



**HAL**  
open science

# The integration of young workers into the labour market in France

Stephen Bazen, Khalid Maman Waziri

► **To cite this version:**

Stephen Bazen, Khalid Maman Waziri. The integration of young workers into the labour market in France. *International Journal of Manpower*, 2020, 41 (1), pp.17-36. 10.1108/IJM-07-2018-0204 . hal-02551770

**HAL Id: hal-02551770**

**<https://amu.hal.science/hal-02551770>**

Submitted on 14 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The integration of young workers into the labour market in France

Stephen Bazen and Khalid Maman Waziri

*Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE, Marseille, France*

## Abstract

**Purpose** – Using a representative survey of young persons having left full-time education in France in 1998 and interviewed in 2001 and 2005, the purpose of this paper is to examine the process of their integration into normal employment (a stable job with a standard employment contract) 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. The latter are determined principally by diploma level and educational specialisation, although other forms of training and labour market experience are relevant.

**Design/methodology/approach** – A stochastic earnings frontier approach is used in order to examine workers' ability to capture their full potential earnings in labour markets where there is inefficient job matching (due to the lack of information, discrimination, over-education or the process of integration).

**Findings** – The results suggest that young workers manage to obtain on average about 82 per cent 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.

**Originality/value** – The stochastic earnings frontier is a useful and appropriate tool for modelling the process of labour market integration 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.

**Keywords** Qualifications, Integration, Labour market, Pay, Stochastic frontier, Youth

**Paper type** Research paper

## 1. Introduction

A particularity of the labour market for young people in France is that it is no longer the norm for an individual leaving full-time education 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]. We refer to this status as “normal employment”. There are several descriptive studies of this process of integration (e.g. Mazari and Recotillet, 2013), and concern among French policy-makers has led to the implementation of regular cohort surveys. 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 CEREP (Centre d'études et de recherches sur les qualifications – the French Centre for the Research on Qualifications), the organisation that undertakes the survey. The current paper uses this survey for the 1998 cohort to examine their integration 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 *et al.*, 1977). It was first used in a labour economics context by Hofler and Polachek (1985) in order to examine the extent to which unemployed persons attain the earnings potential of their human capital investment in the process of their

---

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

reintegration 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 result is due to the existence of the costs involved with additional job search and the lack of perfect information.

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 solely by their human capital (defined in terms of education level and specialisation, 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 inefficiency diminishes over time. Our results suggest that young workers manage to obtain on average about 82 per cent 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 appear to experience discrimination as they have greater earnings inefficiency (as in Diaz and Sanchez 2011). 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 2, in order to place the subsequent analysis in context. We then set out the basic stochastic frontier (SF) model and its uses in labour market analysis. In Section 4, 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 5, where we assess the efficiency of the process of integration of young persons into the labour market. The final section draws together the main conclusions.

## **2. The transition from education to work in France**

### *2.1 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. The lack of work experience and the extension of higher education have meant that even a university level qualification does not guarantee that an individual will obtain a permanent employment contract at the time of leaving full-time education. 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 simple possession of the baccalaureat (high school diploma) entitles a young person to go to university (where fees are minimal – currently less than €500 a year) and where grant-based financial support is available to those from households with incomes below a certain threshold.

Although it is true that most young people entering the labour market in France pursue the objective of obtaining a permanent job contract, achieving employment stability is not the end of the story. Anxiety and uncertainty related to their professional career is lower for employed individuals and even lower for those who hold a permanent job contract (Bédoué *et al.*, 2018). The transition towards job stability is long but smoother for those with higher human capital endowment (Couppié *et al.*, 2006), whereas individuals without any educational qualifications do not manage to stabilise in employment even seven years after

entering the labour market. However, the analysis of school-to-work transitions cannot be reduced to the sole focus of achieving job stability. Dupray (2005) showed that even when young people succeed in obtaining stable employment, the beginning of their working life is characterised by an exceptionally larger number of mobilities compared to those of other labour market participants. These external and internal mobilities are undertaken with the aim of improving on an initial unsatisfactory job match that although providing job security does not correspond to the individual's human capital endowment.

