Explaining the Supply of Apprenticeships

Wolfgang Schwerdt*

- Preliminary Version -

First version 5.5.99
This version 1.7.99

Abstract

This paper develops a model to explain the supply of apprenticeships by firms through endogenising monopsony creating characteristics of post training labour markets to training programs. It applies the formulation of training effects introduced by STEVENS (1994, 1996). Abolishing the « classical dichotomy » of general versus specific training in favour of implementing value vectors with varying transferability, the model is able to show which characteristics of training programs make firms actually pay for non-firm-specific parts of apprenticeship training. Using a multistage sequential Bayesian bargaining model it is able to capture and analyse the two main reasons for firm-sponsored training: Negative net costs during training and a positive expected share of the increased marginal product after training.

*Université de Paris I and CREST

1 Why Do Firms Pay For Training?

There are two reasons why economists study apprenticeship systems: Youth unemployment and training. Youth unemployment is seen widely as one of the most severe forms of failure in the labour market. Several papers indicate that being unemployed at an early stage of one's labour market experience has strong negative effects on the ability to find a stable working relationship in subsequent years[1]. Early unemployment very often is caused by a discontinuance in human capital formation or by sending corresponding signals to future employers, leading to a non-smooth transition from to work. This interconnection between succesful completion of intial post-secondary professional training and early labour market experience switches the focus on Youth Training Schemes (YTS) as well as on youth unemployment rates, by which indicator the efficiency of YTS will have to be measured.

YTS generally consist of two elements: Full Time Vocational Education (FTVE) and Apprenticeship Training (AT). An Apprenticeship is a time limited training contract offered to initially untrained secondary school graduates. Its characteristics are part time schooling and vocational training in the same program. The former provides apprentices with theoretical information about their profession as well as advanced general 'culture skills'. While being trained on-the-job they are taught standard technologies, working discipline and social behaviour as well as more firm specific knowledge as the firm's organisational structure, the specifities of the working place and non-standard technologies used by the firm. The time ratio between schooling and work is usually about 1:4. On completion of the apprenticeship training the apprentice is awarded a nationwide accepted certificate which acknowledges the profession-specific skills of this education.

FTVE is organized as a public vocational training provision and provided in full-time educational establishments where young people are thaught theoretical background and learn practical skills in in-house laboratories. Nearly all european countries employ these two elements in their youth training scheme with varying emphasis. Whereas, e.g., in Germany the overwhelming number of youths take part in AT and even those who undergo a FTVE afterwards accomplish an AT, in France or Sweden the reverse is true. But for a limited number of areas/regions AT virtually does not exist and the absolute majority of youths undergo a FTVE. Britain takes on an intermediate position actually trying to establish a system similar to that in Germany[2]. Further difference comes from the level of school attainments of the young

[1] See MARGOLIS et al. (1999), INKMANN et al. (1998a, 1998b), FRANZ et al. (1997) and ELLWOOD (1982).

[2] See for an analysis STEEDMAN (1993).

people entering the YTS as well as the signals which the completion of one of its elements send to future employers in the respective countries. Compared to Germany the completion of a FTVE in France sends a much better signal to the labour market than AT, so better school-leavers self-select to this kind of training. In Germany AT has a very high labour market value so even about 15% of university graduates do an AT.

With any type of policy, economists have to ask for the division of gains and costs of a programme. This is especially the case with human capital formation where it is not always obvious to determine the benficiant of the training and externalities may easily arise. Classical human capital theory, as it was introduced by BECKER (1964), makes the distinction between general and firm-specific human capital[3] and predicts that in the case of completely general training the trained worker receives all the return from this training and will therefore bear the costs. In case of firm-specific training Becker makes the informal statement that, as they will both have a return to that training, the training firm and the trained worker will share the cost depending on their respective bargaining power. This argument was later formalized by HASHIMOTO (1981). He predicts that whether or not the investment is shared depends on the existence of post-investment costs of evaluating the worker's productivities in the firm and elsewhere. The relationship between the 'transferability' of human capital accumulated in a training programme to others than the training firm and the type of the labour market resulting from a certain type of training was emphasized by STEVENS (1996). She argues that in a broad sense completely general training may be seen as resulting in a completely competitive labour market, where the workers will be paid their increased marginal product. Analogously, a completely firm-specific training will result in a strong monopsonistic power of the training firm, enabling it to obtain a share of the worker's increased marginal product, which is along Hashimoto's argument. While taking the classical distinction as extreme cases, Stevens stresses that most observable training programmes are 'intermediate' in the sense that they provide human capital with varying degrees of transferability. Other recent authors did also recognize that it must be an imperfection of the resulting labour market that makes firms pay for the training of their workers even if it is 'general', as they say, in fact meaning highly transferable. However, most of them model that imperfection by adding a market failure exogenous to training schemes whose contents are

[3]Becker defines a training to be 'perfectly general,' if it increases a worker's marginal product by exactly the same amount in many firms (Becker, 1964, p. 20), whereas he views a training program as 'completely specific', if it increases a worker's productivity only in the firm providing the training but has no effect on his productivity in any other firm (Becker, 1964, p. 26).

assumed to consist of completely general or completely specific elements[4]. It is Stevens' merit to have explicitely formalized the gradual specifity of training progammes, thus emphasizing the endogeneity of the market failure to the type of training programme chosen: Specifity of training contents always result from the inability to transfer knowlegde fully for use with other firms than the training one. In the open labour market the trained worker will get paid only for the part of his skills which he can transfer to other firms.

With this in mind one may now proceed to analyse the distribution of costs and gains from the various elements of YTS. It seems clear that with FTVE, as no firm is involved, all the gains from training must accrue to the trainee. The fact that FTVE, as secondary education, in Europe is generally paid for by the government suggests that this type of training is subject to some important sort of market failure. The case is not that simple with apprenticeship training. Most of the schooling and some parts of the vocational training may be regarded as very general. Knowledge of the firm's specific organisation or non-standard technologies is cleary firm-specific and a lot of on-the-job training contents are transferable. Following theory, apprentices themselves should pay for their general schooling. However, in France as well as in Germany the government pays for it. The reason for this intervention may come from market failures leading to investment in training which is suboptimal in the sense of general efficiency, as the private return is not maximised. A first is caused by externalities associated with subperfect transferable training programmes. Because their social return to training is lower than their private return, external firms may recoup some expected value of the training[5]. The second source are credit constraints on behalf of the workers in the case of completely general training. Both of them lead systematically to an underinvestment in training. Especially the second one may be the reason for government provision of primary and secondary schooling as well as of the very general elements in the YTS.

With regards to training on-the-job, as it is largely transferable, apprentices should at last pay a part of it, too. The classical possibility to do so would be by accepting a wage rate lower than their marginal product. In fact, STEEDMAN (1993) reports that apprenticeship wages are on average around one third of those of skilled workers in the same sector. WINKELMANN (1994) reports that apprentices receive 32% of the average monthly earnings of an unskilled worker with less than 10 years of experience. On the other hand, especially large firms incur substanial net costs in training their

[4] See the models of ACEMOGLU PISCHKE (1998a, 1998b, 1998c), HARHOFF KANE (1997), FRANZ SOSKICE (1995) and SOSKICE (1994).

[5] For a proof of this result see STEVENS (1994).

apprentices even for the general and transferable parts of it[6]. These findings seem to hint that on behalf of the apprentices there is at last partly a cost bearing. Additionally, they provide the first argument why profit maximizing firms would pay for part of the apprenticeship training. As the apprentices are willing to accept a very low wage rate it may simply be the case that the marginal product of the apprentices during their training is higher than the marginal cost incurred by the training to the firm[7]. A second series of arguments is concerned with the expected revenue firms may draw from future ex-apprentices and which may exceed the net costs of training. Applying this notion ACEMOGLU PISCHKE (1999) point out, that in order to see firms paying for the training of apprentices it is necessary that the wages they will have to pay after the training are *compressed* in the sense, that a firm will be able to get a share of the increased marginal product of the trained worker. The post-training wage structure may be compressed by factors which are endogenous to the training programme and others being exogenous. An endogenous wage compression results from a limited transferability of the accumulated human capital to others firms, thus increasing a priori the bargaining power of the firm. In the case of completely firm-specific training, this follows directly from Becker's classical argument. Similar reasoning might apply to transferable training elements. Exogenous compressing factors may be worker specific mobility costs[8], minimum wages and unions' influences. Wage compression always has the effect of reducing the competitivity of the post-training labour market, giving the training firm monopsony power of the worker. Interpreting wage compression as a distortion from the benchmark competitive wage structure one may distinguish two types of distortion, which result in rather different motivations for the existence of firm sponsored training:

- *Downward distorted wage structure:* Firms manage to pay higly skilled workers less than their actual marginal product. Depending on the

[6] A very useful compilation of studies to that issue may be found in HARHOFF KANE (1997), table 2. Additionally it is worth mentioning, that firms have expenses even for the general schooling of their apprentices in giving them paid leave to attend school.

[7] This corresponds to the notion that apprentices do 'unskilled work' (at least at the beginning of the relationship) but for less than the ordinary unskilled wage. FRANZ/SOSKICE (1995) consider this as a main reason why there is an excess supply of apprenticeships in small German (craft) firms. In addition, v. BARDELEBEN et al. (1995) find evidence that under certain assumptions one may conclude that for small firms in Germany the net costs of training are negative.

[8] Which may result from transportation costs, firm-to-worker subsidies (like firm sponsored mortgages in favorable conditions) as well as risk aversion through workers' familiy statuses, number of children and so on.

dynamics of this gap[9] the firm will have an incentive to pay for training.

- *Upward distorted wage structure:* Firms may have to pay a minimum wage to workers which is above their marginal product. By training them the firm is able to increase their marginal product without having to pay higher post-training wages because the worker will not be able to obtain a higher compensation at the open labour market by transferring the increased marginal product to other firms.

