The Effectiveness of Workplace Training in Germany

(Preliminary draft)

Elena Mazza*

University College Dublin

June, 2015

Abstract

This paper looks at the effectiveness of workplace training across quantiles of the wage distribution. It looks at the variation in the returns to training for workers at different skills and salary levels. The estimation of these distributional effects of training has never been addressed in the literature despite the concerns about the currently accepted mean estimates. We apply two unconditional quantile regression models and two mean-regression ones to investigate the distributional and average effects of workplace training. The case under consideration is the one of Germany over the time period comprised between 2000 and 2012. We show that training spells have a positive and significant effect on wages which varies across the wage distribution although none of the quintiles has disproportionate returns when compared to the others.

Keywords: workplace training, returns, quantiles, Germany

JEL Classification: J31, J21, L25, M53

1 Introduction

Given the accelerated pace and pervasiveness of technological development over the last few decades, workplace training likely plays a crucial role in updating skills across the whole wage and skill distribution. Employers investments in training provision are crucial to sustain productivity and competitiveness. It is therefore important to analyse the returns accrued by those undergoing training. Through the estimation of the distributional returns to training we plan to analyse the efficacy of training investments by looking at the effects of training on wages. Knowing that training is associated with significantly higher productivity, increases in wages have to be paid out of productivity gains [1]. For this reason we can interpret the observed wage increases as a lower bound on the likely size of productivity increasesThis lower bound has always been found to be greater or equal to zero in the literature..

*I am thankful to Dr. Christopher Jepsen, Prof. Paul Devereux and Dr. Frank Walsh for their excellent feedbacks and advice. I am also thankful Prof. Peter Mueser and Judith Delaney for their comments and advice on an earlier version of this work. I am grateful to Prof. David Powell for the long-distance support he offered in the use of the Stata programs.

[1]dearden2000gains

Returns to training obtained through the use of the Mincerian framework are normally high, not to say huge. Their estimated magnitude looks even less likely to be an accurate estimate of the true effects of a training spell when they are confronted with coefficients on education and experience. Most of the evidence which has been presented so far shows in fact that one extra year of education has an impact on wages which is comparable to that of a training spell, but training spells last on average way less than a year. These training's payoffs are clearly beyond the rosiest expectations and leave us doubtful about their measurement accuracy.

Some of the hypotheses advanced to explain these results are hard to test: for example, it is true that the role of unmeasured informal training could boost the effects of formal courses but many datasets lack measures of informal training. Selection into training based on individuals motivation and ability is also upward biasing coefficients; empirical literature typically overcomes this problem by using fixed effects models to study the effects of training on wages in longitudinal studies. This strategy is good to get rid of the differences in the levels of wages but does not account for the individual differences in wage growth, and this could influence the estimates of the impact of training especially over short time periods (Pischke, 2001).

There are other possible factors, which are more easily testable, like the bias introduced by the different returns for long and short spells, which is supposed to push results for the effects of "any given" training spell higher. Alternatively, there is the hypothesis that since selection into training is mostly based on the steepness of wage profiles, the estimated average effects of training are mostly influenced by the contribution of people reporting comparatively steeper wage profiles.

In this work we contribute to investigating the causes of the payoffs' overstatement by looking at the variation of returns to training across the quintiles of the wage distribution. We assume that training entails different practices and has a different impact on workers productivity across the wage distribution. In light of this assumption and due to the non-substitutability of workers across the wage/skill distribution we expect to observe different degrees of investment in training and different returns to it. Our results will clarify whether average estimates are driven by the disproportionately high returns to certain quintiles or whether they simply reflect similar returns across quantiles. As we mentioned earlier on, they will also provide insights about the effectiveness of training investment with regard to productivity.

We study the mean and quantile effects of workplace training participation on wages by the estimation of four models: two of them estimate the mean and quantile effects of training at the cross-sectional level whereas the remaining two models estimate the mean and quantile effects of training using panel data. To the best of our knowledge, this study is the first to provide empirical evidence concerning the cross-sectional and longitudinal returns to training across quantiles of the wage distribution. We use two newly-developed estimators introduced in Powell (2010, 2014) for implementing Unconditional Quantile Treatment Regression and Unconditional Quantile Regression with Fixed Effects (Powell, 2010). These estimators successfully overcome some of the most common problems relating to the use of quantiles for studying the welfare effects of a treatment: they provide unconditional estimates of the effects of a treatment. Individual characteristics are not conditioned upon but they are only used to inform the probability distribution of the outcome variable. This work brings new evidence on two branches of the empirical literature on workplace training: firstly it looks at the distributional effects of training and, secondly, it does it by using a recently developed estimator that allows to estimate unconditional quantile effects with longitudinal data.

Our analysis avails of a long survey panel dataset tracking a wide sample of German workers over the last thirty years, the German Socio-Economic Panel(GSOEP). These data provide detailed information about five waves of trainings spells: 1989, 1993, 2000, 2004 and 2008. Our study critically looks into the effect of one or multiple training spells and their length confining the period of interest to 1997-2012: our cross-sectional and longitudinal models consider the three waves of 2000, 2004 and 2008 accounting for the possibility of training accumulation over time.

2 Relation to Previous Literature

Recent literature concerning workplace training has challenged the current thinking about employers' incentives to provide training. In contrast to Becker's framework, the more recent one considers labour markets as not being working in a perfect competitive fashion. Workers are paid below their marginal productivity and there are rents in the employment relationship accruing to the employer. The following two sections provide an overview of the main assumptions lying behind Becker's classical framework and the extensions made by Acemoglu and Pischke.

2.1 Perfectly Competitive Labour Markets

The debate concerning workplace training investments is grounded in the ideas of Pigou (1912) [2] and Rosenstein-Rodan (1943)[3] but it has been mostly shaped by the seminal contribution of Becker's *Human Capital Theory* (1962).
In Becker's framework, training can be either *general* or *specific* (Malcomson, 1999). In both these two cases, training enhances workers productivity albeit at a cost comprising the opportunity cost of foregone earnings and the direct cost of training (Bassanini, Booth, Brunello, De Paola, & Leuven, 2005). The accumulation of productive skills will also make it easier to learn new productive skills, but will also raise the opportunity cost of learning. Becker's framework expects employers to subsidize training in its firm-specific form, whereas general training is financed by employees' through the receipt of lowered wages during the training.

Let us consider the case of a firm investing τ on its workers training. Assuming that there are two periods, t_0 and t_1, the total profits that the firm derives from subsidizing the training are given by:

$$\prod = [y_0 - w_0 - c_0(\tau)] + \frac{y_1(\tau) - w_1(\tau)}{1+r}$$

where y_0 and w_0 are respectively the worker's productivity and wage at t_0, $c_0(\tau)$ is the cost of training incurred by the firm at t_0 and $y_1(\tau)$, $w_1(\tau)$ correspond to productivity and wages at t_1.
At t_1 the firm's profits will be equal to the present value of the difference between $y_1(\tau)$ and $w_1(\tau)$. Firms maximise their profits when training's marginal revenue equals training's marginal cost. By taking the first derivative of $\prod$ with respect to τ we have that:

$$\frac{y_1'(\tau) - w_1'(\tau)}{1+r} = c_0'(\tau)$$

[4]

The marginal revenue is given by the present value of the difference between the marginal product of training, $y_1'(\tau)$, and the corresponding increase in the level of salaries, $w_1'(\tau)$.

If the market operates in a perfectly competitive fashion and training is of the *general* type, then the salary increase following from the training spell equals the increase in workers' productivity. It follows that $y_1'(\tau) = w_1'(\tau)$, meaning that the firm will maximise its profits when the costs incurred are equal to zero, or rather, when it does not pay for the training and lets workers bear the costs through lowered wage rates.
The lack of incentives in general training provision also stems from the fact that workers have the possibility to employ their "general" skills outside of the training firm. In a perfectly competitive market, separation is made easier by the absence of frictions and discourages firms from investing on workers.
The case of a *specific* type of training differs from that of general training. Workers do not reap the benefits of their higher productivity after moving across firms. Firms are therefore expected to offer the trainee a wage which is higher than the *outside* wage rate. Firms are likely to retain their trainees and have therefore incentives to provide them with training. In Becker's framework, in the case of firm-specific training, employer and employee are expected to share both the cost of training and the returns to it.

