calculate gross costs as the sum of payroll costs, training personnel, material, equipment, and structures used in the training as well as direct costs of any external training. However, in many firms trainers are not engaged in training full-time but also work in productive activities. The German study for 1991 takes two approaches to this problem. The first is to prorate time spent on training by part-time personnel (A). The second is to exclude the costs of part-time trainers completely from the calculation of costs (B). The latter approach serves as a lower bound for the training costs born by the firms.

In a second step, the output of apprentices is estimated. A measure of output is designed by multiplying time spent in productive activities by payroll costs of a skilled worker and the relative efficiency of apprentices (C).[14] However, this calculation implicitly assumes that wages of skilled workers are set competitively and reflect marginal product. If the labor markets are not perfect, the marginal product may exceed wages so that the output of apprentices is underestimated. For this reason, an alternative measure of apprentice output is constructed by assuming imperfect labor markets with a markdown of 50% ($2C$).[15] Table 1 illustrates the role of these assumptions using data from VonBardeleben, Beicht, and Fehér (1995) for Germany.

It is evident that at least large firms bear significant costs in providing general training to their apprentices. As a consequence, many firms do not offer apprenticeship training positions in order to avoid this financial burden. However, at least for small and middle-sized firms, there may be net benefits from apprenticeship training depending on the method of calculating training costs and apprentice output.

[13] The German system of apprenticeship training is described e.g. by Soskice (1994) and Harhoff and Kane (1995).

[14] Note that the fraction of time spent in productive activities corresponds to the parameter χ in our formal analysis (cf. section 4.1).

[15] Note that this markdown of 50% corresponds to a worker's bargaining share of $\beta = \frac{1}{2}$ in our formal analysis (cf. section 4.1).

	All firms	Firm size (number of employees)			
		0 - 9	10 - 49	50 - 499	500 +
1) Costs (A)	29,573	27,473	28,176	30,344	35,692
2) Costs (B)	18,051	13,867	15,074	20,283	28,197
Perfect Markets					
Apprentice Output (C)	11,711	12,221	11,465	12,099	10,311
1) Net costs ($A - C$)	17,862	15,252	16,711	18,245	25,381
2) Net costs ($B - C$)	6,340	1,646	3,609	8,184	17,886
Imperfect markets					
Apprentice Output ($2C$)	23,422	24,442	22,930	24,198	20,622
1) Net costs ($A - 2C$)	6,151	3,031	5,246	6,146	15,070
2) Net costs ($B - 2C$)	-5,371	-10,575	-7,856	-3,915	7,575

Table 1: The Costs of Apprenticeship Training in Germany

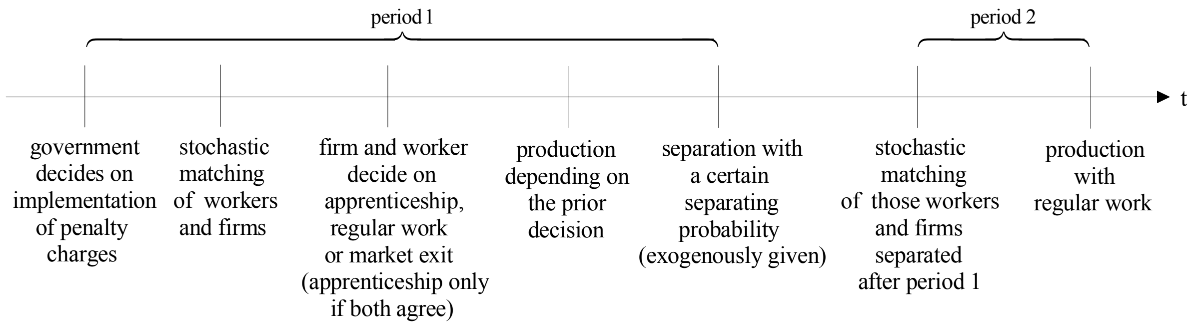

Figure 2: Evolution over Time

4 The Model

We consider a discrete-time model with two types of agents, namely workers and firms. In line with Acemoglu and Pischke (1998a), there are two periods, a training period (period 1) and a working period (period 2). Production takes place in worker-firm pairs and no capital is needed. The initial ability of workers is exogenously given.

At the beginning of period 1, each firm meets one worker whose individual ability is drawn randomly from a distribution that is common knowledge. Firms and workers decide whether to engage in apprenticeship training, to produce with regular work or to stay in the market at all. Apprenticeship training only takes place if both parties prefer it to regular work. If one of them does not agree, the worker is employed regularly. In the second period, all workers are employed regularly, but only those workers have increased productivity who were trained in period 1. In line with Malcomson, Maw, and McCormick (2003), the two periods are connected by an exogenous separating probability ρ. Hence, the status of period 1 is maintained in period 2 with probability $(1 - \rho)$. Altogether, the economy evolves over time as shown in figure 2.

In our model, firms may provide general training both because they are looking forward to future returns in period 2 (as in Acemoglu and Pischke (1999b)) and because they currently benefit from doing so in the first period (as in Becker (1964)). Especially the latter argument has to be considered in the analysis of penalty charges because the implementation of penalty charges affects the current training decision by altering the opportunity costs of refusing to offer apprenticeship training positions. The model assumptions and the training decision of firms and workers are described in the following subsections.

4.1 Model Assumptions

At the beginning of period 1, workers differ by their individual ability that is assumed to be uniformly distributed on the interval $[\theta_L, \theta_H]$.[16] Firms can unambiguously observe the workers' abilities.[17] In line with Malcomson, Maw, and McCormick (2003) and Hanushek, Leung, and Yilmaz (2003), the mass of workers is normalized to unity by defining $\theta_L \equiv 0$ and $\theta_H \equiv 1$. Hence, the cumulative distribution function of individual abilities is $F(\theta) = \theta$.

In the second period, the productivity of all workers employed regularly in period 1 is unchanged. For all trained workers, productivity increases to $\theta_2 = (1 + \alpha)\,\theta$. The parameter $\alpha > 0$ represents the productivity-enhancement of apprenticeship training: the higher θ, the higher is the absolute productivity gain generated by the apprenticeship.[18] We assume that the productivity-enhancement α is identical for all trained workers. This asumption is motivated by the literature dealing with the training decision at the intensive margin.[19] On the other hand, the productivity of workers unemployed in period 1 is reduced to $\theta_2 = (1 - \sigma)\,\theta$ because a fraction $\sigma > 0$ of skills not employed in period 1 is lost and thus no longer available in period 2.[20]

In line with Malcomson, Maw, and McCormick (2003), workers are risk-neutral and maximize their

[16] This heterogeneity of workers is an important extension compared to the model by Malcomson, Maw, and McCormick (2003). The continuous distribution of abilities allows to obtain a smooth participation decision at the individual level. Mincer (1958) and Becker (1962) assume that abilities are normally distributed. Without changing the general results, we assume a uniform distribution of abilities in order to keep the following calculations as simple as possible.

[17] This assumption is in line with Boone and Bovenberg (2006). Furthermore, it is implicitly included into the whole literature on human capital and the life-cycle of earnings. Each worker offers his individual stock of human capital to the firms and is rewarded by a rental price per unit of human capital. Hence, we rule out asymmetric information (hidden knowledge) as source of wage compression as described in section 2.2. If the worker's productivity were not observed by the firm there would be adverse selection as modeled e.g. by de Meza and Webb (2001).

In Malcomson, Maw, and McCormick (2003), the productivity of workers is hidden knowledge and learned by the current employer during the apprenticeship. If the monitoring of the training process is perfect, the ex-post problem of hidden knowledge would be solved (i.e. also potential future employers would learn the worker's productivity).

[18] Formally, this means $\frac{\partial(\theta_2 - \theta)}{\partial\theta} = \alpha > 0$. Intuitively, the accumulation of new skills is easier when more skills are already available. This relationship is also suggested by Ben-Porath (1967) and Mincer (1997). Because the parameter α determines productivity and thus the wage in period 2, it constitutes the key determinant of the return to education as analyzed in the theory of human capital (Mincer (1974)).

[19] In conformity with the literature, we assume the productivity-enhancement unambiguously depending on the amount of training. Because the optimal amount of training per apprentice is independent of θ, also the productivity-enhancement α is independent of θ and thus identical for each worker. Formally, this characteristic is derived in appendix A.

In almost the same manner, Ben-Porath (1967) and Heckman (1976) assume that the absolute (and not the relative) increase in human capital depends on the existing stock of human capital (which may be interpreted as initial ability).

[20] In line with Heckman (1976), the parameter σ describes the depreciation rate of skills.

expected individual income over both periods:[21]

$$U(\theta, \theta_2) = w(\theta) + \delta E[w(\theta_2)] \tag{1}$$

By assumption, the mass of firms is also one, so that each firm meets one worker whose ability is uniformly distributed with $\theta \sim u[0,1]$. Firms are risk-neutral and maximize the sum of expected profits over both periods:

$$\pi(\theta, \theta_2) = \pi(\theta) + \delta E[\pi(\theta_2)] \tag{2}$$

In both periods, the firm's profit is equal to the difference between revenue and costs per worker.[22] By defining the output good as numéraire and assuming an identical, linear one-to-one production function for the connection of output and labor (which is the only factor of production), the profit in each period j depends on the worker's productivity θ_j in the following manner:[23]

$$\pi(\theta_j) = \chi_W \theta_j - w(\theta_j) \tag{3}$$

χ_W is a relative efficiency parameter of workers with $\chi_W = 1$ for regular full-time workers and $0 < \chi_W = \chi < 1$ for apprentices. The parameter χ refers to the allocation of time between training and work in the first period and represents the fraction of time spent in productive activities.[24] $w(\theta_j)$ is the wage depending on the worker's productivity in period j. If a workers is not trained, he can be employed regularly and earns a wage corresponding to the Nash bargaining solution of oligopsonistic labor markets, i.e. $w(\theta_j) = \beta\theta_j$. According to Acemoglu (1997), the parameter $0 < \beta < 1$ indicates the (identical) bargaining power of workers concerning the division of output. In the case of apprenticeship training, the worker receives the training wage $w_A \geq 0$ that is identical and independent of productivity for all apprentices.[25]

In period 2, the wage of all workers corresponds to the Nash bargaining solution. Hence, the profit of

[21] The parameter δ expresses the preference for current and future welfare. The higher δ, the higher ist the weighting of period 2 and the lower is the preference for period 1.

Note that the wage $w(\theta)$ corresponds to the income of the worker with individual ability θ because labor supply is implicitly normalized to unity.

In line with Ben-Porath (1967), we do not analyze a more general utility function of workers. Models of human capital accumulation over the life-cycle can be attributed to two different branches: earnings maximizing models and utility maximizing models. Earnings maximizing models abstract from the labor-leisure choice problem and only analyze the trade-off between investment and income. Utility maximizing models also incorporate the labor-leisure choice so that labor supply becomes endogenous to the model (for example Heckman (1976)).

[22] From the firm's point of view, the worker's ability can be interpreted as individual productivity.

[23] The production side is modeled like in Malcomson, Maw, and McCormick (2003), i.e. the production function exhibits constant returns to scale.

To hold calculations simple, we assume the fixed costs of production to be zero. Without this assumption, our analytical results in section 5 remain the same. The only difference is that there is unemployment also in the laissez-faire equilibrium (cf. section 4.3).

[24] The reduced efficiency of apprentices is due to external schooling, internal seminars, et cetera (cf. section 3.3). Note that the productivity-enhancement α unambiguously depends on χ. The higher χ, the lower the time spent for training activities and the lower α. Because χ is no decision variable of the firm, we neglect this relationship in the following analysis.

[25] Cf. section 3.2. Note that the fixed training wage analytically works like fixed costs of apprenticeship training. There is only one difference: with direct costs of apprenticeship training it would not be optimal to train all workers in the first-best optimum.

We restrict the training wage to satisfy $w_A \leq \chi - (1-\beta)$. This assumption can be justified by economic intuition: the training wage should not exceed the difference between the output of the most productive apprentice and the firm's profit by regularly employing the most productive worker. Because of $w_A \geq 0$ this restriction implies $\chi \geq 1 - \beta$.