These two types of distortion basically explain the incentives for firms to train[10]. To summarize, the arguments that have been put forth by the literature for the existence of endogenously compressed wage structures are:

Asymmetric Information between the current employer and other firms in the economy about

- *Amount of training and human capital*: Potential employers may not be able to judge the exact quantity and quality of the training the worker has received and may therefore be unwilling to compensate workers for these uncredentialled skills[11].

- *The ability of young workers* (screening argument). The early years of a worker's career reveal valuable information about whether a he is suited to the occupation that he has chosen[12]. Workers who are laid off have, on average, lower ability, thus the value of training is relativeley low for them. Since a high ability trained worker cannot quit and signal

[9] The firm will orientate itself on the relation marginal increase of the gap with marginal costs of training.

[10] HARHOFF KANE (1997) as well as FRANZ SOSKICE (1995) report a third motivation for offering apprenticeships which questions this rational approach and which was often mentioned by human resource managers in their interviews: A norm might have developed within the business community to provide such training and so to comply with social expectations. A similar argument might be applied to small firms on the countryside. However, it may also be argued that in behaving so it is very likely that firms secure themselves additional revenues through informal advantages, which have to be taken into account in calculating the net cost of training, so that the first general argument applies.

[11] This explanation for the presence of firm spondored training was first suggested by KATZ and ZIDERMAN (1990) and was formalized by CHANG and WANG (1996); it may be less relevant for the German apprenticeship system where the content of apprenticeship programmes is regulated by the government and follows a well-specified curriculum.

[12] ACEMOGLU PISCHKE (1998a) develop a model on this issue. They show that, when ability and training are complements, this type of asymmetric information leads to a compressed wage structure. Analogous models have been proposed by MALCOMSON et al. (1997), SU (1997) and STATT (1998).

his ability, the employter can keep him while paying less than the full value of his skills.

Asymmetric Information between worker and the employer (efficiency wage argument). This concerns the exact level of effort and diligence exerted by the employee: Wages have to satisfy the worker's incentive constraint which compresses again the wage structure[13].

Interaction of Specific and General Skills: General skills may often only be taught in interconnection with specific skills. As training workers in specific skills yields an increased monopsony power for the firm, it will be willing to pay for the general training, too. In practice, such mixed training programmes in practice usually lead to a limited degree of transferability so the monopsony power of the firm relies by large on the endogenous probability of workers to change jobs after training. Unfortunately this characteristic is mostly neglegted by models based upon this argument[14].

Exogenously compressed wage structures are primarily the result of structural market characteristics:

Labour Market Institutions

- *Minimum wages*: Increases the pay of less skilled workers while leaving the wages of skilled workers largeley unaffected. This leads to the reasoning of downward wage distortion[15].
- *Unions:* Force employers to pay higher wages to less skilled workers. The effect is the same as with minimum wages[16].

Mobility and Transaction Costs: Transaction costs result from matching and search frictions. The presence of search costs in the labour market creates a bilateral monopoly situation in wage determination. There is a match specific surplus, created by costs of finding new partners, and this surplus will be shared by bargaining[17]. It typically implies that firms obtain

[13] Models by ACEMOGLU PISCHKE (1998b), LOEWENSTEIN SPLETZER (1998).

[14] Arguments in this direction where proposed by models from BISHOP (1996), STEVENS (1994) and FRANZ/SOSKICE (1995).

[15] Models by ACEMOGLU PISCHKE (1998c), HASHIMOTO (1982).

[16] This argument is proposed by ACEMOGLU PISCHKE (1998b), HARHOFF/KANE (1997) and SOSKICE(1994).

[17] See DIAMOND (1982), MORTENSEN (1986) and PISSARIDES (1990) for analysis of the standard search and matching model.

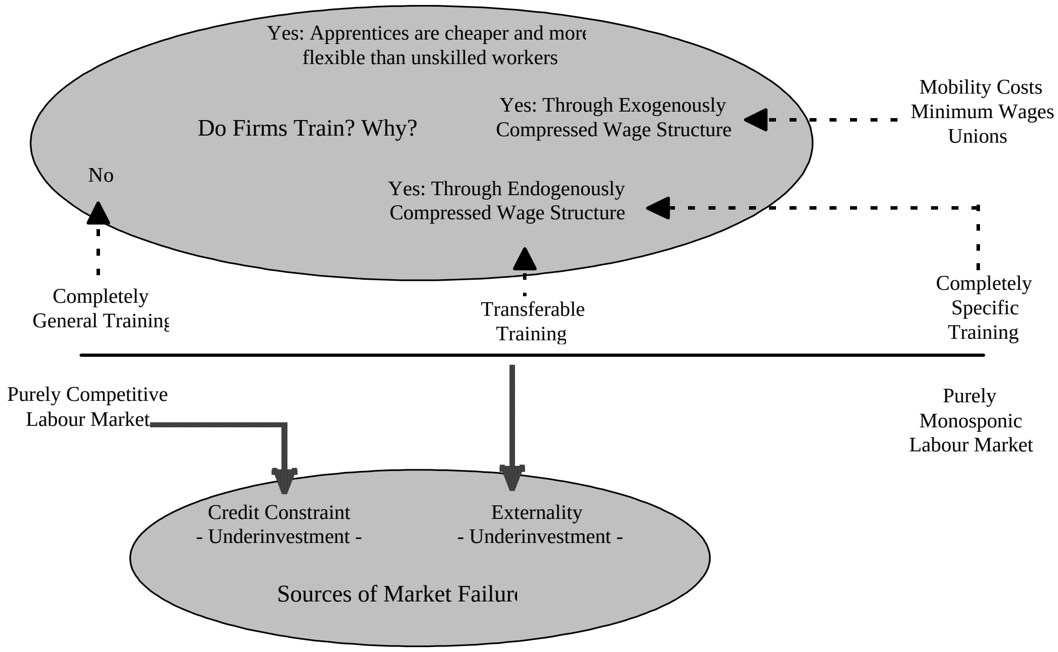

Figure 1: Sources of labour market imperfection and firms' incetnives to pay for training.

a fraction of the productivity of the worker as profits. Bargaining therefore compresses the wage structure and creates incentives for firm-sponsored training[18]. Subjective mobility costs may additionally increase these frictions[19].

Figure (1) summarizes the two basic arguments, negative net costs of training and compressed post-training wage structures. It shows the interconnections of labour market situation and incentives for firms to train. Government intervention usually is the result of the market failures depicted in the lower part of the picture. It is clear that in order to give a reasonable basis to test the (relative) significance of these arguments a comprehensive model should be able to incroporate as many of them as possible. The model which is developed in the following section tries to explain why firms may be willing to pay for apprenticeship training using a game theoretic setting in order to motivate the specific formalization of the endogeneity of the wage compression to the training programme. It incorporates the value vectors of

[18] This story is proposed in detail by ACEMOGLU (1997).

[19] This argument was incorporated in the models of ACEMOGLU/PISCHKE (1998b), HARHOFF/KANE (1997, 1993), STEVENS (1994).

training introduced by STEVENS (1994). In Section 3 the predictions of the model are investigated as well as its ability to represent the aforementioned arguments.

2 The Model

We start by looking at a stylized situation where a firm has to decide at a point t in time whether or not to offer jobs to the following three types of workers:

- Skilled workers: A priori they dispose of the standardized marketable qualification g. Being hired they have yet to be trained in firm specific human capital, q^a. The gain for the firm from hiring them is their marginal product resulting from this firm specific human capital. In order not to incur expected losses the firm will offer them their expected marginal product without firm specific component, which is equal to their marketable value g.
- Unskilled workers: All workers who do not dispose of the industry-specific training. Skill elements which are transferable across all industries are assumed to be zero, reflecting the fact that they do not influence the relative production value of the different types of workers[20]. They receive no training in firm specific technologies so their marginal product is essentially constant over time.
- Apprentices: A mixture of unskilled and skilled workers. Initially untrained workers who obtain the standardized marketable qualification of skilled workers as well as firm-specific human capital during their training. At the end of their training they have the profile of a fully trained skilled worker. During training they are paid a legal wage which may be below the wage for unskilled workers. After training they are paid their marketable value minus the cost they might incur from changing employers.

As for the firms' characteristics we make the following assumptions:

[20] This reflects the notion that a gardener may not transfer a lot of his training to a job in the oil industry but for his very general schooling. Essentially he will enter that industry as an unskilled worker. This does not reflect similarities between some industry-specific skills, whose effects are implicitely assumed to sum up to an expected value of zero.

- Firms in the same industry are assumed to be largely homogenous with regards to their needs in qualification profiles of skilled workers. These profiles consist in knowledge of the firm-specific standard technologies. In addition, every firm has a proprietary degree of specifity (non-standard technology, organisational and communication structures). It is up to the firm, how much of this specifity it will include in the training of apprentices and that of skilled workers respectively.

2.1 Analysis of the Setting

To get a more abstract formulation we assume the temporal structure of the firm's decision to be as depicted in figure (1). A firm i decides a point t in time whether or not to offer a positive amount of apprenticeships. With τ being the uniform length of an apprenticeship the firm knows that at a point $t + \tau$ the apprentice hired in point t will receive simultaneous wage offers by all the firms in the market. Every firm will offer its expected firm-specific marginal product given the information of a successfully completed apprenticeship training. Alternatively, the firm may offer jobs to unskilled workers in the open labour market or try to directly poach skilled workers from other firms in the industry. The situation can be defined as a sequential decision process at each time t. An incumbent firm compares the marginal payoffs from including additional input factors, i.e. making job offers to the various types of workers. A process as it is know from the decomposition of linear programmation under restrictions[21]. However, in the present case there is a substantial interaction between potentially training firms' expected payoffs and the expected payoffs of other firms in the market: An externality to training exists as long as any non-training firm in the labour makret may obtain a positive expected payoff from that training. So, as workers will generally go to the firm which offers them the highest compensation, an incumbent firm basically plays a game with all the other firms in the market in determining the numbers of the different jobs it offers. Additionally, it sets the amount of firm specific information it will include in the training programmes, as it increases its span of possible wage offers. At a time t a firm will play this game iteratively, as long as it obtains a positive expected return from a job offer to one of the three types of workers.

