I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10841

The Assimilation of Young Workers into the Labour Market in France: A Stochastic Earnings Frontier Approach

Stephen Bazen
Khalid Maman Waziri

JUNE 2017

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10841

The Assimilation of Young Workers into the Labour Market in France: A Stochastic Earnings Frontier Approach

Stephen Bazen
Aix-Marseille University, CNRS, EHESS and IZA

Khalid Maman Waziri
Aix-Marseille University, CNRS and EHESS

JUNE 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Assimilation of Young Workers into the Labour Market in France: A Stochastic Earnings Frontier Approach*

Stochastic earnings frontiers have been used in a relatively small number of papers to analyse workers' ability to capture their full potential earnings in labour markets where there is inefficient job matching (due to lack of information, discrimination, over-education or during process of assimilation of migrants). Using a representative survey of young persons having left full-time education in France in 1998 and interviewed in 2001 and 2005, this paper examines the process of their assimilation into normal employment and the extent to which job matches are inefficient in the sense that the pay in a job is below an individual's potential earnings (determined by education, other forms of training and labour market experience). Our results suggest that young workers manage to obtain on average about 82% of their potential earnings three years after leaving full-time education and earnings inefficiency had disappeared four years later. The results are robust to the treatment of selectivity arising from the exclusion of the unemployed in the estimation of the frontier.

JEL Classification: J30, J13, J24, C24

Keywords: stochastic earnings frontier, job matching, youth employment, earnings determination, sample selection stochastic frontier

Corresponding author:
Stephen Bazen
Aix-Marseille School of Economics
Aix-Marseille University
2 Rue de la Charité
13002 Marseille
France

E-mail: stephen.bazen@univ-amu.fr

* Earlier versions of this paper were presented at the Applied Microeconomics Seminar, University of California, Irvine, April 2015, the Institute for Advanced Labour Studies, University of Amsterdam, June 2015, the European Workshop on Efficiency and Productivity Analysis (EWEPA), Helsinki, June 2015 and the Banque de France-Aix-Marseille School of Economics Conference on Labour Markets, Paris, December 2015. We are grateful to participants for helpful comments.

It is no longer the norm for an individual leaving full-time education in France to obtain a job on a standard employment contract. Typically a young person will spend a number of years moving between fixed-term contracts, special employment measures and, in some cases, unemployment[1]. However young persons do not remain young, and at some point they will usually come to occupy a stable job with a standard employment contract[2]. There are several descriptive studies of this process of assimilation (for example, Mazari and Recotillet, 2013) and concern among French policy-makers has led to the implementation of regular cohort surveys in which a group of individuals leaving full-time education in France in a given year are interviewed at three or more later dates. The so-called Generation Survey interviews the same cohort of individuals at three, five, seven and ten years after leaving education. The outcomes are documented in a number of articles produced by the CEREQ, the organisation that undertakes the survey. The current paper uses this survey for the 1998 cohort – a representative sample of those leaving full-time education in France in that year – to examine their assimilation into the labour market from a new angle.

We use the concept of a stochastic earnings frontier, originally used in the analysis of production efficiency (Aigner, Lovell and Schmidt, 1977). It was first used in a labour economics context by Hofler and Polachek (1985) in order to examine the extent to which young persons attain the earnings potential of their human capital investment in the process of assimilation in the labour market. The estimated gap between the earnings which are actually observed and potential earnings[3] for a given human capital endowment, can be conceived as "earnings inefficiency". The relevance of such an approach in a labour market context can be motivated with reference to a sequential job search model, in which an individual accepts the first job proposition for which the offered wage is not less than their reservation wage. This suggests that earnings could be lower than the maximum attainable earnings corresponding to the individual's human capital, and that higher earnings are possible if search is continued on-the-job.

This paper examines the extent to which initial employment is inefficient in the sense that the pay is below the potential earnings of an individual, where the latter is determined by their human capital (defined in terms of education, other forms of training and labour market experience). Such inefficiency can be viewed as the consequence of a bad job match. Using the longitudinal dimension of the survey we also assess the extent to which this earnings

[1] The unemployment rate for under 25s has risen from 10% in 1980 to 25% in 2015.

[2] For the cohort that left full-time education in 1998, for example, 23% of those employed were on fixed-term contracts three years later in 2001, and 14% seven years after entering the labour market (Mazari and Recotillet, 2013).

[3] The term "potential earnings" is used interchangeably with "maximum attainable earnings" and in later sections "frontier earnings".

inefficiency diminishes over time. Our results suggest that young workers manage to obtain on average about 82% of their potential earnings three years after leaving full-time education, and there is no evidence of earnings inefficiency five years later. Factors that give rise to earnings inefficiency include: expressing a desire for stability, being geographically immobile, living outside of an urban area, or being responsible for dependent younger children. Young female workers have greater earnings inefficiency. Our results are robust and in line with job search theories.

We begin by describing the transition from education to work in France in institutional and statistical terms in section 1, in order to place the subsequent analysis in context. We then set out the basic stochastic frontier model and its uses in labour market analysis. In section 3 we present the stochastic earnings frontier model and discuss specification issues such as the possibility of selectivity bias due to the exclusion of the unemployed. The results are presented in section 4, where we assess the efficiency of the process of assimilation of young persons into the labour market. The final section draws together the main conclusions.

1 The transition from education to work in France

(a) General trends

There are several features of the education system and labour market in France that need to be taken into account when examining the decisions made by young persons after the age of eighteen. High and persistent unemployment has made direct access to standard forms of employment difficult for young persons in France. Lack of work experience and the extension of higher education have meant that even a university level qualification is not a guarantee of a permanent employment contract at the time of leaving full-time education. Faced with the prospect of finding employment in difficult labour market conditions, many young persons remain in full-time education after high school, partly to avoid entering unemployment directly but also because the extra education enhances their job prospects. This is possible because in France possession of the baccalaureat (high school diploma) entitles a young person to go to university (where fees are minimal – currently less than 500 euros a year) and where grant-based financial support is available to those from households with incomes below a certain threshold. As a consequence there was a decline in the labour force participation rate for those aged 15 to 24 from 54% in 1982 to around 37% in 1994 after which it stabilised.

(b) The labour market assimilation of 1998 cohort

In the rest of this paper, we will use the CEREQ Generation 1998 longitudinal survey to examine the assimilation of young workers in the French labour market. This particular cohort was selected since the assimilation process was not affected by the 2008 financial crisis and the subsequent economic slowdown. The Generation survey is a sample selected initially

from records provided by different education and training institutions. Participants are interviewed by telephone on their education, personal background, employment history since leaving full-time education (or since the previous survey – this cohort was interviewed in 2001, 2003, 2005 and 2008), current situation and personal judgements in terms of satisfaction and experience of discrimination. The first interviews took place in 2001 and 55,345 individuals responded. This sample is representative of the 600,000 or so individuals leaving full-time education in 1998. Subsequent waves were also carried out by telephone survey, but these are necessarily subject to attrition. We use the retrospective employment history provided in the 2001 survey to establish labour market status twelve months after leaving full-time education, and current status at the time of interview for 2001 and 2005.

The macroeconomic context for the labour market assimilation of this cohort is one of unemployment falling from a high of 10.7% in 1997 to a low of 7.4% in 2001, before rising slightly to stabilise at around 8.5% in the mid 2000s. Employment in general also increased over the same period, from around 61% to 64% of the population aged 16 to 64, in spite of a slowdown in economic growth. An overall view of the assimilation of the 1998 cohort of young persons into the labour market is provided by the Generation survey and is presented in Figure 1. Compulsory national service for young men was phased out at around this time (see Granier, Joseph and Joutard, 2011 for details) and so some 7% of the 1998 cohort are in this status twelve months after leaving full-time education. Around 13% are unemployed after twelve months, although some of them will have worked and then been made unemployed.