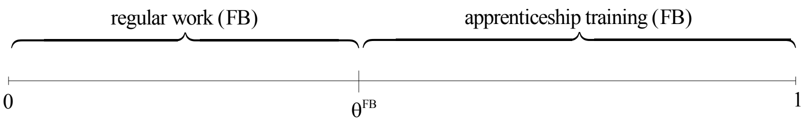

Figure 3: The First-Best Optimum

each firm in period 2 depends on the worker's previous status of employment:[26]

$$\pi(\theta_2) = (1-\beta)\,\theta_2 = \left\{ \begin{array}{ll} (1-\beta)(1+\alpha)\,\theta & if\ apprenticeship\ in\ t=1 \\ (1-\beta)\,\theta & if\ regular\ work\ in\ t=1 \\ (1-\beta)(1-\sigma)\,\theta & if\ unemployment\ in\ t=1 \end{array} \right\} \tag{4}$$

4.2 The First-Best Optimum

In the first-best optimum, the total surplus of workers and firms is maximized. Overall welfare is equal to the sum of the aggregate profits of all firms and the aggregate utility of all workers over both periods. Obviously, there is no unemployment because each unemployed worker would be equivalent to lost productivity, i.e. $u^{FB} = 0$. Each trained worker generates output equal to his productivity $\chi\theta$ during the apprenticeship and $(1+\alpha)\,\theta$ after the apprenticeship has been completed. In the case of regular work, each worker generates output θ in both periods.

In the following, we assume that θ^{FB} describes the welfare maximizing pivotal productivity between apprenticeship training and regular work. Hence, the optimal number of apprentices in the first period is $n^{FB} = 1 - \theta^{FB}$. This situation is illustrated in figure 3. The overall welfare in the first-best optimum (FB) unambiguously depends on θ^{FB}:[27]

$$W^{FB}(\theta^{FB}) = \underbrace{\int_{\theta^{FB}}^{1} (\chi + \delta(1+\alpha))\,\theta d\theta}_{apprenticeship\ training} + \underbrace{\int_{0}^{\theta^{FB}} (1+\delta)\,\theta d\theta}_{regular\ work} \tag{5}$$

In order to determine θ^{FB} we have to maximize with respect to θ^{FB}:

$$\max_{\theta^{FB}} W^{FB}(\theta^{FB}) = \frac{1}{2}(\chi + \delta(1+\alpha)) - \frac{1}{2}v_1(\theta^{FB})^2 \tag{6}$$

Note that the overall welfare does not depend on θ^W because it is not optimal to employ workers regularly if any worker with lower productivity is offered an apprenticeship training position.

[26] The probability of a match in period 2 may explicitly depend on the worker's status of employment in period 1. Different matching probabilities can be justified by different frictions in searching for employment. Mincer (1989) empirically confirms that the probability of unemployment decreases with education. In this context, Brown and Kaufold (1988) stress that the possibility of unemployment reduces expected returns to education. Hence, the return to education is based on higher productivity as well as higher employment probability (Bloch and Smith (1977)). For simplicity, we assume no search frictions and thus the same matching probability for all workers.

[27] Note that the density function of individual abilities is $f(\theta) = 1$.

The first-order condition with respect to θ^{FB} is

$$\frac{\partial W^{FB}(\theta^{FB})}{\partial \theta^{FB}} = -v_1 \theta^{FB} \left\{ \begin{array}{ll} > 0 & if \;\; \alpha < \frac{1-\chi}{\delta} \\ \leq 0 & if \;\; \alpha \geq \frac{1-\chi}{\delta} \end{array} \right\} \tag{7}$$

For $\alpha < \frac{1-\chi}{\delta}$, overall welfare is strictly increasing in θ^{FB} and thus decreasing in the number of apprenticeship training positions. Hence, overall welfare is maximal for $\theta^{FB} = 1$, i.e. no workers should be trained: $n^{FB} = 0$. In this case, overall welfare is equal to $W^{FB}(1) = \frac{1}{2}(1+\delta)$.

For $\alpha \geq \frac{1-\chi}{\delta}$, overall welfare is decreasing in θ^{FB}. Therefore, overall welfare is maximal for $\theta^{FB} = 0$ (i.e. $n^{FB} = 1$), which means $W^{FB}(0) = \frac{1}{2}(\chi + \delta(1+\alpha))$. This corresponds to the product of average ability of all workers and the discounted sum of all parameters influencing the total productivity of apprentices in both periods. Both cases are summarized in the following proposition.[28]

Proposition 1 *Depending on the productivity-enhancement α, overall welfare in the first-best optimum is equal to*

$$W^{FB} = \left\{ \begin{array}{ll} \frac{1}{2}(1+\delta) & if \;\; \alpha < \frac{1-\chi}{\delta} \\ \frac{1}{2}(\chi + \delta(1+\alpha)) & if \;\; \alpha \geq \frac{1-\chi}{\delta} \end{array} \right\} \tag{9}$$

4.3 The Benchmark without Penalty Charges

4.3.1 The Training Decision of Firms

At the extensive margin, each firm decides whether to offer an apprenticeship training position, to employ the worker regularly, or to leave the market. The firm only has this discrete choice but cannot decide on the amount of general training provided to the worker at the intensive margin.[29] The length of the apprenticeship (which is equal to period 1) and the relative efficiency parameter χ do not constitute decision variables of the firm.[30]

No firm will leave the market because it is always possible to make positive profits by employing the worker regularly. In the following, we will analyze the training decision of firms depending on the individual abilities of workers. In this context, we define θ^{LF} to be the worker's productivity that makes a firm just indifferent between apprenticeship training and regular work.

[28] The same result is obtained by considering the discounted sum of wages and profits for a given productivity. The additional surplus generated by a productivity θ with apprenticeship training is greater than with regular work if

$$\chi\theta + \delta(1+\alpha)\theta \geq (1+\delta)\theta \Leftrightarrow v_1\theta \geq 0 \tag{8}$$

[29] This assumption is analytically equivalent to the implication that the amount of training provided at the intensive margin is independent of the worker's ability and thus identical for each firm. Formally, this assumption is justified in appendix A.

[30] These assumptions are justified by the institutional features of the German apprenticeship system as described in section 3. This is the main difference to the framework of Malcomson, Maw, and McCormick (2003) where the length of the apprenticeship is the key determinant of apprenticeship training. In their approach, firms choose the optimal amount of general training depending on the length of the apprenticeship contract.

Definition 1 *The pivotal productivity θ^{LF} is defined to be the lowest productivity that firms decide to offer an apprenticeship training position. A firm prefers apprenticeship training to regular work if its profits over both periods solve*[31]

$$\begin{aligned}
&\chi\theta - w_A + \delta(1-\rho)(1-\beta)(1+\alpha)\theta + \delta\rho(1-\beta)E[\theta_2] \\
&\geq (1-\beta)\theta + \delta(1-\rho)(1-\beta)\theta + \delta\rho(1-\beta)E[\theta_2] \\
&\Leftrightarrow \quad \theta \geq \theta^{LF} \equiv \frac{w_A}{\chi - (1-\beta) + \delta(1-\rho)(1-\beta)\alpha}
\end{aligned}$$

with $E[\theta_2] = \left[1 - F(\theta^{LF})\right](1+\alpha)\frac{1+\theta^{LF}}{2} + F(\theta^{LF})\frac{\theta^{LF}}{2}$.

The pivotal productivity θ^{LF} increases with the training wage (w_A) and the separating probability (ρ) and decreases with the relative efficiency of apprentices (χ) and the productivity-enhancement of apprenticeship training (α). From the firms' point of view, the optimal number of apprenticeship training positions is equal to $1 - \theta^{LF}$.[32]

4.3.2 The Training Decision of Workers

Additionally, we have to analyze the training decision of workers. Note that workers never prefer to stay unemployed because they receive zero income in this case.[33] Workers prefer to be trained in period 1 if the discounted sum of their expected earnings over both periods is bigger with apprenticeship training than with regular work. Hence, the following participation constraint must be satisfied:

$$w_A + \delta\beta(1+\alpha)\theta \geq (1+\delta)\beta\theta \quad \Leftrightarrow \quad \beta(-1+\delta\alpha)\theta \geq -w_A \tag{10}$$

Proposition 2 *For $\alpha \geq \frac{1}{\delta}$, the participation constraint is satisfied for all abilities $\theta \in [0,1]$.*[34] *However, If the productivity-enhancement of apprenticeship training is low, workers with high productivity prefer to remain unskilled. For $\alpha < \frac{1}{\delta}$, a worker prefers apprenticeship training to regular work if*

$$\theta \leq \theta^W \equiv \frac{w_A}{\beta(1-\delta\alpha)} \tag{11}$$

The pivotal productivity θ^W is defined to be the highest productivity that decides to accept an apprenticeship training position if the productivity-enhancement of training is low.

If the productivity-enhancement of apprenticeship training exceeds the lower bound $\frac{1}{\delta}$, all workers prefer apprenticeship training to regular work because their wages in period 2 increase by more than they

[31] The calculation would be more complicated if the probability of maintaining the status of period 1 or the matching probability in period 2 depend on the status in the first period.

[32] Note that $0 \leq \theta^{LF} \leq 1$ is implied by $0 \leq w_A \leq \chi - (1-\beta)$.

[33] Note that there are no unemployment benefits. If there were unemployment benefits greater than zero, low-ability workers may prefer to stay unemployed.

[34] In Malcomson, Maw, and McCormick (2003), there is no worker heterogeneity in initial ability. In this case, incentice compatibility means that the training wage must exceed some lower bound.

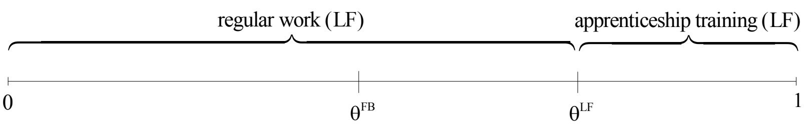

Figure 4: The Laissez-Faire Equilibrium 1

have to forgo in the first period. However, for $\alpha < \frac{1}{\delta}$, only low-ability workers will adopt an apprenticeship training position. Workers with individual ability above θ^{W} prefer to be employed regularly because the productivity-enhancement in period 2 is too low to compensate for the low training wage in period 1.

4.3.3 The Laissez-Faire Equilibrium

In a first step, we concentrate on the case $\alpha \geq \frac{1}{\delta}$, i.e. we assume that all workers prefer to receive apprenticeship training. In the first period, $n^{LF} = 1 - \theta^{LF}$ workers are trained and earn the training wage w_A while all other workers are employed regularly and earn a wage equal to the Nash bargaining solution, i.e. $w(\theta) = \beta\theta$. There is no unemployment because firms and workers always prefer regular work to market exit. This situation is illustrated in figure 4. Hence, in the laissez-faire equilibrium (LF), the welfare of workers (W) in period 1 is equal to

$$W_{1W}^{LF} = \underbrace{\int_{\theta^{LF}}^{1} w_A d\theta}_{apprenticeship\ training} + \underbrace{\int_{0}^{\theta^{LF}} \beta\theta d\theta}_{regular\ work} \tag{12}$$

While the first integral in equation (12) is equal to the aggregate wage sum of apprentices, the second one describes the aggregate wage sum of regular workers.

In the second period, the productivity of trained workers is increased by the factor $(1+\alpha)$. All workers are employed regularly and the wages correspond to the Nash bargaining solution so that wages of skilled workers are increased by the same factor $(1+\alpha)$ compared to the wages of unskilled workers:

$$W_{2W}^{LF} = \underbrace{\int_{\theta^{LF}}^{1} \beta(1+\alpha)\theta d\theta}_{skilled\ work} + \underbrace{\int_{0}^{\theta^{LF}} \beta\theta d\theta}_{unkilled\ work} \tag{13}$$

The first integral in equation (13) is equal to the aggregate wage sum of trained workers and the second one describes the aggregate wage sum of unskilled workers. Hence, aggregate welfare of workers over both periods is equal to the discounted sum of aggregate wages

$$\begin{aligned} W_{W}^{LF} &= W_{1W}^{LF} + \delta W_{2W}^{LF} \\ &= \underbrace{\int_{\theta^{LF}}^{1} \left[w_A + \delta\beta(1+\alpha)\theta\right] d\theta}_{skilled\ work} + \underbrace{\int_{0}^{\theta^{LF}} (1+\delta)\beta\theta d\theta}_{unkilled\ work} \end{aligned} \tag{14}$$

The first integral is equal to the discounted aggregate wage sum of trained workers who earn w_A in the first period and $w(\theta) = \beta(1+\alpha)\theta$ in the second period. The second integral describes the discounted aggregate wage sum of workers without apprenticeship training position.

