On the Link Between On-the-Job Training and Earnings Dispersion.

by

Said Hanchane[1]

and

Jacques Silber[2]

September 2007

Not to be quoted without the authors permission.

[1] Institut d'Economie Publique (IDEP), Marseille, and Laboratoire d'Economie et de Sociologie du Travail (LEST), Aix-en-Provence, France. Email: hanchane@ univ-aix.fr

[2] Department of Economics, Bar-Ilan University, 52900 Ramat-Gan, Israel, and Institut d'Economie Publique (IDEP), Marseille, France. Email: silberj@ ail.biu.ac.il. This paper was started when Jacques Silber visited the Laboratoire d'Economie et de Sociologie du Travail (LEST) in Aix-en-Provence, France. It was revised and completed when he visited the Fundación de Estudios de Economía Aplicada (FEDEA) in Madrid and the Laboratorio Riccardo Revelli at the Collegio Carlo Alberto in Moncalieri (Torino). Jacques Silber thanks these institutions for their warm hospitality.

Abstract

This paper is a first attempt to devise a methodology that allows estimating the exact impact of training on the dispersion of wages. It uses an approach originally proposed by Fields (2003) but extends it to the breakdown of inequality by population subgroups as well as to the case where the earnings function that is at the base of the analysis has to be adjusted for selectivity bias. The empirical illustration is based on a survey conducted in France at the end of the twentieth century.

J.E.L. Classification: J24

Key Words: earnings' dispersion – France - labour market segmentation - on-the-job training – overlapping - selectivity bias – unobserved heterogeneity

1 Introduction

Numerous studies have shown the importance of a skilled workforce both at the individual level, since higher levels of human capital increase earnings and the probability of finding a job, and at more aggregated levels since a more competent labor force improves the chances of success of firms as well as of nations. Human capital investments however do not include only education but encompass also continuing vocational training. A recent study of training among adult workers (see, O. E. C. D., 1999) has shown that the levels of training differ significantly across (OECD) countries, that although men and women participate at fairly equal rates in job-related training, men may receive more financial support from their employers and that training tends to fall off with age, although there are big differences between countries. This study stressed also the fact that training reinforces skill differences resulting from unequal participation in schooling (with again important inter-country differences), that workers receive more training in countries with higher average levels of education and that the wage premium associated with training differs between educational and gender groups (for other studies stressing the unequal distribution of training see, for example, Croquey, 1995, Aventur and Hanchane, 1999, Blundell et al., 1999, Ariga and Brunello, 2003).

Béret and Dupray (2000) emphasized additional aspects of this unequal access to training such as the fact that training seems to be positively correlated with the professional status of the individual in the firm, the nature of his work contract and the size of the firm and that it increases with seniority in the firm. Such conclusions may however imply that the main goal of training is not to increase productivity but to "keep the workers in the firm" (see, Goux and Maurin, 1997). This could also explain why training has often a weak impact on earnings (see, Béret and Dupray, 2000, and Goux and Maurin, 1997) unless the individual participates frequently in training programs (see, Béret and Dupray, 1998).

The role played by training may in fact be even more complex. The OECD report (1999) mentioned earlier stressed thus that "earnings growth after training and the event of training may not be independent variables. Unobserved individual characteristics may determine both the probability that someone is trained and the fact that they earn higher-than-average wages after the training". On the basis of the comparative study that had been conducted this report concluded that half of the earnings gap between those who received training and those who did not is due to the fact that firms providing training pay higher salaries in any case, the second half of the gap being related to factors that have a simultaneous impact on the probability of access to training and on earnings.[3] If however the two populations of individuals, those who received and on-the-job training and those who did not, have unobserved characteristics that are very different, the impact of on-the-job training on earnings should be computed, net of the effect of this unobserved heterogeneity.

Assume we indeed find first that there is a selectivity bias, second that there still remains a net (of the role played by the unobserved heterogeneity) effect of on-the-job training on earnings. If we then divide our sample of workers into two groups, the first one including those who did not receive training (say, group A), the second one, those who did (group B), we will necessarily observe that the between groups (A and B) variance of (the logarithms of)

[3] It is important to understand that the role of training may vary from one country to another. Thus in Germany the educative system is such that the knowledge accumulated at school has a high productive value and there is little uncertainty about the skills of those who hold a diploma. Continued training may then be considered as an additional way of improving the quality of the human capital of the workers and hence has a clear impact on earnings. In France on the contrary there is a lot of uncertainty about the skills of those who hold a diploma, specially at low and intermediate levels, so that firms will choose a strategy that progressively reveals the productive capacities of the workers. Such a matching process explains why access to training has to be selective and is mainly reserved to those workers who succeeded in overcoming the barriers to entry into internal markets.

earnings is significantly different from zero. There are then two possibilities. Either the within groups variance (of the logarithms) of earnings is important, or it is not. In the latter case this would imply that the unobserved heterogeneity that was found to have a significant impact on the probability to receive training and on the earnings themselves is in fact the "hidden" criterion for labor market segmentation.

If however the within groups variance turns out to be important and in particular if it is much greater than the between groups variance, one would have to conclude that there is a great degree of overlapping between the two distribution of earnings, those of groups A and B. It should then be clear that the division of the sample in two groups based on a distinction between those who received and those who did not receive on-the-job training is not relevant any more because the between groups variance turns out to be small compared to that of the within groups. As a consequence on-the-job training (unless the unobserved heterogeneity has also an important effect on the within groups variance) cannot be in such a case a relevant criterion of labor market segmentation.

Testing such hypotheses remained a difficult task until very recently. The main goal of this paper is to show that new developments in income inequality decomposition techniques and in the application of such techniques to regression analysis (see, Fields, 2003) allow us today to implement such tests because it has become possible to determine the exact impact of each variable not only on the overall variance of earnings but also on both the between and within groups dispersion, the groups referring here to those who received and did not receive on-the-job training. Our study may thus shed new light on the link between training and earnings dispersion[4].

We will proceed in three stages. First, as has often been done in the past, in estimating an earnings function that makes a correction for the selectivity bias related to the fact that an individual received or did not receive on-the-job training, we will be able to check the net effect (once this selectivity bias is taken into account) of such a training on earnings.

Second by comparing the relative importance of the between and within groups dispersions of earnings we will find out whether there is a significant degree of overlapping between the distribution of earnings of the two groups previously mentioned (those who received and did not receive on-the-job training).

Third by finally applying Fields' (2003) technique, we will be able to quantify the exact contribution of the observed (the explanatory) variables and of the unobserved individual characteristics to the variance of earnings.

The paper is organized as follows. Section 2 describes the data sources and section 3 gives the estimates of the coefficients of the earnings function. In section 4 we show how it is possible to determine the exact impact of training and other variables on earnings dispersion between and within groups while in section 5 we present the results of such a breakdown. Concluding comments are given in section 6.

2 The data sources

The analysis is based on a survey called «Continued Training 2000». Whereas most of the information on continued training in France comes from firms, this survey was addressed to individuals, whether they received some kind of training or not. This survey covered

[4] For general studies of the causes of increasing wage dispersion, see, for example, Levy and Murnane, 1992, or Karoly, 1992. For studies emphasizing the role of skill biased technological change, see, for example, Bound and Johnson, 1992, Katz and Murphy, 1992, and more recently, Heckman and Lochner, 1998, and Krusell et al., 2000.

individuals who were less than 65 years old, had completed their initial formation and were not in the army at the time of the survey.
The originality of this survey is that it covers all types of training, whatever their goal, whether they had a (direct or indirect) professional objective or whether their aim was more personal. These types of training included therefore:

- Practical training: courses taken within the framework of continued training, seminars and conferences attended. In other words this type of training implies the presence of a specialized "trainer" in a location that is different from the workplace
- On the job training: this training takes place on the job with the help of a tutor and it implies the utilization of the usual work tools.
- Self-formation: here the individual trains himself/herself, eventually with the help of specific tools. This type of training includes what is called open formation or distance learning.
- Alternative periods of training: this type refers to work contracts that request a period of training, mainly what is called in France "qualification contracts".

The list of variables we used as well as more detailed information on the population analyzed in the survey are given in Appendix 1.

3 Estimating the Earnings Function

3.1 Summary Statistics:

Table 1 gives the mean values of the various variables. On average individuals who underwent training earn 29.6% more than those who did not undergo any form of training. Several characteristics can explain such a difference. First the proportion of those who have a Higher Education Diploma (11.9% versus 4.5%) as well as of those who have the "baccalauréat" (17.4%versus 11.1%) is higher among those who received a form of training. On the contrary individuals with a low level of education represent a higher share among those who did not receive any training.
Table 1 indicates also that the proportion of manual workers (whether "specialized" or "qualified") is much higher among those who did not receive any training. On the contrary the share of engineers or of those having a managerial position is much higher among those who received training. Age is higher on average among those who did not receive any training (41.3 versus 39.8). The proportion of those working full time is higher among those who received training (88.6% versus 82.3) while that of those who work part time (between 15 and 30 hours) is somehow higher among those who did not receive any training (2.3% versus 0.5%).

3.2 The Determinants of On-the-Job Training

In order to make a correction for the selectivity bias, we first estimated a Probit model which gives us the determinants of the access to the training[5] that took place during the 14 months preceding the date at which the survey took place[6].