[2] (Pigou, 1920)
[3] (Rosenstein-Rodan, 1943)
[4] The higher is r, the lower is the present value of investments. Their level will therefore be lower.

2.2 Imperfectly Competitive Labour Markets

Becker's theory rests on the assumption that the labour market works in a competitive fashion. Over the last few decades, any of the clashes between theoretical framework and empirical evidences about training investment choices and outcomes has been justified by labour market imperfections such as credit market constraints (Bassanini et al., 2005). Recent literature (Acemoglu and Pischke, 1999) integrated the non-competitiveness of markets into the framework of reference.

In this new setting, limited worker mobility, asymmetric information, search frictions and institutional wage compression make the labour supply curve not completely elastic and give employers the (monopsony) power to set wages (Bassanini et al., 2005). Building on these considerations, Acemoglu and Pischke (1999) provide an explanation to the empirically observed tendency of firms to finance general training: in a non-competitive labour market workers do not get paid according to their marginal product and salaries are equal to a fraction of the overall productivity,

$$w(\tau) = \beta\, y(\tau)$$

where β is between 0 and 1 and can be seen as employees' bargaining power. If β is equal to 1 then wage is equal to productivity and the employee has full bargaining power; if β is between 0 and 1 the salary paid to the employee is only a share of his productivity. Acemoglu and Pischke (1999) consider β to be comprised between 0 and 1.

The total profits that the firm derives from subsidizing the training under the hypothesis of imperfectly competitive markets are given by:

$$\prod = [y_0 - \beta y_0 - c_0(\tau)] + \frac{y_1(\tau) - \beta y_1(\tau)}{1+r}$$

where total profits are equal to the sum of profits at t_0 and t_1 and profits at t_1 are discounted by a factor of $(1+r)$. Firms maximise profits when training's marginal revenue equals training's marginal cost, or rather:

$$\frac{(1-\beta)y_1'(\tau)}{1+r} = c_0'(\tau)$$

Assuming that β is comprised between 0 and 1, the optimal level of investment is positive. In other words, the wage-productivity gap gives employers the chance of recouping the training costs (Acemoglu & Pischke, 1999). The compression of the wage structure brings in fact the increases in the level of skills (and therefore training) to be not as steep as the increases in the productivity generated by the training. As we said earlier, this gap between wage and productivity is expected to be wider at higher skills levels (Acemoglu & Pischke, 1999).

The fundamental conclusion to be drawn from this revision of Becker's framework is that the more the wage structure is compressed, the higher is the amount of training which is sponsored by the firm especially for high-skilled workers (Acemoglu & Pischke, 1999; Bassanini & Brunello, 2008, 2003).

2.3 Previous Empirical Findings

Despite the differences in countries' labour markets and the way they reward individual characteristics they almost unanimously seem to reward training spells in a fairly generous way. American and Canadian literatures (Lynch & Black, 1998; Veum, 1995; Loewenstein & Spletzer, 1999; Parent, 1999, 2003; Frazis & Loewenstein, 2005) provide rich evidence of the positive wage effects of workplace training on wages, finding returns to vary between 0.2% and 3.9%.

European research (Blundell, Dearden, Meghir, & Sianesi, 1999; W. N. Arulampalam & Booth, 2001; Gerfin, 2004) reports positive wage effects comprised between 1% and 11% for male workers.

Blundell, Dearden, Meghir and Sianesi (1999) find that returns to training on the job are, on average, between 3.6% and

4.8%. They also find that employer-provided training has the largest impact on earnings and that its effects are the most long-lasting (13 years against 8–10 years for training from other sources, using US data).

German data provide evidence of large firm-sponsored investments in employees' general skills especially during work hours, starting from the apprenticeship system. The literature investigating the monetary returns to training for the case of Germany has so far reported mixed results, ranging from essentially zero (Pischke, 2000) to 6% (Muehler, Beckmann, & Schauenberg, 2007). Muehler, Beckmann, and Schauenberg (2007) apply non-parametric techniques (i.e. nearest neighbour matching, kernel matching, and local linear regression matching) to evaluate the wage effects of continuous training programs in Germany. Muehler et al. account for the role played by unobservable characteristics in the selection into training by combining matching with a difference-in-differences estimation strategy (MDiD). They find different returns to general and specific training. General training is associated with a significant 5 to 6% increase in wages; the effects of firm-specific training are instead mostly insignificant.

Pischke (2000) studies the relationship between workplace training and earnings growth using GSOEP data over the period 1986-1989. The study presents the results of standard fixed effects earnings regressions and models accounting for individual-specific growth rate of earnings to account for heterogeneity in the wage levels and in wage growth rates. Throughout the two models estimates of the effects of training on earnings are statistically insignificant; nonetheless, it is worth reporting that the fixed-effects models reveal a positive effect of training on earnings (around the 1%) but provides lower coefficients for the "training duration" variable (around the 0.1%). This is clearly hard to reconcile with human capital theory: intuitively, it should be the length and not the incidence of training spells to produce an effect on wages. As individual wage growth is included in the model specification, the magnitude of the estimates on training and training duration changes: specifically, the effect of training duration grows larger whereas that of a single training spell falls close to zero. The coefficient of training incidence in the plain fixed effects model is therefore likely to be driven by the higher participation rates of high earnings growth individuals, whereas the growth in the coefficient on duration between the first and the second model reveals that longer spells are more common among low-earners.

Differences in the estimated returns could be due to the measurement of the dependent variable (wage levels or wage growth), the choice of parametric or non-parametric models or to the study of different periods.

Lang (2012) uses GSOEP data to study the effects of training for employees belonging to different age-groups. Results show that whereas younger workers enjoy wage growth following workplace training, those belonging to older age-groups do not realize any wage rise despite reporting an increased perception of job security [5].

Kuckulenz and Zwick (2003) use BiBB/IAB administrative data to estimate the returns to training on the job at the cross-sectional level. They deal with the potential endogeneity of the treatment by using "supply-side sources of variation in training as instruments" [6], as proposed by Card (2000).

Firms are expected to offer more training after restructuring; this is due to the need to update the skills of their employees to match new skill demands. Restructuring is therefore used as an instrument for training provision. Estimates of the IV model show that training participation has a highly positive and significant effect on wages, around the 16%.

Interestingly, they address the heterogeneity training mark-up by including a set of interaction terms between training and all other explanatory factors in the earnings equation. They find that high-skilled employees profit more from training than low-skilled workers and the training earnings mark-up increases with professional experience but decreases with company tenure. Employees with previous unemployment spells and employees with temporary contracts profit less from training. Finally, smaller firms, firms in a good economic situation and firms that share profits with their employees pay a higher training earnings mark-up.

In the OLS specifications where training is not instrumented the coefficient on training participation falls to the 6%.

Kuckulenz (2006) estimates returns to training in the German labour market using extended Mincer equations over three datasets: Qualification and Career Survey (BiBB/IAB), Micro-Census (MZ) and German Socio-Economic Panel (GSOEP)

[5] (Lang, 2012)

[6] (Kuckulenz & Zwick, 2003), page 6

data. BiBB/IAB and MZ data show a positive correlation between training and wage growth whereas GSOEP data provide no significant correlations between training participation and net or gross wages changes (Kuckulenz, 2006).

