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

Saïd Hanchane, Jacques Silber

DOCUMENT DE TRAVAIL- L.E.S.T.- CNRS – UMR 6123
Université de Provence (U1) et Université de la Méditerranée (U2) - 35 avenue J. FERRY - 13626 AIX-EN-PROVENCE CEDEX
Tél. : Standard : 04.42.37.85.00 - Fax : 04.42. 26.79.37 email lest@univ-aix.fr

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

by

Said Hanchane[1]

and

Jacques Silber[2]

February 2004

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 written while Jacques Silber was visiting the LEST.

1 Introduction

Among the various factors mentioned to explain the increased wage dispersion in several Western countries in the 1980s and 1990s[3] the most popular reason seems to be the demand shift in favor of highly skilled workers (see, Bound and Johnson, 1992, or Katz and Murphy, 1992). This shift could explain the growing wage gap between groups of workers with different levels of education and experience. But it could also be a determinant of the increase in within groups (such as those defined on the basis of age and education) inequality (see, Juhn, Murphy and Pierce, 1993).

This increase in the demand for skills is likely to lead to an upgrading of the skills, whether it is initiated by firms or workers. What would then be the impact on the wage structure of such an upgrading of skills? Very few studies seem to have taken a look at this issue. Constantine and Neumark (1996), using the January 1983 and 1991 Current Population Survey (CPS), found for example that in the United States training was quite valuable though its effect on between educational groups wage dispersion was not important. They argue that "the changes in the distribution of training were not sufficiently large to have substantial effects on the wage structure".

Another interesting study is that of Marcotte (2000). Using data from cohorts of the National Longitudinal Surveys he found that young men entering the labor market beginning in the late 1960s and those entering in the early 1980s differed little in the average incidence of training or earnings premium associated with training. He concluded, however, that shifts in training towards more educated workers could account for more than 40% of the observed increase in college-high school earnings differences among young men. Changing patterns in continuing learning nevertheless do not help explain the growth in earnings differences within equally educated groups.

Blundell et al. (1999) looked at the impact of training on the wage profile of individuals in the United Kingdom over the period between 1981 and 1991, controlling for two types of endogeneity. The first one is related to permanent differences among individuals in their propensity to participate in training, which are correlated with earnings and returns to training. The second source of endogeneity arises when firms that do well and hence are able to pay higher wages are also increasing their training levels or when individuals who "received a bad productivity shock" become eligible for training and find it less costly. They analyzed data from the British National Child Development Surwey, a continuing longitudinal survey of persons living in Great Britain, who were born in 1958. They checked who actually receives training and concluded that individuals who had undertaken Employer-provided training courses in 1981 were more likely to obtain training between 1981 and 1991. Moreover the probability of undertaking such an employer-provided training course increases with school qualification. As far as the returns to training are concerned, they found that employer-provided training leading to a qualification has positive returns, regardless of whether it was obtained with the current or a former employer. Informal employer- provided training on the contrary had much lower returns when a person moves, this clearly indicating that this type of training is rather firm specific. Blundell et al. (1999), however, did not look at the implications of their findings on the wage structure.

In France access to training seems to be quite selective as it is much easier for individuals holding the "baccalauréat"[4] or with two years of studies beyond it (see, Croquey, 1995, or Aventur and Hanchane, 1999). In fact the probability to receive on-the-job training appears to depend on the initial level of the human capital of the individual and to be positively correlated with his professional status in the firm, the nature of his work contract and the size of the firm. It seems that this probability increases also with seniority in the firm (see, Béret and Dupray, 2000). If this

[3] See Levy and Murnane, 1992, or Karoly, 1992, for a review of the evidence in the United States.

[4] National Diploma delivered at the end of High School.

is the case one could argue that there is a matching process in so far as training is first of all aimed at those who have proven they are well integrated in the firm. Training would then serve to "keep the workers in the firm" (see, Goux and Maurin, 1997) and have a weak impact on earnings unless the individual participates frequently in training programs (see, Béret and Dupray, 1998). In short the goal of training would be to select and keep the most performing workers. The inequality in continued training that is observed between the firms would then be a testimony of the existence of segmented labor markets (see, Hanchane and Joutard, 1998).

Past studies have estimated that the net impact of continued training on earnings is around 2%, once the set of characteristics related to the professional history of the individual and of the firm is controlled for. This was the result obtained by Béret and Dupray, 2000, on the basis of the 1993 I.N.S.E.E.[5] Survey on "Formation and Professional Qualification". Similar findings appear in the study of Goux and Maurin (1997) who managed to neutralize the impact of the diversity in wage policies among firms.

Such conclusions should be compared with what is observed in Germany. The breakdown of the wage gap, between those who received training and those who did not, shows that personal characteristics play a much more important role in France. But it turns out that this is a consequence of the selectivity bias, while in Germany this bias is not statistically significant (see, Béret and Dupray, 2000). In other words observed and unobserved individual characteristics play a much greater role in France as far as the probability to have access to training is concerned. But once this fact is taken into account, these characteristics, at the difference of what happens in Germany, do not have any significant impact on individual earnings (see, Goux and Maurin, 1997, for the French case). A study conducted by OECD (see, OECD, 1999, pages 179-181) reached the same conclusion. Half of the earnings gap between those who received training and those who did not is due to the fact that the firms providing training pay higher salaries in any case. The second half of the gap is related to factors that have a simultaneous impact on the probability of access to training and on the earnings.[6]

The importance of the second component of this earnings gap[7] should be related to the fact that the presence of segmentation in the labor market is often explained by the existence of barriers to entry in the (good) jobs that belong to the primary sector[8]. It may be argued that these jobs are characterized by a pattern of skill formation that is based on on-the-job training policies that are very selective. As a consequence the fact that a worker receives on-the-job training may be a signal of his integration in the internal market. For the economist such a participation in on-the-job training programs has therefore become a statistical criterion of segmentation because the two populations of the individuals who received and did not receive on-the-job training have unobserved characteristics that are very different. As a consequence the impact of on-the-job training on earnings should be computed, net of the effect of this unobserved heterogeneity.