In almost the same manner, the welfare of firms (F) in both periods depends on the pivotal productivity θ^{LF}:

$$W_{1F}^{LF} = \underbrace{\int_{\theta^{LF}}^{1} (\chi\theta - w_A)\, d\theta}_{apprenticeship\ training} + \underbrace{\int_{0}^{\theta^{LF}} (1-\beta)\,\theta d\theta}_{regular\ work} \tag{15}$$

$$W_{2F}^{LF} = \underbrace{\int_{\theta^{LF}}^{1} (1-\beta)(1+\alpha)\,\theta d\theta}_{skilled\ work} + \underbrace{\int_{0}^{\theta^{LF}} (1-\beta)\,\theta d\theta}_{unkilled\ work} \tag{16}$$

In period 1, the first integral is equal to the aggregate profits of training firms. These profits are determined by the efficiency parameter χ and the training wage w_A. In period 2, the first integral describes the aggregate profits of firms producing with a trained worker. These profits correspond to the Nash bargaining solution, i.e. the fraction $(1-\beta)$ determines the profit of the firm. In both periods, the second integral is equal to the aggregate profits of firms employing unskilled workers regularly. In a nutshell, the aggregate welfare of firms over both periods adds up to

$$\begin{aligned} W_F^{LF} &= W_{1F}^{LF} + \delta W_{2F}^{LF} \\ &= \underbrace{\int_{\theta^{LF}}^{1} \left[(\chi + \delta(1-\beta)(1+\alpha))\,\theta - w_A\right] d\theta}_{skilled\ work} + \underbrace{\int_{0}^{\theta^{LF}} (1+\delta)(1-\beta)\,\theta d\theta}_{unkilled\ work} \end{aligned} \tag{17}$$

Altogether, overall welfare is determined by the aggregate welfare of workers and the aggregate welfare of firms as described by equations (14) and (17). Hence, overall welfare in the laissez-faire equilibrium for $\alpha \geq \frac{1}{\delta}$ is equal to[35]

$$\begin{aligned} W^{LF} &= W_W^{LF} + W_F^{LF} \\ &= \underbrace{\int_{\theta^{LF}}^{1} (\chi + \delta(1+\alpha))\,\theta d\theta}_{skilled\ work} + \underbrace{\int_{0}^{\theta^{LF}} (1+\delta)\,\theta d\theta}_{unkilled\ work} \end{aligned} \tag{18}$$

Simplifying yields

$$\begin{aligned} W^{LF} &= (1-\theta^{LF})(\chi + \delta(1+\alpha))\frac{1+\theta^{LF}}{2} + \theta^{LF}(1+\delta)\frac{\theta^{LF}}{2} \\ &= \frac{1}{2}(\chi + \delta(1+\alpha)) - \frac{1}{2}v_1(\theta^{LF})^2 \end{aligned} \tag{19}$$

In equation (19), the integrals of equation (18) are calculated as the product of mass and average pro-

[35] Note that overall welfare in this case is very similar to equation (??). The only difference is the pivotal productivity between apprenticeship training and regular work. $\theta^{LF} > \theta^{FB}$ implies that the number of apprenticeship training positions is too low compared to the first-best optimum (cf. section 4.4).

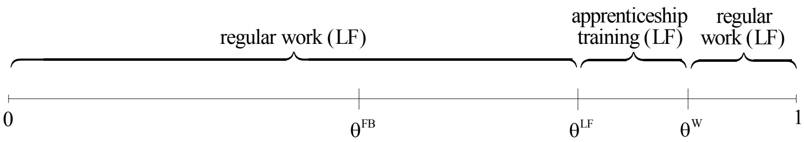

Figure 5: The Laissez-Faire Equilibrium 2

ductivity in the corresponding ability interval. Because individual abilities are uniformly distributed, average productivity in the interval $[\theta^{LF}, 1]$ is calculated by averaging the limit values of this interval: $\frac{1+\theta^{LF}}{2}$. Note that the parameter β cancels out because the bargaining power only determines how the output is divided between workers and firms.

In a second step, we have to analyze the laissez-faire equilibrium for $\alpha < \frac{1}{\delta}$. In this case, all workers with individual ability $\theta > \theta^W$ prefer regular work to apprenticeship training. This situation is illustrated in figure 5. In a nutshell, there are two different subcases that have to be considered. For $\frac{1-\chi}{\delta(1-\rho(1-\beta))} \leq \alpha < \frac{1}{\delta}$, the pivotal productivity θ^W is greater than θ^{LF}.[36] Hence, less workers are trained compared to the case $\alpha \geq \frac{1}{\delta}$ because high-ability workers do not adopt apprenticeship training. Overall welfare is equal to

$$\begin{aligned} W^{LF} &= \underbrace{\int_{\theta^W}^{1} (1+\delta)\,\theta d\theta}_{regular\ work} + \underbrace{\int_{\theta^{LF}}^{\theta^W} (1+\delta\,(1+\alpha))\,\theta d\theta}_{apprenticeship\ training} + \underbrace{\int_{0}^{\theta^{LF}} (1+\delta)\,\theta d\theta}_{regular\ work} && (20) \\ &= \frac{1}{2}(1+\delta) + \frac{1}{2}v_1(\theta^W)^2 - \frac{1}{2}v_1(\theta^{LF})^2 && (21) \end{aligned}$$

Note that the labeling in equation (20) refers to the worker's status of employment in period 1. If the productivity-enhancement α is very small (i.e. $\alpha < \frac{1-\chi}{\delta(1-\rho(1-\beta))}$) the pivotal productivity θ^W is smaller than θ^{LF}. Hence, there are no workers that both prefer apprenticeship training and are offered a training place. All workers are employed regularly so that overall welfare is equal to

$$W^{LF} = \frac{1}{2}(1+\delta) \tag{22}$$

The three different cases of laissez-faire are summarized in the following proposition.

Proposition 3 *Depending on the productivity-enhancement α, overall welfare in the laissez-faire equilibrium is equal to*

$$W^{LF} = \left\{ \begin{array}{ll} \frac{1}{2}(1+\delta) & if\ \ \alpha < \frac{1-\chi}{\delta(1-\rho(1-\beta))} \\ \frac{1}{2}(1+\delta) + \frac{1}{2}v_1[(\theta^W)^2 - (\theta^{LF})^2] & if\ \ \frac{1-\chi}{\delta(1-\rho(1-\beta))} \leq \alpha < \frac{1}{\delta} \\ \frac{1}{2}(\chi+\delta(1+\alpha)) - \frac{1}{2}v_1(\theta^{LF})^2 & if\ \ \alpha \geq \frac{1}{\delta} \end{array} \right\} \tag{23}$$

with $v_1 \equiv (\chi - 1 + \delta\alpha)$.

[36] Note that $\frac{1-\chi}{\delta(1-\rho(1-\beta))} < \frac{1}{\delta}$ because $\chi > \rho(1-\beta)$.

In the following section, this outcome with laissez-faire will be compared to the first-best optimum of section 4.2 in order to evaluate the possibilites of welfare-enhancing government interventions.

4.4 Inefficiencies in the Training Decision of Firms

Comparing overall welfare in the laissez-faire equilibrium with the first-best optimum shows that only for $\alpha < \frac{1-\chi}{\delta}$ the former is efficient because it is equal to the first-best optimum with no training. For $\alpha \geq \frac{1-\chi}{\delta}$, the number of apprenticeship training positions n is inefficiently low.

Proposition 4 *Depending on α, the deviation from the first-best number of apprenticeship training positions is*

$$\triangle n = n^{LF} - n^{FB} = \left\{ \begin{array}{ll} 0 & if \ \ \alpha < \frac{1-\chi}{\delta} \\ -1 & if \ \ \frac{1-\chi}{\delta} \leq \alpha < \frac{1-\chi}{\delta(1-\rho(1-\beta))} \\ \theta^{W} - \theta^{LF} - 1 & if \ \ \frac{1-\chi}{\delta(1-\rho(1-\beta))} \leq \alpha < \frac{1}{\delta} \\ -\theta^{LF} & if \ \ \alpha \geq \frac{1}{\delta} \end{array} \right\} \tag{24}$$

For $\alpha \geq \frac{1}{\delta}$, the deviation from the first-best optimum is driven by the pivotal productivity $\theta^{LF} = \frac{w_A}{\chi-(1-\beta_1)+\delta(1-\rho)(1-\beta_2)\alpha}$ which characterizes the training decision of firms. The higher θ^{LF}, the lower the number of apprenticeship training positions $n^{LF} = 1-\theta^{LF}$ and the larger the inefficiencies in the training decision $\triangle n$. Note that the inefficiencies become smaller if the productivity of trained workers in the second period is increased (α).

Altogether, the training decision of firms bears three different kinds of inefficiencies which are in line with the results of Acemoglu and Pischke (1999b). In order to distinguish between these different kinds of inefficiencies we label the workers' bargaining power by an index for period 1 and period 2 respectively. First, firms take into account only their own gains from higher productivity and neglect the gains for the workers by higher wages in the second period. Hence, firms in aggregate underinvest in the workers' human capital by offering an insufficient number of apprenticeship training positions. This kind of inefficiency could only be eliminated if firms were the sole beneficiaries of apprenticeship training by choosing $\beta_2 = 0$ and thus ruling out any bargaining power of workers in period 2. The higher the bargaining power of workers in the second period, the higher the pivotal productivity θ^{LF} and the larger the difference to the number of apprenticeship training positions in the first-best optimum. The second kind of inefficiency stems from the fact that there is a positive probability of (exogenous) separation after the first period. The pivotal productivity θ^{LF} is increasing in ρ because the firm bears the risk of not participating in the worker's higher productivity after the apprenticeship has been completed. Hence, the training firm does not take into account higher profits of potential future employers in period 2. The higher the separating probability after period 1, the larger the inefficiency in the provision of apprenticeship training. This kind of inefficiency could only be eliminated if there were no exogenous separation after the first period ($\rho = 0$) or if future employers could be identified in advance and included in the apprenticeship contract at the beginning of period 1.

Beyond the approach of Acemoglu and Pischke (1999b), our model covers a third source of inefficiency that is generated by the fixed training wage during the apprenticeship. Firms consider the training wage as costs of providing apprenticeship training but do not take into account that it is also equivalent to the income of apprentices in period 1. Hence, θ^{LF} is increasing in w_A. From the point of view of policy analysis, this is an interesting feature of the training decision of firms because the government can reduce this inefficiency by regulating the training wage towards zero. For example, the government can move the training wage towards zero by statutorily lowering the bargaining power of the unions.[37] However, the first-best optimum is only achieved for $w_A = 0$. Otherwise, all firms with a negative profit would rather leave the market.[38]

Proposition 5 *Depending on α, the inefficient low number of apprenticeship training positions yields the following welfare deviation from the first-best optimum:*

$$\triangle W = \left\{ \begin{array}{ll} 0 & if \;\; \alpha < \frac{1-\chi}{\delta} \\ -\frac{1}{2}\left(\chi - 1 + \delta\alpha\right) \leq 0 & if \;\; \frac{1-\chi}{\delta} \leq \alpha < \frac{1-\chi}{\delta(1-\rho(1-\beta))} \\ -\frac{1}{2}\left(\chi - 1 + \delta\alpha\right) + \frac{1}{2}v_1[(\theta^{W})^2 - (\theta^{LF})^2] < 0 & if \;\; \frac{1-\chi}{\delta(1-\rho(1-\beta))} \leq \alpha < \frac{1}{\delta} \\ -\frac{1}{2}v_1(\theta^{LF})^2 < 0 & if \;\; \alpha \geq \frac{1}{\delta} \end{array} \right\} \quad (25)$$

The inefficiencies in the training decision constitute the necessary condition for welfare-enhancing government interventions. Penalty charges are one possible policy instrument to move the economy towards its first-best optimum by reducing the third kind of inefficiency and thus increasing the number of apprenticeship training positions. The welfare implications of penalty charges are analyzed in the following section.