[5] Whatever the type of training and the closest it was to January 1999.
[6] Actually, as pointed out by a referee, there is another selectivity bias related to the decision to participate in the labor force. We did not take this second bias into account because it would have required us to estimate a bivariate probit in a first stage. This in itself is not a problem but undertaking the type of decomposition that we present in section 4 with a bivariate probit would have enormously increased the complexity of the decomposition. Our goal was to illustrate the kind of breakdown that can be derived when there is a selectivity bias. We hope in future work to be able to derive a similar decomposition when two selectivity biases are taken into account.

In selecting the variables at this stage we took account of the findings of the literature that appeared on France and on Europe on the inequality of access to training. This is why we introduced variables such as gender, nationality, qualification, type of work contract, size of the firm and presence of other workers trained in the firm[7]. Previous work proved the impact on training of the size of the firm in France (see, Dupray and Hanchane, 2003a and 2003b, and Joseph and Lochet, 2001) and countries such as Greece, Italy, Portugal and Spain, at the difference of what happens in Northern Europe (see, Aventur and Möbus, 1999). Fournier et al. (2001) have also shown that the proportion of individuals receiving training depends on the socio-professional category. However since the level of qualification is highly correlated with the socio-professional category we did not introduce the socio-professional category in this probit regression. We introduced also the gender of the individual in order to take into account the likely unequal access of women to training, which may be due to the difficulty women face when asked to reschedule their various activities to receive training. Nationality was introduced to check whether there is discrimination against foreigners in access to training. In addition since previous work (see, Aventur and Hanchane, 1999) had shown that the firms that were the most likely to provide training did it also in the most equal way, we decided to include the variable "Other categories of trainees" in the probit regression.

As is well known, the main contribution of this first stage of the estimation procedure is to give the value of Mill's ratio for each individual. The results of this Probit model nevertheless shed also some interesting light on the factors influencing access to training (see, Table 2). They are in fact similar to those obtained in previous studies in France and show for example that the closer the links between the employer and the employee, the higher the probability of getting access to training. Therefore individuals working under a contract of fixed duration are less likely to receive training than those employed under a contract of indeterminate duration. Naturally those who have the highest probability of receiving training are the "trainees" ("stagiaires" in French). These employees receive an intensive training in order to be integrated in the firm as quickly as possible.

Note also (see, Table 2) that men are more likely than women to benefit from training. Actually the reality is more complex. For full time contacts men have indeed a higher probability of receiving training but the situation is the opposite for part-time contracts (see, Hanchane and Lambert, 2003). It appears that such a training period imposes various monetary and non-monetary costs on the family (such as finding substitutes to take care of the children) and these costs are not identical for men and women. What happens for part-time contracts seems in fact to indicate that women working part-time are employed in professions or sectors where the prevalence of training is usually important.

One may also observe that the managerial staff and the professions that are at the intermediate level of the hierarchy (henceforth "intermediate professions") are those who are the most likely to undertake a program of continued training. A more detailed analysis of the survey, making a distinction between the three kinds of training (standard training period[8], training received while working and self-formation), shows however differences. Thus one finds proportionally more manual workers ("ouvriers") among those receiving training while working. In fact this kind of training is the most common in industries where manual workers are over-represented. Nevertheless, despite what characterizes the category called earlier "training received while working", there is still a significant degree of inequality between the socio-professional groups[9] in the access to training.

[7] Following a referee´s suggestion we tried also to introduce age in the probit model but it did not have any effect on the results of the decomposition technique presented in section 4 below.

[8] «Stage» in French.

[9] «Catégorie socio-professionnelle» in French.

It is interesting to note that these differences are also present at the perception level. Thus, when asked whether in their firm the other employees undertake training, 78% of the managerial staff and the "intermediate professions" answered "yes" and 74% answered "yes", whatever the level of the worker. Manual workers are those who are the most likely to answer "no, never" (37% of them). However 9% of them say that the most qualified workers are those who receive training, the corresponding percentage being 7% in the whole surveyed population.
The impact of the size of the firm is as expected: firms with more than 500 workers are those that proportionally invest the most in training. These firms are those where the type of training labeled earlier "training received while working" is the most frequent, hence the fact that these firms are those in which the low-skilled workers have the highest probability of receiving training. Note that what was called previously "training received while working" is in certain ways an innovative type of training and it seems to be easier to manage in big firms. This kind of training requires for example the presence of an individual personally responsible for the training of the trainee and such a requirement is evidently less constraining in big firms.

3.3 The Returns to Training and the Coefficients of the Earnings Function

The Probit model whose results have just been analyzed allowed us to estimate Mills' ratios that have then been introduced in the earnings function in order to make a correction for the selectivity bias. In addition, following Barnow et al. (1980), we have introduced in the regression a dummy variable equal to 1 for those who received training and to zero otherwise. The coefficient of this variable in the regression will then give us the impact on earnings of receiving training, once the selectivity bias is neutralized and, naturally, other things constant. The other variables introduced in the regression are as follows. There are five dummy variables giving the educational level (six categories) of the individual, five dummy variables giving the qualification level of the job (six categories) and two dummy variables describing the type of contract (three categories). We also introduced[10] age and its square, gender and nationality, seniority in the firm and its square, three dummy variables giving the weekly duration of work (four categories) and finally two variables giving information on the type of work schedule (three categories).
The individual who serves as reference in the regression has a higher education diploma corresponding to two years of study after the "baccalauréat", is a technician, has a fixed duration work contract, works between 15 and 30 hours per week, has an "alternate"[11] work schedule, is male and French.
Table 3 indicates that earnings grow with the level of human capital. One may observe that those who hold a higher education diploma, *ceteris paribus,* earn 14% more than those who studied two years after the "baccalauréat" (the category of reference for the educational level), 20% more than those who have only the "baccalauréat", 29% more than those who have a CAP or BEP, 31.4% more than those who have a BEPC and 41% more than those who have a CEP.
As far as the socio-professional category is concerned, technicians, engineers and individuals who are part of the managerial staff earn significantly more than the other categories.

[10] Clearly many variables (e.g. educational level) which have an impact on access to training have also an effect on wages. Following a referee´s suggestions we modified an earlier specification of our model and introduced nationality and gender in both the probit and the earnings function. The main variable which we believe affects training but may be excluded from the wage equation is that which refers to «other categories of trainees».
[11] In the sense that he works a fixed number of hours per week but the timing of his work (e.g. morning versus afternoon) varies every week but on a regular basis.

Technicians are often individuals who studied two years beyond the "baccalauréat" in a technological section while engineers and managers are often graduate of the prestigious engineering schools or hold what is called a diploma of higher specialized studies (five years beyond the " baccalauréat").
Note also that job security seems to play a discriminating role in so far as those having a work contract of undetermined duration earn more, ceteris paribus, than those having a fixed duration work contract or than trainees.
Age and seniority have, as expected, a non linear effect but note the weak impact of seniority. Similar findings about the effect of seniority in France may be found in the works of Béret (1992), Goux and Maurin (1994) and Hanchane and Joutard (1998). These results are an illustration of the transformations that occurred on the French labor market as well as of its specificity when compared with other industrial countries. Before what is known in France as the "crisis", which started in the mid 1970s, there was a close link between the worker and his job. Qualification was thus acquired progressively while working. The "crisis", which led to a stronger emphasis on competitiveness, displayed the rigidity of internal markets so that external markets became the preferred choice of those individuals who had acquired a minimal level of investment in education. As a consequence, though seniority increased, its return decreased, becoming sometimes even nil. Various studies such as those of Maurice, Sellier and Silvestre (1982), Silvestre (1986), Verdier (1997) and Béret (1992) have actually emphasized these transformations of the French labor market .
We had no exact information on the number of hours worked by an individual during a given month. This is why we used the monthly and not the hourly wage as dependent variable. We were however able to partly control for the number of hours worked since, as indicated earlier, we knew whether the individual worked full time, between 30 and 40 hours, between 15 and 30 hours or less than 15 hours per week. In addition information was available on the schedule of work: whether every week the individual had the same schedule, whether he had an "alternate" schedule (working for example one week during the day, the other at night) or whether his schedule was completely flexible. The results (see, table 3) indicate, as expected, that the more hours an individual works, the higher his earnings. It appears also that those who work according to a fixed schedule earn less than those who work under an "alternate" schedule. Such a compensating scheme however does not seem to exist for those who have a flexible schedule.
Access to training has a significant impact but here, as mentioned previously, caution is required in the interpretation of the results. The impact of training on earnings has evidently to be computed "net of the selectivity effect". The presence of a selectivity bias (the coefficient of Mills´ ratio is significant) implies that unobserved heterogeneity affects both the probability to receive on-the-job training and the earnings themselves[12]. The "net" effect of on-the-job training on earnings is therefore to increase the latter by 24%, the difference between the values of the coefficients of the variable "on-the-job training" (0.50) and of Mills´ ratio (0.26).
In order to better understand the various channels through which on-the-job training may have an impact on the earnings of those who received training we present in Table 4 the results of a regression, where we include only those individuals who received on-the-job training and the dependent variable is the residual of the earnings function (whose estimates are given in Table 3).

[12] As mentioned by a referee the concept of unobserved heterogeneity is evidently cannot be assimilated only to the decision to undertake training. This was however the only type of heterogeneity we could control for.