By 2001, three years after leaving the education system, well over half of the cohort were in stable employment with a standard labour contract, and by 2005 this figure had risen to 73%. The role of fixed term contracts had diminished: 19% in 2001 and 9% four years later. By 2005, hardly any of the cohort were beneficiaries of special employment measures, in part due to the fact that many of these applied solely to those aged under 25 and but also because the scope of the measures had been greatly reduced in general, due to policy changes. Over the period 2001 to 2005, unemployment in the cohort falls and stabilises at 8%, and just under 5% declare that they are out of the labour force.

This overall picture of labour market assimilation is fairly reassuring from a policy-maker's point of view. Young persons may have a difficult time initially but over time their labour market experience converges to a normal working life, within a regulated labour market affording a high degree of protection to incumbent employees. However, there are major differences by education level. While all members of the cohort left full-time education in the same calendar year, there is substantial heterogeneity in the education level obtained. In France, number of years of schooling, or age at the time of finishing education, is not a very useful measure of educational attainment, since around half of the cohort retake one or more years of study in order to attain a certain level or diploma. There are basically four relevant

levels of attainment: basic secondary ending at the age of 16 with or without a diploma, some kind of vocational qualification (such as hairdresser or butcher), the baccalaureat and some form of higher education. In view of the increasing numbers entering higher education in the last thirty years, the latter is sub-divided into technical qualifications obtained after a two year programme, a bachelor's degree and postgraduate study. In the 1998 cohort, two thirds had a baccalaureat or higher, and 44% had completed some form of higher education, nearly half of whom had two years post-baccalaureat education in more vocational and often technical subjects. One in seven had only basic secondary education, and as will be seen below, this group have a particularly difficult time in the French labour market.

Figures 2 to 4 provide a more detailed picture of labour market assimilation in terms of education level. Twelve months after leaving education, it is immediately clear that those with a post-baccalaureat, higher education qualification have a high rate of employment (an unemployment rate of 10%), with more than 45% in jobs with standard employment contracts. Between 17% and 22% are in temporary jobs and around 10% are doing national service. At the other end of the education spectrum, for those with only secondary education (14% of the cohort), the rates of unemployment, temporary work and standard employment are almost the same (around 25% in each case).

The main features of this snapshot at twelve months after leaving education are present in the subsequent trends for 2001 and 2005. By 2001, around two thirds of the higher education groups were in 'permanent' employment, and around 15% on fixed term contracts. Less than 6% were unemployed. This contrasts with those at the lower end: a third in permanent employment, 25% on fixed term contracts and 20% unemployed. By 2005, seven years after having left full-time education, assimilation into 'normal' employment was largely complete (80% or more) for high education groups: 4% were unemployed and 6% on temporary contracts. At the extreme, for those with secondary education only, 20% were unemployed and 15% were on fixed term contracts. Only half had permanent jobs. The degree of assimilation into normal employment for those with a baccalaureat or secondary vocational qualifications was quite high: 8% were unemployed and around 10% in fixed term jobs. Around two thirds were on standard employment contracts.

(c) Transitions between labour market states

These trends mean that there are a larger number of transitions taking place between the different labour market statuses. These will be in part the consequence of prospection for a better job match, but given the nature of the French labour market for young persons, involuntary transitions will occur for those coming to the end of a fixed term contract and for beneficiaries of special employment measures, which are also of fixed duration. A full transition matrix of those employed in some way (permanent, fixed term, special measures) after twelve months after leaving education and their status after 3 years is given in Table 1.

While there is a certain degree of immobility with a majority having the same status in 2001 as in 1999, there is a substantial amount of movement towards jobs with standard labour contracts. Well over a third of those on non-standard contracts had moved to a more permanent employment status. Downward movement is very limited, and not inconsistent with the normal functioning of the labour market.

Turning to the unemployed (Table 2), two thirds of those unemployed in 1999 had moved into some form of employment by 2001, and under a third were unemployed. Again these individuals may have had left unemployment for some time between the two dates. Of those unemployed in 2001, a quarter were unemployed in 2005, and more than 40% had moved into permanent employment. One in six was either in a fixed term job or on a special measure. This is less reassuring from a policy-maker's point of view as a small minority of the cohort would appear to be spending their lives moving between short-term employment contracts and unemployment.

(d) Satisfaction with earnings and over-education

The general picture that emerges is one in which young persons do not all move directly from education into permanent jobs. The process of labour market assimilation involves, to varying degrees, transitions between fixed-term contracts and possibly unemployment. After seven years this process appears to have stabilised and only those with basic secondary education have a non negligible chance of not being in stable employment with a standard labour contract. A further dimension of the process of labour market assimilation is the quality of job matches, in terms of whether individuals with a certain endowment of human capital work in a capacity that corresponds to their skills and competences.

In the Generation survey respondents are asked a series of questions which elicit subjective views and declarations on their labour market experience and current situation. These questions are addressed to individuals in employment, and the answers for 2001 and 2005 are presented in Table 3. The responses in general indicate that those in the cohort who are in employment are satisfied with their job, earnings and job match and the proportion expressing satisfaction increases slightly between 2001 and 2005. However, around a quarter state that they are looking for another job in 2001 (60% of whom consider their current job unsatisfactory) and this falls to 17% in 2005. In both years only two thirds consider that they are working in a job that corresponds to their qualifications: again a significant minority (around 30%) consider that they are not in a good job match. In terms of remuneration, only a small proportion (4%) consider that they are well paid. Over half reply that they are quite well paid, and this rises slightly between 2001 and 2005.

However, a third feel that they are underpaid and 5% say they are very poorly paid. While there is a tendency to give emotive answers on this issue – there is no reference pay level

given, or indication as to whether the respondent feels that they could reasonably expect to obtain higher earnings – the pay dimension of the job match seems to be the least satisfactory of those elicited. In order to pursue this, we next examine the extent to which earnings are consistent with individuals' qualifications and experience using a stochastic frontier approach.

2. The use of stochastic frontiers in labour market analysis

The stochastic frontier approach has been applied to earnings in various labour market contexts. The notion of a frontier corresponds to the maximum potential earnings for a given stock of human capital. Divergences from potential earnings occur because of bad job matches, due to imperfect information or immobility, as a consequence of possible discrimination or due to compensating factors such as working conditions. These divergences correspond to the inefficiency term in the production literature.

(i) The stochastic frontier model

The stochastic frontier model was developed by Aigner, Lovell and Schmidt (1977) to analyse production efficiency for a sample of firms whose output (y_i) is defined as a function of a vector of inputs (x_i) and two error components :

$$y_i = g(x_i ; \beta) + v_i - u_i \qquad (1)$$

The first error term (v_i) is the standard random disturbance present in all econometric models picking up measurement errors, unobserved components and shocks. The second error term (u_i) is what makes the model special as an econometric specification. It captures the distance from the frontier due to inefficiency, notably in the management of resources. Since output for a given combination of inputs cannot be higher than what is technically feasible, as defined by the production function, $g(x_i ; \beta)$, the second error term is necessarily negative and is interpreted as technical inefficiency in production. It is conventional to specify the error term as being non-negative and preceded by a minus sign. The parameters of the production function β are usually, but not necessarily, estimated by maximum likelihood. The presence of two error terms, one of which is non-negative, requires two distributional assumptions to be made in order to specify the likelihood function. For example, if $v_i \sim N(0, \sigma_v^2)$ and $u_i \sim \left| N(\mu_i, \sigma_u^2) \right|$ (the truncated-normal distribution), then the density of the error term $\varepsilon_i \equiv v_i - u_i$ is skew-normal (see Azzalini, 2005) and given by:

$$f(\varepsilon_i) = \frac{\frac{1}{\sigma}\phi\left(\frac{\varepsilon_i + \mu_i}{\sigma}\right) \times \Phi\left(\frac{\mu_i^*}{\sigma_*}\right)}{\Phi\left(\frac{\mu_i}{\sigma_u}\right)} \qquad (2)$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\mu_i^* = \frac{1}{\sigma^2}\left(\mu_i\sigma_v^2 + \varepsilon_i\sigma_u^2\right)$ and $\sigma_* = \frac{\sigma_v^2\sigma_u^2}{\sigma^2}$. $\phi(.)$ and $\Phi(.)$ are the density and cumulative distribution functions of the standard normal distribution, respectively. In fact, the inefficiency error term, u_i, is positively skewed and this implies that the combined error term, ε_i, will be negatively skewed. An additional feature of the stochastic frontier approach and particularly useful in a labour economics context is that the determinants of inefficiency (the vector z_i) can be modelled. This involves setting $\mu_i = \delta' z_i$ in the formula for the density function (2).