One may however wonder whether the significant impact of the unobserved heterogeneity on the average earnings gap between the two groups of individuals who receive and do not receive on-the-job training is a sufficient condition for concluding that it serves as a criterion of segmentation. Assume we indeed find first that there is such a selectivity bias, second that there

[5] I.N.S.E.E. is the National Institute of Statistics and Economic Studies.

[6] 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.

[7] The data sources we used did not allow us to take into account the first component.

[8] The idea according to which barriers to entry represent a test of the hypothesis of segmentation is stressed by Dickens and Lang (19XX), Magnac (1991), Taubman and Wachter (19XX).

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 not (group B), we will necessarily observe that there the between groups (A and B) variance of (the logarithms of) 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. Assume that at least some of the explanatory (the observed) variables that have been taken into account in the earnings function have a significant impact on the earnings of both groups. One may imagine, for example, that some of these variables have an important impact on the earnings of one of the groups while some others have an effect on those of the other group. What should be clear however is 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 really relevant because, under such an hypothesis, the between groups variance is assumed 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 any more a relevant criterion of labor market segmentation.

Testing such hypotheses remained a difficult task until very recently. New developments in income inequality decomposition techniques and in particular in the application of such techniques to regression analysis (see, Fields, forthcoming) allow us now 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.

We will therefore 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' (forthcoming) technique, we will be able to quantify the exact contribution of the observed (the explanatory) variables and of the unobserved individual characteristics on the variance of earnings. In particular we will be able to determine the impact of the unobserved heterogeneity on such a dispersion.

The paper is organized as follows. Section 2 describes the data sources while section 3 gives the estimates of the coefficients of the earnings function, after correcting for the selectivity bias. Section 4 presents the decomposition technique allowing the estimation of the exact impact of training and other variables on the various components of the earnings dispersion and gives the results of such a breakdown. Concluding comments are finally given in section 6.

2 The data sources

The survey

The analysis is based on a survey called « Continued Training 2000 ». This survey represents in fact an additional questionnaire that was part of the « Employment Survey » realized by the French Statistical Institute (I.N.S.E.E.) in March 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. Its main interest is that it allows to observe the attitude of individuals vis-a-vis continued training, the constraints that may limit the access to training and the practices of those who receive this type of training.
This survey covered 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. These individuals were surveyed in a face to face interview on the training they received during their active life, that is, since the end of their studies. More precisely these individuals were allowed to describe six training periods, the information they provided being the most detailed for the training received that was closest to January 1999. This latter information was indeed the one that formed the basis of our empirical analysis.

The concept of training in 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 training 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 other variables

Educational Level
Six categories of diploma have been considered: Higher education diploma, two years of study beyond the "baccalauréat", "baccalauréat", CAP or BEP[9], BEPC[10] and CEP[11].

Socio-Professional Category
Six categories have been distinguished. Engineers or individuals having a managerial position, "intermediate professions", "specialized" manual workers[12] , "qualified" manual workers[13], employees and other professions.

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

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

[10] 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.

[11] This optional exam certifies that the individual completed primary school.
[12] « ouvriers spécialisés » in French.
[13] « ouvriers qualifiés » in French.

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.
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.

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 (this is the value of the difference between the logarithms of wages in the two groups). Several characteristics can explain such a difference. First the proportion of those who have a Higher Education Diploma is much higher (11.9% versus 4.5%) among those who received a form of training. The share of those who have the "baccalauréat" is also higher among the "trained" (17.4%versus 11.1%). On the contrary the shares of those with a low level of education is higher among those who did not receive any training (see, the percentages for the variables "holds a CAP or BEP", holds a "BEPC", holds a CEP). Since earnings tend to increase with the level of education (see, below the results of the regression in Table 3) we have a first reason why earnings are higher among those who received 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, a category expected to have higher earnings, is much higher among those who received training. Differences between the two categories are small as far as the shares of employees, of other professions, of those having a

contract of undetermined duration, of other categories of trainees and of seniority in the firm are concerned. Age that is expected to have a positive effect on earnings is quite higher on average among those who did not receiwe any training (41.3 versus 39.8 for those who received training). The proportion of those working full time, a variable that certainly has a positive impact on monthly earnings, is slightly higher among those who received training (88.6% versus 82.3 among those who did not receive any training). The proportion of those who work part time between 15 and 30 hours on the contrary is somehow higher among those who did not receive any training (2.3% versus 0.5% among those who received training).

3 Estimating the Earnings Function

3.1 The Determinants of On-the-Job Training

In order to make a correction for the selectivity bias, we first estimated a Probit model[14] which gives us the determinants of the access to the training[15] that took place during the 14 months preceding the date at which the survey took place. 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 the 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, 2002). 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[16], 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[17] in the access to training.