3.4 The role played by the side that took the initiative of the training or/and financed it

The explanatory variables in the regression whose results are given in Table 4 give information on the side that took the initiative of the training (individual, firm, both sides) and the way this training was financed. This type of information allows us to identify the heterogeneity of the impact of training, depending on whose side took the initiative of the training and how the latter was financed. These two dimensions are in fact well connected to human capital theory.[13] In the regression given in Table 4 the individual of reference is one whose training was initiated by employment services or delegates of the workers and financed by the employer.

It appears that the forms of training that provide the highest rate of return are those that were initiated by the individual alone or by a common decision of the employer and the employee. These findings are consistent with the fact that general training gives higher returns. Such an individual initiative represents also a signal for the employer that the employee will do his/her best to make use of the training received.

The fact that when the training was undertaken at the initiative of the individual, it has a greater impact on earnings raises naturally the issue of a possible ex post rationalization of the decision-making. Some econometric tests that were undertaken to check such an hypothesis tend however to invalidate such a possibility (see Appendix 2).

The fact that the impact on earnings was higher when the employer financed the training seems to lead to a rejection of an explanation in terms of specific training. First it turns out that even when financed by the employer, some forms of training may be general, in the sense that they can be used elsewhere on the market place.[14] Second it is possible that those who receive training are a more homogeneous group and are subject to greater selection when the employer finances the training. The idea is that the "productive impact" of the training depends first on the capacity shown by the employees to make use of this training, second on the probability that they will work in positions where such training is profitably used. If this is so, one may expect a better selection of the trainees when the employer finances the training. Such an observation does not contradict the fact that training is more efficient when it was initiated by the employee or jointly by the employer and the employee. After all, nothing guarantees that all the individual or joint demands for training will be implemented and we have no data to estimate the probability that this happens.

To better understand the complex links that may exist between the side taking the initiative of the training and that financing it, we made some additional tests. We checked in particular whether, among those who received a training that was financed by the employer, the cases where the initiative of such training was taken by the employer alone, had specific characteristics. Table 5 indicates that as far as the field of training is concerned there is no real difference between the two cases where the employer is involved. It appears however that in the fields of "Hygiene and Security" and of "Industrial Techniques" training is less common when it was initiated by the individual. This is also true for "Trade, Sales and Marketing". It is possible that, at least for the two first cases, the productive context of the firm plays a role and thus the specificity of the firm would play a greater role. As far as the objective of the training is concerned, it seems that when the goal of the training is to obtain a diploma,

[13] This distinction between general and specific training goes back to Becker (1993) and will not be recalled here. As is well known these are extreme cases. Very often the two sides share the financing of the training and hence its benefits and this sharing is a kind of insurance against any opportunistic behavior of the other side (a « quit » by the worker or a « layoff » by the employer). It is also expected that general training should give a higher return since in the case of specific training the employer may have the power to limit the sharing of the rent.

[14] Various theoretical models have been proposed to understand the rationality of such a decision taken by the employer (see, Dupray and Hanchane, 1999, for a review of this literature).

individual initiative is more common, this being also true for the cases where a certification is to be received at the end of the training period. Note finally that the duration of the training period is almost triple when the initiative was taken by the employee.
In short it seems that the individual is more involved when the duration of the training is longer and when it is validated by a diploma or a certification. This would seem to confirm that individuals look more for a form of training that is general rather than specific. This evidently implies that employees are not indifferent to the impact that their training may have on their mobility.

4 Estimating the Contribution of the Explanatory Variables to the Variance of Earnings: The Methodology

4.1. Estimating the Contribution of the Explanatory Variables to the Overall Variance:

To estimate these contributions we use a recent contribution of Fields (2003).
Let us first write the earnings function as

$$y_k = \Sigma_{k=1 \text{ to } (K+3)} \; b_k z_{kj} \qquad (1)$$

where
y_j is the logarithm of the wage of individual j,
$z_{kj} = x_{kj}$ $\forall k=1$ to K, where x_{kj} refers to the value taken by the explanatory variable k for individual j. Note that these K variables do not include that referring to the participation (F_j) in the training program and the impact of the selectivity bias (λ_j). We therefore have also
$z_{K+1,j} = F_j$, $z_{K+2,j} = \lambda_j$ and finally $z_{K+3,j} = u_j$ where u_j is the value taken by the disturbance for individual j. Note also that we will assume below that
$b_{K+1} = c$, $b_{K+2} = d$, $b_{K+3} = 1$

Fields (2003) has proven that

$$\sigma(y_j) = \Sigma_{k=1 \text{ to } (K+3)} [(b_k) \text{Cor} (z_{kj}, y_j) (\sigma(z_{kj}))] \qquad (2)$$

The relative contribution $s_k(y_j)$ of factor k to the dispersion $\sigma(y_j)$ may therefore be expressed as

$$s_k(y_j) = [(b_k) \text{Cor} (z_{kj}, y_j) (\sigma(z_{kj}))] / \sigma(y_j) \qquad (3)$$

Expression (3) may also be written, after simplifying, as

$$s_k(y_j) = [(b_k) \text{Cov} (z_{kj}, y_j)] / V(y_j) \qquad (4)$$

where $V(y_j)$ denotes the variance of the logarithms of wages y_j.
As a consequence the relative contribution of factor x_h (h=1 to k) to this variance is equal to

$$s_h(y_j) = [(b_h) \text{Cov} (z_{hj}, y_j)] / V(y_j) \qquad (5)$$

Similarly the relative contribution of the participation to an «on-the-job» training program may be expressed as

$$s_F(y_j) = [(c) \text{Cov} (F_j, y_j)] / V(y_j) \qquad (6)$$

while that of the Mills ratio will be written as

$$s_\lambda(y_j) = [(d) \text{Cov} (\lambda_j, y_j)] / V(y_j) \qquad (7)$$

Finally the relative contribution of unobserved variables (the disturbance u_j) is equal to

$$s_u(y_j) = \text{Cov} (u_j, y_j) / V(y_j) \qquad (8)$$

While expressions (5) to (8) give the contribution of the various explanatory factors and of the disturbance to the overall variance of the (logarithms of) wages, it is also possible to compute the contribution of these elements to the between and within groups variance.

4.2. Contribution of the Explanatory Variables to the Within-Groups Variance

Introducing the Mills ratio we may rewrite (1) separately for an individual belonging to group A and one belonging to group B. In the first case we write

$$y_{jA} = \Sigma_{k=1 \text{ to } K} b_k x_{kjA} + \rho\sigma_u [(-\phi_j)/(1-\Phi_j)] + w_{jA} \quad (9)$$

where $\Phi(.)$ refers to the distribution function corresponding to the decision to participate in training and ρ is the correlation between the disturbances of the probit model and the earnings function, ϕ_j is the density function corresponding to Φ_j and σ_u is the standard deviation of the error term of the earnings function. Finally note that $[(-\phi_j)/(1-\Phi_j)]$ is the expression for Mill's ratio in group A and $\rho\sigma_u$ that for the coefficient d (see, Green, 2000).
For group B expression (1) will be written as

$$y_{jB} = \Sigma_{k=1 \text{ to } K} b_k x_{kjB} + c + \rho\sigma_u [\phi_j / \Phi_j] + w_{jB} \quad (10)$$

since $[\phi_j / \Phi_j]$ is the expression for Mill's ratio in group B.
As is well-known, the within groups variance is equal to the weighted sum of the variance within each of the two groups A and B, the weights being the population shares (f and (1-f)) of the two groups, so that the contribution $s_{k,WITH}(y_j)$ of each of the (K+2) factors[15] to the within groups variance may then be written as

$$s_{k,WITH}(y_j)=\{(1-f)[(b_k)Cov(z_{kj,j\in B}, y_{j,\, j\in B})/V_B(y_j)]+(f)[(b_k)Cov(z_{kj,\, j\in A}, y_{j,\, j\in A})/ V_A(y_j)] \quad (11)$$

4.3 Contribution of the Explanatory Variables to the Between-Groups Variance

To compute the between-groups variance $V_{BET}(y_j)$ of the (logarithms of) earnings one has evidently to neutralize the within groups dispersion and thus to assume that every worker who received on-the-job training receives the mean (logarithm of) earnings $y_{M,B}$ of those who received such training while those who did not receive any on-the-job are assumed to receive the mean earnings $y_{M,A}$ of those who did not receive any training.
The contribution $s_{k,B}(y_j)$ of each of the (K+3) factors to the between groups variance, using again Fields' (2003) approach, will then be expressed as

$$s_{k,BET}(y_j) = [(b_k) Cov (z_{kM}, y_M)] / V_{BET}(y_j) \quad (12)$$

It is easy to show that

$$Cov (z_{kM}, y_M) = f (1-f) (x_{kMB} - x_{kMA}) (y_{MB} - y_{MA}) \quad (13)$$

and that

$$V_{BET} = f (1- f)(y_{mA} - y_{mB})^2 \quad (14)$$

We may now combine (12), (13) and (14) to derive that

$$s_{k,BET}(y_j) = [(b_k) (x_{kMB} - x_{kMA})] / (y_{MB} - y_{MA}) \quad (15)$$

For the contribution of the variable F_j to the between groups dispersion, one will obtain similarly, remembering that in this case $x_{kMB} = 1$ and $x_{kMA} = 0$,

$$s_{F,BET}(y_j) = [c / (y_{MB} - y_{MA})] \quad (16)$$

The contribution of the ratio of Mill λ_j to the between groups dispersion will be expressed as

$$s_{\lambda,BET}(y_j) = (\rho\sigma_\varepsilon)[(\phi_j / \Phi_j)_m - ((-\phi_j)/(1-\Phi_j))_m] / (y_{MB} - y_{MA}) \quad (17)$$

Finally the contribution of the disturbances to the between groups dispersion will be written as

$$s_{u,BET}(y_j) = [(w_{MB} - w_{MA})] / (y_{MB} - y_{MA}) \quad (18)$$

where w_{MB} and w_{MA} are respectively the mean values of the disturbances in groups B and A.
It is then easy to show that the sum of all the contributions to the between groups dispersion is equal to 1.