Other studies compare the effect of workplace training for both employees and firms. Jones et al. (2012) (Jones, 2012) use data for a large sample of Finnish co-operative banks and compare the effects of general and firm-specific workplace training for both employees and firms. They find measures of training intensity to be positively associated with wages. In particular, general training is associated with somewhat higher wages than firm-specific training, although the difference in coefficients typically is not statistically significant. A key finding is that training has a more robust effect on wages than firm performance. This result differs somewhat from previous literature and could be due to the fact that the analysis is based on a case study. Lopes and Teixeira (2011) (Sofia Lopes & Teixeira, 2013)

We contribute to the literature by providing cross-sectional and longitudinal estimates of the effectiveness of training across the quintiles of the wage distribution. To the best of our knowledge, this is the first attempt to provide distributional estimates of the effectiveness of workplace training. We will not distinguish between general and specific training [7] and only consider those training spells taking place during work hours and being financed by the employer.
Our identification strategy makes use of a newly-developed estimator which has been introduced by Powell (2010; 2014). This new technique allows estimating unconditional quantile regressions both cross-sectionally and with the inclusion of fixed effects for longitudinal analyses. Thanks to the richness of our data we measure the impact of both the incidence and the duration of training.

3 Data

We use data from the German Socioeconomic Panel (GSOEP), a long survey panel dataset tracking a sample of 481,159 individuals (aged 17 or more and living in private households) over the last thirty years. The survey contains a wide range of information about individual and household characteristics.
Extra modules covering particular topics in details are surveyed alongside the core questionnaire. Of particular interest to us are the ones concerning ”training and vocational education”.
Cross-sectional samples range between 9,663 respondents in 1999 and 24,419 in 2000. Our cross-sectional models consider the waves of 2000, 2004 and 2008. For our longitudinal studies we focus on the time period comprised between 1997 and 2012. [8]

The occupational training and vocational education module is included in the 1989, 1993, 2000, 2004 and 2008 surveys. It provides details about training spells that have taken place over the three years before the survey was conducted. We make use of information regarding the duration of the spells and the total number of courses taken over the three years period. The value of our training variable primarily relates to the answer to the following survey question:

There are different opportunities available if one wants to educate oneself further. Think back on the last three years. Have you in that time period done any of the following to further your professional education? The answer categories are:
-Regularly read scientific or professional publications
-Attended professional conventions or congresses
-Participated in professionally oriented courses, including those which are still in progress.

The last category is the one used to construct the training variable. As it was mentioned earlier on training participation is only measured at given points in time, the ones included in this study are 2000, 2004 and 2008.
In the cross-sectional models, *training* is a binary variable taking value “1 = yes” if the employee has attended at least one

[7] As in Acemoglu and Pischke (1999), skills can be considered as industry specific but it is hard to think of skills which could be firm specific. Industry specific skills could still be considered as general at the industry level, and portable across firms in the industry.

[8] Further restrictions apply depending on the model specification as I illustrate in the results’ paragraphs

course in the three years before the interview and "0 = no" if he has not. In the longitudinal models *training* variables switch from 0 to 1 in 2000, 2004 and 2008 if respondents report any training spells over the past three years; training is then coded as 0 or 1 for the following years.

In addition to this training dummy, we construct a training duration variable carrying information about the number of hours of training. The coding of this duration variable in cross-sectional and longitudinal models follows the one used for the generic training variable. In the cross-sectional models training duration corresponds to the length of the most relevant training spell over the past three years. In longitudinal models training duration switches from 0 to the actual duration of training in 2000, 2004 and 2008; the value of this variable is then unchanged over the following years.

We only consider training spells taking place during work hours and being financed by employers. Moreover, due to the retrospective nature of the training variable, our sample includes respondents who are fully employed and whose age is comprised between 18 and 64[9] three years before the training survey is run. We adjust the level of the current minimum wage The gross minimum wage effective on 1/1/2015 is equal to 1,473.33 euros per month for full-time workers. by inflation and retain those respondents whose wage is above the corresponding minimum wage level.

4 Descriptives: Incidence and Participation

The vast majority of those reporting to attend courses with job-related content seems to be doing it thanks to employers' support. In Table 1 we illustrate the number of those who report a training spell and we break down the group into categories according to the training source of financing.

Table 1: Workplace Training financing

	Training	Employer	JobCentre	AnotherSource	Soc. Service	No Support
2000	2,142	49	30	0	773	
2004	1,810	26	31	0	716	
2008	1,626	21	26	0	678	

[10]

The vast majority of those reporting some training, both during work and leisure time, are financed by employers. The statistics presented in Table 1 stand therefore at odds with the predictions of the traditional Beckerian framework: they clearly show that there are many instances of training programmes where the firms bear the entire cost of the course. The way the Beckerian framework would explain the figures in Table 1 is by considering all that training as "firm specific". We follow Acemoglu and Pischke (1999) and consider skills as possibly industry specific but not firm specific. Industry specific skills could still be considered as general at the industry level, and portable across firms in the industry; accordingly, the Beckerian framework would expect firms not to finance any training. As we illustrated earlier on, we expect firms to sponsor training due to the fact that the wage-productivity gap gives them the chance of recollecting the money invested in training. From a theoretical point of view, Table 1 shows that due to the compression of the wage structure, firms have incentives to sponsor training and there are many instances of employer financed training programmes.

In Table 2 we present training incidence across demographic groups. We do so considering a slightly smaller sample, that of the workers reporting employers' financed training during working hours. Table 2 illustrates the composition of the sample of trained and not trained respondents across the three waves of interest. We show average net monthly wage for the two groups and we study participation by gender, age, education categories and firm size.

[9]The training question is posed to all respondents aged 16-64
[10]Training is not necessarily taking place during work hours

Table 2: Descriptive statistics 2000, 2004, 2008

Stats	2000		2004		2008	
	Trained	Not trained	Trained	Not trained	Trained	Not trained
$Avg.income(ineuros)$	1541	1367	1830	1560	2113	1828
$Women$	.28	.27	.25	.27	.27	.28
Age						
$18-30$	.18	.19	.11	.13	.08	.10
$30-40$	.39	.34	.37	.33	.28	.27
$40-50$	.29	.29	.36	.32	.34	.36
50+	.12	.17	.16	.21	.29	.26
$Education$						
Elementary	.06	.13	.05	.09	.06	.07
Secondary	.43	.60	.42	.57	.39	.53
Tertiary	.49	.25	.53	.33	.54	.39
$Firmsize$						
≤ 5	.04	.07	.02	.08	.02	.08
$5-20$	.11	.14	.10	.15	.08	.13
$20-200$	.19	.30	.21	.31	.22	.29
$200-2000$	.24	.24	.24	.24	.31	.24
≥ 2000	.41	.22	.42	.21	.36	.25
Number Obs.	194	1507	279	2336	308	1947

Net monthly income wages are higher for the treatment group across the three waves. The number of female trainees is fairly stable and slighty less than 30% across the waves. The percentage of women in the control group is also about the same in all years.

Trainees tend to bunch in low-middle age groups although the treated sample grows older over time. Human capital theory predicts that younger workers are more likely than older ones to be offered training since the period over which the training investment can be amortized is longer; it may nonetheless be in agents' interests to train workers of any age due to the rapid skills obsolescence caused by the fast technological advances (W. Arulampalam, Booth, & Bryan, 2004). GSOEP data provide evidence of the relentless ageing of the German workforce over the time period under consideration. The percentages of workers in high age groups grow for both the trainees' sample and the controls' one.[11] Our descriptives might well be capturing the tendency to provide more training for higher age groups in order to counteract skills obsolescence. In addition to that, in 2007 the German parliament accepted the government proposal to increase the retirement age from 65 to 67 years of age by 2029 and a gradual increase of the retirement age started to be enacted in 2012. An older workforce with an increasing amount of time to be spent on the labour market calls for higher investments in training for older age groups.

The share of trainees remains noticeably high into tertiary education. Figures for the share of tertiary graduates in the control group are remarkably lower. The biggest share of respondents belonging to the control group holds a secondary education degree. Data point to the existence of a strong complementarity between the various types of human capital

[11] According to the Cologne Institute for Economic research (IW Koln), in 2010 the 56% of the employees in working age and liable to social security contribution fell into the 50-54 age bracket. Moreover, between 2000 and 2010 the percentage of employees liable to social security contribution aged 60-64 has increased from the 11% to the 24.6%.

investments: early achievement and qualifications are important determinants of future educational attainment, individuals with higher educational attainment in turn undertake, and are offered, more training on the job (Blundell et al., 1999).