On the basis of the parameter estimates ($\hat{\beta}$, $\hat{\delta}$ and estimates of the variances of the two error terms) and the distributional assumptions made, the extent of inefficiency is estimated using a formula for the conditional expectation of the inefficiency error term, given the component $\varepsilon_i \equiv v_i - u_i$, where the latter is estimated as $y_i - g\left(x_i;\hat{\beta}\right)$. An estimate of inefficiency for each observation *i* in the skew-normal case, can be obtained by replacing the parameters in the following formula (see Kumbhakar et al., 2015, p.177) with their estimated values:

$$E(u_i|\varepsilon_i) = \mu_i^* + \sigma_* \frac{\phi\left(\frac{\mu_i^*}{\sigma_*}\right)}{\Phi\left(\frac{\mu_i^*}{\sigma_*}\right)} \qquad (3a)$$

Efficiency can be calculates using:

$$EFF_i = E\left(exp(-u_i)|\varepsilon_i\right) = exp\left(-\mu_i^* + \frac{1}{2}\sigma_*^2\right)\frac{\Phi\left(\frac{\mu_i^*}{\sigma_*}\right)}{\Phi\left(\frac{\mu_i^*}{\sigma_*} - \sigma_*\right)} \qquad (3b)$$

The parameters are estimated by maximum likelihood. The parameters in the vector β are used to obtain the marginal effects of the variables that define the frontier and the $\delta's$ determine the marginal effect of the z variables on the extent of inefficiency which are given by:

$$\frac{\partial E(u_i)}{\partial z_{ki}} = \delta_k\left[1 - \tilde{\mu}_i\frac{\phi(\tilde{\mu}_i)}{\Phi(\tilde{\mu}_i)} - \left(\frac{\phi(\tilde{\mu}_i)}{\Phi(\tilde{\mu}_i)}\right)^2\right] \qquad \text{where } \tilde{\mu}_i = \frac{\delta' z_i}{\sigma_u} \quad (4)$$

(see Wang, 2002, and Kumbhakar et al., 2015).

(ii) Stochastic earnings frontiers

To the best of our knowledge, the first published[4] application of the method of stochastic frontiers in a labour economics context was by Hofler and Polachek (1985) in an attempt to measure wage ignorance. Their approach is a direct application of the stochastic frontier (SF) method in relation to job search. Herzog, Hofler and Schlottman (1985) use the SF model to analyse the effect on earnings of inter-state migration again in the context of limited information about the wage offer distribution for first-time migrants compared to repeat migrants, where the former are likely to have less information than the latter. Other applications of the SF model include the analysis of discrimination (Robinson and Wunnava, 1989; Bishop et al., 2008; Diaz and Rosario Sanchez, 2011; Garcia-Prieto and Gomez-Costilla, 2017), the earnings of immigrants (Daneshvary et al, 1992; Lang, 2005) and over-education (Hofler and Murphy, 1992; Jensen, 2010).

Daneshvary et al (1992) provide an analysis of immigrant assimilation in the United States which would be relevant as a parallel to the case of young persons. Assimilation involves relatively low productivity workers learning about a new labour market. In the case of the US, it is not so much an issue of obtaining employment, but more a question of a good job match. Immigrants cannot fully benefit from their home country-specific human capital, and with experience and training increase their US-specific human capital. The process of assimilation is one whereby earnings potential will increase and as it does, the kind of jobs available will evolve. Individuals seek a job match in a context of imperfect information as in Hofler and Polackek (1985). The difference here is that the acquisition of human capital is part of the process of assimilation. Applying this in the context of school-leavers, it is more the case of persons lacking on-the-job experience (rather than country-specific human capital) who learn about the possibilities available in the labour market. The difference in France (and in certain other countries in Europe) is the existence of different contract types, with one which is very advantageous to incumbent workers and a second which acts as a means for firms to adjust their employment levels in the face of uncertain product demand and lock-in labour contracts.

3. A stochastic earnings frontier model for young persons in France

(a) The baseline specification

The dependent variable is the logarithm of monthly earnings. In order to specify the frontier, potential (log) earnings are determined solely by an individual's general human capital in

[4] An earlier version was presented at a conference in 1982. .

terms of education, apprenticeships, internships and actual labour market experience. The first is defined as a dummy variable for the highest diploma obtained: there are five levels corresponding to those used in the descriptive analysis above. Apprenticeships are a form of on-the-job training combined with some college attendance, and along with having done one or more internship is represented by a dummy. Finally, and in view of the nature of the sample, actual labour market experience is measured as the number of *months occupied* and drawn from the retrospective calendar provided by the respondent during the interview. Unlike the standard Mincer model where human capital may depreciate over time, experience enters as a linear term since it corresponds to the beginning of the career. The frontier thus defines the maximum potential earnings attainable for a given level of human capital. Descriptive statistics are presented in Tables A.1 and A.2 in the appendix.

As the distribution of the non negative error term will be positively skewed, the composite error will be negatively skewed. Before proceeding to estimate stochastic frontier, it is customary to first undertake a skewness test. Schmidt and Lin (1984) and Coelli (1995) have proposed tests for this purpose. Both are based on ordinary least squares residuals obtained for the equation for the dependent variable as if there was a single error term. In the former case the test statistic is :

$$\sqrt{b_1} = \frac{m_3}{\left(m_2\sqrt{m_2}\right)} \qquad (5)$$

where m_i is the *i*th moment of the OLS residual. This statistic has a non-standard distribution and critical values can be obtained from D'Agostino and Pearson (1973). Coelli (1995) shows that $\frac{m_3}{\left(\sqrt{\frac{6}{n}m_2^3}\right)}$, where n is the sample size, asymptotically has a standard normal distribution. Both can be regarded as tests of the existence of a stochastic frontier and in the case of negative skewness the statistic would have to be negative. Both pre-tests are undertaken below.

We estimate stochastic earnings frontier models using the truncated normal distribution ($u_i \sim \left| N\left(\mu_i, \sigma_u^2\right)\right|$, $\mu_i = \delta' z_i$), where the vector z contains systematic determinants of the extent of inefficiency. This distributional assumption is preferred to the half-normal alternative firstly because it is more general and contains the half-normal as a special case, and secondly because it is not possible to model the exogenous determinants of differences in inefficiency with a half-normal distribution, since $u_i \sim \left| N\left(0, \sigma_u^2\right)\right|$ or in other words $\mu = 0$ (although it is possible to model heteroscedasticity $\left(\sigma_{ui}^2\right)$ in terms of observable factors). In what follows, a truncated-normal specification is estimated along with a function for inefficiency in terms of a certain number of explanatory variables.