We may therefore conclude, using all the previous results, that the contribution of a given factor k (k = 1 to K) to the total variance V_{TOT} of the logarithms of wages is the sum of three elements:

[15] Here evidently there is no contribution of factor F to the within groups variance.

- its impact via its contribution to the within group A variance V_A; this effect will be expressed as

$$\{[(b_k)Cov(z_{kj,\,j\in A},y_{j,\,j\in A})/\,V_A(y_j)]\}\{(f)V_A(y_j)/V_{WITH}(y_j)\}\,\{V_{WITH}(y_j)/V_{TOT}(y_j)\} = \{(f)[(b_k)Cov(z_{kj,\,j\in A},\,y_{j,\,j\in A})]/V_{TOT}(y_j)\} \quad (19)$$

- its impact via its contribution to the within group B variance V_B; this effect will be expressed as

$$\{[(b_k)Cov(z_{kj,\,j\in B},y_{j,\,j\in B})/V_B(y_j)]\}\{(1\text{-}f)V_B(y_j)/V_{WITH}(y_j)\}\,\{V_{WITH}(y_j)/V_{TOT}(y_j)\} = \{(1\text{-}f)[(b_k)Cov(z_{kj,\,j\in B},\,y_{j,\,j\in B})]/V_{TOT}(y_j)\} \quad (20)$$

- its impact via the between groups variance V_{BET} ; this effect will be expressed as

$$\{[(b_k)\,Cov(z_{kM}\,,\,y_M)]\,/V_{BET}(y_j)\}V_{BET}(y_j)/V_{TOT}(y_j)\}=\{[(b_k\,)\,Cov\,(z_{kM}\,,\,y_M)]\,/\,V_{TOT}(y_j\,)\} \quad (21)$$

Thus we end up with a total impact of the variable k expressed as

$$\{[(f)(b_k)Cov(z_{kj,j\in A},y_{j,j\in A})]+[(1\text{-}f)(b_k)Cov(z_{kj,j\in B},y_{j,j\in B})]+[(b_k)\,Cov\,(z_{kM}\,,\,y_M)]\}/V_{TOT}(y_j) \quad (22)$$

Similar results may be derived for the contribution of the variable on-the-job training F_j , of Mills' ratio and of the disturbances (see, Appendix 3)

5. Decomposing the Variance of Earnings: The Results

The empirical results of this investigation are given in tables 6 to 10.

Table 6 gives the decomposition of the total variance of the logarithm of wages into two components, the between groups (the groups being those who received training and those who did not) and the within groups variance. It appears that most of the dispersion (94.5% of the variance) takes place within groups while the between groups variance represents only 5.5% of the total variance.

Tables 7 and 8 give the contribution of the different explanatory variables to the between and within groups variances respectively. These two tables should be interpreted differently.

5.1 Contributions of the various variables to the within groups variance:

To analyze the results we have to remember the formulations given in section 4 where the contribution of a given variable k (in percentage) to the variance of the regression is shown to be a function of the coefficient b_k of this variable in the regression (earnings function), its correlation Cor (z_{kj} , y_j) with the dependent variable (the logarithm of earnings) and its dispersion, relative to the dispersion of the dependent variable $\sigma(z_{kj}) / \sigma(y_j)$. Such a general interpretation evidently holds also in the case of the within groups variance.

The results in table 8 are given separately for those who received (group B) and those who did not receive (group A) any training. Table 8 indicates that the three variables that contribute most to the dispersion of (the logarithms of) earnings in group A are respectively the dichotomous variables "works full-time" (explains 27.7% of the variance of the logarithm of earnings) and "works part time, 0 to 15 hours" (explains 9.48% of the variance of the logarithms of earnings) and the residuals (36.72% of the variance of the logarithms of earnings). Since the dependent variable represents the monthly earnings, because we had no way of estimating the hourly wages, the role played by the number of hours of work is not surprising. Once this factor is taken into account, it thus appears that approximately half of the remaining variance is explained by the unobserved heterogeneity of the individuals.

The two other variables that have a somehow significant contribution to the variance of the (logarithms of) earnings in group A are the dichotomous variables "holds a CEP" and "employee". In group B (those individuals that received training) the two most important contributions to the variance of earnings are again those of the residuals and of the variable "worked full time". The explanation is similar to that given earlier for group A and will not be repeated here. The contribution (in percentage) of these two variables is however smaller because for the group of individuals who received training other variables play a role. First note the important contribution of the dummy variable "Engineer or Managerial Position" (14.5%). Second the other variable whose contribution should be mentioned is Mill's ratio but we will discuss it below in a separate section. Let us now turn to an analysis of the contribution of the different variables to the between groups variance.

5.2 Contributions of the various variables to the between groups variance

In section 4 we have shown that the contribution of a given variable to the between groups variance is positively related to the coefficient of this variable in the regression and to the difference between the two groups in the means of this variable and negatively related to the difference between the two groups in the means of the dependent variable (logarithms of earnings). Four of the five variables that have an important impact on this between groups variance were already mentioned when we analyzed the determinants of the within groups variance. These variables are by decreasing order of importance (in their contribution to the between groups variance) as follows: "Holds a CEP" (17%), "Works Full Time" (15%), "Is an Engineer or Has a Managerial Position" (13%), "Is a Specialized Worker" (6.1%) and finally "Works less than 15 hours per week" (5.1%). The impact of Mills ratio and of the variable "received vocational training" will be discussed below in a separate section.

Equation (15) may also be used to derive some additional intuitive interpretation of the results. Let us rewrite (15) as $(y_{MB} - y_{MA}) / (x_{kMB} - x_{kMA}) = (b_k) / s_{k,BET}(y_j)$

Since the variables y_{MB} and y_{MA} are logarithms, their difference may be interpreted as the percentage difference in the average earnings of the two groups. Moreover for all the dummy variables, the expression $(x_{kMB} - x_{kMA})$ refers in fact to the difference between the percentage of individuals in group A who have characteristic k (e.g. are "Employees") and the corresponding percentage in group B. The ratio $[(b_k) / s_{k,BET}(y_j)]$ is therefore in this case a kind of elasticity and tells us by how much the percentage difference between the average earnings in the two groups will increase (in absolute, not relative terms) when the gap between the percentages of individuals who have characteristic k in the two groups increases by 1% (here also in absolute, not relative terms). Let us see what this implies for the five variables mentioned previously by looking at the data of Table 1, 3 and 7. Table 1 indicates for example that 34.5% of those who did not receive training had a "CEP" while the corresponding percentage among those who received training is 15.8%. The difference between these two percentages is hence equal to 18.7%. Remembering that the difference between the average values of the logarithms of earnings in the two groups is equal to –0.296, we derive that the ratio $[(b_k) / s_{k,BET}(y_j)]$ for this variable is equal to (-0.296/0.187) = -1.58. In other words assume that this gap between the two groups in the percentage of those having a "CEP" decreases by 1%, from 18.7% to 17.7%. This then implies that the average gap in earnings between the two groups will decrease by 1.6%, from 29.6% to 28%.

Using the same kind of computations, one can find out that, ceteris paribus, the average gap in earnings between the two groups will increase by 2.7% if the percentage difference between

the two groups in the number of individuals who are engineers or have a managerial position increases by 1%. Similarly it will increase by 4.7% if the percentage difference between the two groups in the number of individuals who work full time increases by 1% but will decrease by 2.7% if the percentage difference between the two groups in the number of individuals who are specialized workers increases by 1%. On the contrary it will decrease by 16.2% if the percentage difference between the two groups in the number of individuals who work less than 15 hours increases by 1%. Note that the strong reaction (16.2%) that was just mentioned is due to the fact that, as indicated in Table 1, the actual gap between the two groups in the percentage of those working less than 15 hours is rather small since it is equal to 0.5 – 2.3 = -1.8%.

5.3 The contributions of Mills´ ratio and of the variable "Received Vocational Training" to the variance of earnings

Let us fist look at the contribution of Mills' ratio to the within groups variance of earnings. It appears (see, Table 8) that 3% of the variance of earnings among those who did not receive any training (group A) is due to differences among the individuals in the value taken by the Mills ratio, that is to that part of the unobserved heterogeneity that has an impact on the a priori probability to receive such a training. Among those who did in fact receive such a vocational training, the contribution of Mills ratio to the variance of earnings is even equal to 8%.