The decision tree resulting from this situation is partially represented in figure (3). At any iteration the deciding firm chooses to offer an apprenticeship, an unskilled position or to make a poach offer to skilled workers.

[21] Of whom the widely used Kuhn-Tucker optimization procedure is but a special case. See CHIANG (1984).

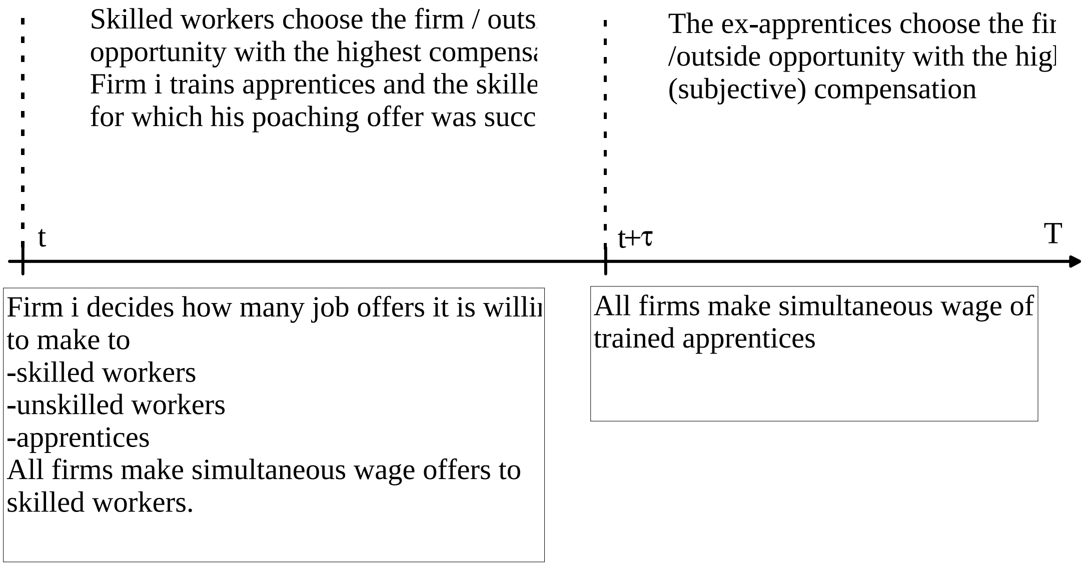

Figure 2: Decision structure of a firm at a point t.

The coloured nodes depict the firms sequential choice, the uncolored nodes are simultaneous moves by all concurring firms in the labour market. The simultaneous moves for skilled workers take place at time t after all firms have made their job offer decisions, and thus are uncertain by their outcome. The simultaneous moves for apprentices takes place at time τ, with uncertain outcome either. To identify formally the outcome of this decision process we start by identifying the elements of this game. Let N be the overall number of firms in the market, $I(x, l, t, \mathbf{n})$ be the information set of player i at time t at the decision node[22] x, l be the iteration counter of the different job offers that successively 'fill up' the vector

$$\mathbf{n} = \begin{pmatrix} n^a \\ n^u \\ n^s \end{pmatrix}.$$

n^a is the number apprenticeships offered, n^u and n^s are the numbers of unskilled and skilled jobs offered respectively. We denote σ_{xl} as the strategy x chosen by a firm at iteration l and $\boldsymbol{\sigma}_{-x}$ as the profile of mixed strategies of the $N-1$ other firms in the market. Further we define σ_{Al} as the strategy

[22] With $x \in \{0, A, U, S\}$ and 0 : Did not previously offer a job (first round), A : Previous job offer was an apprenticeship, U : Previous job offer was an unskilled position, S : Previous job offer was a poaching offer.

to offer an apprenticeship, σ_{Ul} and σ_{Sl} as the strategies to offer a job to an unskilled worker and to make a poaching offer to a skilled worker respectively. The set of possible strategies at iteration l is denoted by $\Delta(S_l)$. The choice of a strategy is done for every single job to be offeres successively by a respective updating of the information set $I(x, l, t, \mathbf{n})$. That is, in the first round the firm takes $\mathbf{n}$ to be the zero vector. It starts by comparing its utilities from choosing one of the elements of $\mathbf{n}$ to be set to one. It the next iteration it compares the utilities given that one of the elements in $\mathbf{n}$ is non-zero and so on. To illustrate this iteration process look again at figure (3). At time t a firm starts by comparing the payoffs from employing a worker of one of the three qualification types, given its information set $I(0, 1, t, \mathbf{n})$. It has to be read as the information firm i has in the first round having employed 0 workers yet. In the picture, the firm concludes that to start by offering an apprentice yields the highest expected utility. In deciding that the firm does not know whether it will succeed or not to make a competitive wage to the apprentice after his training. It explicitely takes this possibility into account by assigning a positive probability to all possible outcomes of a choice and then by comparing the expected utilities from making a job offer to one of the three qualification types. In the second iteration the firm again compares the expected utilities given the information that it has already offered an apprenticeship, which changes at least the net marginal product of second apprenticeship offer. The result of this choice is left undetermined in the figure. For any firm in the market the game consists of L_i successive simultaneous moves with $L_i = n_i^a + n_i^u + n_i^s$ being the overall final number of its job offers.

To identify formally where the firms end up in the equilibrium, first notice that this is a sequential game of imperfect information. The incumbent firm is uncertain about the amount of specifity other firms in the market will include in their training of apprentices and skilled workers. This will influence the respective probabilities of successfully retaining an apprentice after training as well as of successfully poaching skilled workers form other firms. To eliminate possible unreasonable equilibriums it is important that the firm's action is optimal for some belief that it might have about the strategy used by the other players. This solution concept is called a weak Bayesian equilibrium[23]. It requires that at any point in the game, a player's strategy prescribes optimal actions from that point onwards, given his opponent's

[23] See FUDENBERG TIROLE (1996, 1991) for a discussion of Bayesian Equilibria as well as Myerson (1991) who refers to the concept as a *weak sequential equilibrium*). It extends the principle of sequential rationality by formally introducing the players' beliefs. The sequential rationality of a player's strategy is tested via his beliefs about what may have happened prior to her move. Essentially lets the player end up maximising their

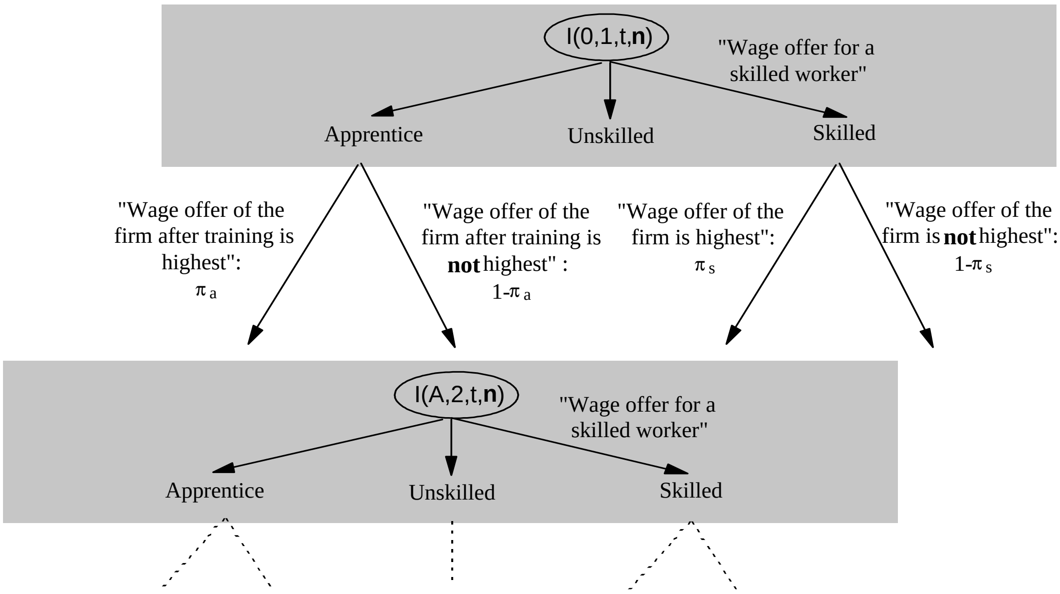

Figure 3: Decision tree of a firm i in a in a game consisting of L successive simultaneous move games in which the N firms (players) determine their optimal number of employees of three qualification types.

strategies and his beliefs about what has happened so far in the game. These beliefs have to be consistent with the strategies being played. By defining a system of beliefs[24], π_i, and implementing the principle of sequential rationality[25], we may identify for firm i a strategy profile $\sigma_i = (\sigma_{1x}, \ldots, \sigma_{L_i x})$ as

expected utility using their beliefs about the probability that they are at a certain node in the game.

[24] This may be defined as a specification of a probability $\pi(x) \in [0,1]$ for each decision node x in Γ such that

$$\sum_{x \in I(x,t,l,\mathbf{n})} \pi(x) = 1$$

for all information sets I. A system of beliefs can be thought of as specifying, for each information set, a probabilistic assessment by the player moving at that set of the relative probabilities of being at each of the information set's various decision nodes, conditional on play having reached that information set.