The share of trainees working in big companies is higher than the share of control group respondents. The share of trainees grows together with the firm size whereas the control group concentrates in firms with less than 200 employees. looking into the relationship between wages and firm sizes, Lehmer and Moeller (Lehmer & Moller, 2010) find a large firm size premium of 11% for the case of Germany. this is likely to be due to: decreased costs of monitoring through matching of productive workers; the likelihood of shirking being higher in big than in small firms and causing large establishments have to pay efficiency wages; big firms being more able to share rents because of greater market power and lower prices for non-labor inputs; the so-called productivity hypothesis assuming that the required performance standards are higher in big firms which have to be compensated by higher wages and that more productive employees are needed to operate high-technology machines. Urban wage premia might be well operating on top of firm size wage differentials. To this regard Lehmer and Moeller (Lehmer & Moller, 2010) find an urban wage premium of 8% for the case of Germany. Firm size and urban-agglomeration effects operate independently proving the role played by externalities in the urban environment and not only within firms. Firm size and regional factors could therefore bear an effect on the incidence and the returns to training.

Overall, the average demographic characteristics for the group of trainees reflect both the self-selection of certain groups into training and the selection of these groups by employers. To clarify what the incidence of training across quintiles is, Table 3 reports the percentage of trained employees by quintile of the wage distribution. For each of the three waves we report the number of trainees per quintile (first row) and the percentage of training participants for each quintile (second row).

Table 3: Training incidence by quintiles

Training wave	q1	q2	q3	q4	q5	Num. trainees
2000	77	140	167	246	322	952
q % 2000	4.31	7.72	10.16	15.08	18.73	-
2004	24	109	181	194	369	877
q % 2004	4.31	6.44	9.27	11.55	16.52	-
2008	80	108	153	235	213	789
q % 2008	6.48	8.67	12.46	18.33	18.05	-

As we see from Table 3 the vast majority of trainees is concentrated in the top quintiles of the distribution. This statistics might reflect the higher incidence of training in larger firms and the firm size premia. For these two reasons trainees might seem to cluster at the top of the distribution. Alternatively, this could mean that firms provide formal training to everyone at the top of the distribution whereas they resort to less formal and more informal training at the bottom.

As it has been pointed out in the literature, training's length is among the key factors determining returns to training. Due to the nature of tasks performed in the occupation and the varying opportunity cost of time across the skills distribution we should expect training spells to have an average different duration at different quintiles of the wage distribution. Table 4 shows the average length (in hours) of the most relevant training spell reported by quintile of the wage distribution.

Table 4 only shows a slight decline in the average length of training at the very top of the distribution; nonetheless the length of training grows higher or remains high across the four quintiles. We normally consider highly skilled workers to have a higher opportunity cost of time; for this reason we expect the average length of training to decrease towards the top of the wage distribution. This assumption is confirmed if we look at the very top quintiles. Moreover, we can reasonably assume that the scarce incidence and the shorter length of training in the bottom quintiles of the wage distribution could

Table 4: Workplace Training length by quantiles

Training avg. length	q1	q2	q3	q4	q5
2000	11.35	16.75	16.49	16.90	15.15
Std. dev.	(12.89)	(28.70)	(29.23)	(25.79)	(14.78)
2004	10.28	8.92	14.12	13.91	13.84
Std. dev.	(16.71)	(11.16)	(20.95)	(16.44)	(15.28)
2008	11.33	11.17	8.64	12.42	9.82
Std. dev.	(21.63)	(17.19)	(14.60)	(19.84)	(10.13)

be due to the higher turn over registered in this part of the wage distribution. All in all training's length remains high until the third quintile.

5 Methods

Estimation by mean regression assumes that the estimated marginal effect of training on the dependent variable is constant over the dependent variable distribution. Under this framework, the effect of having a training spell could be represented by a rightward shift of the conditional wage distribution. Quantile regressions measure instead the payoffs across the wage distribution. The effects of training spells over the deciles describe the changes both in the location and also in the shape of the wage distribution.

We present the estimates of the returns to on-the-job training obtained through the use of mean-regressions applied to cross-sectional waves and to the panel version of the data. In addition to that we illustrate the estimates calculated by the use of cross-sectional and panel quantile regressions. We analyse whether the estimates provided by the mean-regression models are due to homogeneous returns across the wage distribution or prompted by the results on some of the quantiles. Given that low, medium and high skilled workers are not perfectly substitutable in the tasks they perform, we expect their returns to human capital to be valued in an idiosyncratic way. Put differently, the interest rate on their training investments could be different. For these reasons it could be that mean estimates do a poor job in characterizing the returns to training; this is the rationale behind our interest for quantile treatment effects.

Selection into training might be affected by personal motivation, work ethic and work dedication. These characteristics are supposed to be non-randomly distributed in the sample. In addition to that, we need to deal with the selection of firms into training provision, which is likely to be composed by the largest ones or those operating in rapidly expanding sectors. These two main avenues of selection, at the individual and firm-level, constitute potential biases for our estimates. If individuals are not randomly assigned to treatment, the effect of the treatment on the outcome variable will be confounded by the biasedness of the characteristics which are specific of the treated group.

5.1 Mean Regression Models

We first run OLS earnings regressions on four waves of the GSOEP dataset and later look at the longitudinal mean effects of reporting a training spell by the use of Fixed Effects.

Omitted time-invariant individual characteristics may be correlated with the trainee status, bringing trainee status *per se'* to be spuriously correlated with the outcome variable.

In the longitudinal part of this study, we use fixed effects models. In this setting, our identification will precisely originate from changes in the training status and individual characteristics whereas time-invariant factors will be dropped. Fixed effects can be therefore a remedial technique in cases where we want to extrapolate the effects of wage levels and individual fixed characteristics; they could nonetheless still fall short of providing proper identification due to the possible correlation between training and wage growth rates, as highlighted in Pischke (2000). Due to the fact that our training variable tends to remain constant across years we do not apply a first difference fixed model(see Pischke, 2000). Our fixed effects model has the following functional form:

$$lnw_{it} = X_{it}\beta + \gamma Training_{it} + \alpha_i + \epsilon_{it}$$

where X_{it} is a set of regressors like tenure with the current employer, ISCO and firm size; $Training_{it}$ is the training variable and α_i is a fixed person-specific constant affecting all time invariant determinants of the level of earnings. In order to analyse whether training has a permanent or temporary effect on productivity and earnings, we firstly include the 2000 training variable into the model and later add variables identifying further training spells (in 2004 and 2008). The training dummy "switches" from zero to one as individuals participate into training.
In addition to this specification we provide evidences concerning the duration of training by the inclusion of a training variable carrying information about the number of weeks spent in training.

5.2 Quantile Regression Models

Through the use of quantile regression models, we study the impact and the effects of training participation across the income distribution. [12] In this study, we present the estimates of the unconditional quantile treatment effects for three cross-sections (2000, 2004, 2008) and for longitudinal data. We do it through the implementation of unconditional quantile regression models and two newly-developed estimators introduced by Powell (2010a, 2010b, 2013, 2014).