(b) The treatment of selectivity

In the survey used here, there are significant numbers of individuals who do not have earnings. Since the earnings equation is estimated using data on persons who are employed ($y_i > 0$), those who do not have a job are excluded from the analysis. This exclusion, if not random, could lead to biased estimates of the stochastic earnings frontier equation. In many labour economics applications the method proposed by Heckman (1976, 1979) is used to address sample selection issues. However, the method is inappropriate for nonlinear models (Greene (2010)). Kumbhakar et al. (2009), Greene (2010), Lai et al. (2012) and Lai (2015) have presented different approaches to estimating a stochastic frontier model with sample selection. The main issue involved is the fact that the likelihood function of the stochastic frontier with sample selection is not of closed form. Kumbhakar et al. (2009) assume a selection process which depends on the inefficiency component, and use the quadrature method to evaluate the likelihood function. Greene (2010) uses a simulated maximum likelihood approach to evaluate the likelihood function of the stochastic frontier model with sample selection, where sample selection is assumed to occur as a correlation between the unobservables of the selection model and the two-sided standard random disturbance of the stochastic frontier model. Lai (2015) proposes a model which assumes a similar selection mechanism to that in Greene (2010), but uses the closed skew-normal distribution to derive a *closed* form of the likelihood function of the stochastic frontier model with sample selection. The selection model estimated on the entire sample (working and not working) is:

$$d_i = 1\left(w_i^T \gamma + e_i > 0\right) \qquad (6)$$

The model above is a Probit model where d_i is a dummy for being in employment at the date of the survey. The vector w_i contains individual characteristics which may affect the probability of being in employment at the date of the interview, and includes: being responsible for young children, dummy variables for the highest diploma obtained, and dummy variables representing female workers, being born of immigrant parents, living in couple, past geographic mobility, living in an urban area, the delay to obtaining the higher diploma and having as a main objective obtaining permanent employment. e_i is a normalized error term (with variance equal to one).

y_i is observed only when ($d_i = 1$) where $y_i = x_i^T \beta + \varepsilon_i$, with $\varepsilon_i = v_i - u_i$

Lai (2015) makes the following assumptions in order to derive the likelihood function of the SF with sample selection :

(a) the two symmetric error terms (v_i, e_i) are assumed to be independent of the explanatory variables in the frontier (x_i), and those in the selection model (w_i).
(b) the errors are assumed to follow a bivariate normal distribution :

$$\begin{pmatrix} e_i \\ v_i \end{pmatrix} \sim N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{pmatrix} \right] \qquad (7)$$

where ρ is the correlation coefficient between the two error terms. There is no sample selection bias when $\rho = 0$.
(c) in the subsample for which y_i is observed ($d_i = 1$), the inefficiency component u_i is assumed to be truncated–normal, and independent of the error terms (v_i, e_i).

The likelihood function for the whole sample is then:

$$L = \prod_{i=1}^{n} \left[f\left(\varepsilon_i | d_i = 1\right) Pr(d_i = 1) \right]^d Pr(d_i = 0)^{1-d} \qquad (8)$$

To write the likelihood function, we need to know the conditional probability density function $f\left(\varepsilon_i | d_i\right)$. The latter is derived by from the assumption that $v_i | d_i$ and $u_i | d_i$ are both closed skew-normal (CSN) random vectors. A linear combination of CSN random variables is also a CSN random variable, and so $\varepsilon_i | d_i$ is also a CSN random variable. The properties of CSN distribution allow one to straightforwardly derive the conditional probability density function of $\varepsilon_i | d_i$.

In practice, Lai (2015) uses a two-step estimation method as a means of simplifying the procedure. The probit model of the selection process is fitted in the first step, and the parameters γ are estimated (the results used here are given in Table A.3 in the appendix). In the second step, the remaining parameters are estimated conditional on the estimated γ. Because of the two-step procedure, the variance-covariance matrix of the parameters estimated in the second step need to be adjusted to take account of the variability of $\tilde{\gamma}$. Thus, the Murphy-Topel variance estimator is recommended by Lai (2015) to adjust the asymptotic variance of the estimates from the SF model with sample selection.

4. Results

Using data on those who are in employment, we first undertake skewness tests since if the composite error term is not negatively skewed there can be no stochastic frontier. The tests indicate that the residual is negatively skewed in 2001: (Table 4). The Schmidt and Lin statistic is -0.249 (critical value -0.057) and the Coelli statistic is -19.61 (critical value -1.96).

Therefore an earnings frontier can be estimated for 2001. This is not the case for 2005, where the tests indicate that the residual is positively skewed and this is incompatible with the kind of stochastic frontier approach[5]. We thus proceed to estimate a stochastic earnings frontier model for 2001 only and the parameters are estimated by maximum likelihood (see Table 5).

The coefficient on the selectivity term is not significant ($t = 0.87$) which means that the necessary exclusion of non wage-earners from the model of the stochastic earnings frontier has no effect on the estimated coefficients (which numerically are almost the same in the model with and without correction for selectivity). The coefficient estimates of the determinants of the earnings frontier have predictable sizes and signs, and all are highly significant. The constant term plays an important role in situating the frontier and contains the effect of leaving full-time education with a secondary only education level as this serves as the reference level for the education coefficients. Someone on the frontier in 2001 with a postgraduate diploma will earn 83% more than someone with secondary education other things being equal. For a bachelor's degree and a second year college leaver, the differentials on the frontier are 42% and 32% respectively. Someone leaving school with only basic education will earn 10% less than an individual who has the baccalaureat on the frontier, around 4% less than individual having completed an apprenticeship. Vocational qualifications and having undertaken at least one internship both entail higher earnings than the reference category. An additional month of actual labour market experience raises earnings by an estimated 0.5%, corresponding to an increase of 6.2% a year for fully occupied individuals at this early stage in their careers.

Using the formula above (equation 3b), the average estimated efficiency of earnings is 81.7% (the median is 83.4%). The effect of the factors that determine the gap between actual earnings and frontier earnings is measured in terms of their marginal effect on the extent of inefficiency and calculated using the formula given above (equation 4). The marginal effect is proportional to the coefficient, and has the same sign. The results indicate young women in particular tend to earn substantially less than the maximum attainable earnings on the frontier given their human capital (Table 6). Those having pursued their studies outside the Paris region or living outside of the main urban areas have smaller gaps. The marginal effects of the other influences are small. There is no earnings gap for individuals born of immigrant parents other things being equal. Factors likely to constrain job mobility also play a significant role. Young persons having stability as a priority, being geographically immobile, and having young children tend to earn slightly less than their potential earnings.

A possible source of misspecification concerns the heterogeneity of the diploma obtained. In the estimated model various forms of qualification are aggregated by level, but within each

[5] Polachek and Yoon (1996) have considered the case where there may be two-sided inefficiency.

level there are different areas of specialisation. It is likely that possession of master's degree in finance entails higher earnings in general than a master's degree in early modern history. The previous models are augmented to incorporate in addition to the general level of human capital, nine areas of diploma specialisation in terms of sector (agriculture, industry or tertiary), business or engineering school, humanities, scientific or other form of degree, and teaching. Obviously the effect of these forms of specialisation is not relevant to all diploma levels, and so we have created interaction terms of the five diploma levels and compatible forms of specialisation. The marginal effect of a given level of education on earnings will therefore depend on the type of specialisation, with no particular specialisation being the reference category (along with secondary only education). Different forms of specialisation add to or diminish the effect of the level of human capital reflected in a given level of diploma. The residual tests, as in the baseline case, confirm that in this augmented model that a stochastic frontier approach is valid for 2001 but not 2005 (see Table 4). Furthermore, while there is some weak evidence of selectivity (the coefficient is significant at a 10% level) the difference between the estimated coefficients with and without correction for selectivity is negligible (see Table 7). Average efficiency is slightly higher at 83% (a median of 85%).