[25] Equilibrium strategies should specify optimal behaviour from any point in a game onwards, i.e. a player's strategy should specify optimal actions at every point in the game tree.

sequentially rational if we have for any iteration l [26]:

$$E[u_{xl}|I_l, \pi_i, \sigma_{xl}, \boldsymbol{\sigma}_{-x}] \geq E[u_{x'l}|I, \pi_i, \sigma'_{x'l}, \boldsymbol{\sigma}_{-x'}] \tag{1}$$

for all $\sigma'_{x'l} \in \Delta(S_l)$. If strategy profile σ_i satisfies this condition for all information sets I_l, then we can say it to be sequentially rational given a firm i's belief system π_i[27]. To identify a system of beliefs that satisfies this specific consistency requirement we may use the fact that in the current problem each player assigns strictly positive probability to each possible action at every one of her information sets. So for each node x_l a player will compute the probability of having reached that node given play of strategies σ_{xl}, $P(x_{xl} \mid \sigma_{xl})$ by assigning conditional probabilities of being at each of these nodes using Bayes' rule

$$P(x_l \mid I_l, \sigma_{xl}) = \frac{P(x_l \mid \sigma_{xl})}{\sum_{x'_l \in I_l} P(x'_l \mid \sigma_{xl})}.$$

With this expression it is possible to define a profile of strategies and a system of beliefs (σ_i, π_i) to be a weak perfect Bayesian equilibrium if the strategy profile σ_i is sequentially rational given belief system π_i, i.e. (1) is fulfilled, and the system of beliefs π_i is derived from strategy profile σ_i through Bayes' rule. So, finally, we may derive equilibrium conditions for the firm's decision with regards to apprenticeships. From the argumentation sequential rationality at any iteration l a firm will offer jobs of the different types until

$$\begin{aligned} u_{Al} \stackrel{!}{=} u_{Ul} \stackrel{!}{=} u_{Sl} \stackrel{!}{=} 0 \\ u_{q^a} \stackrel{!}{=} u_{q^s} \stackrel{!}{=} 0 \end{aligned} \tag{2}$$

is valid. u_{q^a} is the marginal utility the firm receives from increasing the firm-specific training content in apprenticeship training and u_{q^s} is the corresponding marginal utility with initial training of skilled workers. Form (2) and (1) follows that a firm will offer at least one apprenticeship, if

$$E[u_{Al}|I_l, \pi_i, \sigma_{Al}, \boldsymbol{\sigma}_{-A}] > \text{ for some } l. \tag{3}$$

[26] $E[u_{xl}|I_l, \pi, \sigma_{xl}, \boldsymbol{\sigma}_{-x}]$ denotes player i's expected utility starting at his information set I_l if his beliefs regarding the probabilities of being at the various nodes in I_l are given by π, if he follows strategy σ_{xl} and if his rivals use strategies $\boldsymbol{\sigma}_{-x}$.

[27] That is, at a point (n, l) given the belief of firm i about the probability π that it is a node "apprentice will stay" the firm will have no incentive to revise the previous strategy "hire an apprentice". The condition that this is going to be satisfied is that the expected utility from the strategy "hiring an apprentice" was higher than any other strategy at point $(n, l-1)$.

Notice that our argumentation did not only yield the classical optimality conditions but also gave a reasonable way to model the interaction of the incumbent firms' decicions through the probabilities determining the expected marginal returns using Bayes' rule.

2.2 The Rules of Play

Now, in order to specify the relevant elements of the setting outlined above, some technical assumptions are needed to define the rules of play and the possible actions for firms in the market. Concerning the behavioral structure of the participants we assume that firms are risk neutral with regards to the job offers they make. So, their utility maximization reduces to an expected profit maximization. All firms *know* that all the other firms in the market behave rational with regards to the hiring of employees (common knowledge of rationality). Further, the decision to become an apprentice is taken to be the result of as a lifetime utility maximization. Thus, apprentices are willing to work for a very low wage. Firms may fulfill future apprentices' participation constraints by simply offering the externally (legally) determined wage rate. There is a sufficient and nonelastic supply of apprentices and unskilled workers available to the firm. Firms are able to detect without uncertainty (e.g. from an individual's CV), whether a worker with another firm or in the open labour market is skilled or unskilled in the sense that he does not dispose of the industry-specific skill requirements. Firms always offers skilled workers their expected transferable marginal product. Unskilled workers are always offered a minimum wage. Apprentices are offered the official apprenticeship wage. The probability that a skilled worker will join the firm is identical for poached workers or workers on the open labour market[28]. To model the marginal products a firm has to calculate with we apply the value vector formulation introduced by Stevens[29]. Thus, the ex post effect of any training can be expressed in terms of a value vector whose general form is:

$$\mathbf{v} = (v_1, v_2, \ldots, v_m, \ldots, v_M, v_{M+1}, \ldots, v_N).$$

The completion of an apprenticeship at a given firm results in a homogenous degree of training for any apprentice being trained with that firm. The

[28] To include all reasons of leaving, "probability of leaving" in this context has to be interpreted as any external incentive which motivates the ex-apprentice to leave her firm after training. It comprises direct poaching by other firms, leaving the labour force but also the probability that the training firm itself will not make a competitive wage offer to the apprentice after her training.

[29] See for example STEVENS (1994).

type of apprenticeships provided are industry specific. It follows that the number of firms in the market having a non-random utility (in form of an expected marginal product) is approximately equal to the number of firms in the industry the deciding firm is in. Training effects which are valuable to all Firms in the market are standardized to zero. The expected value of an apprenticeship training is equal to all external firms (M) who can make non-random use of that training, but has a firm specific mark-up q_1^a for the training firm 1 itself[30]. The constant transferable element in training is called g and the degree of transferability is determined by the number of firms in the market for this transferable element M. Note, that this formulation implies constant returns to scale:

$$E[\mathbf{v}^a] = (g + q_1^a, g, \dots, g, 0, \dots, 0)$$

The expected value of a worker skilled for a certain industry is identical with the ex post effect of an apprenticeship training in that industry, but has the specific mark up with the firm actually employing the worker

$$E[\mathbf{v}^s] = (g, \dots, g + q_m^s, \dots, g, 0, \dots, 0)$$

The average (expected overall) marginal product of an apprentice during his training period is equal to that of an unskilled worker. The motivation for this assumption is illustrated in figure 4. The steep ascending curve represents the marginal product of an apprentice during training. He starts with a very low value he ends up essentially at the value of a fully trained skilled worker. A successfully poached worker begins with his transferable value. Being trained in firm-specific human capital he reaches the level of an in-house fully trained ex-apprentice[31]

Finally, the actual value vector of the two types is defined by adding firm specific shocks. These are represented by random variables ε_m, which are independent, identically distributed with zero mean, variance σ_m^2. Thus, for the apprenticeship training the ex-post value vector becomes

$$\mathbf{v}^a = (g + q_1^a + \varepsilon_1, g + \varepsilon_2, \dots, g + \varepsilon_M, \varepsilon_{M+1}, \dots, \varepsilon_N)$$

[30] This corresponds to specific human capital like "knowledge of the inner circles", non-standard firm technology or simply the fact of already knowing the future workplace after the apprenticeship [introduction cost]. A similar but mulitplicative formulation of the value of training to the training firm is done by FRANZ SOSKICE (1995). The multiplicative fashion has some economic appeal and the additive way used here is rather arbitrary. However, they do not model the transferability of the general or certificated element of the training.

[31] Note, that this implies tha assumption that the incumbent firm decides to train apprentices and poached skilled workers the same level of firm-specific information. However, this does not need to be the case.

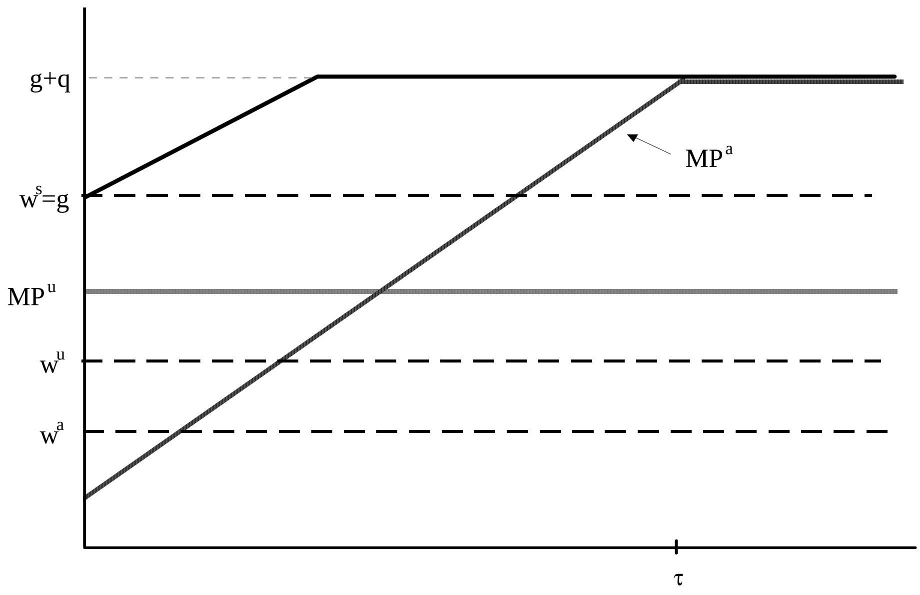

Figure 4: Marginal products and wages of skilled workers ($g{+}q, w^s$), unskilled workers (MP^u, w^u) and apprentices during training (MP^a, w^a).

As these shocks do play an important role in the proceeding of the paper it may be worth deepening the discussion a bit on them. Most critical seems to be the assumption that they are independently distributed. An argument for this may be the multiplicity of the influence factors determining the use of qualification in a given firm. They result from strategic firm decisions as well as technology (productivity), supply or demand shocks. Additionally these shocks incorporate the effect of screening of apprentices during training. With the assumption that the individual abilities of trainees at a firm have a non-degenerate distribution with an expected value of zero, the detection of a very 'unable' apprentice would result in a strong negative shock on the value (marginal product) he has for the training firm. In combination with the 'condition to stay', to be discussed in the next section, such negative shocks may result in wage offers by the training firom at time τ which are in effect too low to match the worker's outside opportunity. This would be equivalent to lay off the worker after training, or to 'let him be poached intentionally' by another firm. Finally, these shocks do represent simply a factor of incertitude as well as proprietary information of the worker/ ex-apprentice facing job offers of the respective firm[32]. With regards to skilled

[32] Thus, it captures unobserved variation in tastes with regards to the attributes of alternatives, errors in the perception and optimization by the ex-apprentice.

workers they formalize time specific changes in the marginal products of skilled workers with a given firm.

2.3 How to Play

In this section the rules are defined which a firm has to follow in order to make successful moves in the relevant labour markets.