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

[14] See, section I in the methodological Appendix, for a short presentation of why such a procedure has to be used.

[15] Whatever the type of training and the closest it was to January 1999.

[16] « Stage » in French.

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

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.

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

The Probit model whose results have just been analyzed allowed us to estimate the 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 regsession 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 age and its square, seniority in the firm and its square, three dummy variables giving the weekly duration of work (4 categories) and finally two variables giving information on the type of work schedule (3 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 and has an "alternate"[18] work schedule. 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, engineess and individuals who are part of the managerial staff earn significantly more than the other categories. 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

[18] 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.

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 the Mills ratio is significant) implies that the unobserved heterogeneity affects both the probability to receive on-the-job training and the earnings themselves. 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 the 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).

3.3 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.[19] 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

[19] The 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.

making. Some econometric tests that were undertaken to check such an hypothesis tend however to invalidate such a possibility.[20]

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.[21] 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 have 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 a 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, 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.

---

[20] 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.

[21] 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).

4 The Impact of Training on Earnings Dispersion

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

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

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

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 is the value taken by the disturbance u_j for individual j. Note also that we will assume below that

$b_{K+1} = c$, $b_{K+2} = d$, $b_{K+3} = 1$

Following Fields(forthcoming) it can be shown (see, section II in the methodological Appendix) that the relative contribution of factor k to the variance $V(y_j)$ may be expressed as

$$s_k(y_j) = [(b_k) \text{Cov}(z_{kj}, y_j)] / V(y_j) \quad (2)$$

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

$$s_h(y_j) = [(b_h) \text{Cov}(z_{hj}, y_j)] / V(y_j) \quad (3)$$

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

$$s_F(y_j) = [(c) \text{Cov}(F_j, y_j)] / V(y_j) \quad (4)$$

while that of the Mills ratio will be written as

$$s_\lambda(y_j) = [(d) \text{Cov}(\lambda_j, y_j)] / V(y_j) \quad (5)$$

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

$$s_u(y_j) = \text{Cov}(\varepsilon_j, y_j) / V(y_j) \quad (6)$$

While expressions (3) to (6) 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 Separating the Between and Within Groups Effects

4.2.1 Contribution of the Explanatory Variables to the Within-Groups Variance

Introducing the Mills ratio (see, the methodological Appendix and in particular equation (A-6) we may rewrite (1) separately for an individual belonging to group A and one belonging to group B. In the first case we write

$$y_{jA} = \Sigma_{k=1 \text{ to } K} b_k x_{kjA} + \rho\sigma_\varepsilon [(-\phi_j)/(1-\Phi_j)] + w_{jA} \qquad (7)$$

Since $[(-\phi_j)/(1-\Phi_j)]$ is the expression for Mill's ratio in group A and $\rho\sigma_\varepsilon$ that for the coefficient d (see, Green, 2000).

For group B expression (1) will be written as

$$y_{jB} = \Sigma_{k=1 \text{ to } K} b_k x_{kjB} + c + \rho\sigma_\varepsilon [\phi_j / \Phi_j] + w_{jB} \qquad (8)$$

since $[\phi_j / \Phi_j]$ is the expression for Mill's ratio in group B.

Contribution to the within groups variance:

Since the within groups variance is equal to the weighted sum of the variance within each of the two groups A and B (see, section III of the methodological Appendix) we ma write that

$$V_{WITH} = f\, V_A(y_j) + (1-f)\, V_B(y_j) \qquad (9)$$

Combining (2) and (9) we derive that the contribution $s_{k,WITH}(y_j)$ of each of the (K+2) factors[22] to the within groups variance may be written as

$$s_{k,WITH}(y_j) = \{(1-f)[(b_k)\text{Cov}(z_{kj,j\in B}, y_{j,\, j\in B})/V_B(y_j)] + (f)[(b_k)\text{Cov}(z_{kj,\, j\in A}, y_{j,\, j\in A})/\, V_A(y_j)] \qquad (10)$$

One may however further expand (1) by assuming 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

$$w_j = \Sigma_{g=1 \text{ to } G} e_g r_{g,j} + \nu_j \text{ for } j\in B \qquad (11)$$

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

$$V(w_{j,j\in B}) = \Sigma_{g=1 \text{ to } G+1}[(e_g)\,Cov\,(r_{g,j\,j\in B},\, w_{j\,j\in B})] \qquad (12)$$

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

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

$$s_g(y_{j,\,j\in T}) = [(e_g)\,Cov\,(r_{g,j,\,j\in B},\, y_{j,\,j\in B})/\ V_B(y_j)] \qquad (13)$$

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

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. Using (7) and (8) the mean values y_{mA} and y_{mB} of the logarithms of wages in groups A and B may be expressed as

$$y_{mA} = \Sigma_{k=1 \text{ to } K}\, b_k x_{kmA} + \rho\sigma_\varepsilon [((-\phi_j)/(1-\Phi_j))_m] + w_{mA} \qquad (14)$$

and

$$y_{mB} = \Sigma_{k=1 \text{ to } K}\, b_k x_{kmB} + c + \rho\sigma_\varepsilon [(\phi_j / \Phi_j)_m] + w_{mB} \qquad (15)$$

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

$$y_m = \Sigma_{k=1 \text{ to } K} b_k (f\, x_{kmA} + (1-f)\, x_{kmB}) + (1-f)\, c + \rho\sigma_\varepsilon [f\, ((-\phi_j)/(1-\Phi_j))_m + (1-f)\, (\phi_j / \Phi_j)_m] \qquad (16)$$

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

Contributions 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, where $y_{M,A}$ and $y_{M,B}$ are given in expressions (14) and (15).