5 The Welfare Analysis of Penalty Charges

5.1 The Pivotal Productivities with Penalty Charges

If identical penalty charges $T \geq 0$ are imposed in the case of regular work, the opportunity costs of refusing to offer apprenticeship training positions are increased.[39] Hence, the pivotal productivity between apprenticeship training and regular work is decreased.

Definition 2 *The pivotal productivity θ_A^{PC} is defined to be the lowest productivity that firms decide to offer an apprenticeship training position if they are exposed to penalty charges. With penalty charges, a*

[37] Note that the productivity-enhancement of apprenticeship training must be high enough to satisfy the participation constraint of workers (cf. section 4.3.2).

[38] The first-best optimum would also be achieved if it were possible to compensate firms for their losses. However, we assume that this policy instrument is not available.

[39] The training quota for the identical firms is one, i.e. every firm is assigned to train the worker it meets. Consequently, the penalty charges are identical for each untrained worker. Note that penalty charges are analytically the same as fixed costs of production that only accrue in the case of regular work.

firm prefers apprenticeship training to regular work if its profits over both periods solve

$$\chi\theta - w_A + \delta(1-\rho)(1-\beta)(1+\alpha)\theta \geq (1-\beta)\theta - T + \delta(1-\rho)(1-\beta)\theta$$
$$\theta \geq \theta_A^{PC} \equiv \frac{w_A - T}{\chi - (1-\beta) + \delta(1-\rho)(1-\beta)\alpha}$$

With respect to θ_A^{PC}, penalty charges work like a reduction in the training wage. However, with penalty charges it could be beneficial for firms to leave the market in period 1 (with zero profits) and reenter in period 2 in order to avoid the unprofitable apprenticeship or the financial burden of penalty charges, respectively.[40] Hence, all workers below the pivotal productivity θ_U^{PC} stay unemployed in period 1.

Definition 3 *The pivotal productivity θ_U^{PC} is defined to be the lowest productivity that firms decide to employ regularly compared to zero profits in the case of market exit. With penalty charges, a firm prefers regular work to market exit if its profits over both periods solve*

$$(1-\beta)\theta - T + \delta(1-\rho)(1-\beta)\theta \geq 0$$
$$\theta \geq \theta_U^{PC} \equiv \frac{T}{(1-\beta)(1+\delta(1-\rho))}$$

Note that both pivotal productivities depend positively on the separating probability ρ:

$$\frac{\partial\theta_A^{PC}}{\partial\rho} > 0 \;\; ; \;\; \frac{\partial\theta_U^{PC}}{\partial\rho} > 0 \tag{26}$$

Like θ^{LF}, the pivotal productivity θ_A^{PC} is increasing in ρ (i.e. there are less apprenticeship training positions) because separation becomes more likely and thus the expected return to apprenticeship training is reduced for the training firm. In the same way, θ_U^{PC} is increasing in ρ (i.e. there are more workers unemployed) because it becomes less profitable for the firm to pay the penalty charges and retain a regular worker.

[40] If firms are bared from reentering in period 2, the welfare losses due to penalty charges would be more complicated. On the one hand, welfare would decrease because of missing employment places in period 2. On the other hand, the number of firms that leave the market in period 1 would decline because they have to account for expected profits in period 2.

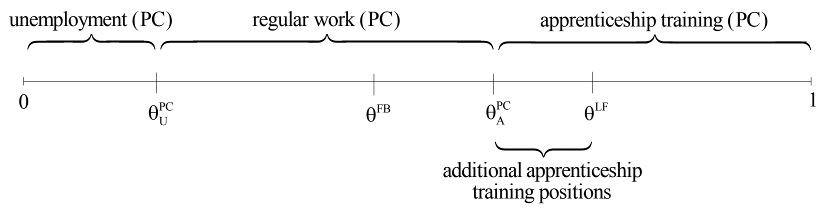

Figure 6: The Equilibrium with Penalty Charges

Proposition 6 *The relationship of the pivotal productivities is the following:*[41]

$$1 \geq \theta^{LF} \geq \theta_A^{PC} \geq \theta_U^{PC} \geq 0 \tag{29}$$

This relationship is implied by the penalty charges satisfying $T \leq \bar{T} \equiv \frac{(1-\beta)(1+\delta(1-\rho))}{\chi+\delta(1-\rho)(1-\beta)(1+\alpha)} w_A$.

Altogether, $n^{PC} = 1 - \theta_A^{PC}$ workers are trained in period 1. Hence, there are $n^{PC} - n^{LF} = \theta^{LF} - \theta_A^{PC}$ additional apprenticeship training positions compared to the laissez-faire equilibrium. All workers with individual ability between θ_A^{PC} and θ_U^{PC} are employed regularly while all workers with ability below θ_U^{PC} stay unemployed. This situation is illustrated in figure 6.

5.2 The Equilibrium with Penalty Charges

With penalty charges, there is an increased number of apprenticeship training positions. On the other hand, there is unemployment among low-ability workers because firms prefer market exit to regular work. Altogether, the welfare of workers with penalty charges is equal to

$$\begin{aligned} W_W^{PC} = & \underbrace{\int_{\theta_A^{PC}}^{1} \left[w_A + \delta\beta\left(1+\alpha\right)\theta\right] d\theta}_{apprenticeship\ training} \\ & + \underbrace{\int_{\theta_U^{PC}}^{\theta_A^{PC}} \left(1+\delta\right)\beta\theta d\theta}_{regular\ work} + \underbrace{\int_{0}^{\theta_U^{PC}} \delta\beta\left(1-\sigma\right)\theta d\theta}_{unemployment} \end{aligned} \tag{30}$$

[41] Assume that $\dot{\theta}$ is defined to be the pivotal productivity between apprenticeship training and market exit:

$$\dot{\theta} = \frac{w_A}{\chi + \delta\left(1-\rho\right)\left(1-\beta\right)\left(1+\alpha\right)} \tag{27}$$

$\dot{\theta}$ lies between θ_A^{PC} and θ_U^{PC} (cf. appendix B) and thus has no impact on the labor market outcome.

Note that with respect to the relationship of the pivotal productivities there is a second case which is mathematically feasible:

$$1 \geq \theta^{LF} \geq \theta_U^{PC} > \dot{\theta} > \theta_A^{PC} \;\; or \;\; \theta_U^{PC} > \theta^{LF} > \dot{\theta} > \theta_A^{PC} \quad if \;\; T > \bar{T} \tag{28}$$

In this case (II), penalty charges would have to satisfy $T > \bar{T}$. Intuitively, $\bar{T}$ maximizes the number of apprenticeship training positions at the cost of suppressed regular work. If the administration costs of penalty charges exceed some lower bound, this second case never accrues in the optimum (cf. appendix C.2) so that we can neglect it. In more detail, case (II) is presented in appendix D. The formal proof why only these two cases are feasible is shown in appendix B.

The first integral is equal to the discounted aggregate wage sum of trained workers who earn w_A in period 1 and $w(\theta) = \beta(1+\alpha)\theta$ in period 2. The second integral in equation (30) describes the discounted aggregate wage sum of workers who are employed regularly in both periods. And the third integral is equal to the discounted aggregate wage sum of low-ability workers who are employed regularly in period 2 but stay unemployed in the first period. Note that all workers with $\theta < \theta_U^{PC}$ earn nothing in period 1. In the second period, there is no unemployment. All workers are employed regularly and compensated according to the Nash bargaining solution.

Concerning the welfare of firms, we must also consider the penalty charges that are imposed on all $\theta_A^{PC} - \theta_U^{PC}$ firms that employ workers regularly in period 1. Hence, compared to the laissez-faire equilibrium, the profits of firms with regular work are reduced by T. In consequence, all firms that meet a worker with productivity below θ_U^{PC} decide to leave the market and thus make zero profits in period 1. Formally, the welfare of firms with penalty charges is equal to

$$W_F^{PC} = \underbrace{\int_{\theta_A^{PC}}^{1} \left[(\chi + \delta(1-\beta)(1+\alpha))\theta - w_A\right] d\theta}_{apprenticeship\ training} + \underbrace{\int_{\theta_U^{PC}}^{\theta_A^{PC}} \left[(1+\delta)(1-\beta)\theta - T\right] d\theta}_{regular\ work} + \underbrace{\int_0^{\theta_U^{PC}} \delta(1-\beta)(1-\sigma)\theta d\theta}_{unemployment} \quad (31)$$

The first integral in equation (31) is equal to the discounted aggregate profits of training firms (period 1) and firms meeting a trained worker (period 2), respectively. The second integral describes the discounted aggregate profits of firms that decide to employ workers regularly in both periods. And the third integral is equal to the discounted aggregate profits of those firms that meet low-ability workers in period 2. In the second period, the profits of all firms correspond to the Nash bargaining solution.

In order to calculate the overall welfare with penalty charges we have to consider the total amount of penalty charges which is not lost but may be spent for the provision of public goods. Hence, we have to respect for public revenues $\tau(T) \equiv \int_{\theta_U^{PC}}^{\theta_A^{PC}} T d\theta$ in the overall welfare equation (34). Unfortunately, the implementation of penalty charges causes administration costs $C(T)$. We assume that these administration costs depend on the level of penalty charges as described by equation (33):[42]

$$C(T) = \frac{c}{2}T^2 \quad with\ C'(T) > 0\,;\ C''(T) > 0 \quad (33)$$

Administration costs are increasing in T because firms try to reduce their training costs by lowering their compulsory training quota. This effort is assumed to increase disproportionately in the level of penalty charges. Furthermore, public spending has to be financed by distortionary taxation which generates

[42] As mentioned above, we assume the cost parameter c to exceed some lower bound (cf. appendix C.2):

$$c \geq \bar{c} \equiv \frac{1-(1-\rho)(1-\delta(1-\sigma))}{(1-\rho)(v_4)^2(1+\lambda)} \quad (32)$$

Without changing our results in general, this assumption guarantees $T^* \leq \bar{T}$. In appendix D.1, the case $c < \bar{c}$ is discussed in detail.

efficiency losses of λ per unit of money. Altogether, the overall welfare with penalty charges is equal to

$$\begin{aligned} W^{PC}(T) &= W_W^{PC} + W_F^{PC} + \tau(T) - (1+\lambda)\, C(T) \\ &= \underbrace{\int_{\theta_A^{PC}}^{1} (\chi + \delta(1+\alpha))\, \theta d\theta}_{apprenticeship\ training} + \underbrace{\int_{\theta_U^{PC}}^{\theta_A^{PC}} (1+\delta)\, \theta d\theta}_{regular\ work} \\ &\quad + \underbrace{\int_0^{\theta_U^{PC}} \delta(1-\sigma)\, \theta d\theta}_{unemployment} - (1+\lambda)\frac{c}{2}T^2 \end{aligned} \tag{34}$$

In order to determine the optimal penalty charges we have to maximize W^{PC} with respect to T:

$$\begin{aligned} \max_T W^{PC}(T) &= \frac{1}{2}(\chi + \delta(1+\alpha))\,[1 - (\theta_A^{PC})^2] + \frac{1}{2}(1+\delta)\,[(\theta_A^{PC})^2 - (\theta_U^{PC})^2] \\ &\quad + \frac{1}{2}\delta(1-\sigma)\,(\theta_U^{PC})^2 - (1+\lambda)\frac{c}{2}T^2 \end{aligned} \tag{35}$$

With respect to overall welfare, the implementation of penalty charges solves a trade-off. On the one hand, penalty charges are welfare-enhancing because they increase the number of apprenticeship training positions and thus the fraction of skilled workers in the second period. Hence, there are higher wages and higher profits after the apprenticeship because the output is shared between worker and firm. On the other hand, penalty charges are costly because some firms will leave the market so that there is unemployment among workers with lowest ability. Additionally, the implementation of penalty charges causes administration costs which have to be financed by distortionary taxation.