2.3.1 Holding apprentices

Recall that the auction process at time τ which determines the future wage of an ex-apprentice may be interpreted as a simultaneous move game. All players (firms) move only once and at the same time by giving their wage offer. The ex-apprentice will work for the firm which offers him the highest compensation, i.e. to which his training has the highest value. Let c be the (pecuniary and non-pecuniary) non-occupational opportunity costs of mobility that arise in case an individual changes employers. It may be interpreted as the money value of the individual's subjective distutility as well as systematically arising poaching costs for the poaching firm. The 'condition to stay', i.e. the sufficient incentive to stay with the firm after training, then is for an ex-apprentice:

$$
\begin{aligned}
w^* \overset{!}{\leq} w_1^A + c &= (1-\theta_1^a)v_1 + c = (1-\theta_1^a)(g + q_1^a + \varepsilon_{1\tau}) + c. \\
0 &< \theta_1^a < 1, \qquad g \ , \ q_1^a >> 0
\end{aligned} \tag{4}
$$

Where w^* is to interpreted as a the value of the 'outside opportunity' that is offered to the worker. It may consist of an external wage offer, i.e. $w^* = w_m^A$, but also an indirect utility value he might have from leaving the labour force or from being unemployed. Recall that the training firm may decide to offer a wage too low to meet the outside opportunity utility because of bad screening results. This is equivalent to laying off the worker. v_1 is the marginal product the ex-apprentice will produce with the training firm 1 after training and $(1-\theta_1^a)$ is the share the training firm will be able to retain from the ex-apprentice's marginal product. As firms will a priori try to offer the highest possible wage this transforms to

$$
\begin{aligned}
v_m &\overset{!}{\leq} v_1 + c \\
\Leftrightarrow g + \varepsilon_{m\tau} &\overset{!}{\leq} g + q_1^a + \varepsilon_{1\tau} + c \\
\Leftrightarrow \varepsilon_{m\tau} - \varepsilon_{1\tau} &\overset{!}{\leq} q_1^a + c
\end{aligned}
$$

or, for a non-industry firm

$$\varepsilon_{m\tau} - \varepsilon_{1\tau} \overset{!}{\leq} g + q_1^a + c$$

2.3.2 Poaching Skilled workers

Similar reasoning applies to the poaching of skilled workers at time t. The poaching firm will have to offset the workers' mobility costs as well as the employing firm's own concurring wage offer. On the other hand, as the poaching firm does not value the firm specific skills of the worker to be poached, it may rationally only offer him the expected marginal product. With firms in the same industry this is equivalent to the transferable element of the skilled worker's training, g, as well as the random value the firm has for workers of this type at time t.

$$w^* + c \overset{!}{\leq} w_1^S = g + \varepsilon_{1t}. \tag{5}$$

2.3.3 Constructing Beliefs

To formalize the probability that a worker (ex-apprentice or skilled worker) will accept a wage offer we apply the property of his indirect utility function to be strictly increasing in w. Thus, a worker follows an utility maximisation by accepting the highest wage offer available to her in the market[33].

Probability of Holding an Ex-Apprentice To determine this probability for an ex-apprentice we notice that at point τ an ex-apprentice will receive the following offer vector

$$\begin{pmatrix} g + a + \varepsilon_1 & \text{from it's former training firm} \\ g + \varepsilon_m - c & \text{from any of the remaining } M-1 \text{ firms in the industry} \\ \varepsilon_n - c & \text{from any of the remaining } N-M \text{ firms in the market} \end{pmatrix}$$

The condition of the ex-apprentice to stay with his training firm is then

$$1[v_1 > v_2 - c,\ v_1 > v_3 - c,\ \ldots\ , v_1 > v_N - c]$$

or equivalently

$$v_1 + c = v^{\max}. \tag{6}$$

As it has a priori no information about the firm specific shocks in general the question for the training firm is to get a reasonable concept of the firm

[33] For a proof see appendix A.2.

specific shocks' distribution. In section 2.1 it was argued that the players use Bayes' rule to determine the probability that the game has reached a certain decision node. Recall that having played the strategy σ_{Al} at the last iteration, the probability that the firm will find itself at node x_l : 'the apprentice will accept the wage offer after training', which is equivalent to (6) to be valid, then can be expressed as

$$P(x_l \mid I_l, \sigma_{Al}) = \frac{P(x_l \mid \sigma_{Al})}{\sum_{x_l' \in I_l} P(x_l' \mid \sigma_{Al})} \tag{7}$$

One implication of this expression is the possibility to express the Odd's ratio of the probabilities to end at a decision node by choosing a strategy from a given information set as

$$\frac{P(x_l \mid I_l, \sigma_{Al})}{P(x_l' \mid I_l, \sigma_{Al})} = \frac{P(x_l \mid \sigma_{Al})}{P(x_l' \mid \sigma_{Al})}. \tag{8}$$

With x_l' as the possible alternative nodes following σ_{Al}, which are the result from (6) not to be valid. This ratio is a useful measure of a firm' s uncertainty about the outcome of a strategy σ_{Al}. Using the probability expression in (7) we may model the firms' beliefs of the relative outcomes of a strategy as independent of the information set the firm may have at any iteration of a game played at t. This useful property may be retained if we assume an i.i.d. exponential, logistic or normal distribution with $E[\varepsilon_i] = 0$ for the firm specific error terms in (**??**). Thus, the expression in (6) will be extreme value distributed of type one[34]. Using the value vector (**??**) we obtain the following expression for the probability that a training firm may retain an apprentice after training:

$$P(v_1 = v^{\max} | \sigma_a) = \frac{\exp(g + q_1^a + c)}{\exp(g + q_1^a + c) + (M-1)\exp(g) + (N-M)} \tag{9}$$

To see the relation to (8) note that the odds ratio of the training firm and a firm which has a structural use for the ex-apprentice's qualifications is defined as

$$\frac{P(v_1 = v^{\max} | \sigma_{Al})}{P(v_m = v^{\max} | \sigma_{Al})} = \frac{\exp(g + q_1^a + c)}{\exp(g)}.$$

That is, the ratio of the probability that the training firm will be able to retain the apprentice after training with respect to the probability that it

[34]For a derivation of this result see appendix A.3.

will loose it to another firm in the industry is independent of the value the apprentice has for any other firm in the market. Taking logs yields

$$\log\left[\frac{P(v_1 = v^{\max}|\ \sigma_{Al})}{P(v_m = v^{\max}|\ \sigma_{Al})}\right] = q_1^a + c.$$

This may be interpreted as the rate at which the training firm may be able to increase it's relative probability with regards to another firm in the industry. It is equal to amount of specific training it gives to the apprentice plus the apprentice's mobility costs. The respective rate with regards to a firm which has only a random value for the apprentice is

$$\log\left[\frac{P(v_1 = v^{\max}|\ \sigma_{Al})}{P(v_n = v^{\max}|\ \sigma_{Al})}\right] = g + q_1^a + c.$$

This rate is shifted by the non-random value an apprentice has for the training firm.

Probability of Successfully Poaching a Skilled Worker Analogously, we may determine the probability that a firm will succesfully poach a skilled worker from another firm as

$$P(v_i = v^{\max}|\ \sigma_S) = \frac{\exp(g)}{\exp(g + q_m^s + c) + (M-1)\exp(g) + (N-M)}. \tag{10}$$

From this expression it is clear that the poaching firm has to offset the specific value the skilled worker has for his actual employing firm, q_M^s, plus the cost he incurs, in case he switches firms. This notion is equivalent to the poaching condition in (5) for skilled workers.

2.3.4 Expected Bargaining Powers

In case the training firm can make the highest bid its share from the ex-apprentice's marginal product is ex post:

$$\begin{aligned}
\theta_1^a &= \frac{v_1 - v_{m'}}{v_1} \qquad\qquad | \ v_1 = v^{\max}, v_{m'} - c = v^{\sec ond} \\
&= \frac{g + q_1^a + \varepsilon_1 - g - \varepsilon_{m'} + c}{v_1} \\
&= \frac{c + q_1^a + \varepsilon_1 - \varepsilon_{m'}}{g + q_1^a + \varepsilon_1}
\end{aligned} \tag{11}$$

It raises with increasing mobility costs of the worker and decreases the higher the highest alternative wage offer is. At the time of the training decision, however, the firm will have to rely upon an expected value of θ_1^a. Recognizing that on average the external firms' specific shocks are zero, (11) becomes

$$\begin{aligned} E[\theta_1^a] &= \overline{\theta}_1^a = \frac{v_1 - E[v_{m'}]}{v_1} \qquad\qquad | \; v_1 = v^{\max}, v_{m'} - c = v^{\sec ond} \\ &= \frac{g + q_1^a + \varepsilon_1 - g + c}{v_1} \\ &= \frac{c + q_1^a + \varepsilon_1}{v_1} \end{aligned} \tag{12}$$

The difference between (11) and (12) arises during the auction process and is not predictable for the training firm at the time it decides upon its offers its given informational state. The incumbent firm will determine its training decision upon the expected wage offer to ex-apprentices

$$\begin{aligned} w_1^a &= (1 - \overline{\theta}_1^a) v_1 \\ &= \left[1 - \frac{c + q_1^a + \varepsilon_1}{v_1}\right] v_1 \\ &= g - c \end{aligned} \tag{13}$$

This means that the ex-apprentice will be effectively offered his transferable value of training minus his mobility costs. Similar reasoning lead the wage offer to skilled workers to be simply

$$w_1^a = g$$

And the firm's expected share after hiring them to be

$$\overline{\theta}_1^s = \frac{q_1^s + \varepsilon_1}{v_1} \tag{14}$$

Comparing (12) and (14) shows an interesting feature. Firms will apriori be able to retain a higher share from the marginal product of ex-apprentices than that of skilled workers as they do not have the ex-apprentices to pay for their mobility costs.