We estimate unconditional quantile treatment effects, rather than conditional ones, following a stream of the literature which has illustrated the limitations of conditional quantile estimation and encouraged the implementation of unconditional quantile techniques (Powell, 2014a, 2014b, 2013, 2010; Borah & Basu, 2013; Fort, 2012; Andrews, Li, & Lovenheim, 2012; Firpo, Fortin, & Lemieux, 2009; Frölich & Melly, 2010; Firpo, 2007). [13] Over the last few years, a considerable amount of consensus has been gathered about the inappropriateness of conditional quantile regression as a mean to study welfare effects.
Intuitively, conditional quantile regression defines "high quantiles" as observations with high earnings given the levels of the covariates included in the model but does not identify how the unconditional expectation of Y changes if the unconditional distribution of X changes (Borah and Basu, 2013)[15].
Information regarding the level of education or background characteristics of the trained or not trained workers are useful for the model identification. Nonetheless, by simply including these characteristics as regressors in our model alters our estimates' meaning.
A regression conditioning on characteristics defines in fact, e.g. "low quantiles" as observations with low earnings given job training status *and* education. This classification may include people who are actually at the top of the wage distribution to be considered as part of the lowest quantiles. In other words, observations at the top of the conditional distribution could be at the bottom of the unconditional distribution. To this regard, Andrews, Li and Lowenheim (2012), conclude that conditional quantiles are hard to interpret for policy purposes since they are *unobserved*. For this reason, when look-

[12]In the Appendix we provide an overview of the quintile estimator as opposed to the mean least squares one.
[13]Quantile regression was firstly introduced by Koenker and Basset (1982) in the form of Conditional Quantile Regression as an alternative and a compliment to ordinary least squares regression (Ker, 2011). Conditional quantile regression can be used to test for the presence of heteroskedasticity in the regression residuals [14] and can be used to overcome the violation of the errors normality assumption in the OLS framework. Conditional quantiles also allow identifying those cases in which the conditional mean is non-linear in x. If the slope of the quantiles regression lines increases across the quantiles, then the outcome variable and the predictor will not be linearly related.
Besides that, conditional quantile regressions are not subject to the distortions produced by the presence of outliers in the traditional OLS framework. Their statistic of interest is in fact other than the mean, which is heavily influenced by the presence of outliers.
[15]Importantly, if the conditional effect of the covariates does not change over the quantiles (i.e. if the data generating process follows a linear-in-parameters additive model structure), then the conditional and the unconditional treatment effects of X on the outcome variable are identical for any quantile of Y. If this is not the case and conditional effects are heterogeneous and vary over the values of other covariates, then conditional quantile models differ from the unconditional quantile ones.

ing at the wage effects of training we should not use conditional quantiles since they would not allow us to identify the unconditional quantiles of the outcome variable distribution; through the use of a multivariate conditional quantile model we would estimate the effects of training on subgroups of individual whose position in the wage distribution is high or low given the set of covariates. It is therefore very difficult to derive policy implications from conditional quantile estimates.

In addition to this substantial difference in the results' interpretation, the law of iterated expectations does not hold for the case of conditional quantiles, making it impossible to infer the effects of a policy change on the unconditional distribution of the outcome variable when starting from the conditional ones (Firpo et al., 2009).

In the case of OLS mean regression, the conditional mean averages up to the unconditional mean; in other words, the law of iterated expectations [16] holds and we can go from conditional to unconditional.

This property is not valid for the case of quantile regression whereas instead the law of iterated expectations does not hold and conditional quantiles do not average up to their population (i.e. unconditional) counterparts.[17] This means that with quantile models there is a substantial difference between conditional and unconditional estimation and conditioning on regressors in the traditional "additive" way does not allow us to infer what the effect on the unconditional outcome variable distribution is.

In conclusion, the estimates of conditional quantile models are not only conceptually different from those of unconditional quantile ones. They also provide no indication of the effects of the treatment on the unconditional distribution of the outcome variable. In this study we overcome these structural limitations by resorting to unconditional quantile regression methods and applying the estimators introduced by Powell (2010a, 2010b, 2013, 2014).

5.3 Unconditional Quantile Regression Models

Powell develops two Unconditional Quantile Regression estimators, for cross-sectional and for longitudinal models. The two estimators are derived through the use of Generalized Method of Moments (GMM): parameters are estimated through the minimization of residuals in sample moments conditions representing the population ones, on the basis of the "analogy principle". In the Appendix we provide a thorough description of the derivation of the cross-sectional and longitudinal estimators. The two subsections' contents are based on the recent work by D. Powell (2010, 2010b, 2013, 2015), and we refer the reader to these papers for a more detailed discussion of the econometric model.

Unconditional quantile treatment effects describe the effect of treatment variables on quintiles of the dependent variable distribution without reference to the control variables (Powell, 2013). Intuitively, the easiest way to obtain unconditional quantile treatment effects is to avoid conditioning on covariates. This is however not possible in cases where the treatment is not exogenous and in cases where, despite having an exogenous treatment, we want to base the identification on the use of additional covariates.

Powell presents models where covariates are not conditioned upon in the traditional way but they only inform the probability that an observation falls into a given quintile. This probability does therefore change across individuals. Explanatory variables are divided into "control" variables and "treatment" variables and "treatment" variables shift the distribution of the outcome variables. Control variables are instead considered as partial determinants [18] of the rank in the distribution. Formally, control variables are excluded from the Structural Quantile Function (henceforth SQF) and are therefore not studied in this framework. The outcomes of the process at the cross-sectional level are modelled as:

[16]
$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y \mid X))$$

[17] For instance, for the case of quantiles, the 25th quantile of the conditional distribution does not average up to the 25th unconditional quantile.

[18] We call them partial because there is still an unobserved disturbance given by, broadly speaking, *ability* that we cannot observe.

$$Y_i = D_i' \beta (U_i^*), \;\; U_i^* \sim U(0,1)$$

The analogous outcomes in the longitudinal case are:

$$Y_{it} = D_{it}' \beta (U_{it}^*), \;\; U_{it}^* \sim U(0,1)$$

In both these two cases, D_{it} represents the treatment variables and Y_{it} is the distribution of the outcome. U_{it}^* represents ability or proneness to the outcome (Doksum, 1974) and may be a function of several disturbance terms. U_{it}^* is an unknown function of both an individual fixed effect and an observation-specific disturbance term.

We firstly illustrate the estimator used for the cross-sectional analysis and later move on to illustrating the longitudinal one.

5.3.1 Cross-sectional Unconditional Quantile Regression Models

Sample moments are generically defined by the following expression:

$$\hat{g}(b) = \frac{1}{N} \sum g_i(b)$$

Covariates are allowed to inform the probability that an observation falls under a given quantile through the following expression:

$$\hat{\tau}_{x_i}(b) = \hat{P}(y_i \leq \hat{\gamma}(\tau, \tilde{b}) + \tilde{d}_i' \tilde{b} \mid x_i)$$

Where $\hat{\gamma}(\tau, \tilde{b})$ is the τ^{th} quantile of the distribution of $y_i - \tilde{d}_i' \tilde{b}$ and solves:

$$\frac{1}{N} \sum \mathbf{1}(y_i - \tilde{d}_i' \tilde{b} \leq \hat{\gamma}(\tau, \tilde{b})) = \tau$$

The probability $\hat{\tau}_{x_i}(b)$ that an observation is below or above a given quantile function can be estimated through the use of a logit or probit regression (Powell, 2013). The estimation is straightforward due to the fact that $\hat{\gamma}(\tau, \tilde{b})) + \tilde{d}_i' \tilde{b}$ is known. As we noticed earlier on, this probability is dependent on x_i.

The sample moment condition, derived on the basis of the analogy principle with the population moment is:

$$g_i(b) = d_i[\mathbf{1}(y_i \leq \tilde{\gamma}(\tau, \tilde{b}) + \tilde{d}_i' \tilde{b}) - \hat{\tau}_{x_i}(b)]$$

and the estimator is finally:

$$\hat{\beta(\tau)} = argmin \, \hat{g}(b)' \, \hat{A} \, \hat{g}(b)$$

As we said $\hat{g}(b) = \frac{1}{N} \sum g_i(b)$ and $\hat{A}$ is in this case the identity matrix.