---

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

The contribution $s_{k,B}(y_j)$ of each of the (K+3) factors to the between groups variance, using again Fields' (forthcoming) approach, will then be expressed as

$$s_{k,BET}(y_j) = [(b_k) \text{Cov}(z_{kM}, y_M)] / V_{BET}(y_j) \qquad (17)$$

For example to determine the contribution of a given characteristic x_k to the between groups variance of the logarithms of wages we have to compute, as shown in (17), the covariance

$\text{Cov}(x_{kM}, y_M)$. The latter may be written as

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

$$+ [\sum_{i \in B}(x_{kMB} - (f\, x_{kMA} + (1\text{-}f)\, x_{kMB}))(y_{MB} - y_M)]\} \qquad (18)$$

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

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

$$\text{Cov}(z_{kM}, y_M) = f(1\text{-}f)(x_{kMB} - x_{kMA})(y_{MB} - y_{MA}) \qquad (19)$$

Since the between groups variance may be expressed (see, section III of the methodological Appendix) as

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

we can combine (17), (19) and (20) to derive that

$$s_{k,BET}(y_j) = [(b_k)(x_{kMB} - x_{kMA})] / (y_{MB} - y_{MA}) \qquad (21)$$

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 $x_{kMA} = 0$,

$$s_{F,BET}(y_j) = [c / (y_{MB} - y_{MA})] \qquad (22)$$

The contribution of the ratio of Mill λ_j to the between groups dispersion will be expressed, using (7) and (8), as

$$s_{\lambda,BET}(y_j) = (\rho\sigma_\varepsilon)[(\phi_j / \Phi_j)_m - ((-\phi_j)/(1-\Phi_j))_m] / (y_{MB} - y_{MA}) \qquad (23)$$

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

$$s_{u,BET}(y_j) = [(w_{MB} - w_{MA})] / (y_{MB} - y_{MA}) \qquad (24)$$

where u_{MB} and u_{MA} are respectively the mean values of the disturbances in groups B and A.

Combining expressions (21) to (24) it is easy to show that the sum of all the contributions to the between groups dispersion is equal to 1 since

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

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

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

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

4.2.3 Summary :

We may therefore conclude, using all the previous results, that

1) 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:

- its impact via its contribution to the within group A variance V_A (see, (10)); this effect will be expressed as

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

$$\{(f)[(b_k)\text{Cov}(z_{kj, j\in A}, y_{j, j\in A})]/V_{TOT}(y_j)\} \qquad (25)$$

- its impact via its contribution to the within group B variance V_B, (see again (10)); this effect will be expressed as

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

$$\{(1-f)[(b_k)\text{Cov}(z_{kj, j\in B}, y_{j, j\in B})]/V_{TOT}(y_j)\} \qquad (26)$$

- its impact via the between groups variance V_{BET} (see, (21)); this effect will be expressed as

$$\{[(b_k)\text{Cov}(z_{kM}, y_M)] / V_{BET}(y_j)\}V_{BET}(y_j)/V_{TOT}(y_j)\} = \{[(b_k)\text{Cov}(z_{kM}, y_M)] / V_{TOT}(y_j)\} \quad (27)$$

Combining (25), (26) and (27) we end up with a total impact of the variable k expressed as

$$\{[(f)(b_k)\text{Cov}(z_{kj,j\in A}, y_{j,j\in A})]+[(1-f)(b_k)\text{Cov}(z_{kj,j\in B}, y_{j,j\in B})]+[(b_k)\text{Cov}(z_{kM}, y_M)]\}/V_{TOT}(y_j) \quad (28)$$

2) the contribution of the variable on-the-job training F_j which, evidently, will have an impact only via its contribution to the between groups variance V_{BET} (see, (22)); this effect will be expressed as

$$\{[(c)\text{Cov}(F_j, y_M)] / V_{BET}(y_j)\}\{V_{BET}(y_j)/V_{TOT}(y_j)\} = \{[(c)\text{Cov}(F_j, y_M)] / V_{TOT}(y_j)\} \quad (29)$$

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.

3) The contribution of Mills' ratio will be the sum of three elements:

- its impact via its contribution to the within group A variance V_A (see, (10)). This effect will be expressed as

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

$$\{(f)[(d)Cov((\lambda_{j,\, j\in A},\ y_{j,\, j\in A})]/V_{TOT}(y_j)\} \qquad (30)$$

- its impact via its contribution to the within group B variance V_B, (see again (10)). This effect will be expressed as

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

$$\{(1\text{-}f)[(d)Cov((\lambda_{j,\, j\in B},\ y_{j,\, j\in B})]/V_{TOT}(y_j)\} \qquad (31)$$

- its impact via the between groups variance V_{BET} (see, (23)). This effect will be expressed as

$$\{(\rho\sigma_\varepsilon)[(\phi_j / \Phi_j)_m - ((-\phi_j)/(1-\Phi_j))_m] / V_{BET}(y_j)\}\{V_{BET}(y_j)/V_{TOT}(y_j)\} =$$

$$\{(\rho\sigma_\varepsilon)[(\phi_j / \Phi_j)_m - ((-\phi_j)/(1-\Phi_j))_m]\} / V_{TOT}(y_j)\} \qquad (32)$$