As far as the between groups variance of earnings is concerned, we have to take into account the contributions of the dummy variable "Received Vocational Training" a well as that of the Ratio of Mills. The combined contribution of these two variables to the between groups variance of earnings may be considered as the net contribution of training to the between groups variance of earnings. Such a contribution takes into account not only the impact of training on earnings but also that of the unobserved heterogeneity that has an effect on the probability of receiving training. It thus appears that 35% (170 – 135) of the between groups variance of earnings is due to this combined effect. Note that the sign of the contribution of the Mills ratio is negative. This implies that the pure effect of vocational training on the between groups dispersion of earnings would have been stronger, had there be no impact of the unobserved variables on the probability of receiving training.

To better understand the significance of contributions that are greater than 100% we have to go back to the interpretation in terms of elasticities, that has been given previously in the case of between groups dispersion. The value of 170% which is the contribution of the variable "Received Vocational Training" to the between groups variance of earnings should therefore be understood as follows. Assume we divide the population of workers in two groups and observe the percentage difference in the average earnings in these two groups. Decide then to give to one of these two groups only (group B) vocational training. If the difference in size between the group who received training and that who did not increases in absolute value by one percent (e.g. if group B, that receives training is equal to 29.26% of the workers rather than 28.26% so that group A will represent 70.74% rather than 71.74% of the workers) and if the unobserved heterogeneity is assumed not to have any impact on the probability to receive training, then the gap between the average earnings in both groups would, ceteris paribus, be equal to 170% of its present value.

The story in reality is however very different. The allocation of workers to training is not random. On the contrary the unobserved heterogeneity has an impact on the probability to

receive training as well as on the earnings themselves. We have therefore to neutralize this effect of the unobserved heterogeneity before we can say anything on the net impact of training of wages. This is why we have to conclude (see, Table 7) that the net (of the effect of the unobserved heterogeneity on the probability of receiving training) contribution of earnings to the between groups variance is only equal to 35%. Using again our interpretation in terms of elasticities, the interpretation of the results should now be as follows. Assume we divide the population of workers in two groups, observe the percentage difference in the average earnings in these two groups and decide to give to one of these two groups only (group B) vocational training. If then the difference in size between the group who received training and that who did not, increases in absolute value by one percent, we will find out, once we take into account the impact of the unobserved heterogeneity on the probability to receive training (that is, on the way the individuals were allocated to the two groups) that the gap between the average earnings in both groups will, ceteris paribus, be equal to 35% of its present value (remember that the other variables explain 65% of the present gap).

6 Conclusions

This paper is a first attempt to devise a methodology that allows estimating the exact impact of training on the dispersion of wages. It uses an approach originally proposed by Fields (2003) but extends it to the breakdown of inequality by population subgroups as well as to the case where the earnings function that is at the base of the analysis has to be adjusted for selectivity bias. The empirical illustration is based on a survey conducted in France at the end of the twentieth century.

The results of the analysis show first that when a distinction is made between workers who received training and those who did not, the between groups dispersion explains only 5.5% of the overall variance of earnings. We also found that more than one third of this between groups variance was explained by the combined effect of the unobserved heterogeneity and the distinction between those who received and did not receive on-the-job training. We also noted that the unobserved heterogeneity led to a drastic reduction of the impact of training on earnings, since those who received training were also those who had a priori the highest probability of receiving training.

Most of the earnings dispersion however turned out to be a within groups dispersion and more than two thirds of this within groups variance of the logarithms of earnings could be explained by the variables that were taken into account. Ignoring the contribution of the number of hours of work[16] that really played the role of a control variable since we did not have any information on the hourly wages, we may observe that, for those who did not receive any training, the main (other) factors of this dispersion are the level of education (contribution of 6.5%) and the type of occupation (contribution of 15.5%). For those who received training the following variables (in addition to the number of hours of work) played an important role: the type of occupation (contribution of 24%), the level of education (8.5%) and the unobserved heterogeneity (the Mills ratio has a contribution of 8.2%).

It should therefore be clear, given that there is a small between groups and a big within groups dispersion, that there is a lot of overlapping between the distributions of earnings of the two groups, those who received and those who did not receive training. Such findings imply indeed that the unobserved heterogeneity plays a key role in the selection of those who receive training and thus indirectly has an impact on the difference between the *average* earnings of those who receive and do not receive training. It cannot however be considered as

[16] The different dummy variables included in the regression to measure the impact of the hours of work on earnings explain approximately 35% of the variance of earnings for those who did not receive any on-the-job training and 20% for those who received such a training.

a variable that could lie behind market segmentation. This is so because the within groups variance is much higher than that of the between groups so that the distributions of earnings of these two groups show a great degree of overlapping. In other words there is a much greater degree of heterogeneity within than between the two groups corresponding to those who received and did not receive on-the-job training. As a consequence if labor market segmentation exists, it must be based on other criteria.

Bibliography

Ariga, K. and G. Brunello (2003) "Education, Training and Productivity: Evidence from Thailand and the Philippines," Empirical Analysis of Economic Institutions Discussion Paper Series No. 13, Kyoto Univeristy.

Aventur, F, Hanchane S, (1999) "Justice sociale et formation continue dans les entreprises françaises", *Formation Emploi*, 66: 5-20.

Aventur, F., Möbus, M. (1999), *Formation professionnelle initiale et continue en Europe*, Magnard-Vuibert Multimédia, France.

Barnow, B.S., Cain, G.G., Golberger, S. (1980) "Issues in the analysis of selectivity bias" in E.W. Stromsdorfer and G. Farkas (eds.) *Evaluation Study Review Annual*, vol. 5, Beverly Hills, CA., Sage Publications, pp. 43-59.

Becker, G. (1993), *Human Capital : A Theorical and Empirical Analysis*, Third Edition, Columbia University Press.

Béret, P (1992), "Salaires et marchés internes", *Économie Appliquée*, XLV (2).

Béret, P., Dupray, A. (1998) "La formation professionnelle continue: de l'accumulation de compétences à la validation de la performance", *Formation Emploi*, 63: 61-80.

Béret, P., Dupray, A. (2000) "Allocation et effet salarial de la formation professionnelle continue en France et en Allemagne : une approche en terme d'information", *Economie publique*, 5: 221-269.

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

Bound, J. and G. Johnson, 1992, "Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations", *American Economic Review*, 82: 371-92.

Crocquey, E. (1995) "La formation professionnelle continue: des inégalités d'accès et des effets sur la carrière peu importants à court terme", *Travail et Emploi*, 65: 61-68.

Dupray, A., Hanchane, S. (1999) "Identification des paramètres d'intérêt en présence de biais de sélection et de censure: exemple de la modélisation de la probabilité du passage par la formation continue", Paper presented at the ninth meeting of SESAME, Saint-Etienne, September 8,9 and 10.

Dupray, A., Hanchane, S. (2003b) "La formation continue en enterprise et ses effets sur le salaire et la mobilité des jeunes," in J. F. Lochet, ed., *Entreprises et jeunes débutants*, L´Harmattan.

Dupray, A., Hanchane, S. (2003b) "Modalités de participation à la formation continue et effets sur la carrière salariale," in Regards pluriels sur la formation continue, vol.1, De Beock.

Fi.elds, G.S. (2003) "Accounting for Income Inequality and its Change: A New Method, with Application to the Distribution of Earnings in the United States," *Research in Labor Economics*

Fournier, C., Hanchane, S., Lambert M., Perez, C., Thery, M. and Thomas, G. (2001), "Un panorama des formations des personnes en France," *CEREQ bref*, No. 172.

Goux D. et Maurin E. (1994), "Education, expérience et salaire: tendances récentes et évolution de long terme", *Economie et Prévision*, 116.

Goux, D., Maurin, E. (1997) "Les entreprises, les salariés et la formation continue", *Economie et Statistique*, 306: 41-55.

Greene W. H. (2000), *Econometric Analysis*, Fourth edition, Prentice Hall.

Hanchane S, Joutard X. (1998) "Une approche empirique de la structure du marché du travail : Salaires, formes de mobilité et formation professionnelle continue" , *Economie et Prévision*, 135.

Hanchane S. et Lambert M. (2003), «La variété des modes de formation: usages et enjeux», *Formation Emploi*, 81 : 67-80.

Heckman, J. (1979) "Sample specification bias as a specification error", *Econometrica*, 47 (1): 153-161.

Heckman, J. and L.Lochner (1998) "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents," *Review of Economic Dynamics* 1: 1-58.

Joseph, O., Lochet J.-F. (2001) "La formation en cours d´emploi dans le processus d´insertion," in J.-F. Lochet, ed., *Entrepise et jeunes débutants*, L´Harmattan.

Karoly, L. A. (1992) "Changes in the Distribution of Individual Earnings in the United States: 1967-1986," *Review of Economics and Statistics* 74 (1): 107-115.

Katz, Lawrence F., and Murphy, Kevin M. (1992) "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *Quarterly Journal of Economics,* 107.

Krusell, P., L. E. Ohanian, J.-V. Rios-Rull and G. L. Violante (2000) "Capital-Skill Complementarirty and Inequality: A Macroeconomic Analysis," *Econometrica*, 68: 1029-1054.

Levy, F. and R. J. Murnane, "US Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature*, 1992 (30): 1333-81.