It is clear that a non specialist postgraduate diploma corresponds to lower frontier earnings (63% more than secondary education only) than say master's degree in engineering (a premium of 100%) or one obtained at a business school (a premium of 108%). There is no additional premium for postgraduate diplomas or bachelor's degrees in humanities. Scientific degrees at graduate level or bachelors both correspond to higher frontier earnings than equivalent non specialist or humanities qualifications. Someone with a general bachelor's degree will earn 34% more on the frontier than someone with secondary only and this premium is 74% if the degree was obtained from a business school. There is a clear pattern of heterogeneous returns to university qualifications on the earnings frontier for business, scientific and engineering specialisations.

Among the lower education levels, diplomas with a specialisation in agriculture are associated with lower frontier earnings while those with an industrial bent are rewarded in the same way as general non specialist diploma types. There is a key difference with aggregate estimates (Table 5) for vocational qualifications. If the latter are specific to the agriculture or tertiary sectors, then there is a negative premium over someone with secondary only education. The type of education, and not just the level attained, is thus relevant for the determination of frontier earnings.

Obviously educational choices and earnings can be influenced by unobserved heterogeneity. In a cross section, it is impossible to separate the latter from earnings inefficiency. One way in which unobserved heterogeneity can be taken into account is through the use of fixed effects. While these can be incorporated in the stochastic earnings frontier approach in a straightforward manner, the consequence is to prevent the estimation of key parameters of

interest on time-invariant variables such as education and apprenticeships. The only time-varying explanatory variable on the model for the frontier is labour market experience.

5. Conclusions

The assimilation of young persons in the French labour market is a relatively long process which is inversely related to education level. The institutional framework for employment contracts has led to the development of fixed-term contracts and special employment measures, and these are major features of the labour market for young persons. However the young are not forever young: they grow older and the vast majority leave this stratum of the labour market which is characterised by unstable employment and high turnover. Possession of some kind of post-secondary diploma is highly correlated with the movement into a stable job with a standard employment contract. The quality of initial job matches would appear to be unsatisfactory for a significant minority, in particular the correspondence between qualifications and job requirements, and related to this the degree of satisfaction with earnings.

Stochastic frontier models are estimated in which the earnings inefficiency term is a function of various personal and job characteristics. It is found that while the extent of earnings efficiency on average (a measure of job match quality) is around 82% three after leaving the education system, it had disappeared seven years after the cohort had left. Earnings on the frontier are determined not only by education level, but also the specialisation of the diploma obtained. Experience and other forms of training also play a role. Factors influencing the extent of inefficiency are gender and factors that limit an individual's mobility. This means that while informational deficiencies will diminish over time enabling individuals to obtain a potentially better job match, the ability to do so may be hindered by personal preferences (where to live and have a stable employment contract) and constraints (living as a couple and presence of children). The results obtained are robust to selection into employment.

That no earnings frontier appears to exist in 2005, seven years after the cohort had left full-time education, means that these factors are no longer sources of inefficiency and that earnings differences reflect not only productivity related to human capital but also compensating earnings differentials related to the career and personal choices made. The stochastic earnings frontier would therefore seem to be a useful and appropriate tool for modelling the process of labour market assimilation of certain groups (young persons, migrants and the long-term unemployed) where over-education due to inefficient initial job matches occurs. Over time this situation tends to be rectified as job mobility leads to improved matching and less inefficiency.

References

Aigner D., Knox Lovell C. and P. Schmidt, (1977), "Formulation and estimation of stochastic frontier production models", *Journal of Econometrics* 6, 21-37.

Azzalini, A., (2005), "The skew-normal distribution and related multivariate families", *Scandinavian Journal of Statistics*, 32, 159-188.

Belotti, F., Daidone, S., Ilardi, G., Atella, V. (2013), "Stochastic frontier analysis using Stata", *Stata Journal*, 13(4), 719–758.

Bishop J.A., Grodner A., Liu H. and Chiou, J. R. (2007), "Gender earnings differentials in Taiwan: a stochastic frontier approach", *Journal of Asian Economics*, 18, 934-945.

Coelli, Tim. (1995), "Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis", *Journal of productivity analysis*, 6(3), 247-268.

Daneshvary N., Herzog H.W., Hofler R.A., and Schlottmann A.M. (1992), "Job search and immigrant assimilation : an earnings frontier approach", *Review of Economics and Statistics*, vol.74, 482-492.

Dias, M. A. and R Sanchez (2011), "Gender and potential wage in Europe: a stochastic frontier approach", *International Journal of Manpower*, 32, 410-425.

Granier, P., Joseph, O., & Joutard, X. (2011), "Military Service and the Integration of Young Men into the Labor Market Based on their Level of Education", *Revue économique*, 62(4), 651-686.

Greene, W. (2008) "The econometric approach to efficiency analysis", in H. Fried, K. Lovell and S. Schmidt, *The Measurement of Productive Efficiency*, Oxford University Press, Oxford.

Greene, W (2010), "A stochastic frontier model with sample selection", *Journal of Productivity Analysis*, 34, 15-24.

Herzog, H., R. A. Hofler and A. Schlottmann (1985) "Life on the frontier: migrant information, earnings and past mobility, *Review of Economics and Statistics*, 67, 373-382.

Hofler R. A. and K. Murphy (1992) "Underpaid and overworked: measuring the effect of imperfect information on wages", *Economic Inquiry*, 30, 511-529.

Hofler, R. A. and S. Polachek (1985) "A new approach to measuring wage ignorance in labor markets", *Journal of Economics and Business*, 77, 267-276.

Jensen, U., Gartner, H. and S. Rassler (2010), "Estimating German overqualification with stochastic earnings frontiers" *Advances in Statistical Analysis*, 94, 33-51.

Kumbhakar, S. C., Wang, H., & Horncastle, A. P. (2015), "A practitioner's guide to stochastic frontier analysis using Stata", *Cambridge University Press*.

Lai, H.P. (2015), "Maximum likelihood estimation of the stochastic frontier model with endogenous switching or sample selection", *Journal of Productivity Analysis*, 43, 105-117.

Lang G. (2005), "The difference between wages and wage potentials: earnings disadvantages of immigrants in Germany,"*Journal of Economic Inequality*, 3, 21-42.

Mazari Z. and I. Recotillet (2013) "Generation 2004: des débuts de trajectoire durablement marqués par la crise ? " *Bref du CEREQ*, No. 311.

Polachek, S. and J. Robst (1998) "Employee labor market information: comparing direct world of work measures of workers' knowledge to stochastic frontier estimates", *Labour Economics*, 5(2), 231-242.

Polachek, S. and B. J. Yoon (1987) "A Two-Tiered Earnings Frontier Estimation of Employer and Employee Information in the Labor Market", *Review of Economics and Statistics*, 69(2), 296-302.

Robinson M.D. and P.V. Wunnava (1989), "Measuring direct discrimination in labor markets using a frontier approach: evidence from CPS female earnings data", *Southern Economic Journal 56*, 212-218.

Schmidt, P., and Lin, T. F. (1984), "Simple tests of alternative specifications in stochastic frontier models", *Journal of Econometrics*, 24(3), 349-361.

Wang, H. J. (2002), "Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model", *Journal of Productivity Analysis*, 18(3), 241-253.