Combining (30), (31) and (32) we end up with a total impact of Mills' ratio expressed as

$$\{[f\ (d)\ Cov((\lambda_{j,\, j\in A},\ y_{j,\, j\in A})] + [(1\text{-}f)\ (d)\ Cov((\lambda_{j,\, j\in B},\ y_{j,\, j(B)})]$$

$$+ [(\rho\sigma_\varepsilon)((\phi_j / \Phi_j)_m - ((-\phi_j)/(1-\Phi_j))_m)]\}\ /V_{TOT}(y_j)\} \qquad (33)$$

4) The contribution of the disturbances will also be the sum of three elements:

- their impact via their contribution to the within group A variance V_A (see, (10)). This effect will be expressed as

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

$$\{(f)[Cov(w_{j,\, j(A)},\ y_{j,\, j(A)})]/V_{TOT}(y_j)\} \qquad (34)$$

- their impact via their contribution to the within group B variance V_B, (see again (10)). This effect will be expressed as

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

$$\{(1\text{-}f)[Cov(w_{j,\, j(B)},\ y_{j,\, j(B)})]/V_{TOT}(y_j)\} \qquad (34)$$

- their impact via the between groups variance V_{BET} (see, (24)); this effect will be expressed as

$$\{[Cov\ (w_M, y_M)] / V_{BET}(y_j)\}\{V_{BET}(y_j)/V_{TOT}(y_j)\} = \{[Cov\ (w_M, y_M)] / V_{TOT}(y_j)\} \qquad (35)$$

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

$$\{(f)[Cov(w_{j,\,j\in A}, y_{j,\,j\in A})]\} + \{(1\text{-}f)[Cov(w_{j,\,j\in B}, y_{j,\,j\in B})]\} + \{[Cov\,(w_M, y_M)]\} / V_{TOT}(y_j)\} \quad (36)$$

5) Finally the effect of the additional variables r_{gj} introduced in (11) will be expressed, using (34), as

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

$$\{(1\text{-}f)[(e_g)\,Cov\,(r_{g,j,\,j\in B}, y_{j,\,j\in B})]/\,V_{TOT}(y_j)\} \quad (37)$$

4.3 The Empirical Results

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.

Contributions of the various variables to the within groups variance

To analyze the results we have to remember the formulations given in equation (A-18) in the Appendix on the basis of equation (2) where the contribution of a given variable k (in percentage) to the variance of the regression is shown to be a function 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”. Note that although their coefficients in the regression are relatively similar (-0.27 and –0.22), the simple correlation (see, Table 11) with the dependent variable is higher in absolute value for “Employees” (-0.39) than for “Holds a CEP” (-0.17). in addition the standard deviation of the variable “Employees” is higher (0.470) than that of the variable “Holds a CEP” (0.364, see Table 1). This is why in group A the variable “Employees” has a greater contribution to the variance of earnings than the variable “Holds a CEP”.

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. These two variables have a however a smaller contribution (in percentage) because for the group of individuals who received training other variables play a role. Firstly note the important contribution of the dummy variable “Engineer or Managerial Position” (14.5%). This variable has a high coefficient in the regression since Engineers and those having a managerial position earn, ceteris paribus, 35.6% more than technicians, the occupational group that serves as reference. Moreover this variable ”Engineers or Managerial Position” is highly correlated with (the logarithm of) earnings (correlation coefficient equal to 0.55) and its standard deviation (0.385, see Table 1) is also quite high. 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.

Contributions of the various variables to the between groups variance

Equation (21) indicates 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 (21) may also be used to derive some additional intuitive interpretation of the results. Let us rewrite (21) 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%.

The contributions of the 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).

5 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 (forthcoming) 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[23] 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 that even though 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 be considered as 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 distribution 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

[23] 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.

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

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.

Constantine, J. and D. Neumark (1996) « Training and the Growth of Wage Inequality ,» *Industrial Relations*, 35 (4) : 491-510.

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.

Dickens, W. et K. Lang (1985), "A Test a Dual Labor Market Theory", *The American Economic Review*, 75.

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.

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. (1998),"Gender Economic Courses in Liberal Arts Colleges : Futher Result", *Journal of Economic Education*, 29 (4).

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.

Hanchane S, Stankiewicz F (2004), « Propositions pour une théorie organisationnelle de la formation– Les enseignements de l'enquête Formation Continue 2000 », Paper to be presented at the *Journées de Micro-économie Appliquée*, May 2004, Lille.

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

Juhn, C., K. M. Murphy and B. Pierce (1993) "Wage Inequality and the Rise of the Returns to Skill," *Journal of Political Economy*, 101 (3): 410-442.

Heckman J. et V. Hotz (1986) "An Investigation of Labor Market Earnings of Panamian Males Evaluating the Sources of Inequality", *Journal of Human Resources*, 21.

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.

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.

Magnac T. (1991), "Segmented or Competitive Labor Market?", *Econometrica*, 5 (1): 165-188.

Marcotte, D. (2000) "Continuing Education, Job Training and the Growth of Earnings inequality," *Industrial and Labor Relations Review* 53 (4): 602-623.