Maurice M., Sellier F et Silvestre J-J (1982), *Politique de l'éducation et organisation industrielle en France et en Allemagne. Essai d'Analyse sociétale*, PUF.

Mood, A.M., Graybill, F.A., Boes, D.C., 1974. *Introduction to the theory of statistics*. McGraw-Hill, Auckland.

OECD (1999) *Perspectives de l'emploi* – Chapter 3 : Formation des travailleurs adultes dans les pays de l'OCDE: mesure et analyse.

Silvestre J-J. (1986), "Marchés du travail et crise économique : de la mobilité à la flexibilité", *Formation Emploi*, 14.

Verdier E. (1997), "Insertion des jeunes à la française : vers un ajustement structurel ?", *Travail et Emploi,* 69.

Wooldrige J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Appendix 1: List of variables and information on the population analyzed

List of variables

Educational Level

Six categories of diploma have been considered: Higher education diploma, two years of study beyond the "baccalauréat", "baccalauréat", CAP or BEP[17], BEPC[18] and CEP[19].

Socio-Professional Category

Six categories have been distinguished. Engineers or individuals having a managerial position, "intermediate professions", "specialized" manual workers[20] , "qualified" manual workers[21], employees and other professions.

Seniority

For all the individuals seniority in the firm is that observed in March 2000.

Duration of Work

The exact number of hours per week is not known. There is however enough information to make a distinction between those working full-time, between 30 and 40 hours, between 15 and 30 hours and less than 15 hours per week.

Type of Work Schedule

Here a distinction is made between those who have every week the same schedule, those who have an "alternate" schedule (e.g. working one week the day, the other at night) and those who have a flexible schedule.

Training

As mentioned before the training that is described with most details is that which took place the closest to January 1999. This is why the training variable refers to this training period.

The population analyzed in the present study

We have concentrated our attention on the private sector because we believe that in the public sector the selection process is different and the profitability of continued training takes different forms. In addition we have excluded what is called "alternative contracts" and "subsidized contracts" which by their nature imply a minimal number of hours of training, partly financed by the State. It is then clear that for this type of contracts questions relative to the side who took the initiative of the training are totally irrelevant.
Note also that 90% of the individuals in the sample analyzed undertook the training for professional rather than for other reasons. We have therefore excluded individuals who undertook the training for personal or social reasons, whose goal was related to family, arts or sport or whose objectives were linked to responsibilities they have in the political world, in private associations or unions. In some cases the initiative of the training came from the employer. We then took into account, when this was relevant, the fact that the training was more or less imposed on the individual or that the individual was the only one in the firm, because of his/her specific qualifications, to receive such a training.

[17] These are professional diploma whose preparation requires at least two years and towards which one starts studying after 7 to 9 years of school.

[18] In France after primary education (5 years), all pupils follow a common curriculum at the « Collège » at the end of which they take the BEPC exam.

[19] This optional exam certifies that the individual completed primary school.
[20] « ouvriers spécialisés » in French.
[21] « ouvriers qualifiés » in French.

We had also to eliminate the observations for which no data on earnings were available so that we ended up with a sample of 8261 individuals, among whom 2335 received some type of training during the last period covered by the survey.

Appendix 2: Econometric analysis of the training decision

In a first stage a dichotomic variable was introduced which was equal to one if the earnings gap between 1999 and 2000 was greater than 900 francs (30% of the cases). Then a bivariate probit model was estimated on the basis of such a variable and that measuring the side that took the initiative of the training. Such a model allows one to check whether there is a correlation between the error terms of the two equations. In other words we test here the possibility that unobserved characteristics such as motivation have an impact on both the increase in earnings between January 1999 and March 2000 and the fact that the individual said he (she) took the intitiative of the training. Such a correlation term turns out however never to be significant, whether one works with the whole sample of those who received training or with a sample which includes only those whose training was financed by the employer.

Another technique was used to test this type of correlation. Here a conditional bivariate probit model was introduced so that the dichotomous variable indicating whether the increase in earnings was higher than 900 francs became an explanatory variable of that representing the side that took the training intitiative (see, Greene, 1998, and Woolridge, 2002). Here again no significant effect was observed.

In other words the fact that an employee belongs to the group of individuals (31%) who experienced the highest increase in earnings between 1999 and 2000 does not have any influence on the fact that he declares that he took the initiative of the training.

Appendix 3: Additional details on the methodology

I) The contribution of the explanatory variables to the variance of earnings according to Fields (2003)

Recalling that the earnings function is expressed as

$$y_k = \sum_{k=1 \text{ to } (K+3)} b_k z_{kj} \quad \text{(A-1)}$$

we derive

$$\text{Var}(y_j) = \text{Cov}[\sum_{k=1 \text{ to } (K+3)} b_k z_{kj}, y_j] \quad \text{(A-2)}$$

Dividing both sides of (A-2) by Var (y_j), we then derive that

$$1 = \text{Cov}[\sum_{k=1 \text{ to } (K+3)} b_k z_{kj}, y_j]/\text{Var}(y_j) \quad \text{(A-3)}$$

It is however well-known (see, Mood, Graybill and Boes, 1974) that

$$\text{Cov}[\sum_{k=1 \text{ to } (K+3)} b_k z_{kj}, y_j] = \sum_{k=1 \text{ to } (K+3)} \text{Cov}[b_k z_{kj}, y_j] \quad \text{(A-4)}$$

Expression (A-3) may therefore be expressed as

$$1 = \sum_{k=1 \text{ to } (K+3)} \text{Cov}[b_k z_{kj}, y_j]/\text{Var}(y_j) \quad \text{(A-5)}$$

since $\text{Cov}[\sum_{k=1 \text{ to } (K+3)} b_k z_{kj}, y_j] = \sum_{k=1 \text{ to } (K+3)} \text{Cov}[b_k z_{kj}, y_j] = \text{Var}(y_j)$

If we also remember that the correlation coefficient between $b_k z_{kj}$ and y_j may be expressed as

$$\text{Cor}[b_k z_{kj}, y_j] = \text{Cov}[b_k z_{kj}, y_j]/((\sigma(b_k z_{kj}), (\sigma(y_j)) \quad \text{(A-6)}$$

we end up, combining expressions (A-3) to (A-6), with

$$1 = \sum_{k=1 \text{ to } (K+3)} [\text{Cor}(b_k z_{kj}, y_j)(\sigma(b_k z_{kj}))/(\sigma(y_j))] \quad \text{(A-7)}$$

However since

$$\text{Cor}(b_k z_{kj}, y_j) = \text{Cor}(z_{kj}, y_j) \quad \text{(A-8)}$$

expression (A-7) implies that

$$\sigma(y_j) = \sum_{k=1 \text{ to } (K+3)} [(b_k)\text{Cor}(z_{kj}, y_j)(\sigma(z_{kj}))] \quad \text{(A-9)}$$

The determinants of the disturbance among those who received training:

One may assume that the disturbance u_j $(j \in B)$ of those who received on-the-job training is itself a function of G variables such as the mode of financing, the training, the side who took the initiative of the training, etc…

In other words one would estimate a new regression that would be written as

$$w_j = \sum_{g=1 \text{ to } G} e_g r_{g,j} + \nu_j \text{ for } j \in B \quad \text{(A-10)}$$

so that the variance $V(w_{j,j\in B})$ of the disturbance $w_{j,j\in B}$ may be written as

$$V(w_{j,j\in B}) = \sum_{g=1 \text{ to } G+1} [(e_g)\text{Cov}(r_{g,j\, j\in B}, w_{j\, j\in B})] \quad \text{(A-11)}$$

where the $(G+1)^{th}$ factor refers to the new disturbance ν_j (e_{g+1} is evidently equal to 1).

We may therefore write that the contribution $s_g(y_{j,\, j\in T})$ of these additional (G+1) factors to the within group B variance $V_B(y_j)$ will be expressed as

$$s_g(y_{j,\, j\in T}) = [(e_g)\text{Cov}(r_{g,j,\, j\in B}, y_{j,\, j\in B})/V(w_{j,\, j\in B})] \times [(V(w_{j,\, j\in B})/V_B(y_j)]$$

$$s_g(y_{j,\, j\in T}) = [(e_g)\text{Cov}(r_{g,j,\, j\in B}, y_{j,\, j\in B})/V_B(y_j)] \quad \text{(A-12)}$$

The group means

Let y_{mA}, y_{mB} and y_m be the means of the variable y_j in group A, group B and the whole population respectively. The mean values y_{mA} and y_{mB} of the logarithms of wages in groups A and B may be expressed as

$$y_{mA} = \sum_{k=1 \text{ to } K} b_k x_{kmA} + \rho\sigma_u [((-\phi_j)/(1-\Phi_j))_m] + w_{mA} \quad \text{(A-13)}$$

and

$$y_{mB} = \sum_{k=1 \text{ to } K} b_k x_{kmB} + c + \rho\sigma_u [(\phi_j/\Phi_j)_m] + w_{mB} \quad \text{(A-14)}$$

where x_{kmA}, x_{kmB}, w_{mA}, w_{mB}, $((-\phi_j)/(1-\Phi_j))_m$ and $(\phi_j/\Phi_j)_m$ are respectively the mean values of characteristic k, the disturbance u_j and Mill's ratios in groups A and B.