2.4 The Firm's Decision

From the introductory discussion we know that firms may be willing to pay for the training of apprentices out of two reasons:

a) Apprentices yield an expected value which is well above zero, even for the transferable elements of their skills. This results from the limited transferability of the general part of their training (by a limited number of firms valuing it), as well as their mobility costs and the training firms advantage in the bidding for the apprentice to stay through the specific elelment they incuded in the training.

b) Apprentices may be cheaper than unskilled workers. Whereas unskilled workers obtain a priori unlimited contracts, and firms have to pay severance payments in case they lay them off, the contracts with apprentices ar a priori time-limited. Thus, they provide firms who have value for low skilled work with a flexible tool to adapt for demand side shocks.

To determine the firms' decision, we recur to the optimality condition (2). As the game consists of nothing more than a sequential comparision of the marginal contributions the different types of workers yield for the firm's utility which may be interpreted as comparing the respective (expected) marginal products. Thus a firm determines the optimal amount of workers to be employed following that decision rule. To give flesh to these marginal utilities (marginal products as the firms are risk neutral) we introduce the following general profit function:

$$\Upsilon(\mathbf{n},\mathbf{q}) = \left((A(n^a,q^a))^{\alpha}(U(n^u))^{1-\alpha}\right)^{\beta}(S(n^s,q^s))^{1-\beta} \tag{15}$$

$$\text{with } 0<\alpha<1 \text{ and } 0<\beta<1$$

$A()$ as the net product function (production value minus variable and fixed costs) for a number of newly hired apprentices n^a. $U()$ and $S()$ are the respective net product functions for numbers of newly hired unskilled workers n^u and skilled workers n^s. β is the substitutional elasticity between skilled and unskilled labour and α is the substitutional elasticity between the two groups of unskilled workers. Using the **constant returns to scale**[35] **assumption** with respect to the different worker groups, we may model the net product functions using the following linear forms:

- Unskilled workers are simply assumed to yield convex cost functions, $C(n^u)$, with regards to the number of workers employed

$$U(n^u) = n^u MP^u - C(n^u) \tag{16}$$

- Apprentices are assumed to have the same marginal product as unskilled workers in their training phase. After their training phase they

[35] This may be motivated by the fact that usually the number of newly hired people in a firm is relatively little in comparison with the overall number of employes. So it may be reasonable that at the time of the hiring, for this relatively small number of workers the firm assumeds fixed marginal products.

have the full production value of a skilled worker introduced to the firms technology $v_1 = g + q_1^a + \varepsilon_{1\tau}$ minus the wage the firm has to pay for him. The expected share the firm may keep from her production value, (12), is denoted by $\overline{\theta}_1^a$. The probability that the firm will be able to hold the apprentice after training is denoted by π_a.

$$A(n^A, a^A) = n^a \left[r\pi^a \overline{\theta}_1^a v_1 \right] + MP^u - C(q^a, n^a) \qquad (17)$$

We assume $C(q_1^a, n^a)$ to be convex in q_1^a and n^a. It represents the apprentice's training costs and depends on the level of firm-specifity that is included in the apprenticeship training as well as the number of apprenticeships offered.

- Newly hired skilled workers yield the full production value - in case the firm is able to poach them. This probability is denoted by π^s. As they originally have no firm specific knowledge they have to get some initial training which costs a marginal factor $\Sigma'_{q_1^S}$ depending on the amount of firm specificity that the firm decides to include in the worker's training. So, $\Sigma(n^s, q_1^s)$ is assumed to be convex in q_1^S and n^s and represents the training costs as well as non-wage worker costs like expected hiring and / or seperation costs. Note that, contrary to the training costs with apprentices, the firm incurs these costs only in case it succesfully poaches a worker. With π_s being the probability that a firm successfully poaches a skilled the expected net marginal product it receives from this offer writes as

$$S(n^s, q_1^s) = \pi_s \left[n^s \overline{\theta}_1^s v_1 - \Sigma(n^s, q_1^s) \right] \qquad (18)$$

To clear up terms we begin by substituting $\overline{\theta}_1^a$ in (17) with (12) and $\overline{\theta}_1^s$ in (14) with (18) so we get

$$\begin{aligned} A(n^a, q_1^a) &= n^a \left[r\pi_a \frac{c + q_1^a + \varepsilon_1}{v_1} v_1 \right] + MP^u - C(q_1^a, n^a) \\ &= n^a [r\pi_a (c + q_1^a + \varepsilon_1)] + MP^u - C(q_1^a, n^a) \end{aligned}$$

and

$$\begin{aligned} S(n^s, q_1^s) &= \pi_s \left[n^s \frac{q_1^s + \varepsilon_1}{v_1} v_1 - \Sigma(n^s, q_1^s) \right] \\ &= \pi_s \left[n^s (q_1^s + \varepsilon_1) - \Sigma(n^s, q_1^s) \right]. \end{aligned}$$

Because of α and β the firm's objective function is strictly concave. Since the logarithm is a monotonic function, the values that maximize $\Upsilon(\mathbf{n}, a)$ are the same as those that maximize $\ln \Upsilon(\mathbf{n}, \mathbf{q})$.

$$\ln \Upsilon(\mathbf{n}, \mathbf{q}) = \alpha\beta \ln A(n^a, q_1^a) + (1-\alpha)\beta \ln U(n^u) + (1-\beta) \ln S(n^s, q_1^s)$$

The optimization problem of the firm then is

$$\max_{\mathbf{n}\ ,\ \mathbf{q}} \ln \Upsilon(\mathbf{n}, \mathbf{q}) \quad = \quad \alpha\beta \ln A(n^a, q_1^a) + (1-\alpha)\beta \ln U(n^u) \qquad (19)$$

$$+(1-\beta) \ln S(n^S, q_1^S) \qquad (20)$$

$$\text{s.t. } \mathbf{n} \geq 0$$
$$\mathbf{q} \geq 0$$

Notice that the value vectors for ex-apprentices and for skilled workers realise their firm-specific shock components at different time instances (t for unskilled workers, τ for ex-apprentices). As this does not change the argumentation, however, the time-subscripts are omitted. As the firm is price taker with regards to the apprentice's training compensation as well as to the unskilled wage their participation constraints may be assumed to be met. So, the FOC become[36]

$$\frac{\partial \ln \Upsilon(\mathbf{n}, \mathbf{q})}{n^u} = MP^u - C'_{n^u} \stackrel{!}{=} 0 \qquad (21)$$

[36] To differentiate the probabilities notice that from (9) and (10)

$$\pi_a = \frac{\exp(g + q_1^a + c)}{\exp(g + q_1^a + c) + (M-1)\exp(g) + (N-M)}$$

with its derivation with repsect to a, using $[...] = \exp(g + q_1^a + c) + (M-1)\exp(g) + (N-M)$:

$$\pi'_A = \frac{\exp(g + q_1^a + c)[...] - \exp(2(g + q_1^a + c))}{[...]^2} = \pi_a - (\pi_a)^2$$

and

$$\pi_s = \frac{\exp(g)}{\exp(g + q_m^s + c) + (M-1)\exp(g) + (N-M)}$$

with the derivative with respect to q_1^s:

$$\pi'_S = 0$$

$$\frac{\partial \ln \Upsilon(\mathbf{n},\mathbf{q})}{n^a} = r\pi_a(c+q_1^a+\varepsilon_1)+MP^u - C'_{n^a} \stackrel{!}{=} 0 \tag{22}$$

$$\frac{\partial \ln \Upsilon(\mathbf{n},\mathbf{q})}{n^S} = \pi_s(q_1^S+\varepsilon_1) - \Sigma'_{n^S} \stackrel{!}{=} 0 \tag{23}$$

$$\frac{\partial \ln \Upsilon(\mathbf{n},\mathbf{q})}{q_1^a} = n^a[r\pi_a + r\pi_a(c+q_1^a+\varepsilon_1)(1-\pi_a)] - C'_{q_1^a} = 0 \tag{24}$$

$$\frac{\partial \ln \Upsilon(\mathbf{n},\mathbf{q})}{q_1^S} = n^s - \Sigma'_{q_1^s} = 0 \tag{25}$$

The SOC for all types are given by the principal minors of the Hessian[37]:

$$\begin{pmatrix} -C''_{n^u n^u} & & & & \\ 0 & -C''_{n^a n^a} & & & \\ 0 & 0 & -\Sigma''_{n^S n^S} & & \\ 0 & & & \begin{array}{l} n^a r[2(\pi^a-(\pi^a)^2)+ \\ (1-\pi^a)(\pi^a-(\pi^a)^2)(c+q_1^a+\varepsilon_1)] \\ -C''_{q_1^a q_1^a} \end{array} & \\ 0 & 0 & 0 & 0 & -\Sigma''_{q_1^s q_1^s} \end{pmatrix}$$

By using the First Order Conditions above we may derive the following relationships: First notice that

$$r\pi_a(c+q_1^a+\varepsilon_1) = C'_{n^a} - C'_{n^u} = \Delta. \tag{26}$$

That is, the differential between the per worker expenses of apprentices and unskilled workers is equal to the discounted after-training-income the firm expects from training an apprentice. We denote it by Δ. Inserting this in condition (23) we may derive the optimal number of apprenticeships offered.

$$n^a = \frac{C'_{q_1^a}}{[r\pi_a + \Delta(1-\pi_a)]}$$

[37] As by definition $C''_{q_1^a q_1^a} > 0$, with (26) the SOC for a global optimum are met in case that

$$\begin{aligned} r2(\pi_a - \pi_a^2) &< (1-\pi_a)^2\Delta \\ r2\pi_a &< (1-\pi_a)\Delta \\ \frac{r2\pi_a}{(1-\pi_a)} &< \Delta \end{aligned}$$

To determine the relative influence factors we use the definition of π^A in (9) to get

$$n^a = \frac{C'_{q_1^a}\left[(M-1)+(N-M)\exp(-g)+\exp(q_1^a+c)\right]}{\left[((M-1)+(N-M)\exp(-g))\,\Delta+\exp(q_1^a+c)r\right]} \tag{27}$$

From (3) follows a condition that n^a is positive

$$n^a > 0 = 1\left[\frac{C'_{q_1^a}}{[r\pi_a+\Delta(1-\pi_a)]} \mid n^a = 1\right] > 0$$

which recommends a somple tobit analysis for estimation...