Table 1 Transition matrix for those employed in both 1999 and 2001 (percentages)				
	After 3 years			
After 12 months	Permanent	Fixed-term	Special Measures	Total
Permanent	93.9	5.2	0.8	62
Fixed-term	39.5	56.4	4.1	30
Special Measures	25.4	13.4	61.1	8
Total	72	22	6	100

Table 2 Transitions for those unemployed in 1998 and 2001 (percentages)					
	Position in 2001 or 2005				
Unemployed in:	Permanent job	Fixed-term contract	Special Measures	Unemployed	Out of labour force
1998	34	17	7	29	3
2001	41	14	3	26	7

Table 3 Satisfaction with current job (percentages)		
	2001	2005
Looking for another job?	26	17
Current job:		
Satisfied	72	79
Not satisfied	26	21
Current job in line with aspirations:		
Fully	53	39
Reasonably	26	41
In current job in terms of qualifications:		
Good match	66	67
Over-qualified	28	29
Not qualified enough	6	5
Satisfaction with earnings		
Very	4	3
Quite	57	60
Not very	32	32
Not at all	7	5

Table 4 Left skewness test (validity of the stochastic frontier specification)				
Skewness statistics	2001		2005	
	Without specialization	With specialization	Without specialization	With specialization
Schmidt and Lin (1984)	-0.249	-0.184	0.055**	0.090
Coelli (1995)	-19.608	-14.501	2.252**	3.695
Conclusion	Left skewness	Left skewness	Right skewness	Right skewness
All estimates are significant at 1% except those marked ** (significant at 5%), * (significant at 10%) or ns (not significant at 10%)				

Table 5 Stochastic frontier estimates 2001 truncated-normal specification with sample selection correction		
Dependent variable : log earnings	**Correction for sample selection**	**No correction for sample selection**
Post graduate	0.607	0.607
Bachelor's	0.354	0.353
Two year diploma	0.280	0.280
Baccalaureat – High School diploma	0.098	0.098
Vocational diploma	0.026	0.026
Secondary only	Reference	
Experience	0.005	0.005
Internship	0.045	0.045
Apprenticeship	0.044	0.044
Constant	6.888	6.890
Determinants of Inefficiency		
Desire for stability	0.058	0.058
Female	0.264	0.264
Responsible for child under 5	0.042	0.042
Couple	-0.013**	-0.013*
Immigrant origins	-0.011 ns	-0.011 ns
Urban area	-0.066	-0.066
Immobility	0.079	0.079
Delay in school	-0.004**	-0.004**
Education in Ile-de-France	-0.183	-0.184
Constant	0.003 ns	0.003*
sigma_u	0.171	0.171
sigma_v	0.212	0.212
rho	0.014 ns	-
Average Efficiency (Median)	82.09 (83.90)	81.65 (83.42)
Log L	-3764.414	-616.611
Observations	37,087	
All estimates are significant at 1% except those marked ** (significant at 5%), * (significant at 10%) or ns (not significant at 10%)		

Table 6 - Average marginal effects for inefficiency determinants (only statistically significant influences are presented)		
Determinants	No controls for education specialization	With controls for education specialization
Desire for stability	0.033	0.031
Female	0.149	0.141
Responsible for child under 5	0.024	0.024
Couple	-0.007	-0.007
Urban area	-0.037	-0.033
Immobility	0.045	0.041
Delay in school	-0.002	-0.005
Education in Ile-de-France	-0.104	-0.108

Table 7 Stochastic frontier estimates 2001 – truncated-normal specification with and without sample selection correction inclusion of educational specialization		
Dependent variable : log earnings	**Correction for sample selection**	**No correction for sample selection**
Post graduate	0.490	0.486
Humanities and social sciences	0.049 ns	0.049 ns
Business school	0.240	0.240
Sciences and technology	0.138	0.138
Teaching	-0.050 ns	-0.050 ns
Engineering	0.206	0.205
Bachelor's	0.289	0.286
Humanities and social sciences	0.032 ns	0.032 ns
Business school	0.265	0.266
Sciences and technology	0.115	0.114
Teaching	0.088	0.087
Tertiary	0.001 ns	0.001 ns
Two year diploma	0.266	0.263
Agriculture	-0.128	-0.129
Industry	0.004 ns	0.003 ns
Tertiary	0.037	0.037
Baccalaureat – High School diploma	0.100	0.098
Agriculture	-0.069	-0.070
Industry	0.015 ns	0.014 ns
Tertiary	-0.007 ns	-0.008 ns
Vocational diploma	0.095	0.093
Agriculture	-0.119	-0.118
Industry	-0.048 ns	-0.047 ns
Tertiary	-0.094	-0.092
Secondary only	Reference	
No particular specialisation	Reference	
Experience	0.005	0.005
Internship	0.038	0.038
Apprenticeship	0.038	0.037
Constant	6.869	6.886

Determinants of Inefficiency		
Desire for stability	0.059	0.060
Female	0.272	0.269
Responsible for child under 5	0.047	0.046
Couple	-0.014**	-0.013*
Immigrant origins	-0.006 ns	-0.007 ns
Urban area	-0.063	-0.064
Immobility	0.080	0.079
Delay in school	-0.009	-0.009
Education in Ile-de-France	-0.207	-0.208
Constant	-0.026 ns	-0.025 ns
sigma_u	0.171	0.171
sigma_v	0.211	0.211
rho	0.095*	-
Average Efficiency (Median)	83.33 (85.25)	82.93 (84.81)
Log L	-3312.782	-166.414
Observations	37,087	37,087
All estimates are significant at 1% except those marked ** (significant at 5%), * (significant at 10%) or ns (not significant at 10%)		