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.

OCDE (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.

Taubman P. and M.L. Wachter, (1986). "Segmented Labor Market," in Ashenfelter O. and R. Layard (eds.), *Handbook of Labour Economic*, volume 2, North Holland.

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

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

Methodological Appendix

I) The Empirical Specification of the Earnings Function

The problem here is to be able to identify separately the impact on earnings of the selectivity bias on one hand and the effect of vocational training on the other hand. The identification of the impact on earnings of a policy aiming for example at easing the professional insertion of youth in the labor force or at providing continued vocational training can be based on OLS only if the individuals participating in such programs have been randomly selected. In other words one would estimate an earnings function of the type

$$Y = Xb + cF + \varepsilon \qquad \text{(A-1)}$$

where Y is the logarithm of monthly earnings, X is a matrix of explanatory variables, F a dummy variable indicating whether the individual participated (F=1) or not (F=0) in the training program and u is a random variable of mean 0 and standard deviation σ_u.
However when the participation to the training program is subject to prior selection, equation (A-1) above cannot be estimated by OLS because the parameter c would be influenced by both the selectivity bias and the structural effect linked to the participation in the training program.

Let us therefore express the conditional expectation of earnings as

$$E(Y/X,F) = Xb + cF + E(\varepsilon/X,F) \qquad \text{(A-2)}$$

We need therefore to estimate the expectation of the residual u. But the distribution of these residuals is truncated by the values of F which cannot be considered as exogenous. We need therefore in a first step to specify the determinants of F and use a Probit model written as

$$F^* = Z\gamma + v \qquad \text{(A-3)}$$

with F=1 if F*>0 and F=0 otherwise.

In other words we calculate

$$\text{Prob}\{F=1\} = \Phi(Z\gamma) \qquad \text{(A-4)}$$

and

$$\text{Prob}\{F=0\} = 1 - \Phi(Z\gamma) \qquad \text{(A-5)}$$

where $\Phi(.)$ refers to the distribution function of the normal law.

If we assume that the joint distribution (v, ε) is a bivariate normal distribution expressed as $N(0, 0, 1, \sigma_\varepsilon, \rho)$, it is easy to determine the Residual of equation (2) on the basis of the properties of a truncated bivariate normal distribution (see, Greene, 2000).

Finally it is possible to separately identify the impact of the selectivity bias and that of the participation (F) in the training program by estimating the following OLS regression (see, Barnow et al., 1980):

$$Y = Xb + cF + (F \times \rho\lambda) + ((1-F) \times \rho\lambda^*) + w \qquad \text{(A-6)}$$

where

$$\lambda = \phi(Z\gamma)/\Phi(Z\gamma) \qquad \text{(A-7)}$$

and

$$\lambda^* = -\phi(Z\gamma)/[1-\Phi(Z\gamma)] \qquad \text{(A-8)}$$

ϕ being the normal density function while λ and λ^* are Mill's ratios when F=1 and F=0 respectively. Finally w is the residual of the regression.

Correcting the variance-covariance matrix for heteroskedasticity (because of the presence of an estimated variable λ and of the sensitivity of the residuals to the selectivity bias) is a standard procedure (see, Heckman, 1979). One should just remember in the estimation procedure that the ratios of Mills are not the same when F=1 and F=0.

II) The contribution of the explanatory variables to the variance of earnings: Fields' (forthcoming) approach.

Using (1) we may write that

$$\text{Var}(y_j) = \text{Cov}[\Sigma_{k=1 \text{ to } (K+3)}\, b_k z_{kj}, y_j] \qquad \text{(A-9)}$$

Dividing both sides of (A-9) by Var (y_j), we then derive that

$$1 = \text{Cov}[\Sigma_{k=1 \text{ to } (K+3)}\, b_k z_{kj}, y_j]/\ \text{Var}(y_j) \qquad \text{(A-10)}$$

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

$$\text{Cov}[\Sigma_{k=1 \text{ to } (K+3)}\, b_k z_{kj}, y_j] = \Sigma_{k=1 \text{ to } (K+3)} \text{Cov}[b_k z_{kj}, y_j] \qquad \text{(A-11)}$$

Expression (A-10) may therefore be expressed as

$$1 = \Sigma_{k=1 \text{ to } (K+3)} \text{Cov}[b_k z_{kj}, y_j]\ ]/\ \text{Var}(y_j) \qquad \text{(A-13)}$$

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

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

$$\text{Cor}[b_k z_{kj}, y_j] = \text{Cov}[b_k z_{kj}, y_j]/\ ((\sigma(b_k z_{kj}), (\sigma(y_j)) \qquad \text{(A-14)}$$

we end up, combining expressions (A-11) to (A-14), with

$$1 = \Sigma_{k=1 \text{ to } (K+3)} [\text{Cor}(b_k z_{kj}, y_j)\ (\sigma(b_k z_{kj}))/\ (\sigma(y_j))] \qquad \text{(A-15)}$$

However since

$$\text{Cor}(b_k z_{kj}, y_j) = \text{Cor}(z_{kj}, y_j) \qquad \text{(A-16)}$$

expression (A-15) implies that

$$\sigma(y_j) = \Sigma_{k=1 \text{ to } (K+3)}[(b_k)\,Cor\,(z_{kj}, y_j)\,(\sigma(z_{kj}))] \qquad \text{(A-17)}$$

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

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

$$\Leftrightarrow s_k(y_j) = [(b_k)\,Cor\,(z_{kj}, y_j)\,(\sigma(z_{kj}))] \,/\, \sigma(y_j) \qquad \text{(A-18)}$$

Using (A-14) expression (A-18) may be also expressed, after simplifying, as

$$s_k(y_j) = [(b_k)\,Cov\,(z_{kj}, y_j)] \,/\, V(y_j) \qquad \text{(A-19)}$$

where $V(y_j)$ denotes the variance of the logarithms of wages y_j.