Finally with (25) we get

$$\frac{n^a}{n^s} = \frac{C'_{q_1^a}}{\Sigma'_{q_1^a}[r\pi_a+\Delta(1-\pi_a)]}$$

3 Predictions of the Model

The following tables summarizes the predictions of the model of the respective influence factors on the optimal number of apprentices as well as the apprentice/employee ratio[38]:

	$\Delta = 0$	$\Delta > 0$	$\Delta < 0$
	First Derivative		
$C'_{q_1^a}$	+	+	
q_1^a			
c			
g			
Δ			
C'_{n^a}			
C'_{n^u}			
M			
r			

[38] For their derivation see appendix A.1.

References

[1] ACEMOGLU, D. (1997), 'Training and Innovation in an Imperfect Labor Market', *Review of Economic Studies*, vol. 64, 445-464.

[2] ACEMOGLU, D., PISCHKE, J.-S. (1999), 'Beyond Becker : Training in Imperfect Labour Markets', *The Economic Journal*, vol. 109, F112 - F142.

[3] ACEMOGLU, D., PISCHKE, J.-S. (1998a), 'Why do firms train ? Theory and Evidence', *Quarterly Journal of Economics*, vol. 113, 79 - 119.

[4] ACEMOGLU, D., PISCHKE, J.-S. (1998b), 'The Structure of Wages and Investment in General Training', NBER Working Paper No. 6357; Forthcoming Journal of Political Economy, June1999.

[5] ACEMOGLU, D. , PISCHKE, J.-S. (1998c), 'Minimum Wages and On-The-Job Training', MIT and Princeton University : mimeographed.

[6] BARDELEBEN, R. V., BEICHT, U., FEHER, K. (1995), 'Betriebliche Kosten und Nutzen der Ausbildung. Repräsentative Ergebnisse aus Industrie, Handel und Handwerk', *Berichte zur beruflichen Bildung*, Heft 187. Bielefeld : Bertelsmann.

[7] BECKER, G.S. (1964), *Human Capital*, Chicago, New York : The University of Chicago Press, 3rd ed.

[8] BISHOP, J.H. (1996), 'What We know About Employer-Provided Training : A Review of the Literature', Cornell University, Center for Advanced Human Ressources Studies : Working Paper 96-09.

[9] CHANG, C., WANG, Y. (1996), 'Human Capital Investment under Asymmetric Information: The Pigovian Conjecture Revisited', *Journal of Labour Economics*, vol. 14, 505-519.

[10] CHIANG, A.C. (1984), *Fundamental Methods of Mathematical Economics*, Singapore: McGraw-Hill Book Co.

[11] DIAMOND, P. (1982), 'Wage Determination and Efficiency in Search Equilibrium', Review of Economic Studies, vol. 49, 217 - 227.

[12] ELLWOOD, D.T. (1982) 'Teenage Unemployment: Permanent Scars or Temporary Blemishes?'. In: FREEMAN, R.B., WISE, D.A. (eds.), *The Youth Labor Market Problem : Its Nature, Causes and Consequences.* Chicago : The University of Chicago Press.

[13] FRANZ, W., INKMANN, J., POHLMEIER, W., ZIMMERMANN, V. (1997). 'Young and Out in Germany : On the Causes and Consequences of Young Outsiders on the German Labor Market'. NBER - Discussion Paper, nr. 6212.

[14] FRANZ, W., SOSKICE, D. (1995), 'The German Apprenticeship System'. In: BUTTLER, F., FRANZ, W., SCHETTKAT, R., SOSKICE, D. (Hrsg.), Institutional Frameworks And Labor Market Performance. Comparative Views on the German and U.S. Economies. London/New York : Routledge, 208 - 234.

[15] FUDENBERG, D., TIROLE, J. (1996), *Game Theory*, Cambridge Mass.: The MIT Press.

[16] FUDENBERG, D., TIROLE, J. (1991), 'Perfect Bayesian and Sequential Equilibrium', *Journal of Economic Theory*, vol. 53, 236 - 260.

[17] HARHOFF, D., KANE, Th. (1997), 'Is the German Apprenticeship System a Panacea for the U.S. Labor Market ?' , Journal of Population Economics, vol. 10, 197 - 217.

[18] HARHOFF, D., KANE, Th. J. (1993), 'Financing Apprenticeship Training: Evidence from Germany', Working Paper No. 4557, NBER, Cambridge: Mass.

[19] HASHIMOTO, M. (1982), 'Minimum Wage Effects of Training on the Job', *American Economic Review*, vol. 72, 1070 - 1087.

[20] HASHIMOTO, M. (1981), 'Firm-Specific Human Capital as a Shared Investment', *American Economic Review*, vol. 71(3), 475 - 482.

[21] INKMANN, J., KLOTZ, S., POHLMEIER, W. (1998a) 'Growing Into Work', ZEW - Discussion Paper, No. 98-47

[22] INKMANN, J., KLOTZ, S., POHLMEIER, W. (1998b) 'Permanente Narben oder temporäre Blessuren?' Universität Konstanz : mimeographed.

[23] KATZ, E., ZIDERMAN, A. (1990), 'Investment In General Training : The Role of Information and Labour Mobility', *Economic Journal*, vol. 100, 1147 - 1158.

[24] LOEWENSTEIN, M.A. SPLETZER, J.R. (1998), 'Dividing the costs and returns to general training', *Journal of Labour Economics*, vol. 16, 142 - 171.

[25] MALCOMSON, J.M., MAW, J.W., McCORMICK, B. (1997), 'General Training by Firms, Contract Enforcability, and Public Policy', University of Southhampton and University of Swansea : mimeographed.

[26] MARGOLIS, D.N., PLUG, E., SIMONNET, V., VILHUBER, L. (1999), 'The Role of Early Career Experiences in Determining LAter Career Success: An International Comparison', CREST: mimeographed.

[27] MORTENSEN, D.T. (1986), 'Job Search and Labor Market Analysis'. In: ASHENFELTER, O., LAYARD, R. (eds.), *Handbook of Labor Economics*, vol. 2, Chap. 15.

[28] PISSARIDES, C.A. (1990), *Equilibrium Unemployment Theory*, Oxford : Basil Blackwell.

[29] RYAN, P. (1980), 'The Cost of Job Training for a Transferable Skill', *British Journal of Industrial Relations*, vol. 8, 334 - 352.

[30] SOSKICE, D. (1994), 'The German Training System : Reconciling Markets and Institutions'. In: LYNCH, L. (1994) Training and the Private Sector : International Comparisons Chicago : University of Chicago Press for NBER, 25 - 60. !!! LESEN

[31] STATT, A.-L. (1998), 'Great Prospects : Employer Provided Training as a Credible Screening Device', University of Stirling : mimeographed.

[32] STEEDMAN, H. (1993), 'The Economics of Youth Training in Germany', Economic Journal, vol. 103, 1273 - 1291.

[33] STEVENS, M. (1996), 'Transferable Training and Poaching Externalities'. In: BOOTH, A.L., SNOWER, D.J. (1996), *Acquiring Skills - Market Failures, Their Symptoms And Policy Responses,* Cambrige UK : Cambridge University Press.

[34] STEVENS, M. (1994), 'A Theoretical Model of On-The-Job Training With Imperfect Competition', *Oxford Economic Papers*, vol. 46, 537 - 562.

[35] SU, L.-J. T. (1997), 'General Training in a Labour Market with Adverse Selection', UCLA : mimeographed.

A Mathematical Annex

A.1 Derivation of the Model Predictions

The following relationship results for the number of apprentices:

$$\begin{aligned} n_a &= \frac{C'_{q_1^a}}{\left[(1-\pi^a)\,\Delta+\pi^a r\right]} \\ &= \frac{C'_{q_1^a}}{\left[\begin{array}{c}\left(\frac{\exp(g+q_1^a+c)+(M-1)\exp(g)+(N-M)-\exp(g+q_1^a+c)}{[...]}\right)\Delta \\ +\frac{\exp(g+q_1^a+c)}{[...]}r\end{array}\right]} \\ &= \frac{C'_{q_1^a}\left[\exp(g+q_1^a+c)+(M-1)\exp(g)+(N-M)\right]}{\left[\begin{array}{c}((M-1)\exp(g)+(N-M)\exp(g))\,\Delta \\ +\exp(g+q_1^a+c)r\end{array}\right]} \\ &= \frac{C'_{q_1^a}\left[(M-1)+(N-M)\exp(-g)+\exp(q_1^a+c)\right]}{\left[((M-1)+(N-M)\exp(-g))\,\Delta+\exp(q_1^a+c)r\right]} \end{aligned}$$