Figure 1 The assimilation of the 1998 cohort into the labour market 1998-2005

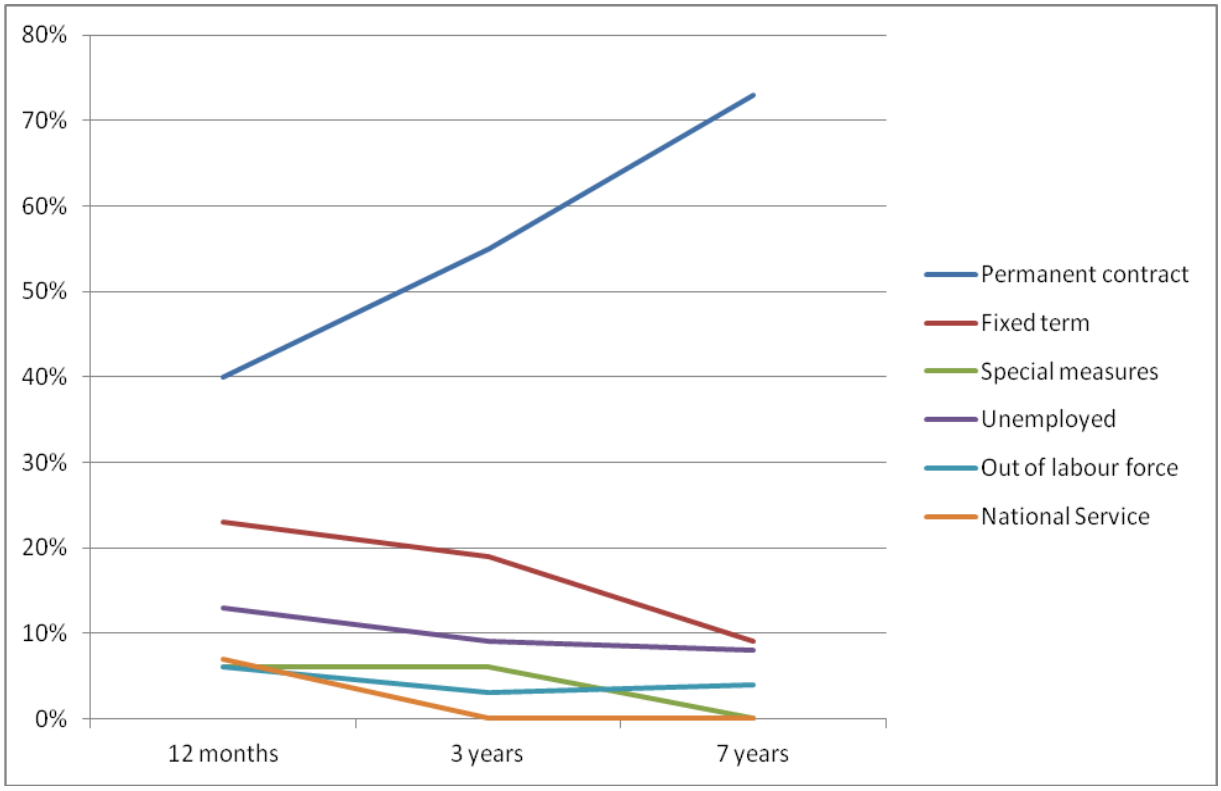


Figure 2 Labour market status of 1998 cohort after 12 months

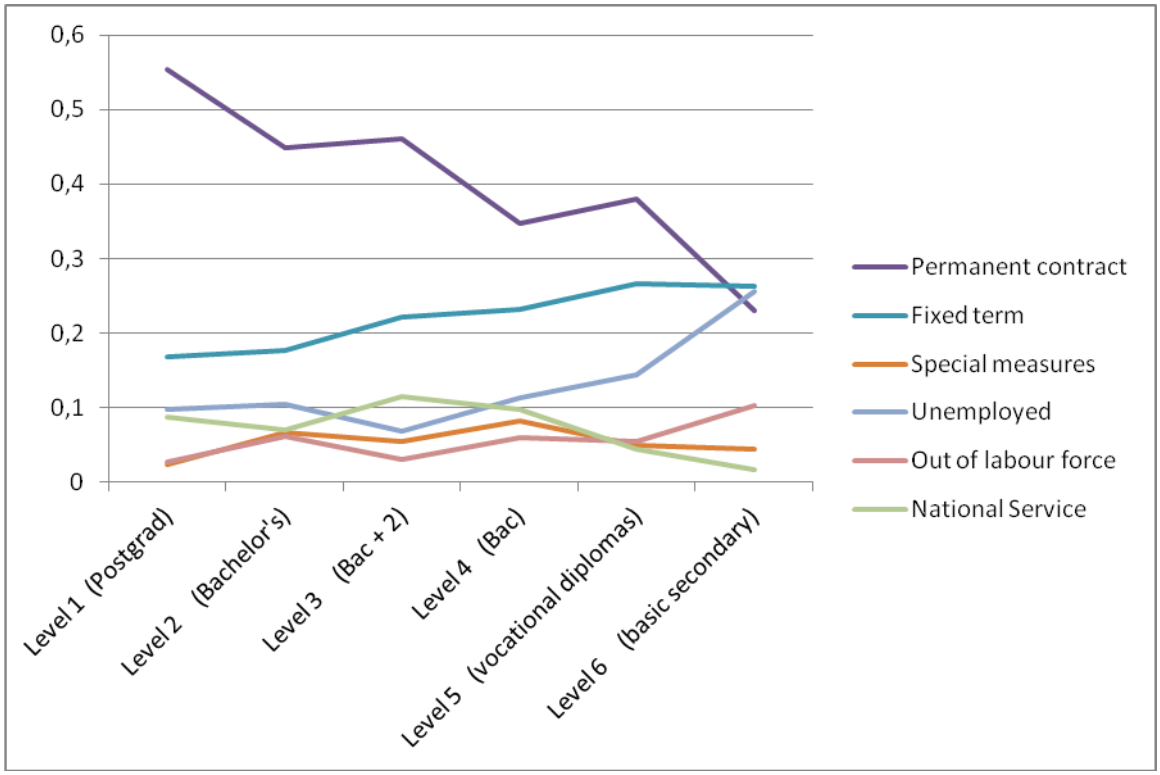


Figure 3 Labour market status of the 1998 cohort after three years

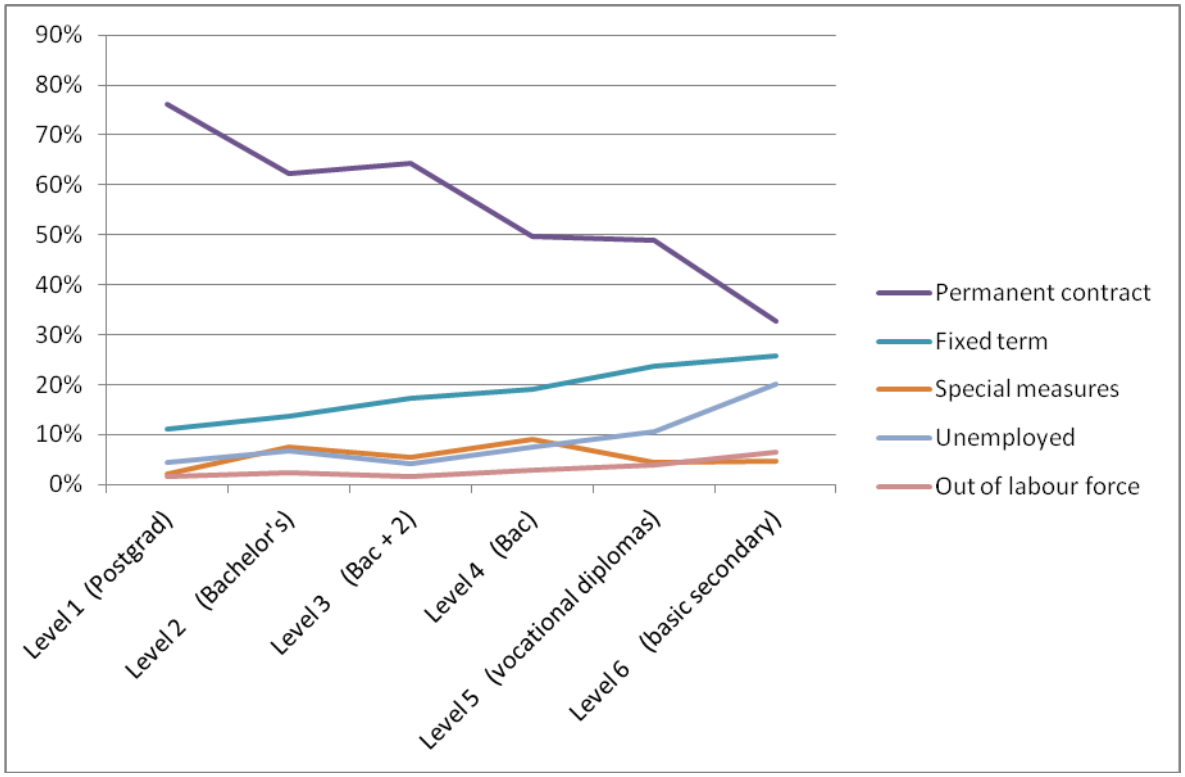
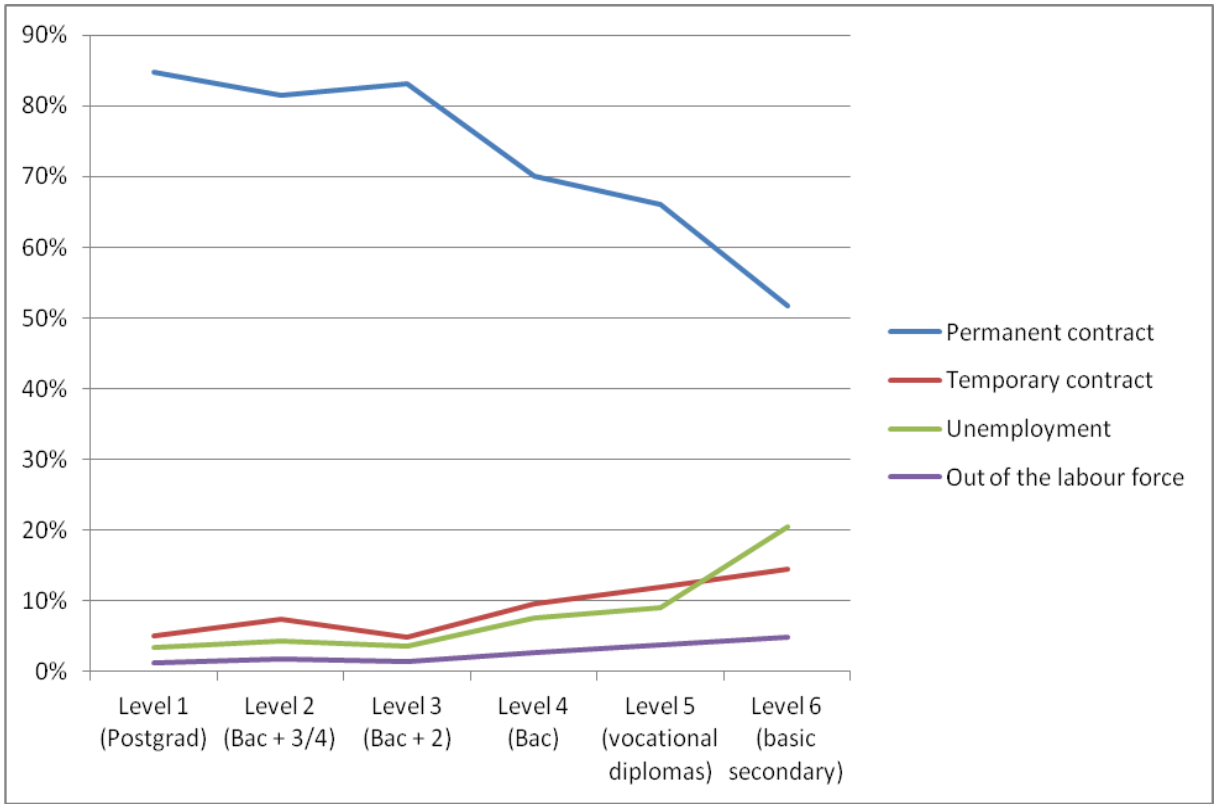