The overall mean y_m will then be expressed as

$$y_m = \sum_{k=1 \text{ to } K} b_k (f x_{kmA} + (1-f) x_{kmB}) + (1-f) c + \rho\sigma_u [f((-\phi_j)/(1-\Phi_j))_m + (1-f)(\phi_j/\Phi_j)_m] \quad \text{(A-15)}$$

since by definition $f w_{mA} + (1-f) w_{mB} = 0$.

On the covariance Cov (x_{kM}, y_M):

This covariance may be written as

$$\text{Cov}(z_{kM}, y_M) = (1/N)\{[\sum_{i\in A}(x_{kMA} - (f x_{kMA} + (1-f) x_{kMB}))(y_{MA} - y_M)] + [\sum_{i\in B}(x_{kMB} - (f x_{kMA} + (1-f) x_{kMB}))(y_{MB} - y_M)]\} \quad \text{(A-16)}$$

where x_{kMA} and x_{kMB} are the mean values of x_k in groups A and B.

It is then easy to show that, after some simplifications, one ends up with

$$\text{Cov}(z_{kM}, y_M) = f(1-f)(x_{kMB} - x_{kMA})(y_{MB} - y_{MA}) \quad \text{(A-17)}$$

The between groups variance V_{BET}:

It may be expressed as

$$V_{BET} = f(y_{mA} - y_m)^2 + (1-f)(y_{mB} - y_m)^2 \quad \text{(A-18)}$$

$$\Leftrightarrow V_{BET} = f(y_{mA})^2 + (1-f)(y_{mB})^2 - (f\, y_{mA} + (1-f)\, y_{mB})^2 \quad \text{(A-19)}$$

since $y_m = f\, y_{mA} + (1-f)\, y_{mB}$

Expression (A-38) may then be easily simplified to derive finally that

$$V_{BET} = f(1-f)(y_{mA} - y_{mB})^2 \quad \text{(A-20)}$$

The sum of all the contributions to the between groups dispersion:

It is easy to show that it is equal to 1 since

$$\sum_{k=1 \text{ to } K} b_k (x_{kMB} - x_{kMA})] + c + (\rho\sigma_u)[(\phi_j / \Phi_j)_m - ((-\phi_j)/(1-\Phi_j))_m] + (w_{MB} - w_{MA})$$

$$= \{[\sum_{k=1 \text{ to } K} b_k x_{kMB}] + c + [(\rho\sigma_u)(\phi_j / \Phi_j)_m] + w_{MB}\}$$

$$- \{[\sum_{k=1 \text{ to } K} b_k x_{kMA}] + (\rho\sigma_u)[((-\phi_j)/(1-\Phi_j))_m] + w_{MA}\}$$

$$= (y_{MB} - y_{MA})$$

Overall contribution of the variable on-the-job training F_j:

It will have evidently an impact only via its contribution to the between groups variance V_{BET}; this effect will be expressed as

$$\{[(c)\,\text{Cov}(F_j, y_M)] / V_{BET}(y_j)\}\{V_{BET}(y_j)/V_{TOT}(y_j)\} = \{[(c)\,\text{Cov}(F_j, y_M)] / V_{TOT}(y_j)\} \quad \text{(A-21)}$$

where F_j as before, will be equal to 0 for group A and to 1 for group B, while y_M will be equal to y_{MA} gor group A and to y_{MB} for group B.

Overall contribution of Mills' ratio:

It will be the sum of three elements:

- its impact via its contribution to the within group A variance V_A. This effect will be expressed as

$$\{[(d)\text{Cov}(\lambda_{j,\, j\in A}, y_{j,\, j\in A})/ V_A(y_j)]\}\{(f)V_A(y_j)/V_{WITH}(y_j)\}\{V_{WITH}(y_j)/V_{TOT}(y_j)\} =$$
$$\{(f)[(d)\text{Cov}((\lambda_{j,\, j\in A}, y_{j,\, j\in A})]/V_{TOT}(y_j)\} \quad \text{(A-22)}$$

- its impact via its contribution to the within group B variance V_B. This effect will be expressed as

$$\{[(d)\text{Cov}(\lambda_{j,\, j\in B}, y_{j,\, j\in B})/ V_B(y_j)]\}\{(1-f)V_B(y_j)/V_{WITH}(y_j)\}\{V_{WITH}(y_j)/V_{TOT}(y_j)\} =$$
$$\{(1-f)[(d)\text{Cov}((\lambda_{j,\, j\in B}, y_{j,\, j\in B})]/V_{TOT}(y_j)\} \quad \text{(A-23)}$$

- its impact via the between groups variance V_{BET}. This effect will be expressed as

$$\{(\rho\sigma_u)[(\phi_j / \Phi_j)_m - ((-\phi_j)/(1-\Phi_j))_m] / V_{BET}(y_j)\}\{V_{BET}(y_j)/V_{TOT}(y_j)\} =$$
$$\{(\rho\sigma_u)[(\phi_j / \Phi_j)_m - ((-\phi_j)/(1-\Phi_j))_m]\} / V_{TOT}(y_j)\} \quad \text{(A-24)}$$

We therefore end up with a total impact of Mills' ratio expressed as

$$\{[f\,(d)\,\text{Cov}((\lambda_{j,\, j\in A}, y_{j,\, j\in A})] + [(1-f)\,(d)\,\text{Cov}((\lambda_{j,\, j\in B}, y_{j,\, j(B)})]$$
$$+ [(\rho\sigma_u)((\phi_j / \Phi_j)_m - ((-\phi_j)/(1-\Phi_j))_m)]\} /V_{TOT}(y_j)\} \quad \text{(A-25)}$$

Overall contribution of the disturbances:

It will also be the sum of three elements:

- their impact via their contribution to the within group A variance V_A. This effect will be expressed as

$$\{[\text{Cov}(w_{j,\, j(A}, y_{j,\, j(A)})/ V_A(y_j)]\}\{(f)V_A(y_j)/V_{WITH}(y_j)\}\{V_{WITH}(y_j)/V_{TOT}(y_j)\} =$$
$$\{(f)[\text{Cov}(w_{j,\, j(A}, y_{j,\, j(A)})]/V_{TOT}(y_j)\} \quad \text{(A-26)}$$

- their impact via their contribution to the within group B variance V_B. This effect will be expressed as

$$\{[\text{Cov}(w_{j,\, j(B}, y_{j,\, j(B)})/V_B(y_j)]\}\{(1-f)V_B(y_j)/V_{WITH}(y_j)\}\{V_{WITH}(y_j)/V_{TOT}(y_j)\} =$$
$$\{(1-f)[\text{Cov}(w_{j,\, j(B}, y_{j,\, j(B)})]/V_{TOT}(y_j)\} \quad \text{(A-27)}$$

- their impact via the between groups variance V_{BET}; this effect will be expressed as

$$\{[\text{Cov}(w_M, y_M)] / V_{BET}(y_j)\}\{V_{BET}(y_j)/V_{TOT}(y_j)\} = \{[\text{Cov}(w_M, y_M)] / V_{TOT}(y_j)\} \quad \text{(A-28)}$$

where $w_M = w_{MA}$ if individual j belongs to group A and $w_M = w_{MB}$ if individual j belongs to group B. Similarly $y_M = y_{MA}$ if individual j belongs to group A and $y_M = y_{MB}$ if individual j belongs to group B.

The overall effect of the disturbances may hence be expressed as

$$\{(f)[\text{Cov}(w_{j,\, j\in A}, y_{j,\, j\in A})]\} + \{(1-f)[\text{Cov}(w_{j,\, j\in B}, y_{j,\, j\in B})]\} + \{[\text{Cov}(w_M, y_M)]\} /V_{TOT}(y_j)\} \quad \text{(A-29)}$$

The effect of the additional variables r_{gj} introduced earlier:

It will be expressed, as

$$\{[(e_g)\,\mathrm{Cov}\,(r_{g,j,\,j\in B}\,,\,y_{j,\,j\in B})/\,V_B(y_j)]\}\{(1\text{-}f)\,V_B(y_j)/\,V_{WITH}(y_j)\}\{V_{WITH}(y_j)/V_{TOT}(y_j)]\} =$$
$$\{(1\text{-}f)[(e_g)\,\mathrm{Cov}\,(r_{g,j,\,j\in B}\,,\,y_{j,\,j\in B})]/\,V_{TOT}(y_j)\} \qquad \text{(A-30)}$$

Table 1: Mean[22] Values of Variables

Variable	Group of Individuals who Received Training	Group of Individuals who did not Receive any Training
Higher Education Diploma	0.119	0.045
Holder of "Baccalauréat"	0.174	0.111
Holds a CAP or BEP	0.3247	0.360
Holds a BEPC	0.058	0.064
Holds a CEP	0.158	0.345
"Specialized" Worker	0.048	0.156
"Qualified"Worker	0.1921	0.290
Engineer or Managerial Position	0.181	0.071
Employee	0.330	0.359
Other Professions	0.048	0.031
Has a Work Contract of Undetermined Duration	0.946	0.926
Other Categories of Trainees	0.0266	0.032
Seniority in Firm	11.9 (9.57)	11.4 (9.62)
Square of Seniority in firm	232 (307)	223 (315)
Age	39.8 (8.97)	41.3 (9.74)
Square of Age	1663 (730)	1798 (810)

[22] For the quantitative variables we also give the standard deviation in parenthesis.