Differentiating equation (34) with respect to T yields the first-order condition

$$\frac{\partial W^{PC}(T)}{\partial T} = 0$$

$$\underbrace{(\chi - 1 + \delta\alpha)\, \theta_A^{PC} \left(-\frac{\partial \theta_A^{PC}}{\partial T}\right)}_{more\ apprenticeship\ training\ positions} = \underbrace{(1+\delta\sigma)\, \theta_U^{PC} \frac{\partial \theta_U^{PC}}{\partial T}}_{more\ unemployment} + \underbrace{(1+\lambda)\, cT}_{administration\ costs} \tag{36}$$

Note that the partial derivatives of the pivotal productivities have the following sign: $\frac{\partial \theta_A^{PC}}{\partial T} < 0$ and $\frac{\partial \theta_U^{PC}}{\partial T} > 0$. Hence, both sides of equation (36) are unambiguously positive. Equation (36) compares the marginal benefits (LHS) and the marginal costs (RHS) of an increase in T. On the one hand, the welfare gains on the LHS arise because the number of apprenticeship training positions is increased by $(-\frac{\partial \theta_A^{PC}}{\partial T})$ which generates additional productivity of $(\chi - 1 + \delta\alpha)$ per unit of initial ability. On the other hand, there are two negative welfare effects which are represented by the RHS. First, the number of unemployed workers is increased by $\frac{\partial \theta_U^{PC}}{\partial T}$ which is equal to reduced productivity of $(1 + \delta\sigma)$ per unit of initial ability. Second, there are additional welfare losses of $(1 + \lambda)\, cT$ due to the administration costs.

The optimal penalty charges T^* are obtained by substituting the pivotal productivities θ_A^{PC} and θ_U^{PC}

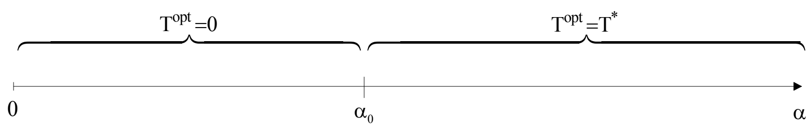

Figure 7: Optimal Penalty Charges

into equation (36) and solving for T:[43]

$$T^{*}=\frac{v_1\left(v_4\right)^2}{v_1\left(v_4\right)^2+v_2\left(v_3\right)^2}w_A \tag{37}$$

with $v_1 \equiv (\chi-1+\delta\alpha)$, $v_2 \equiv 1+(v_4)^2(1+\lambda)c+\delta\sigma$, $v_3 \equiv (\chi-(1-\beta)+\delta(1-\rho)(1-\beta)\alpha)$ and $v_4 = (1-\beta)(1+\delta(1-\rho))$. The optimal penalty charges have to satisfy $0 \leq T^{*} \leq \bar{T}$. In order to guarantee $T^{*} \geq 0$, α has to exceed $\frac{1-\chi}{\delta}$. However, T^{*} never reaches the upper limit $\bar{T}$.[44]

Additionally, we have to consider the participation constraint of workers as explained in section 4.3.2. Although the overall welfare in the laissez-faire equilibrium deviates from the first-best optimum for $\alpha \geq \frac{1-\chi}{\delta}$ (cf. equation (25)), the optimal penalty charges are greater than zero not until the productivity-enhancement of apprenticeship training exceeds the lower bound $\frac{1-\chi}{\delta(1-\rho(1-\beta))}$ which lies above $\frac{1-\chi}{\delta}$.[45] Hence, the implementation of penalty charges only increases the overall welfare if the participation constraint is satisfied for a sufficient number of workers. These results are summarized in the following proposition and illustrated in figure 7.

Proposition 7 *Optimal penalty charges are equal to*

$$T^{opt}=\left\{\begin{array}{lll} 0 & if & \alpha<\frac{1-\chi}{\delta(1-\rho(1-\beta))} \\ T^{*} & if & \alpha \geq \frac{1-\chi}{\delta(1-\rho(1-\beta))} \end{array}\right\} \tag{38}$$

The optimal level of penalty charges explicitly depends on the productivity-enhancement of apprenticeship training. If α is low, it is optimal to reject the implementation of penalty charges, i.e. it is optimal to set $T^{opt}=0$. However, if α exceeds the critical level $\frac{1-\chi}{\delta(1-\rho(1-\beta))}$ it is optimal to implement penalty charges according to $T^{opt}=T^{*}$. Note that even for very high administration costs there is some critical level of productivity-enhancement which makes optimal penalty charges greater than zero. The implementation of penalty charges becomes more likely the lower the critical level of productivity-enhancement. This critical level $\frac{1-\chi}{\delta(1-\rho(1-\beta))}$ is decreasing in the relative efficiency of apprentices, the discount factor and the bargaining power of workers. However, it is increasing in the separating probability after the first period.

[43] The calculation of T^{*} is shown in appendix C.1. Note that T^{*} describes a maximum because $\alpha>0$ guarantees that the second derivative is negative, i.e. $\frac{\partial^2 W^{PC}(T)}{\partial T^2}<0$.

[44] These considerations are presented in appendix C.2.

[45] Cf. appendix C.3.

Proposition 8 *For* $\alpha \geq \frac{1-\chi}{\delta(1-\rho(1-\beta))}$, *the comparative statics of the optimal penalty charges are as follows:*[46]

$$\frac{\partial T^*}{\partial \alpha} \geq 0 \qquad if \;\; \alpha \leq \alpha' \tag{39}$$
$$\frac{\partial T^*}{\partial \sigma} < 0 \tag{40}$$
$$\frac{\partial T^*}{\partial \rho} > 0 \tag{41}$$
$$\frac{\partial T^*}{\partial c} < 0 \tag{42}$$
$$\frac{\partial T^*}{\partial \lambda} < 0 \tag{43}$$
$$\frac{\partial T^*}{\partial w_A} > 0 \tag{44}$$

For $\alpha \leq \alpha'$, the optimal penalty charges are increasing in α because a greater productivity-enhancement increases the benefits of additional apprenticeship training positions. However, for high values of α the opposite is true because firms already take into account the benefits of trained workers in the second period. On the other hand, T^* is decreasing in σ because a greater depreciation of skills during unemployment increases the welfare loss of higher unemployment. Furthermore, the optimal penalty charges are increasing in ρ because firms consider future gains in productivity to a lesser extent. Hence, the optimal penalty charges must go against this inefficiency by increasing the opportunity costs of regular work. Quite intuitively, the optimal penalty charges are decreasing in c and λ because these parameters define the costs of administration and distortion. For one of these two parameters going towards infinity, optimal penalty charges are converging to zero. Finally, the partial derivative of T^* with respect to the training wage is positive because penalty charges reduce the inefficiencies in the number of apprenticeship training positions by affecting θ_A^{PC} like a reduction in w_A.

By substituting the optimal penalty charges (38) into equation (34) we obtain the optimal overall welfare with penalty charges.

Proposition 9 *With optimal penalty charges, the overall welfare is equal to*[47]

$$\left(W^{PC}\right)^* = \left\{ \begin{array}{ll} W^{LF} & if \quad \alpha < \frac{1-\chi}{\delta(1-\rho(1-\beta))} \\ \frac{1}{2}\left(\chi + \delta\left(1+\alpha\right)\right) - \frac{1}{2}\frac{v_1 v_2}{v_1 (v_4)^2 + v_2 (v_3)^2}(w_A)^2 & if \quad \alpha \geq \frac{1-\chi}{\delta(1-\rho(1-\beta))} \end{array} \right\} \tag{45}$$

Altogether, the overall welfare is increased by the implementation of penalty charges if the productivity-enhancement exceeds some lower bound $\frac{1-\chi}{\delta(1-\rho(1-\beta))}$ which guarantees that the participation constraint is satisfied for a sufficient number of workers.

Proposition 10 *For* $\alpha \geq \frac{1-\chi}{\delta(1-\rho(1-\beta))}$, *the comparative statics of the optimal overall welfare with penalty*

[46] These partial derivatives are calculated in appendix C.4.
[47] The calculation is shown in appendix C.5.

charges are as follows:[48]

$$\frac{\partial (W^{PC})^*}{\partial \alpha} > 0 \quad (46)$$
$$\frac{\partial (W^{PC})^*}{\partial \sigma} < 0 \quad (47)$$
$$\frac{\partial (W^{PC})^*}{\partial \rho} < 0 \quad (48)$$
$$\frac{\partial (W^{PC})^*}{\partial w_A} < 0 \quad (49)$$
$$\frac{\partial (W^{PC})^*}{\partial c} < 0 \quad (50)$$
$$\frac{\partial (W^{PC})^*}{\partial \lambda} < 0 \quad (51)$$

The effects of the comparative statics operate via two channels. First, there is a direct effect on overall welfare which is identified by equation (35). Additionally, there is an indirect effect because the optimal penalty charges are altered according to equations (39) to (44). Altogether, the optimal overall welfare is increasing in α because trained workers are more productive in period 2. On the other hand, it is decreasing in σ because formerly unemployed workers have lower productivity in the second period. Furthermore, $(W^{PC})^*$ is negatively affected by the separating probability ρ and the training wage w_A because both higher fluctuation and higher wage costs of apprentices lead to increased inefficiencies in the provision of apprenticeship training positions. Finally, the optimal overall welfare is decreasing in c and λ which determine the costs of penalty charges.

5.3 Individual Welfare Consequences

While the previous section analyzes the welfare consequences of penalty charges in aggregate, it is also important to investigate the implications for different groups of workers. High-ability workers are not affected because they are trained anyway.[49] While workers with middle ability above average benefit from penalty charges because firms now offer them apprenticeship training positions, low-ability workers suffer from unemployment in period 1 and thus reduced wages in period 2. These different implications are summarized in table 2. In a nutshell, the increased number of apprenticeship training positions is achieved at the costs of unemployment among workers with lowest ability.

6 Conclusion

This paper presents a two-period partial-equilibrium model that systematically compares the costs and benefits of penalty charges. As discussed in section 2, there are two theoretical explanations in the

[48] These partial derivatives are calculated in appendix C.6.
[49] In a general-equilibrium analysis, also implications for skill prices would have to be considered.

Worker Type	**Welfare Difference**
Workers with low ability ($0 \leq \theta < \theta_U^{PC}$)	$-\beta\,(1+\delta\sigma)\,\theta < 0$
Workers with ability below average ($\theta_U^{PC} \leq \theta < \theta_A^{PC}$)	± 0
Workers with ability above average ($\theta_A^{PC} \leq \theta < \theta^{LF}$)	$w_A + \beta\,(-1+\delta\alpha)\,\theta > 0$
Workers with high ability ($\theta^{LF} \leq \theta \leq 1$)	± 0

Table 2: Welfare Effects for Workers Depending on Individual Ability

literature for firms providing general training. Our formal analysis is based on the recent training literature with oligopsonistic labor markets but the model is adaped to the German system of apprenticeship training with a fixed length of the apprenticeship and an identical training wage independent of individual productivity. Furthermore, the model incorporates worker heterogeneity in ability which allows to analyze the welfare implications of penalty charges for different groups of workers.

In the laissez-faire equilibrium, the training decision of firms bears three different kinds of inefficiencies. In consequence, the number of apprenticeship training positions is too low compared to the first-best optimum. In line with the results of Acemoglu and Pischke (1999b), firms do not take into account the benefits from increased productivity accruing both for workers and other employers in the future. These two inefficiencies are increasing in the bargaining power of workers and the separating probability after the training period, respectively. In our model, there is a third kind of inefficiency in the number of apprenticeship training positions which is generated by the fixed training wage in period 1. However, penalty charges can reduce this third kind of inefficiency and thus increase the number of apprenticeship training positions. Hence, the implementation of penalty charges moves the economy towards its first-best optimum with respect to the labor market decision between apprenticeship training and regular work.

The formal analysis shows that the optimal policy depends on the productivity-enhancement of apprenticeship training. If the increase in productivity due to an apprenticeship is low, it is optimal to reject the implementation of penalty charges. In this case, the laissez-faire equilibrium corresponds to the first-best optimum. However, optimal penalty charges increase the overall welfare if the productivity-enhancement exceeds some lower bound which guarantees that the participation constraint is satisfied for a sufficient number of workers. This lower bound of productivity-enhancement is decreasing in the relative efficiency of apprentices, the discount factor and the bargaining power of workers. However, it is increasing in the separating probability after the first period.