5.3.2 Longitudinal Unconditional Quantile Regression Models

As we can see from the way outcomes are modelled [19], Powell (2014) proposes a framework where parameters vary based on a non-additive disturbance term. He argues in fact that the conceptual motivation underlying the use of quantiles is to allow the parameters of interest to vary based on the non-separable disturbance term U_{it}. For the case of panel data, separating α_i partially undermines this original motivation and there is frequently little economic justification to allow the parameters to vary based only on part of the disturbance term and exclude the other part simply because it is fixed over

[19]

$$Y_{it} = D_{it}' \beta (U_{it}^*), \;\; U_{it}^* \sim U(0,1)$$

time (Powell 2014). Observations at the top of the $(Y_{it} - alpha_i)$ distribution may be at the bottom of the Y_{it} distribution. For this reason, additive fixed effect models cannot provide information about the effects of the policy variables on the outcome distribution.

Sample moments are:

$$g_i(b) \;=\; \frac{1}{T} d_{it}[\mathbf{1}(y_{it} \leq d_{it}'b) \;-\; \frac{1}{T}\sum \mathbf{1}(y_{is} \leq d_{is}'b)]$$

and

$$h(b) = \frac{1}{NT}\sum\sum \mathbf{1}(y_{it} \leq d_{it}'b) \;-\; \tau$$

We could define the second part of the first moment condition, i.e. $\frac{1}{T}\sum \mathbf{1}(y_{is} \leq d_{is}'b)$, as $\tau_i(b)$. It makes intuitive sense to think that the individual fixed effect provides information about the distribution of the disturbance (Powell, 2010). [20]

For the estimation, we first need to define $\hat{\gamma_t}(\tau, \tilde{b})$ as the τ^{th} quantile of the distribution of $(y_{it} - x_{it}'\tilde{b}) \;=\; \tau$. For any estimate $\tilde{b}$, the values $\hat{\gamma_t}(\tau, \tilde{b})$ are known. $\hat{\gamma_t}(\tau, \tilde{b})$ solves the following equation:

$$\frac{1}{N}\sum \mathbf{1}(y_{it} \;-\; \tilde{x_{it}'}\tilde{b} \;\leq\; \hat{\gamma_t}(\tau, \tilde{b})) \;=\; \tau$$

This equation brings the second sample moment, $h_t(b) \;=\; 0$ defining year fixed effects to hold for all t. The first step in the estimation process is therefore estimating $\tilde{b}$ and calculate the fixed effects by solving for $\hat{\gamma_t}(\tau, \tilde{b})$.

Finally, $\hat{\beta}$ is obtained as:

$$\hat{\beta(\theta)} \;=\; argmin\, \hat{g}\,(b')\,\hat{A}\,\hat{g}\,(b)$$

where $g_i b$ is defined by the first moment condition and $\hat{A}$ can be the identity matrix. The β minimizing this condition is the $\hat{\beta(\tau)}$.

6 Results

We first estimate the mean returns to training by running cross-sectional earning regressions with net monthly income as the dependent variable. *Training* variables are coded as dummies indicating whether the individual has had at least one training spell in the three years before the survey was conducted. *Extra week* variables are duration-based measures of training intensity; they indicate the impact of an extra week of training on net monthly wage.

Our models include controls for education (in years), tenure, occupational category and firm size.

Table 5 shows the estimates of the coefficients for the workplace training variables across the three cross-sections of interest. Rows identify cross-sectional models for 2000, 2004 and 2008, for simplicity, we only report the value of the coefficient on training and training duration.

The coefficients are all positive although significance is lost over the last cross-section (2008). The magnitude of the coefficients on a given training spell is similar to what has been previously found in the literature. Interestingly, the one on an extra week of training has a lower magnitude, which is still considerably high in absolute values. Overall our results confirm the correlation between training participation and wage levels and the substantial impact of training on wages.

In the next table we investigate whether looking at the quintiles of the wage distribution provides a clarification to the large magnitude of the coefficients in Table 5. Table 6 shows the effects of a training spell and those of an extra week of training.

Firstly we would like to point out that, like conditional mean regression coefficients and conditional quantile regression coefficients, unconditional coefficients are interpreted as the effect of a marginal change in the predictor on the unconditional

[20] Practically, consistent estimation of τ_i is not needed since the first sample moment is equivalent to a series of pairwise comparisons and can be replaced by an alternative formulation, Powell (2010), page 12.

Table 5: OLS: Cross-sectional Training Impact on Net-Wage

	1 Coeff. and Std. Error	Num. Individuals
Training 2000	0.053***	9,301
	(0.016)	-
Extra week 2000	0.028**	9,301
	(0.013)	-
Training 2004	0.04***	7,292
	(0.012)	-
Extra week 2004	0.025***	7,292
	(0.008)	-
Training 2008	0.023	6,170
	(0.014)	-
Extra week 2008	0.005	6,170
	(0.009)	-

Dependent variable is the logarithm of wage

* $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$

Table 6: Unconditional Quantile Regression

	1 20q	2 40q	3 60q	4 80q	5 Num. Ind.
Training 2000	.06**	.06***	.08***	.04	9,301
Std. Error	(.027)	(.019)	(.020)	(.032)	-
Extra week 2000	.02	.04**	.04**	.05*	9,301
Std. Error	(.028)	(.018)	(.019)	(.03)	-
Training 2004	.1***	.07***	.08***	.01	7,292
Std. Error	(.023)	(.016)	(.027)	.024	-
Extra week 2004	.06***	.05***	.05***	.01	7,292
Std. Error	(.013)	(.017)	(.015)	(.014)	-
Training 2008	.07**	.06***	.08***	.03	6,170
Std. Error	(.035)	(.026)	(.021)	.03	-
Extra week 2008	0	.03	.04***	0	6,170
Std. Error	.02	.02	.015	.017	-

* $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$

Dependent variable is the logarithm of wage quantile of interest.

The model captures some variation in the returns to a generic training spell and to an extra week of training. This could be the result of different training incidence and length of the training spells across quantiles combined with a different opportunity cost of time: the ones at the top of the wage distribution are more likely to have slightly longer spells (see descriptive statistics) having a high opportunity cost of time; for the ones at the bottom, the opposite holds. For this reason, an extra week of training for top-earners is likely to yield returns which are as low as those for workers with lower salaries and way longer training spells (bringing their marginal productivity to decrease).

Table 7 presents the results of three fixed effects models where the effects of training are evaluated over time, by using the dataset in its panel form. The first column illustrates the effects of reporting at least one training spell in the three years prior to 2000 on the following years whereas the second model specification considers the three spells separately. The remaining column includes the effects of an extra week of training (reported in 2000) over time.

The sample size in the second model is different from that of the other two specifications. The effect of a training spell in 2000 and the effect of an extra week of training in 2000 are evaluated on a shorter time span 1997 - 2001) to avoid overlaps with the following spells of training.[21]

Table 7: Fixed Effects Models

	(1) income Coef./std.errors	(2) income Coef./std.errors	(3) income Coef./std.errors	(4) income Coef./std.errors
Training 2000	.053***	.066***		
	(.008)	(.010)		
Training 2004		.036***		
		(.007)		
Training 2008		.028***		
		(.007)		
# Weeks of Training			0.028***	
			(.007)	
Constant	6.72***	6.38***	6.692***	
	(.101)	(.034)	(.103)	
No. of Obs.	14285	60739	14285	
No. of Ind.	3960	5764	3960	
R-Squared (overall)	.059	0.07	.055	

Dependent variable is the logarithm of wage. Models include a variable for wage growth, dummies for firm size, occupation divided into 9 broad categories, potential experience, potential experience squared, cubic and quartic.
* $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$

A fixed effects model is a linear model where parameters are identified from within changes in variables, eliminating all time-invariant heterogeneity among the observations. In our models we apply individual fixed effects; the regression's parameters represent therefore the effect of the within-person changes on the dependent variable, i.e. net monthly income. Part of the selection problems due to the heterogeneity among respondents are dealt with by the use of such an identification strategy. The effects of individuals' proneness to success and those of heterogeneous levels of ability on the outcome variable are accounted for by considering within-individual variation and discarding the differences in levels. [22]