Works Full-time (40 hours at least)	**0.886**	**0.823**
Works Part-time (30 to 40 hours)	**0.043**	**0.048**
Works Part-time (less than 15 hours)	**0.0051**	**0.023**
Has the Same Work Schedule Every Day	**0.592**	**0.645**
Has a Variable Work Schedule	**0.257**	**0.217**
Logarithm of Monthly Wage	**9.15 (0.518)**	**8.85 (0.566)**

Table 2: Results of the Probit Analysis

Dependent Variable: Probability to Receive Vocational Training

Explanatory variable	Coefficient	T-values
Constant	-0.583	-5.02
Female	-0.066	-15.7
French	0.318	-12.6
« Specialized » Worker	-1.053	-1.7
«Qualified» Worker	-0.640	4.4
Engineer or managerial position	0.131	2.2
Employee	-0.389	-7.5
Other Professions	-0.080	-0.9
Has a Work Contract of Undetermined Duration	0.151	1.8
Other Categories of trainees	0.221	1.8
Firm's Size : 0-3	-0.294	-4.1
3-9	-0.296	-6.1
10-49	-0.212	-4.6
49-100	-0.0106	-0.2
More than 500	0.328	7.9

Note: -log V = 4496; $\chi 2(14)=844.2$; N=8261

Table 3: Regression Results

Dependent Variable: Logarithm of Monthly Earnings

Explanatory Variables	Coefficient	T-values
Constant	7.94	109.5
Higher Education Diploma	0.143	7.52
Holder of "Baccalauréat"	-0.064	-4.15
Holds a CAP or BEP	-0.164	-11.5
Holds a BEPC	-0.177	-9.31
Holds a CEP	-0.285	-18.6
"Specialized" Worker	-0.156	-6.96
"Qualified"Worker	-0.100	-5.65
Engineer or Managerial Position	0.356	21.1
Employee	-0.146	-9.41
Other Professions	-0.00034	-0.015
Has a Work Contract of Undetermined Duration	0.042	2.15
Other Categories of Trainees	-0.0062	-0.222
Seniority in Firm	0.011	8.03
Square of Seniority in firm	-0.00013	-3.26
Age	0.022	6.81
Square of Age	-0.00022	-5.54

Works Full Time (40 hours at least)	**0.666**	**49.8**
Works Part Time (30 to 40 Hours)	**0.406**	**19.8**
Works Part Time (less than 15 Hours)	**-0.814**	**-27.7**
Has the Same Work Schedule Every Day	**-0.041**	**-3.72**
Has a Variable Work Schedule	**-0.035**	**-2.72**
Female	**-0.148**	**-15.8**
Foreigner	**-0.062**	**-3.94**
Received Vocational Training	**0.479**	**9.91**
Mill's Ratio	**-0.249**	**-8.61**
R-Square	**0.670**	
Adjusted R-Square	**0.669**	
F-Value for Regression	**670**	

Table 4: Regression results

Dependent Variable: Residual of Regression of Table 2

Explanatory Variables	Coefficient of Regression	T-Values
Constant	-0.080	-2.2
Training originated in individual initiative*	0.107	2.7
Training originated in initiative from firm*	0.097	2.6
Training originated in initiative from both the individual and the firm*	0.100	2.6
The individual financed the training**	-0.104	-2.4
Other type of Financing**	-0.134	-5.6

* The reference category is "Other sources of initiative"

** The reference category is « Financing by the firm »

Table 5: Characteristics of the training financed by the employer as a function of the side that took the initiative of the training[23]

SIDE TAKING THE INITIATIVE OF THE TRAINING	TRAINING TAKEN AT THE INITIATIVE OF THE EMPLOYER	TRAINING TAKEN AT THE INITIATIVE OF BOTH THE EMPLOYER AND THE EMPLOYEE	TRAINING TAKEN AT THE INITITATIVE OF THE EMPLOYEE
FIELD OF TRAINING			
Trade, sales, marketing	**11.8**	**10.8**	**7.1**
Industrial techniques	**20.4**	**19.9**	**16.4**
Hygiene and security	**14.5**	**7.3**	**5.5**
Secretarial tasks, office management, computer skills	**25.5**	**31.4**	**27.7**
Other categories	**27.8**	**30.6**	**43.3**
TYPE OF TRAINING			
Practical training, course, seminar	**63.1**	**92.6**	**88.7**
On-the job training	**34.6**	**2**	**4**
Other categories	**2.3**	**5.4**	**7.2**

[23] For each characteristic, the sum of each column is equal to 100%.

GOAL OF TRAINING			
Getting adjusted to the job	**94.9**	**92.6**	**88.7**
Changing jobs	**2**	**2**	**4**
Getting a diploma or a recognized qualification	**3.1**	**5.4**	**7.2**
TRAINING WITH CERTIFICATION			
Yes	**7.7**	**9**	**12.9**
LENGTH OF THE TRAINING PERIOD (IN HOURS)			
Median Value	**16**	**24**	**24**
Mean Value	**46.7**	**75.2**	**150.4**

Table 6: Breakdown of the Total Variance
into the Sum of Between and Within Groups Variances

Type of Variance	Value and Share (in percentage)
Overall Variance	**32.33 (100%)**
Between Groups Variance	**1.78 (5.5%)**
Within Groups Variance	**30.55 (94.5%)**

Table 7: Contributions of the Explanatory Variables to the Between Groups Variance

Explanatory Variables	Contribution (in percentage)
Higher Education Diploma	3.55
Holder of "Baccalauréat"	-1.31
Holds a CAP or BEP	1.82
Holds a BEPC	0.35
Holds a CEP	17.04
"Specialized" Worker	6.10
"Qualified"Worker	2.71
Engineer or Managerial Position	13.13
Employee	2.14
Other Professions	-0.28
Has a Work Contract of Undetermined Duration	0.32
Other Categories of Trainees	0.003
Seniority in Firm	1.53
Square of Seniority in firm	-0.37
Age	9.61
Square of Age	-11
Works Full-time (40 hours at least)	15

Works Part-time (30 to 40 hours)	**-0.61**
Works Part-time (less than 15 hours)	**5.07**
Has the Same Work Schedule Every Day	**0.75**
Has a Variable Work Schedule	**-0.12**
Received Vocational Training	**161.8**
Mill's Ratio	**-127.6**
Residuals	**0**

Table 8: Contributions of the Explanatory Variables to the Within Groups Variance

Explanatory Wariables	Individuals Belonging to Group A (Did not Receive any Vocational Training)	Individuals Belonging to Group B (Received Vocational Training)
Higher Education Diploma	1.31	3.07
Holder of "Baccalauréat"	-0.21	0.065
Holds a CAP or BEP	-0.30	1.96
Holds a BEPC	0.18	0.49
Holds a CEP	5.90	3.36
"Specialized" Worker	1.58	0.88
"Qualified" Worker	-0.75	1.03
Engineer or Managerial Position	6.91	14.67
Employee	4.13	5.12
Other Professions	-0.0002	0.0002
Has a Work Contract of Undetermined Duration	0.20	0.366
Other Categories of Trainees	0.0073	0.026
Seniority in Firm	4.29	5.44
Square of Seniority in firm	-1.60	-1.86
Age	-1.24	-7.78
Square of Age	2.01	10.1
Works Full-time (40 hours at least)	25.97	18.19

Works Part-time (30 to 40 hours)	**-1.83**	**-1.59**
Works Part-time (less than 15 Hours)	**9.40**	**2.66**
Has the Same Work Schedule Every Day	**0.083**	**0.060**
Has a Variable Work Schedule	**0.055**	**-0.15**
Female	**5.00**	**4.29**
French	**0.15**	**0.008**
Mill's Ratio	**3.52**	**7.72**
Residuals	**35.49**	**33.01**

Table 9: Contribution of the Explanatory Variables of the Regression of Table 3 to the Variance of the Within group B Variance

Explanatory Variables	Contribution (in percentage) to the Variance of the Within Group B Variance
Training originated in individual initiative*	0.16
Training originated in inititiative from firm*	-0.63
Training originated in initiative from both the individual and the firm*	1.026
The individual financed the training**	0.178
Other type of Financing**	1.651
Residual of the Regression of Table 3	31.64

Table 10: Breakdown of the Overall Variance of the Logarithm of Wages

Type of Variance	Value
Between Groups Variance of Actual (Logarithms of) Incomes	0.018
- Contribution of Vocational Training (Dummy Variable "Received Vocational Training" Plus Mills Ratio)	0.006
- Contribution of Other Variables	0.012
Variance of Predicted (Logarithms of) Incomes	0.213
- Between Groups Variance of Predicted (Logarithms of) Incomes	0.018
- Within Groups Variance of Predicted (Logarithms of) Incomes	0.195
- Contribution of the group who did not receive any vocational training (group A)	0.145
- Contribution of the group who received vocational training (group B)	0.050
Variance of Actual (Logarithms of) Incomes	0.323
- Between Groups Variance of Actual (Logarithms of) Incomes	0.018
- Within Groups Variance of Actual (Logarithms of) Incomes	0.306
- Contribution of the group who did not receive any vocational training (group A)	0.230
- Contribution of the group who received vocational training (group B)	0.076