With respect to overall welfare, the implementation of penalty charges solves a trade-off. On the one hand, penalty charges are welfare-enhancing because they increase the number of apprenticeship training positions and thus the fraction of skilled workers in the working period. On the other hand, penalty charges are costly because some firms will leave the market which generates unemployment among workers with low ability. Additionally, the implementation of penalty charges causes administration costs that have to be financed by distortionary taxation.

The overall welfare with optimal penalty charges is increasing in the productivity-enhancement of apprenticeship training. However, it is decreasing in the depreciation rate of skills during unemployment, the separating probability after the training period, the efficiency loss of distortionary taxation, the level of administration costs, and the training wage. There are two groups of workers that are affected by the implementation of penalty charges. While workers with middle ability above average benefit from penalty charges because they now receive apprenticeship training, low-ability workers are hurt because they are exposed to unemployment. The welfare of high-ability workers and of workers with middle-ability below average remains unchanged.

In our model, the number of apprenticeship training positions and the number of unemployed workers are endogenously determined depending on the individual ability of workers. This extension advances the analysis of Malcomson, Maw, and McCormick (2003), where the number of trained workers is determined by some fixed costs of training. Nevertheless, our model has been kept simple for expositional and calculational reasons. For example, we assume a uniform distribution of abilities and do not explicitly explain the process of collective wage setting for apprentices. However, the underlying insights into the model presented here are reasonably robust to various types of generalization. Hence, they constitute a promising basis for further research.

A The Productivity-Enhancement of Apprenticeship Training

The productivity-enhancement α is assumed to be identical for all trained workers because it unambiguously depends on the amount of training t which is chosen by the training firms at the intensive margin. Analytically, this training decision at the intensive margin is considered before investigating the firms' decision at the extensive margin (i.e. whether to provide apprenticeship training, to offer regular work, or to leave the market). The parameters $\alpha \equiv \alpha(t)$ with $\alpha'(t) > 0$ and $\chi \equiv \chi(t)$ with $\chi'(t) < 0$ are assumed to depend on the amount of training t. Each firm maximizes the discounted sum of its profits over both periods

$$\pi(\theta, t) = \chi(t)\,\theta - w_A + \delta(1-\rho)(1-\beta)(1+\alpha(t))\,\theta + \delta\rho(1-\beta)\,E[\theta_2] \tag{A1}$$

Suppose that the impact of t on the expected productivity in period 2 is neglected. Hence, the optimal amount of t is defined by the FOC

$$\begin{aligned}\chi'(t)\,\theta + \delta(1-\rho)(1-\beta)\,\alpha'(t)\,\theta &= 0\\ \delta(1-\rho)(1-\beta)\,\alpha'(t) &= -\chi'(t)\end{aligned} \tag{A2}$$

While the LHS of equation (A2) describes the marginal return by an additional unit of t, the RHS defines the marginal costs. The optimal amount of training t^* per apprentice is independent of θ. This result is also justified by economic intuition because the apprenticeship follows a prescribed curriculum (cf. section 3.2). Therefore, also the productivity-enhancement $\alpha(t^*)$ is independent of θ and thus identical for all workers.[50]

[50] Note that δ describes the relative weight of the second period so that we can interpret $(1-\delta)$ as the length of the apprenticeship. Because t^* is decreasing in δ, it is increasing in $(1-\delta)$. This is the main result of Malcomson, Maw, and McCormick (2003): increasing the length of the apprenticeship by government regulation increases the optimal amount of training at the intensive margin.

This situation can be illustrated by an example. Suppose that the functional forms of the parameters are $\alpha(t) = t$ and $\chi(t) = e^{-t}$. In this case, the FOC is equal to

$$\delta(1-\rho)(1-\beta) = e^{-t} \tag{A3}$$

Solving for t yields the optimal amount of training

$$t^{*} = -\ln[\delta(1-\rho)(1-\beta)] \tag{A4}$$

Hence, the productivity of trained workers in period 2 is equal to

$$\theta_2 = (1+\alpha(t^{*}))\,\theta = (1-\ln[\delta(1-\rho)(1-\beta)])\,\theta \tag{A5}$$

In line with Ben-Porath (1967) and Heckman (1976), the stock of human capital in the second period (θ_2) is linearly depending on θ. However, the productivity-enhancement α is independent of the individual ability θ.

B The Relationship of the Pivotal Productivities

With respect to the labor market decision of firms, the four pivotal productivities are the following (cf. sections 4.3.1 and 5.1):

$$\theta^{LF} = \frac{w_A}{\chi-(1-\beta)+\delta(1-\rho)(1-\beta)\alpha} \;;\; \theta_A^{PC} = \frac{w_A - T}{\chi-(1-\beta)+\delta(1-\rho)(1-\beta)\alpha}$$
$$\dot{\theta} = \frac{w_A}{\chi+\delta(1-\rho)(1-\beta)(1+\alpha)} \;;\; \theta_U^{PC} = \frac{T}{(1-\beta)(1+\delta(1-\rho))}$$

By comparing these pivotal productivities we obtain the following relationships:

$$\theta^{LF} > 0 \;;\; \theta_A^{PC} > 0 \quad if \;\; T < w_A \;;\; \dot{\theta} \geq 0 \;;\; \theta_U^{PC} \geq 0 \tag{B1}$$

$$\theta^{LF} \leq 1 \;;\; \theta_A^{PC} \leq 1 \;;\; \dot{\theta} < 1 \;;\; \theta_U^{PC} \leq 1 \quad if \;\; T \leq (1-\beta)(1+\delta(1-\rho)) \tag{B2}$$

$$\theta^{LF} \geq \theta_A^{PC} \;;\; \theta^{LF} > \dot{\theta} \;;\; \theta^{LF} \geq (<)\theta_U^{PC} \quad if \;\; T \leq (>)\frac{(1-\beta)(1+\delta(1-\rho))}{\chi-(1-\beta)+\delta(1-\rho)(1-\beta)\alpha} w_A \tag{B3}$$

$$\theta_A^{PC} \geq (<)\dot{\theta} \quad if \;\; T \leq (>)\bar{T} \equiv \frac{(1-\beta)(1+\delta(1-\rho))}{\chi+\delta(1-\rho)(1-\beta)(1+\alpha)} w_A \tag{B4}$$

$$\theta_A^{PC} \geq (<)\theta_U^{PC} \quad if \;\; T \leq (>)\bar{T} \;;\; \dot{\theta} \geq (<)\theta_U^{PC} \quad if \;\; T \leq (>)\bar{T} \tag{B5}$$

Combining these relationships yields three cases that are mathematically feasible:

$$(1)\quad 1 \geq \theta^{LF} \geq \theta_A^{PC} \geq \dot{\theta} \geq \theta_U^{PC} \geq 0 \qquad if\ \ 0 \leq T \leq \bar{T} \tag{B6}$$

$$(2)\quad 1 \geq \theta^{LF} \geq \theta_U^{PC} > \dot{\theta} > \theta_A^{PC} \qquad if\ \ \bar{T} < T \leq \frac{(1-\beta)(1+\delta(1-\rho))}{\chi-(1-\beta)+\delta(1-\rho)(1-\beta)\alpha} w_A \tag{B7}$$

$$(3)\quad \theta_U^{PC} > \theta^{LF} > \dot{\theta} > \theta_A^{PC} \qquad if\ \ T > \frac{(1-\beta)(1+\delta(1-\rho))}{\chi-(1-\beta)+\delta(1-\rho)(1-\beta)\alpha} w_A \tag{B8}$$

With respect to our formal analysis, we only have to distinguish between the folowing two cases:

$$(I)\quad 1 \geq \theta^{LF} \geq \theta_A^{PC} \geq \dot{\theta} \geq \theta_U^{PC} \geq 0 \qquad if\ \ 0 \leq T \leq \bar{T} \tag{B9}$$

$$(II)\quad 1 \geq \theta^{LF} \geq \theta_U^{PC} > \dot{\theta} > \theta_A^{PC}\ \ or\ \ \theta_U^{PC} > \theta^{LF} > \dot{\theta} > \theta_A^{PC} \qquad if\ \ T > \bar{T} \tag{B10}$$

C The Optimal Penalty Charges

C.1 Calculation of Optimal Penalty Charges

The optimization problem is described in section 5.2. Substituting the pivotal productivities θ_A^{PC} and θ_U^{PC} into the FOC (36) yields

$$\frac{\partial W^{PC}(T)}{\partial T} = 0$$

$$(\chi - 1 + \delta\alpha)\,\theta_A^{PC}\left(-\frac{\partial\theta_A^{PC}}{\partial T}\right) = (1+\delta\sigma)\,\theta_U^{PC}\frac{\partial\theta_U^{PC}}{\partial T} + (1+\lambda)\,cT$$

$$(\chi - 1 + \delta\alpha)\frac{w_A - T}{(v_3)^2} = (1+\delta\sigma)\frac{T}{(v_4)^2} + (1+\lambda)\,cT$$

with $v_3 \equiv (\chi - (1-\beta) + \delta(1-\rho)(1-\beta)\alpha)$ and $v_4 \equiv (1-\beta)(1+\delta(1-\rho))$.

From this we obtain

$$\left(\frac{\chi-1+\delta\alpha}{(v_3)^2} + \frac{1+\delta\sigma}{(v_4)^2} + (1+\lambda)\,c\right)T = \frac{\chi-1+\delta\alpha}{(v_3)^2} w_A$$

$$\frac{v_1(v_4)^2 + (1+\delta\sigma)(v_3)^2 + (v_3)^2(v_4)^2(1+\lambda)\,c}{(v_3)^2(v_4)^2}T = \frac{v_1}{(v_3)^2} w_A$$

with $v_1 \equiv (\chi - 1 + \delta\alpha)$ and $v_2 \equiv (1 + (v_4)^2(1+\lambda)\,c + \delta\sigma)$.

Solving for T yields the optimal penalty charges (cf. equation (37) in section 5.2)

$$T^* = \frac{v_1(v_4)^2}{v_1(v_4)^2 + (1+\delta\sigma)(v_3)^2 + (v_3)^2(v_4)^2(1+\lambda)\,c} w_A$$

$$T^* = \frac{v_1(v_4)^2}{v_1(v_4)^2 + v_2(v_3)^2} w_A \tag{C1}$$

C.2 Requirements for the Productivity-Enhancement

For the case $0 \leq T \leq \bar{T} \equiv \frac{(1-\beta)(1+\delta(1-\rho))}{\chi+\delta(1-\rho)(1-\beta)(1+\alpha)} w_A = \frac{v_4}{v_3+v_4} w_A$, the optimal penalty charges are given by equation (C1). Hence, T^* has to satisfy the following two conditions in order to lie within the required interval:

$$T^* \geq 0 \quad \Leftrightarrow \quad \alpha \geq \frac{1-\chi}{\delta} \tag{C2}$$

and

$$\begin{gathered}
T^* \leq \frac{v_4}{v_3+v_4} w_A \\
\frac{v_1 v_4}{v_1 (v_4)^2 + v_2 (v_3)^2} \leq \frac{1}{v_3+v_4} \\
v_1 v_3 v_4 \leq v_2 (v_3)^2 \\
\delta\alpha \left[v_4 - (1-\rho)(1-\beta) v_2\right] \leq (1-x) v_4 + (\chi - (1-\beta)) v_2 \\
\delta (1-\beta) \alpha \left[1 - (1-\rho)(v_2 - \delta)\right] \leq (1-x) v_4 + (\chi - (1-\beta)) v_2
\end{gathered}$$

We assume that administration costs exceed some lower bound

$$c \geq \bar{c} \equiv \frac{1-(1-\rho)(1-\delta(1-\sigma))}{(1-\rho)(v_4)^2(1+\lambda)} \Leftrightarrow 1-(1-\rho)(v_2-\delta) \leq 0 \tag{C3}$$

so that the inequality is satisfied for

$$\alpha \geq 0 \tag{C4}$$

Altogether, by summarizing the inequalities (C2) and (C4), α must satisfy the following condition:

$$\alpha \geq \frac{1-\chi}{\delta} \tag{C5}$$

C.3 Optimal Penalty Charges if the Participation Constraint of Workers is not Satisfied

The participation constraint of workers may not be satisfied if the productivity-emhancement of apprenticeship training is too low, i.e. if $\alpha < \frac{1}{\delta}$ (cf. section 4.3.2). With respect to the optimal penalty charges as deducted in section 5.2, there are three cases that have to be considered:[51]

(a) $\theta^W \geq \theta^{LF} \;\Leftrightarrow\; \frac{1-\chi}{\delta(1-\rho(1-\beta))} \leq \alpha < \frac{1}{\delta}$

In case (a), the overall welfare is equal to

$$W^{(a)} = \frac{1}{2}(\chi + \delta(1+\alpha)) - \frac{1}{2}v_1(\theta_A^{PC})^2 - \frac{1}{2}v_1[1-(\theta^W)^2] - \frac{1}{2}(1+\delta\sigma)(\theta_U^{PC})^2 - (1+\lambda)\frac{c}{2}T^2 \tag{C6}$$

[51] Note that $\frac{1-\chi}{\delta(1-\rho(1-\beta))} \leq \frac{1}{\delta}$ is implied by $\chi \geq 1-\beta$.