III) Decomposing the overall variance into between and within groups variances:

Let us divide the population into two groups of individuals, those who did not receive on-the-job training (group A) and those who had on-the-job training (group B). Let f and (1-f) be the respective shares of groups A and B in the total population.

The overall variance V_{TOT} of the logarithms of wages in the whole population will then be written as

$$V_{TOT} = (1/N)\,[\,\Sigma_{j\in A}(y_{jA} - y_m)^2 + \Sigma_{j\in B}(y_{jB} - y_m)^2\,] \qquad \text{(A-20)}$$

where N is the size of the whole population.

Calling N_A and N_B the number of individuals in groups A and B, so that $f = (N_A/N)$ and

$(1\text{-}f) = (N_B/N)$, expression (A-1) may also be written as

$$V_{TOT} = \{(N_A/N)[(1/N_A)\Sigma_{j\in A}((y_{jA}-y_{mA})+(y_{mA}-y_m))^2]\} + \{(N_B/N)[(1/N_B)\Sigma_{j\in B}((y_{jB}-y_{mB})+(y_{mB}-y_m))^2]\}$$

$$\Leftrightarrow V_{TOT} = f\,[(1/N_A)\Sigma_{j\in A}((y_{jA}-y_{mA})^2] + f\,(y_{mA}-y_m)^2 + (1\text{-}f)[(1/N_B)\Sigma_{j\in B}((y_{jB}-y_{mB})^2] + (1\text{-}f)(y_{mB}-y_m)^2$$

$$V_{TOT} = f\,V_A(y_j) + (1\text{-}f)\,V_B(y_j) + f\,(y_{mA}-y_m)^2 + (1\text{-}f)(y_{mB}-y_m)^2 \qquad \text{(A-21)}$$

It is then easy to show that (A-21) indicates in fact that the overall variance V_{TOT} is equal to the sum of the within and between groups variances.

The within groups variance:

It is well-known that the within groups variance V_{WITH} is equal to the population weighted sum of the variance within each group:

$$V_{WITH} = f\,[(1/N_A)\Sigma_{j\in A}((y_{jA}-y_{mA})^2] + (1\text{-}f)[(1/N_B)\Sigma_{j\in B}((y_{jB}-y_{mB})^2]$$

$V_{WITH} = f\, V_A(y_j) + (1\text{-}f)\, V_B(y_j)$ (A-22)

where $V_A(y_j)$ and $V_B(y_j)$ refer to the variance of the logarithms of wages in groups A and B.

The between groups variance:

Let us call it V_{BET}. It may be expressed as

$V_{BET} = f\,(y_{mA} - y_m)^2 + (1\text{-}f)\,(y_{mB} - y_m)^2$ (A-23)

$\Leftrightarrow V_{BET} = f\,(y_{mA})^2 + (1\text{-}f)\,(y_{mB})^2 - (f\, y_{mA} + (1\text{-}f)\, y_{mB})^2$ (A-24)

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

Expression (A-24) may then be easily simplified to derive finally that

$V_{BET} = f\,(1\text{-}f)(y_{mA} - y_{mB})^2$ (A-25)

It should then be clear that the sum of V_{BET} in (A-25) and V_{WITH} in (A-22) is equal to the overall variance V_{TOT} in (A-21) so that

$V_{TOT} = = V_{WITH} + V_{BET}$ (A-26)

Table 1: Mean Values of Variables

Variable	Group of Individuals who Received Training	Group of Individuals who did not Receive any Training
Higher Education Diploma	0.119058	0.0448869
Holder of “Baccalauréat”	0.174304	0.110699
Holds a CAP or BEP	0.324197	0.360445
Holds a BEPC	0.0582441	0.0642929
Holds a CEP	0.157602	0.345427
“Specialized” Worker	0.0475375	0.155586
“Qualified”Worker	0.192291	0.289571
Engineer or Managerial Position	0.180728	0.0713804
Employee	0.329764	0.358927
Other Professions	0.0479657	0.0308809
Has a Work Contract of Undetermined Duration	0.946467	0.925582
Other Categories of Trainees	0.0265525	0.0323996
Seniority in Firm	11.8745	11.4096
Square of Seniority in firm	232.455	222.653
Age	39.785	41.2762
Square of Age	1663.33	1798.48

Works Full-time (40 hours at least)	0.886081	0.82349
Works Part-time (30 to 40 hours)	0.0432548	0.0477557
Works Part-time (less than 15 hours)	0.00513919	0.023456
Has the Same Work Schedule Every Day	0.592291	0.645461
Has a Variable Work Schedule	0.256531	0.217179
Logarithm of Monthly Wage	9.1468	8.8506