The first model shows that training has a positive and significant impact on net monthly income over time. Following a generic training spell, the expected increase in the outcome variable is equal to 5.3%.
The second column adds two additional regressors, training in 2004 and training in 2008, to the first baseline specification. These estimates confirm the findings relating to the decreasing returns to successive investments in human capital (Blundell et al., 1999). The effects of each spell on net wages are respectively 3.6% and 2.8%. The coefficient on training in 2000 increases to 6.6% in this second specification. As this model considers the time period 1997-2012 (due to the inclusion of the second and third training survey waves) we could interpret the growth in the coefficient on a generic spell as indicating that through the exercise of the newly acquired skills workers grow more and more productive or by considering the initially lower coefficient as indicating that some of the workers' increased productivity is retained by the employers to pay back for

[21] Pischke considers the three years between 1986 and 1989 in his fixed effects models specifications.
[22] In cases like the one under consideration here, it would be optimal to control for both individual and firm fixed effects. Unfortunately, the lack of data linking individuals to firms prevents us from including firm fixed effects in the model.

their investments. Finally, an extra week of training yields an average payoff of 2.8%. This result shows that the effect of an extra week of training is lower than that of a generic spell regardless of its length.

Table 8 presents the results of six unconditional quantile fixed effects models evaluating the effects of the number of training spells and training duration on the net monthly wages over time. For every training questionnaire's wave, we run two models. The first model's specification includes a training dummy variable for 2000 and controls for individual and year fixed effects. The second model analyses the effects of an extra week of training.

Table 8: Unconditional Quantile Regression - Longitudinal

	1 20q	2 40q	3 60q	4 80q	5 Num. Ind.
Training 2000	.05**	.04	.08***	.06***	3960
Std. Error	(.018)	(.023)	(.017)	(.023)	-
Extra week 2000	.03***	.02	.03***	.00	3960
Std. Error	(.009)	(.019)	(.007)	(.009)	-
Training 2004	.06***	.06***	.05***	.03***	5098
Std. Error	(.019)	.013	.014	.01	-
Extra week 2004	.00	.02***	.03***	.02***	5098
Std. Error	.02	(.007)	(.012)	(.008)	-
Training 2008	.07***	.06***	.07***	.06***	5573
Std. Error	(.023)	(.018)	(.014)	(.018)	-
Extra week 2008	.05***	.05***	.04***	.03***	5573
Std. Error	.015	.009	.012	.005	-

* $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$

Dependent variable is the logarithm of wage

Estimates for the effects of a training spell over time are all positive and most of the times significant.

Again the model captures some variation in the returns to a generic training spell and to an extra week of training. Over time returns to the 2000 and 2004 waves decrease, hinting at the temporary effect of training. Returns to 2008 training are either stable or even increasing in the second year after the survey.

We are only able to compare the coefficients deriving from the fixed effect model with the first two lines in Table 8. Despite the presence of variation in the returns to training there is no evidence of disproportionate returns for a given quintile. As we said earlier, this could be due to the difference in length of training and opportunity cost of time.

7 Summary and Discussion

The prolific literature studying the returns to workplace training has provided evidence of the positive impact of training on salaries even in cases of employer-financed training spells. The magnitude of the estimated coefficients on training tends to be large, raising doubts about the accuracy of the measures used for training. Importantly, all the evidence provided so far refers to mean estimates of the effects of workplace training on earnings. We have analysed whether returns vary across the distribution and if some of the quintiles benefit from disproportionate returns driving the mean estimates higher.

The vast majority of those reporting to attend courses with job-related content seems to be doing it thanks to employers' support. there are many instances of training programmes where the firms bear the entire cost of the course. Descriptive statistics show that trainees report higher wages and that the gender composition of the treated and controls' group is similar. Trainees tend to be concentrated in low-middle age groups although the treated sample grows older over time. As it has been discussed, this could be the result of both the general ageing of the workforce and the statutory increase of the retirement age. Workers reporting training spells are also more educated and work in larger firms. Incidence is higher in top quintiles were the length of training also tends to be high.

Results show that returns to workplace training vary across the wage distribution although there is no dominant quintile which is strongly biasing the mean estimates upwards or downwards. Mean estimates always fall into the quintile estimates range and coefficients are positive throughout the models. Depending on the waves under consideration, coefficients looking at the cross-sectional effects of training increase or decrease as we adopt a longitudinal perspective. To the aim of preserving our coefficients from being “contaminated” by subsequent spells we restrict our longitudinal analysis to the three years before the training variable is surveyed and one year after the training variable is surveyed. For the case of 2004 quintile coefficients decrease longitudinally; for the 2000 wave they only slightly decrease longitudinally; for 2008 they are either unaltered or they even increase over time. Overall our estimates for an extra week in training are lower than the ones associated to a generic spell showing that longer spells of training might play against the accumulation of higher returns.

In future versions of this paper we will further discuss the policy implications that can be drawn from our results. We would like to address both individuals training choices and firms budget allocation decisions in our discussion.

8 Appendix

8.1 Quintiles overview

Given the following data generating process:

$$y = X\beta + \epsilon$$

$X \in R^K$ is a vector of explanatory variables, $\beta \in R^K$ is a vector of unknown parameters and ϵ is an unknown independently and identically distributed error term. The least squares estimator of $\hat{\beta}$ is defined as:

$$argmin\, \Sigma\, (y - X\beta)^2$$

In a conditional quantile regression, $\hat{\beta_\theta}$ is the solution to the following weighted (ρ_θ) sum of residuals:

$$argmin\, \Sigma\, \rho_\theta |y - X\beta|$$

where $\rho_\theta = \theta$ if $Y - X\beta \geqslant 0$ and $\rho_\theta = (1-\theta)$ if $Y - X\beta < 0$.

For quantiles above the median, e.g. $\theta = 0.75$, a higher weight is placed on residuals above the quantile than on residuals below the quantile: this pushes the minimization up above the median, which is where we want it to be in such cases. The reverse holds true for quantiles below the median. For those quantiles, less weight is placed on residuals above the quantile, which pushes the value down below the median.

8.2 Cross-Sectional Unconditional Quantile Regression Models: Estimators

Powell introduces a generalized quantile regression method (GQR) to estimate unconditional quantile treatment effects. The estimator does not assume that treatment variable(s) are orthogonal to the “proneness to the outcome”. GQR allows in fact the covariates to inform the distribution of the disturbance.

The model makes four assumptions, these conditions should hold jointly with probability one:

1) Potential Outcomes and Monotonicity: $y_i = d_i'\beta(u_i^*)$ is increasing in $u_i^* \sim U(0,1)$.

   The assumption models the disturbance as a uniform variable.

2) Conditional Independence:

   (a) $P(u_i^* \leq \tau \mid d_i, x_i) = P(u_i^* \leq \tau \mid x_i)$

   (b) $E[d_i \nu_i] = 0$

According to the first part, the policy variable(s) does not provide additional information about the disturbance once the covariates are conditioned on. Variation in d cannot predict heterogeneity in $u^* \mid x$ [23].

The second part addresses the estimation error problem. The estimation error is given by $\nu_i = \hat{\tau}_{x_i} - \tau_{x_i}$. The assumption states that estimation errors are not a problem as long as they are orthogonal to the policy variables [24].

3) Full Rank: if $P(d'c = 0 \mid x_i) = 1$ for all x_i, then $c = 0$

   This is a full rank condition that requires the variation in the policy variable to be independent of the covariates.

4) Continuity: y_i is continuously distributed conditional on d_i, x_i. This is necessary for identification.