Because θ^W is independent of T, the optimal penalty charges are the same as in the standard case:

$$T^{opt} = T^* = \frac{v_1 (v_4)^2}{v_1 (v_4)^2 + v_2 (v_3)^2} w_A \tag{C7}$$

Although workers with high ability refuse apprenticeship training, the overall welfare with T^* is higher than in the laissez-faire equilibrium (cf. equation (23)) because

$$\begin{aligned} (W^{(a)})^* &= \frac{1}{2}(1+\delta) + \frac{1}{2}v_1 \left(\theta^W\right)^2 - \frac{1}{8}\frac{\beta^2\chi^2 v_1 v_2}{v_1 (v_4)^2 + v_2 (v_3)^2} \\ &> \frac{1}{2}(1+\delta) + \frac{1}{2}v_1 \left(\theta^W\right)^2 - \frac{1}{2}v_1 \left(\theta^{LF}\right)^2 = W^{LF} \end{aligned}$$

if $v_1 > 0 \Leftrightarrow \alpha > \frac{1-\chi}{\delta}$ which is implied by the lower bound of the interval for α.

(b) $\theta^W < \theta^{LF} \wedge \theta^W < \theta_A^{PC} \Leftrightarrow \alpha < \frac{1-\chi}{\delta(1-\rho(1-\beta))} \wedge T < \frac{1-\chi-\delta\alpha(1-\rho(1-\beta))}{2(1-\delta\alpha)}\chi$

In case (b), the overall welfare is equal to

$$W^{(b)} = \frac{1}{2}(1+\delta) - \frac{1}{2}(1+\delta\sigma)(\theta_U^{PC})^2 - (1+\lambda)\frac{c}{2}T^2 \tag{C8}$$

Because $W^{(b)}$ is decreasing in T, the optimal penalty charges are equal to zero (i.e. $T^{opt} = 0$) which implies[52]

$$(W^{(b)})^* = \frac{1}{2}(1+\delta) = W^{LF} \tag{C9}$$

(c) $\theta_A^{PC} \leq \theta^W < \theta^{LF} \Leftrightarrow \alpha < \frac{1-\chi}{\delta(1-\rho(1-\beta))} \wedge T \geq \frac{1-\chi-\delta\alpha(1-\rho(1-\beta))}{2(1-\delta\alpha)}\chi$

In case (c), the overall welfare is the same as in case (a). However, because T^* is not feasible (i.e. $T^* < \frac{1-\chi-\delta\alpha(1-\rho(1-\beta))}{2(1-\delta\alpha)}\chi$) it is optimal to choose $T^{opt} = 0$ as suggested in case (b). Hence, the overall welfare with optimal penalty charges in case (c) is equal to

$$(W^{(c)})^* = \frac{1}{2}(1+\delta) = W^{LF} \tag{C10}$$

Proposition 11 *Altogether, the optimal penalty charges for $\alpha < \frac{1}{\delta}$ are equal to*

$$T^{opt} = \left\{ \begin{array}{lll} 0 & if & \alpha < \frac{1-\chi}{\delta(1-\rho(1-\beta))} \\ T^* & if & \frac{1-\chi}{\delta(1-\rho(1-\beta))} \leq \alpha < \frac{1}{\delta} \end{array} \right\} \tag{C11}$$

[52] For $\alpha < \frac{1-\chi}{\delta(1-\rho(1-\beta))}$, the overall welfare in the laissez-faire equilibrium is equal to $W^{LF} = \frac{1}{2}(1+\delta)$ (cf. equation (23)).

C.4 Comparative Statics of Optimal Penalty Charges

It is important to analyze in which way the optimal penalty charges are affected by changes in the key parameters of the model. For $\alpha \geq \frac{1-\chi}{\delta(1-\rho(1-\beta))}$, the comparative statics with respect to productivity-enhancement, depreciation rate, separation probability, administration costs, efficiency loss, and training wage are as follows:

$$\frac{\partial T^*}{\partial \alpha} = \frac{\delta v_2 v_3 (v_4)^2 \left[v_3 - 2(1-\rho)(1-\beta) v_1\right]}{\left[v_1 (v_4)^2 + v_2 (v_3)^2\right]^2} w_A$$

$$\rightarrow \frac{\partial T^*}{\partial \alpha} \left\{ \begin{array}{ll} \geq 0 & if \;\; \alpha \leq \alpha' \equiv 2\frac{1-\chi}{\delta} + \frac{\chi-(1-\beta)}{\delta(1-\rho)(1-\beta)} \\ < 0 & if \;\; \alpha > \alpha' \end{array} \right\} \tag{C12}$$

$$\frac{\partial T^*}{\partial \sigma} = \frac{-\delta v_1 (v_3)^2 (v_4)^2}{\left[v_1 (v_4)^2 + v_2 (v_3)^2\right]^2} w_A < 0 \tag{C13}$$

$$\frac{\partial T^*}{\partial \rho} = \frac{2\delta (1-\beta) v_1 v_3 v_4 \left[v_2 v_4 \alpha - v_3 (1+\delta\sigma)\right]}{\left[v_1 (v_4)^2 + v_2 (v_3)^2\right]^2} w_A > 0 \tag{C14}$$

$$\frac{\partial T^*}{\partial c} = \frac{-v_1 (v_3)^2 (v_4)^4 (1+\lambda)}{\left[v_1 (v_4)^2 + v_2 (v_3)^2\right]^2} w_A < 0 \tag{C15}$$

$$\frac{\partial T^*}{\partial \lambda} = \frac{-v_1 (v_3)^2 (v_4)^4 c}{\left[v_1 (v_4)^2 + v_2 (v_3)^2\right]^2} w_A < 0 \tag{C16}$$

$$\frac{\partial T^*}{\partial w_A} = \frac{v_1 (v_4)^2}{v_1 (v_4)^2 + v_2 (v_3)^2} > 0 \tag{C17}$$

C.5 Overall Welfare with Optimal Penalty Charges

The optimal overall welfare with penalty charges is determined by substituting the optimal penalty charges into equation (35):

$$
\begin{aligned}
(W^{PC})^* &= \frac{1}{2}(\chi+\delta(1+\alpha)) - \frac{1}{2}(\chi-1+\delta\alpha)(\theta_A^{PC})^2 \\
&\quad -\frac{1}{2}(1+\delta\sigma)(\theta_U^{PC})^2 - (1+\lambda)\frac{c}{2}(T^*)^2 \\
&= \frac{1}{2}(\chi+\delta(1+\alpha)) - \frac{1}{2}v_1\left(\frac{w_A-T^*}{v_3}\right)^2 \\
&\quad -\frac{1}{2}(1+\delta\sigma)\left(\frac{T^*}{v_4}\right)^2 - (1+\lambda)\frac{c}{2}(T^*)^2 \\
&= \frac{1}{2}(\chi+\delta(1+\alpha)) - \frac{1}{2}\frac{v_1(v_4)^2+v_2(v_3)^2}{(v_3)^2(v_4)^2}(T^*)^2 \\
&\quad +\frac{v_1}{(v_3)^2}w_A T^* - \frac{1}{2}\frac{v_1}{(v_3)^2}(w_A)^2 \\
&= \frac{1}{2}(\chi+\delta(1+\alpha)) \\
&\quad -\frac{1}{2}\frac{v_1(v_4)^2+v_2(v_3)^2}{(v_3)^2(v_4)^2}\left(\frac{v_1(v_4)^2}{v_1(v_4)^2+v_2(v_3)^2}w_A\right)^2 \\
&\quad +\frac{v_1}{(v_3)^2}w_A\left(\frac{v_1(v_4)^2}{v_1(v_4)^2+v_2(v_3)^2}w_A\right) - \frac{1}{2}\frac{v_1}{(v_3)^2}(w_A)^2 \\
&= \frac{1}{2}(\chi+\delta(1+\alpha)) \\
&\quad +\frac{1}{2}\frac{(v_1)^2(v_4)^2}{(v_3)^2\left[v_1(v_4)^2+v_2(v_3)^2\right]}(w_A)^2 - \frac{1}{2}\frac{v_1}{(v_3)^2}(w_A)^2
\end{aligned}
$$

Simplifying yields

$$
(W^{PC})^* = \frac{1}{2}(\chi+\delta(1+\alpha)) - \frac{1}{2}\frac{v_1 v_2}{v_1(v_4)^2+v_2(v_3)^2}(w_A)^2 \tag{C18}
$$

C.6 Comparative Statics of Optimal Overall Welfare

It is important to analyze in which way the overall welfare with optimal penalty charges is affected by changes in the key parameters of the model. The comparative statics with respect to productivity-enhancement, depreciation rate, separating probability, training wage, administration costs, and efficiency loss are as follows:

$$\frac{\partial (W^{PC})^*}{\partial \alpha} = \frac{1}{2}\delta \frac{(v_1)^2 (v_4)^4 + 2v_1 v_2 (v_3)^2 (v_4)^2 + (v_2)^2 (v_3)^2 \left[(v_3)^2 - (w_A)^2\right] + 2(1-\rho)(1-\beta) v_1 (v_2)^2 v_3 (w_A)^2}{\left[v_1 (v_4)^2 + v_2 (v_3)^2\right]^2} > 0 \quad \text{(C19)}$$

$$\frac{\partial (W^{PC})^*}{\partial \sigma} = -\frac{1}{2}\frac{\delta (v_1)^2 (v_4)^2}{\left[v_1 (v_4)^2 + v_2 (v_3)^2\right]^2} (w_A)^2 < 0 \quad \text{(C20)}$$

$$\frac{\partial (W^{PC})^*}{\partial \rho} = -\frac{\delta (1-\beta) v_1 \left[(1+\delta\sigma) v_1 v_4 + (v_2)^2 v_3 \alpha\right]}{\left[v_1 (v_4)^2 + v_2 (v_3)^2\right]^2} (w_A)^2 < 0 \quad \text{(C21)}$$

$$\frac{\partial (W^{PC})^*}{\partial w_A} = -\frac{v_1 v_2}{v_1 (v_4)^2 + v_2 (v_3)^2} w_A < 0 \quad \text{(C22)}$$

$$\frac{\partial (W^{PC})^*}{\partial c} = -\frac{1}{2}\frac{(v_1)^2 (v_4)^4 (1+\lambda)}{\left[v_1 (v_4)^2 + v_2 (v_3)^2\right]^2} (w_A)^2 < 0 \quad \text{(C23)}$$

$$\frac{\partial (W^{PC})^*}{\partial \lambda} = -\frac{1}{2}\frac{(v_1)^2 (v_4)^4 c}{\left[v_1 (v_4)^2 + v_2 (v_3)^2\right]^2} (w_A)^2 < 0 \quad \text{(C24)}$$

D Case (II)

D.1 Relevance

Case (II) refers to the following relationship of the pivotal productivities (cf. appendix B):

$$1 \geq \theta^{LF} \geq \theta_U^{PC} > \dot{\theta} > \theta_A^{PC} \quad or \quad \theta_U^{PC} > \theta^{LF} > \dot{\theta} > \theta_A^{PC}$$