Table 2: Results of the Probit Analysis

Dependent Variable: Probability to Receive Vocational Training

Explanatory variable	Coefficient	T-values
Constant	-0.582913	-5.018
Female	-0.066122	-15.673
French	0.318499	-12.612
« Specialized » Worker	-1.05278	-1.739
«Qualified» Worker	-0.639545	4.373
Engineer or managerial position	0.131243	2.225
Employee	-0.389227	-7.453
Other Professions	-0.080194	-0.923
Has a Work Contract of Undetermined Duration	0.150849	1.773
Other Categories of trainees	0.220701	1.801
Firm's Size : 0-3	-0.294301	-4/052
3-9	-0/295934	-6.093
10-49	-0/21176	-4.621
49-100	-0.0105794	-0.170
More than 500	0.327546	7.899

Note: -log V = 4496; $\chi 2(14)=844.1866$; N=8261

Table 3: Regression Results

Dependent Variable: Logarithm of Monthly Earnings

Explanatory Variables	Coefficient	T-values
Constant	7.79499	104.191
Higher Education Diploma	0.141637	7.54541
Holder of “Baccalauréat”	-0.0610305	-3.95364
Holds a CAP or BEP	-0.148738	-10.4908
Holds a BEPC	-0.172615	-9.03481
Holds a CEP	-0.268688	-17.5295
“Specialized” Worker	-0/167361	-6.57283
“Qualified”Worker	-0.0823735	-4.11363
Engineer or Managerial Position	0.355713	18/4884
Employee	-0.217003	-12.5544
Other Professions	-00496812	-1.90557
Has a Work Contract of Undetermined Duration	-0.00137579	-1.9821
Other Categories of Trainees	0.00977079	-0.041973
Seniority in Firm	0.00977079	6.88255
Square of Seniority in firm	-0.000111141	-2.64295
Age	0.0219171	6.51289
Square of Age	-0.000210687	-5.20881

Works Full Time (40 hours at least)	0.708941	52.7373
Works Part Time (30 to 40 Hours)	0.400116	19.1893
Works Part Time (less than 15 Hours)	-0.820346	-26.8967
Has the Same Work Schedule Every Day	-0.0417639	-3.70048
Has a Variable Work Schedule	-0.0310386	-2.3757
Received Vocational Training	0.503431	9.53992
Mill's Ratio	-0.263616	-8.38388
R-Square	0.659	
Adjusted R-Square	0.658	
F-Value for Regression	693.1	

Table 4: Regression results

Dependent Variable: Residual of Regression of Table 2

Explanatory Variables	Coefficient of Regression	T-Values
Constant	-0/0799758	-2.17485
Training originated in individual initiative*	0.107008	2.73129
Training originated in initiative from firm*	0.0965566	2.58234
Training originated in initiative from both the individual and the firm*	0.099925	2.60159
The individual financed the training**	-0.104276	-2.39751
Other type of Financing**	-0.133894	-5.61997

* *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[24]

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

[24] 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	5.07

hours)	
Has the Same Work Schedule Every Day	0.75
Has a Variable Work Schedule	-0.12
Received Vocational Training	170
Mill's Ratio	-135.03

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	3.042
Holder of “Baccalauréat”	-0.20	0.062
Holds a CAP or BEP	-0.027	1.78
Holds a BEPC	0.018	0.48
Holds a CEP	5.57	3.16
“Specialized” Worker	1.70	0.9527
“Qualified”Worker	-0.062	0.851
Engineer or Managerial Position	6.91	14.64
Employee	7.12	7.60
Other Professions	-0.034	0.027
Has a Work Contract of Undetermined Duration	0.22	0.393
Other Categories of Trainees	0.0016	0.0058
Seniority in Firm	3.74	4.74
Square of Seniority in firm	-1.32	-1.53
Age	1.97	9.85
Square of Age	-1.19	-7.50
Works Full-time (40 hours at least)	27.67	18.19

Works Part-time (30 to 40 hours)	-1.80	-1.57
Works Part-time (less than 15 Hours)	9.48	2.68
Has the Same Work Schedule Every Day	0.084	0.060
Has a Variable Work Schedule	0.048	-0.13
Mill's Ratio	3.72	8.17
Residuals	36.72	34.02

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

Explanatory Variables	Contributioo (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

Table 11: Simple Correlation Coefficient Between the Explanatory Variables of the Regression and the Logarithm of Wages

Variable	Group of Individuals who Received Training	Group of Individuals who did not Receive any Training
Higher Education Diploma	0.344	0.254
Holder of "Baccalauréat"	-0.138	0.593
Holds a CAP or BEP	-0.132	0.213
Holds a BEPC	0.000	0.000
Holds a CEP	-0.167	-0.246
"Specialized" Worker	-0.139	-0.158
"Qualified"Worker	-0.136	0.933
Engineer or Managerial Position	0.554	0.426
Employee	-0.386	-0.333
Other Professions	-0.133	0.224
Has a Work Contract of Undetermined Duration	0.199	0.102
Other Categories of Trainees	-0.136	-0.377
Seniority in Firm	0.263	0.225
Square of Seniority in firm	0.232	0.213
Age	0.253	0.394
Square of Age	0.260	0.522

Works Full-time (40 hours at least)	0.419	0.579
Works Part-time (30 to 40 hours)	-0.998	-0.119
Works Part-time (less than 15 hours)	-0.237	-0.432
Has the Same Work Schedule Every Day	-0.153	-0.239
Has a Variable Work Schedule	0.495	-0.214
Mill's Ratio	-0.526	-0.405