The analytical derivatives are:

$$\begin{aligned} \frac{\partial n_a}{\partial C'_{q_1^a}} &= \frac{\left[(M-1)+(N-M)\exp(-g)+\exp(q_1^a+c)\right]}{\left[((M-1)+(N-M)\exp(-g))\,\Delta+\exp(q_1^a+c)r\right]}>0 \\ \frac{\partial^2 n_a}{\partial\left(C'_{q_1^a}\right)^2} &= 0 \end{aligned} \tag{28}$$

as long as Δ is not negative, this expression is positive...YYY

$$\begin{aligned} \frac{\partial n_a}{\partial q_1^a} &= \frac{C'_a\left[(M-1)+(N-M)\exp(-m)+\exp(a+c)\right]}{\left[((M-1)+(N-M)\exp(-m))\,\Delta+\exp(a+c)r\right]} \\ & \frac{u'v-v'u}{v^2} \\ u' &= C'_a\exp(a+c) \\ v' &= \exp(a+c)r \\ &= \frac{\begin{array}{c}C'_a\exp(a+c)\left[((M-1)+(N-M)\exp(-m))\,\Delta+\exp(a+c)r\right]- \\ \exp(a+c)r\left[(M-1)+(N-M)\exp(-m)+\exp(a+c)\right]\end{array}}{\left[((M-1)+(N-M)\exp(-m))\,\Delta+\exp(a+c)r\right]^2} \\ &= \frac{\exp(a+c)\left[\begin{array}{c}(M-1)\Delta C'_a+(N-M)\exp(-m)\Delta C'_a+\exp(a+c)rC'_a \\ -r(M-1)-r(N-M)\exp(-m)-\exp(a+c)r\end{array}\right]}{...>0} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\exp(a+c)\left[(M-1)(\Delta C'_a - r) + (N-M)\exp(-m)(\Delta C'_a - r) + \exp(a+c)(rC'_a - \right.}{... > 0} \\
&= \left[(M-1)(\Delta C'_a - r) + (N-M)\exp(-m)(\Delta C'_a - r) + \exp(a+c)(rC'_a - r)\right] \frac{... > 0}{... > 0} \\
&\gtreqless 0
\end{aligned}
$$

$$
\frac{\partial^2 n^a}{\partial (q_1^a)^2} = ???
$$

As $C'_a > 1$ (siehe entsprechenden Proof), $(rC'_a - r) > 0$. The final sign of the expression depends of the relative importance of the respective terms as well as the sign of $\Delta = w^A - w^U$. However, interpretations are possible. Comsider the case, firms do have no discounting ($r = 0$) and there is essentially no difference betwenn w^A and w^U. Then the effect of augmenting the specifity of the firm's training contents would clearly increase the number of apprenticeships trained. In the datatset we look at, we could not find statistically significant evidence that the medianwages of apprentices and low skilled worker groups are different. However it is reasonabvle that discounting diminuishes the utility a firm receives from a future skilled apprentice. To what extend, however, has to be identified empirically.

$$
\begin{aligned}
\frac{\partial n_a}{\partial c} &= \frac{C'_a\left[(M-1) + (N-M)\exp(-m) + \exp(a+c)\right]}{\left[((M-1) + (N-M)\exp(-m))\,\Delta + \exp(a+c)r\right]} \\
&\gtreqless 0 \\
\frac{\partial^2 n_a}{\partial c^2} &= ???
\end{aligned}
$$

This follows from the same argument as for $\frac{\partial n_a}{\partial a}$.

$$
\begin{aligned}
\frac{\partial n_a}{\partial M} &= \frac{C'_a\left[(M-1) + (N-M)\exp(-m) + \exp(a+c)\right]}{\left[((M-1) + (N-M)\exp(-m))\,\Delta + \exp(a+c)r\right]} \\
& \frac{u'v - v'u}{v^2} \\
u' &= (-1)C'_a(N-M)\exp(-m) \\
v' &= (-1)(N-M)\exp(-m) \\
&= \frac{(-1)C'_a(N-M)\exp(-m)\,[\text{de bas}] - (-1)(N-M)\exp(-m)C'_a\,[\text{de haut}]}{\left[((M-1) + (N-M)\exp(-m))\,\Delta + \exp(a+c)r\right]^2} \\
&= \frac{(-1)C'_a(N-M)\exp(-m)\,[\text{de bas}] - (-1)(N-M)\exp(-m)C'_a\,[\text{de haut}]}{\left[((M-1) + (N-M)\exp(-m))\,\Delta + \exp(a+c)r\right]^2}
\end{aligned}
$$

$$\frac{\partial^2 n_a}{\partial M^2} = ???$$

$$\begin{aligned}
\frac{\partial n_a}{\partial r} &= \frac{C_a' \left[(M-1) + (N-M)\exp(-m) + \exp(a+c)\right]}{\left[((M-1) + (N-M)\exp(-m))\,\Delta + \exp(a+c)r\right]} \\
&= ??? \\
\frac{\partial^2 n_a}{\partial r^2} &= ???
\end{aligned}$$

$$\begin{aligned}
\frac{\partial n_a}{\partial M} &= \frac{C_a' \left[(M-1) + (N-M)\exp(-m) + \exp(a+c)\right]}{\left[((M-1) + (N-M)\exp(-m))\,\Delta + \exp(a+c)r\right]} \\
&\quad ??? \\
\frac{\partial^2 n_a}{\partial M^2} &= ???
\end{aligned}$$

$$\begin{aligned}
\frac{\partial n_a}{\partial \Delta} &= \frac{C_a' \left[(M-1) + (N-M)\exp(-m) + \exp(a+c)\right]}{\left[((M-1) + (N-M)\exp(-m))\,\Delta + \exp(a+c)r\right]} \\
&\quad ??? \\
\frac{\partial^2 n_a}{\partial \Delta^2} &= ???
\end{aligned}$$

with $\Delta = w^A - w^U$

A.2 Indirect Utility

An indirect utility function $v(p, w)$, with w being the individual's wage and p being the price vector of the consumption set X, and is the value of the utility maximization problem of an individual given it's budget constraint. Given the individuals utility function $u()$ is defined as follows:

Suppose that $u()$ is a continuous utility function representing a locally nonsatiated preference relation $\succeq$ defined on the consumtpion set $X = R^L_+$, then the indirect utility function $v(p, w)$ is

1. Homogenous of degree zero.
2. Strictly increasing in w and nonincreasing in p_l for any l (consumption good index).
3. Quasiconvex; that is, the set $\{(p, w) : v(p, w) \leq v_0\}$ is convex for any v_0.

4. Continuous in p and w.

Because $u(x(p,w)) = v(p,w)$ and from property 1 we can say that the individuals utility from a higher wage offer is strictly higher than from a lower, i.e. $u(w) > u(w')$ for any $w > w'$.

A.3 Deriving the Probability of Changing Firms

It was assumed that ex-apprentices as well as skilled workers are rational in the sense that they make choices that maximize their perceived utility (subject to contraints of expenditures). Using the argument of indirect utility functions they do this by accepting the highest wage offer minus mobility costs (subjective overall payoff from a move). In the model they generally face N offers, the number of firms in the market. A priori these consist of the value vector (marginal product) $\mathbf{v}$ they have for the respective firms in the market times a vector of firm specific random shocks $\boldsymbol{\varepsilon}$. $\mathbf{v}$ is the vector of wage offers and $\boldsymbol{\varepsilon}$ is a residual vector that captures unobserved variations in tastes, in the attributes of alternatives, errors in the perception and optimization by the ex-appretnice/skilled worker as well as unobserved firm specific shocks. It is to be shown that (9) is valid, given given i.i.d. logistic, normally or exponentially distributed shocks in the residual vector. For apprentices $\mathbf{v}$ takes the form

$$\begin{pmatrix} v_1 & = g + q_1^a \\ v_2 & = g \\ \vdots & \\ v_M & = g \\ v_{M+1} & = 0 \\ \vdots & \\ v_N & = 0 \end{pmatrix}$$

with the first expression to be the value an ex-apprentice has for his training firm. For skilled workers the argument is roughly the same, with the exception that the term in the value vector which has a firm-specific mark-up belongs to the firm the skilled worker is working at the time of the poaching offer. Then $\mathrm{v}^{\max} = \max(v_1, ..., v_n)$ then has an extreme value distribution of type I[39], so

$$F(\varepsilon_i < \varepsilon) = \exp(-\exp(-\varepsilon)) \tag{29}$$

and density function

$$f(\varepsilon_i) = \exp(-\varepsilon_i \cdot - \exp(-\varepsilon_i)). \tag{30}$$

[39] See JOHNSON et al. (1995).

The wage offer accepted must satisfy the condition

$$v_i^{\max} = \max[v_1, v_2, \ldots, v_M, v_{M+1}, \ldots, v_N]$$

which implies

$$\varepsilon_i + v_i > \varepsilon_j + v_j \quad \forall i \neq j$$

or

$$\varepsilon_j < \varepsilon_i + v_i - v_j \quad \forall i \neq j$$

Then

$$\begin{aligned} \Pr(v_i = v^{\max}) &= \Pr(\varepsilon_j < \varepsilon_i + v_i - v_j) \quad \forall i \neq j \\ &= \int_{-\infty}^{\infty} \prod_{j \neq i} F(\varepsilon_i + v_i - v_j)\ f(\varepsilon_i)\ d\varepsilon_i \end{aligned} \tag{31}$$

As with (29) and (30)

$$\begin{aligned} \prod_{j\neq i} F(\varepsilon_i + v_i - v_j)\ f(\varepsilon_i) &= \prod_{j \neq i} \exp(-\exp(-\varepsilon_i + v_i - v_j))\ \exp(-\varepsilon_i \cdot - \exp(-\varepsilon_i)) \\ &= \exp\left(\varepsilon_i - \exp(-\varepsilon_i)\left(1 + \sum_{j\neq i} \frac{\exp(v_j)}{\exp(v_i)}\right)\right) \end{aligned}$$

and if we write

$$\ell_i = \log\left(1 + \sum_{j \neq i} \frac{\exp(v_j)}{\exp(v_i)}\right) = \log\left(\sum_{j=1}^{N} \frac{\exp(v_j)}{\exp(v_i)}\right)$$

(31) becomes

$$\begin{aligned} \Pr(v_i = v^{\max}) &= \int_{-\infty}^{\infty} \exp\left(-\varepsilon_i - \exp(-\varepsilon_i + \ell_i)\right) d\varepsilon_i \\ &= \exp(-\ell_i) \int_{-\infty}^{\infty} \exp\left(-\varepsilon_i^* - \exp(-\varepsilon_i^*)\right) d\varepsilon_i^* \quad \text{where } \varepsilon_i^* = \varepsilon_i - \ell_i \\ &= \exp(-\ell_i) \\ &= \frac{\exp(v_i)}{\sum_{j=1}^{N} \exp(v_j)}. \end{aligned}$$

For the case of an apprentice this yields:

$$P(v_1 = v^{\max}) = \frac{\exp(g + q_1^a + c)}{\exp(g + q_1^a + c) + (M-1)\exp(g) + (N - M)}$$

And for the case of a skilled worker:

$$P(v_1 = v^{\max}) = \frac{\exp(g)}{\exp(g + q_1^a + c) + (M-1)\exp(g) + (N - M)}$$