Figure 4 Labour market status of the 1998 cohort after seven years



APPENDIX

Table A.1 Descriptive statistics – whole sample

Variables	2001		2005	
	Mean	**Standard deviation**	**Mean**	**Standard deviation**
Level 1 – post graduate	0.114	0.318	0.087	0.283
Level 2 – bachelor's	0.130	0.337	0.116	0.320
Level 3 – two year diploma	0.217	0.412	0.273	0.446
Baccalaureat – High School diploma	0.203	0.402	0.208	0.406
Level 5 – Vocational diploma	0.201	0.401	0.197	0.398
Level 6 – Secondary only	0.134	0.341	0.118	0.322
Experience	26.714	9.427	76.175	15.818
Internship	0.625	0.484	0.643	0.479
Apprenticeship	0.202	0.401	0.203	0.402
Specialization 1 - Agriculture	0.036	0.187	0.048	0.215
Specialization 2 – Industry	0.280	0.449	0.288	0.453
Specialization 3 - Tertiary [a]	0.339	0.474	0.361	0.480
Specialization 4 – Humanities and social sciences	0.113	0.317	0.088	0.284
Specialization 5 – Business schools	0.013	0.113	0.008	0.090
Specialization 6 – Sciences and technology	0.053	0.225	0.039	0.194
Specialization 7 – Engineering	0.023	0.150	0.024	0.152
Specialization 8 – Teaching	0.032	0.176	0.032	0.177
Specialization 9 - General	0.109	0.312	0.111	0.314
Stability	0.657	0.475	0.651	0.477
Female	0.471	0.499	0.464	0.499
Responsible for child under 5	0.111	0.314	0.400	0.490
Couple	0.366	0.482	0.621	0.485
Immigrant origins	0.092	0.289	0.073	0.261
Work in urban area	0.843	0.364	0.814	0.389
Workers immobility	0.612	0.487	0.722	0.448
Delay in school	1.724	1.601	1.680	1.578
Education in Ile-de-France	0.134	0.340	0.104	0.305
Observations	41,597		10,893	

Table A.2 Descriptive statistics – for employed individuals only

Variables	2001		2005	
	Mean	**Standard deviation**	**Mean**	**Standard deviation**
Level 1 – post graduate	0.121	0.327	0.091	0.288
Level 2 – bachelor's	0.134	0.341	0.121	0.326
Level 3 – two year diploma	0.232	0.422	0.286	0.452
Baccalaureat – High School diploma	0.205	0.404	0.209	0.407
Level 5 – Vocational diploma	0.196	0.397	0.193	0.395
Level 6 – Secondary only	0.111	0.315	0.100	0.300
Experience	28.448	7.551	78.602	11.940
Internship	0.626	0.484	0.642	0.480
Apprenticeship	0.206	0.405	0.206	0.405
Specialization 1 - Agriculture	0.036	0.186	0.048	0.215
Specialization 2 – Industry	0.284	0.451	0.289	0.453
Specialization 3 - Tertiary [a]	0.334	0.472	0.358	0.479
Specialization 4 – Humanities and social sciences	0.115	0.319	0.093	0.290
Specialization 5 – Business schools	0.014	0.117	0.008	0.089
Specialization 6 – Sciences and technology	0.056	0.231	0.040	0.197
Specialization 7 – Engineering	0.025	0.157	0.025	0.155
Specialization 8 – Teaching	0.035	0.184	0.034	0.182
Specialization 9 - General	0.101	0.301	0.104	0.306
Stability	0.652	0.476	0.649	0.477
Female	0.461	0.498	0.459	0.498
Responsible for child under 5	0.105	0.307	0.405	0.491
Couple	0.377	0.485	0.638	0.481
Immigrant origins	0.084	0.277	0.063	0.244
Work in urban area	0.844	0.363	0.813	0.390
Workers immobility	0.597	0.491	0.720	0.449
Delay in school	1.704	1.605	1.671	1.585
Education in Ile-de-France	0.135	0.342	0.104	0.305
Monthly wage	1252.387	430.863	1540.256	527.2447
Observations	31,087		10,028	

Table A.3 Probit estimates – (for sample selection correction)	
Dependent variable : being in employment in 2001	**Coefficients**
Post graduate	0.800*
Bachelor's	0.647***
Two year diploma	0.542***
Baccalaureat – High School diploma	0.278***
Vocational diploma	0.364
Secondary only	Reference
Experience	0.080***
Internship	0.011
Apprenticeship	0.095*
Agriculture - two year diploma	0.284
Agriculture - Baccalaureat	0.284*
Agriculture – Vocational diploma	-0.095
Industry – two year diploma	0.195*
Industry – Baccalaureat	0.278***
Industry – Vocational diploma	-0.067
Tertiary – bachelor's	-0.265
Tertiary – two year diploma	0.167*
Tertiary - Baccalaureat	0.156**
Tertiary Vocational diploma	-0.218
Humanities and social sciences – post graduate	-0.103
Humanities and social sciences – bachelor's	-0.191
Business schools – post graduate	-0.161
Business schools – bachelor's	-0.157
Sciences and technology – post graduate	-0.139
Sciences and technology – bachelor's	0.086
Engineering – postgraduate	0.156
Teaching – post graduate	-0.157
Teaching – bachelor's	0.316
General	Reference
Desire for stability	0.095***
Female	-0.266***
Responsible for child under 5	-0.210***
Couple	0.036
Immigrant origins	-0.101**
Urban area	-0.015

Immobility	0.120***
Delay in school	-0.023**
Education in Ile-de-France	0.018
Constant	-0.867***
Log L	-8978.783
Observations	41,597
*** (significant at 1%), ** (significant at 5%), and * (significant at 10%)	