This set of assumptions leads to the following two population moment conditions:

$$E\, d_i[\mathbf{1}(y_i \leq d_i'\beta(\tau)) - \hat{\tau}_{x_i}] = 0$$

and

$$E[\mathbf{1}(y_i \leq d_i'\beta(\tau)) - \tau] = 0$$

If $\hat{\tau}_{x_i}(b) \equiv \hat{P}(y_i \leq d_i'b \mid x_i)$, the sample moments are:

$$g_i(b) = d_i[\mathbf{1}(y_i \leq d_i'b) - \hat{\tau}_{x_i}(b)]$$

and

$$h(b) = \frac{1}{N}\sum \mathbf{1}(y_i \leq d_i'b) - \tau$$

[25]

For comparison, the sample moment that traditional quantile regression relies on is the following:

$$\tilde{g_i}(b) = d_i\,[1(y_i \leq d_i'b) - \tau]$$

In this equation, τ is not determined by x_i; Powell's estimator replaces instead τ with a function of x, allowing the covariates to inform the distribution of the disturbance (Powell, 2013, page 15). The set of x, or rather, the set of covariates, provides information about the proneness to the outcome.

8.3 Longitudinal Unconditional Quantile Regression Models: Estimators

As in the case of the cross-sectional model, we need to make four assumptions which are supposed to hold jointly with probability one:

1) Potential Outcomes and Monotonicity: $y_{it} = d_{it}'\beta(u_{it}^*)$ is increasing in $u_{it}^* \sim U(0,1)$. The assumption models the disturbance as a uniform variable.

2) Conditional Independence: $E[1(u_{it}^* \leq \tau) - 1(u_{is}^* \leq \tau) \mid d_i, \alpha_i] = 0$ for alls,t. This means that the distribution of u_{it}^* can change over time but this change must be independent of α_i, d_i.

3) Full Rank: $E[d_i]$ is rank k. This assumption requires within-individual variation in the policy variables.

4) Continuity: y_i is continuously distributed conditional on d_i, α_i.

No restriction is imposed on the relationship between u_{it}^* and α_i as well as on the relationship between α_i and d_i. These assumptions lead to the following two population moment conditions:

$$E\, d_i[\mathbf{1}(y_{it} \leq d_{it}'\beta(\tau)) - \frac{1}{T}\sum \mathbf{1}(y_{is} \leq d_{is}'\beta(\tau))] = 0$$

[23] Interestingly, this relaxes the standard assumption $P(u_i^* \leq \tau \mid d_i) = \tau$ according to which, for a given value of the treatment effect, the probability that the disturbance falls under a quantile is fixed across individuals.

[24] A different choice of the method to estimate τ_{x_i} (i.e. probit, logit or less parametric techniques), makes the gap $\hat{\tau}_{x_i} - \tau_{x_i}$ grow or shrink.

[25] Notice that if all variables are treatment variables, then $\hat{\tau}_{x_i} = \tau_{x_i} = \tau$.

and

$$E[1(y_{it} \leq d'_{it}\beta(\tau)) - \tau] = 0$$

.

References

Acemoglu, D., & Pischke, J.-S. (1999). Beyond becker: training in imperfect labour markets. *The Economic Journal*, *109*(453), 112–142.

Andrews, R. J., Li, J., & Lovenheim, M. F. (2012). *Quantile treatment effects of college quality on earnings: Evidence from administrative data in texas* (Tech. Rep.). National Bureau of Economic Research.

Arulampalam, W., Booth, A. L., & Bryan, M. L. (2004). Training in europe. *Journal of the European Economic Association*, *2*(2-3), 346–360.

Arulampalam, W. N., & Booth, A. L. (2001). Learning and earning: Do multiple training events pay? a decade of evidence from a cohort of young british men. *Economica*, *68*(271), 379–400.

Bassanini, A., Booth, A. L., Brunello, G., De Paola, M., & Leuven, E. (2005). *Workplace training in europe* (Tech. Rep.). IZA Discussion Papers.

Bassanini, A., & Brunello, G. (2003). Is training more frequent when wage compression is higher? evidence from the european community household panel.

Bassanini, A., & Brunello, G. (2008). Is training more frequent when the wage premium is smaller? evidence from the european community household panel. *Labour Economics*, *15*(2), 272–290.

Blundell, R., Dearden, L., Meghir, C., & Sianesi, B. (1999). Human capital investment: the returns from education and training to the individual, the firm and the economy. *Fiscal studies*, *20*(1), 1–23.

Borah, B. J., & Basu, A. (2013). Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence. *Health Economics*, *22*(9), 1052–1070.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *The Annals of Statistics*, 267–277.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, *75*(1), 259–276.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, *77*(3), 953–973.

Fort, M. (2012). Unconditional and conditional quantile treatment effect: Identification strategies and interpretations.

Frazis, H., & Loewenstein, M. A. (2005). Reexamining the returns to training functional form, magnitude, and interpretation. *Journal of Human Resources*, *40*(2), 453–476.

Frölich, M., & Melly, B. (2010). Estimation of quantile treatment effects with stata. *Stata Journal*, *10*(3), 423.

Gerfin, M. (2004). *Work-related training and wages: an empirical analysis for male workers in switzerland* (Tech. Rep.). IZA Discussion paper series.

Jones, P. . K. A., Derek C. & Kalmi. (2012). The effects of general and firm-specific training on wages and performance: Evidence from banking. *Oxford Economic Papers*, *64*(1), 151 - 175. Retrieved from `http://search.ebscohost.com/login.aspx?direct=truedb=eohAN=1289146site=ehost-live`

Ker, A. (2011). Conditional and unconditional quantile estimation of telecommunications engel curves. *Unpublished manuscript, University of Guelph*.

Kuckulenz, A. (2006). *Continuing vocational training in germany: A comparative study using 3 german data set* (Tech. Rep.). ZEW Discussion Papers.

Kuckulenz, A., & Zwick, T. (2003). *The impact of training on earnings: Differences between participant groups and training forms* (Tech. Rep.). ZEW Discussion Papers.

Lang, J. (2012). The aims of lifelong learning: Age-related effects of training on wages and job security.

Lehmer, F., & Moller, J. (2010). Interrelations between the urban wage premium and firm-size wage differentials: A microdata cohort analysis for germany. *Annals of Regional Science*, *45*(1), 31 - 53. Retrieved from `http://search.ebscohost.com/login.aspx?direct=truedb=eohAN=1124231site=ehost-live`

Loewenstein, M. A., & Spletzer, J. R. (1999). General and specific training: Evidence and implications. *Journal of Human Resources*, 710–733.

Lynch, L. M., & Black, S. E. (1998). Beyond the incidence of employer-provided training. *Industrial & Labor Relations Review*, *52*(1), 64–81.

Malcomson, J. M. (1999). Individual employment contracts. *Handbook of Labor Economics*, *3*, 2291–2372.

Muehler, G., Beckmann, M., & Schauenberg, B. (2007). The returns to continuous training in germany: new evidence from propensity score matching estimators. *Review of Managerial Science*, *1*(3), 209–235.

Parent, D. (1999). Wages and mobility: The impact of employer-provided training. *Journal of Labor Economics*, *17*(2), 298–317.

Parent, D. (2003). Employer-supported training in canada and its impact on mobility and wages. *Empirical Economics*, *28*(3), 431–459.

Pigou, A. (1920). *Wealth and welfare.* Macmillan, London.

Powell, D. (2010). Unconditional quantile regression for panel data with exogenous or endogenous regressors. *Available at SSRN 1722954*.

Powell, D. (2013). A new framework for estimation of quantile treatment effects: Nonseparable disturbance in the presence of covariates.

Powell, D. (2014a). Documentation for quantile regression for panel data (qrpd).

Powell, D. (2014b). Estimation of quantile treatment effects in the presence of covariates.

Rosenstein-Rodan, P. N. (1943). Problems of industrialisation of eastern and south-eastern europe. *The Economic Journal*, 202–211.

Sofia Lopes, A., & Teixeira, P. (2013). Productivity, wages, and the returns to firm-provided training: fair shared capitalism? *International Journal of Manpower*, *34*(7), 776–793.

Veum, J. R. (1995). Sources of training and their impact on wages. *Industrial & Labor Relations Review*, *48*(4), 812–826.