Case (II) becomes relevant if the optimal penalty charges would exceed the critical level $\bar{T}$ (cf. apendix C.2):

$$T^* > \bar{T} \equiv \frac{v_4}{v_3 + v_4} w_A \quad \text{(D1)}$$

This implies

$$\delta (1-\beta) \alpha \left[1 - (1-\rho)(v_2 - \delta)\right] > (1-x) v_4 + (\chi - (1-\beta)) v_2$$

which is equivalent to

$$c < \bar{c} \ \wedge \ \alpha > \alpha_{\bar{T}} \equiv \frac{(1-x) v_4 + (\chi - (1-\beta)) v_2}{\delta (1-\beta) \left[1 - (1-\rho)(v_2 - \delta)\right]} \quad \text{(D2)}$$

D.2 Optimal Penalty Charges in Case (II)

High penalty charges with $T > \bar{T}$ imply that the number of apprenticeship training positions is maximized because all workers with $\theta \geq \dot{\theta}$ are trained. All other workers stay unemployed, i.e. no firm decides to

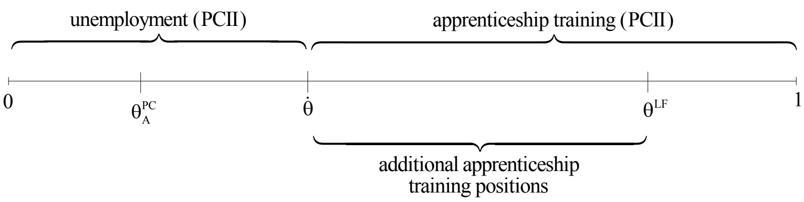

Figure 8: The Equilibrium with Penalty Charges in Case (II)

employ workers regularly because $\theta_A^{PC} < \theta_U^{PC}$.[53] This situation is illustrated in figure 8. Hence, the overall welfare with penalty charges in case (II) depends on the pivotal productivity $\dot{\theta}$ in the following manner:

$$W^{PC(II)} = \underbrace{\int_{\dot{\theta}}^{1} \left(\chi + \delta\left(1+\alpha\right)\right)\theta d\theta}_{apprenticeship\ training} + \underbrace{\int_{0}^{\dot{\theta}} \delta\left(1-\sigma\right)\theta d\theta}_{regular\ work} - \left(1+\lambda\right)C\left(T\right) \tag{D3}$$

Simplifying yields

$$W^{PC(II)} = \frac{1}{2}\left(\chi + \delta\left(1+\alpha\right)\right) - \frac{1}{2}\left(\chi + \delta\left(\alpha+\sigma\right)\right)\dot{\theta}^2 - \left(1+\lambda\right)\frac{c}{2}T^2 \tag{D4}$$

Proposition 12 *In case (II), the optimal penalty charges have to be as small as possible because $W^{PC(II)}$ is strictly decreasing in T. Hence, the optimal penalty charges in case (II) are equal to*[54]

$$T^*_{(II)} = \bar{T} \equiv \frac{v_4}{v_3 + v_4} w_A \tag{D5}$$

D.3 Optimal Penalty Charges in Both Cases

Taken together cases (I) and (II), the optimal level of penalty charges explicitly depends on the productivity-enhancement of apprenticeship training and the level of administration costs. If α is very low, it is optimal to reject the implementation of penalty charges, i.e. to set $T^{opt} = 0$. However, if α exceeds the critical level $\frac{1-\chi}{\delta(1-\rho(1-\beta))}$ it is optimal to implement penalty charges according to $T^{opt} = T^*$. Only if administration costs are low and α is very high (cf. condition (D2)), it is optimal to choose $T^{opt} = \bar{T}$. Note that the optimal penalty charges are bounded above by $\bar{T}$ which maximizes the number of apprenticeship training positions at the cost of suppressed regular work. The result for low administration costs is summarized in the following proposition and illustrated in figure 9.

[53] Case (II) also implies that the number of unemployed workers is maximized. Note that the overall welfare is not depending on θ_U^{PC} being smaller or greater than 1.

[54] Because case (II) becomes relevant only for $T > \bar{T}$, the optimal penalty charges must be slightly bigger than $\bar{T}$. However, to simplify matters, we assume $T^*_{(II)} = \bar{T}$.

Figure 9: Optimal Penalty Charges in Case (II)

Proposition 13 *For $c < \bar{c}$, the optimal penalty charges are equal to*[55]

$$T^{opt} = \left\{ \begin{array}{lll} 0 & if & \alpha < \frac{1-\chi}{\delta(1-\rho(1-\beta))} \\ T^{*} & if & \frac{1-\chi}{\delta(1-\rho(1-\beta))} \leq \alpha \leq \alpha_{\bar{T}} \\ \bar{T} & if & \alpha > \alpha_{\bar{T}} \end{array} \right\} \quad \text{(D6)}$$

References

Acemoglu, D. (1997): "Training and Innovation in an Imperfect Labour Market," *The Review of Economic Studies*, 64(3), 445–464.

Acemoglu, D., and J.-S. Pischke (1998a): "Beyond Becker: Training in Imperfect Labor Markets," *NBER Working Paper*, 6740.

——— (1998b): "Why Do Firms Train? Theory and Evidence," *The Quarterly Journal of Economics*, 113(1), 79–119.

——— (1999a): "Minimum Wages and On-the-Job Training," *NBER Working Paper*, 7184.

——— (1999b): "The Structure of Wages and Investment in General Training," *The Journal of Political Economy*, 107(3), 539–572.

Balmaceda, F. (2001): "Firm-Sponsored General Training in a Frictionless Labor Market," *ILADES-Georgetown University Working Paper*, 134.

Becker, G. S. (1962): "Investment in Human Capital: A Theoretical Analysis," *The Journal of Political Economy*, 70(5), 9–49.

——— (1964): *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Columbia University Press, New York.

Belot, M., J. Boone, and J. V. Ours (2002): "Welfare Effects of Employment Protection," *CEPR Discussion Paper*, 3396.

[55] If administration costs are near to zero, α_0 may exceed $\alpha_{\bar{T}}$. In this case, high penalty charges $\bar{T}$ are set as soon as the participation constraint of workers is satisfied.

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

Bishop, J. H. (1997): "What We Know About Employer-Provided Training: A Review of the Literature," *Research in Labor Economics*, 16, 19–87.

Bloch, F. E., and S. P. Smith (1977): "Human Capital and Labor Market Employment," *The Journal of Human Resources*, 12(4), 550–560.

Blundell, R., L. Dearden, C. Meghir, and B. Sianesi (1999): "Human Capital Investment: The Returns from Education and Training to the Individual, the Firm and the Economy," *Fiscal Studies*, 20(1), 1–23.

BMBF (2004): *Berufsbildungsbericht*. Bundesministerium für Bildung und Forschung, Bonn.

Boone, J., and L. Bovenberg (2006): "Optimal Welfare and in-Work Benefits with Search Unemployment and Observable Abilities," *Journal of Economic Theory*, 126, 165–193.

Bougheas, S., and Y. Georgellis (2004): "Early Career Mobility and Earning Profiles of German Apprentices: Theory and Empirical Evidence," *Labour*, 18(2), 233–263.

Brown, E., and H. Kaufold (1988): "Human Capital Accumulation and the Optimal Level of Unemployment Insurance Provision," *Journal of Labor Economics*, 6(4), 493–514.

Brunello, G. (2002): "Is Training More Frequent When Wage Compression is Higher? Evidence from 11 European Countries," *CESifo Working Paper*, 637.

Cahuc, P., and A. Zylberberg (2004): *Labor Economics*. MIT Press, Cambridge.

Chang, C., and Y. Wang (1996): "Human Capital Investment under Asymmetric Information: The Pigovian Conjecture Revisited," *Journal of Labor Economics*, 14(3), 505–519.

Cooke, L. P. (2003): "A Comparison of Initial and Early Life Course Earnings of the German Secondary Education and Training System," *Economics of Education Review*, 22, 79–88.

de Meza, D., and D. C. Webb (2001): "Advantageous Selection in Insurance Markets," *RAND Journal of Economics*, 32(2), 249–262.

Eckaus, R. S. (1963): "Investment in Human Capital: Comment," *The Journal of Political Economy*, 71(5), 501–504.

Elmeskov, J., J. Marint, and S. Scarpetta (1998): "Key Lessons for Labour Market Reforms: Evidence from OECD Countries' Experiences," *Swedish Economic Policy Review*, 5(2), 205–252.

Freeman, R. B., and J. L. Medoff (1984): *What Do Unions Do?* Basic Books, New York.

Goode, R. B. (1959): "Adding to the Stock of Physical and Human Capital," *The American Economic Review*, 49(2), 147–155.

Hamilton, G. (1996): "The Market for Montreal Apprentices: Contract Length and Information," *Explorations in Economic History*, 33(4), 496–523.

Hanushek, E. A., C. K. Y. Leung, and K. Yilmaz (2003): "Redistribution Through Education and Other Transfer Mechanisms," *Journal of Monetary Economics*, 50, 1719–1750.

Harhoff, D., and T. J. Kane (1995): *Financing Apprenticeship Training: Evidence from Germany.* Office of Educational Research and Improvement (ED), Washington, D.C.

——— (1997): "Is the German Apprenticeship System a Panacea for the U.S. Labor Market?," *Journal of Population Economics*, 10(2), 171 – 196.

Heckman, J. J. (1976): "A Life-Cycle Model of Earnings, Learning, and Consumption," *The Journal of Political Economy*, 84(4), 11–44.

Jones, I. S. (1986): "Apprentice Training Costs in British Manufacturing Establishments: Some New Evidence," *British Journal of Industrial Relations*, 24(3), 333–362.

Kessler, A. S., and C. Luelfesmann (2002): "The Theory of Human Capital Revisited: On the Interaction of General and Specific Investments," *CESifo Working Paper*, 776.

Lazear, E. P. (2003): "Firm-Specific Human Capital: A Skill-Weights Approach," *NBER Working Paper*, 9679.

Loewenstein, M. A., and J. R. Spletzer (1998): "Deviding the Costs and Returns to General Training," *Journal of Labor Economics*, 16(1), 142–171.

——— (1999): "General and Specific Training: Evidence and Implications," *The Journal of Human Resources*, 34(4), 710–733.

Lynch, L. (1991): "The Role of Off-the-Job Vs. On-the-Job Training for the Mobility of Women Workers," *American Economic Review*, 81(2), 151–156.

Malcomson, J. M., J. W. Maw, and B. McCormick (2003): "General Training by Firms, Apprentice Contracts, and Public Policy," *European Economic Review*, 47(2), 197–227.

Masters, A. M. (1998): "Efficiency of Investment in Human and Physical Capital in a Model of Bilateral Search and Bargaining," *International Economic Review*, 39(2), 477–494.

Mincer, J. (1958): "Investment in Human Capital and Personal Income Distribution," *The Journal of Political Economy*, 6(4), 281–302.

——— (1974): *Schooling, Earnings, and Experience*. Columbia University Press, New York.

——— (1989): "Human Capital and the Labor Market: A Review of Current Research," *Educational Research*, 18(4), 27–34.

——— (1997): "The Production of Human Capital and the Life Cycle of Earnings: Variation on a Theme," *Journal of Labor Economics*, 15(1), 26–47.

Nickell, S. (1998): "Unemployment: Questions and Some Answers," *Economic Journal*, 108, 802–816.

Parsons, D. O. (1974): "The Cost of School Time, Foregone Earnings, and Human Capital Formation," *The Journal of Political Economy*, 82(2), 251–266.

Pirenne, H. (1936): *Economic and Social History of Medieval Europe*. Harcourt Brace, New York.

Ryan, P. (1980): "The Costs of Job Training for a Transferable Skill," *British Journal of Industrial Relations*, 18(3), 334–352.

Scarpetta, S. (1996): "Assessing the Role of Labour Market Policies and Institutional Settings on Unemployment: A Cross-Country Study," *OECD Econimic Studies*, 26, 43–98.

Soskice, D. (1994): *Reconciling Markets and Institutions: The German Apprenticeship System*. The University of Chicago Press, Chicago.

Veum, J. R. (1999): "Training, Wages, and the Human Capital Model," *Southern Economic Journal*, 65(3), 526–538.

VonBardeleben, R., U. Beicht, and K. Fehér (1995): *Betriebliche Kosten und Nutzen der Ausbildung - Repräsentative Ergebnisse Aus Industrie, Handel und Handwerk*, Berichte Zur Beruflichen Bildung 187. Bertelsmann, Bielefeld.