## *2.2 The labour market integration of 1998 cohort*

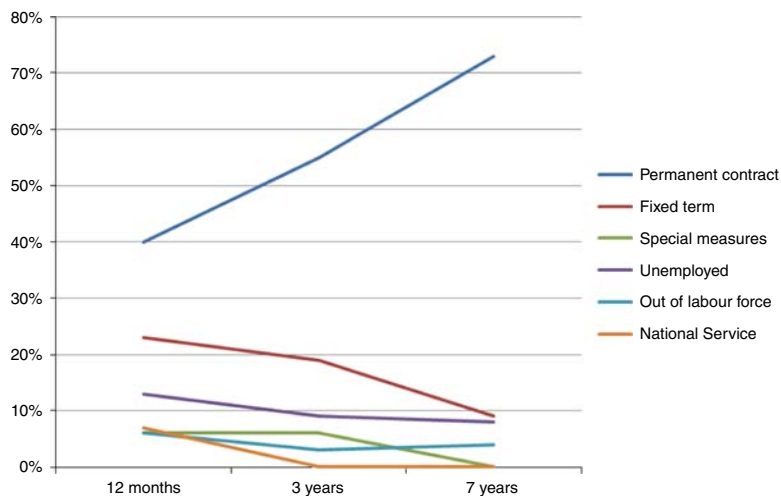
In the rest of this paper, we will use the CEREQ Generation 1998 longitudinal survey to examine the integration of young workers in the French labour market. This particular cohort was selected since the integration process was not affected by the 2008 financial crisis and the subsequent economic slowdown. The generation survey is a sample selected initially from records provided by different education and training institutions. Participants are interviewed on their education, personal background, employment history since leaving full-time education (or since the previous survey), 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. The data are collected by telephone questionnaire. Tracing individuals is, therefore, not straightforward and a high response rate will be unlikely.

For the vast majority the records provided by the education authorities will contain the parents' address and normally their telephone number. When there is no reply to the initial call, up to ten further attempts are made. At the time, the records contained landline numbers and no reply could be due to absence from domicile at the time of the call, for example. A presentation and detailed analysis of the data collection methods are provided in Lopez and Moncel (2006) who reported an overall response rate of 30 per cent varying according to education level between 25 and 33 per cent (Table I, p. 21). Subsequent waves were also undertaken by telephone survey, but these are necessarily subject to attrition. We use the retrospective employment history provided in the 2001 interviews to establish labour market status 12 months after leaving full-time education, and current status at the time of interview for 2001 and 2005.

An overall view of the integration 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 *et al.*, 2011 for details) and so some 7 per cent of the 1998 cohort are in this status 12 months after leaving full-time education. Around 13 per cent 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 per cent. The role of fixed term contracts had diminished: 19 per cent in 2001 and 9 per cent 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 years and but also because the scope of the measures had been greatly reduced in general, due to policy changes. Over the period 2001–2005, unemployment in the cohort falls and stabilises at 8 per cent, and just under 5 per cent declare that they are out of the labour force.

This overall picture of labour market integration 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. Although all members of the cohort left full-time education in the same calendar year, there is substantial heterogeneity in the education level obtained.

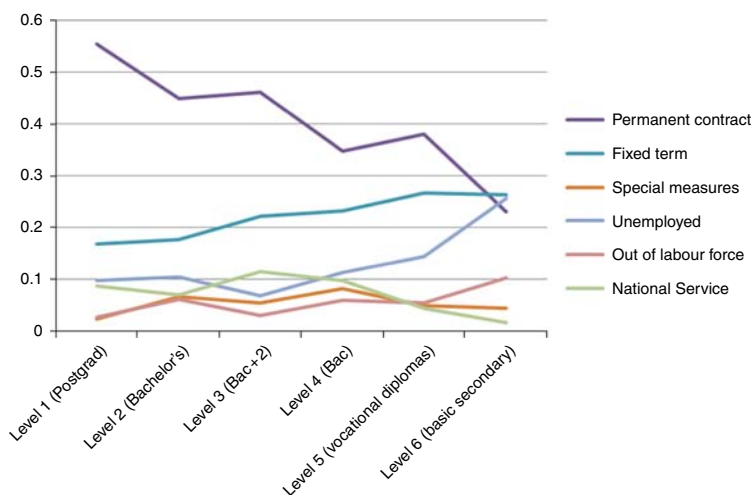
**Figure 1.**  
The integration of  
the 1998 cohort  
into the labour market  
1998–2005



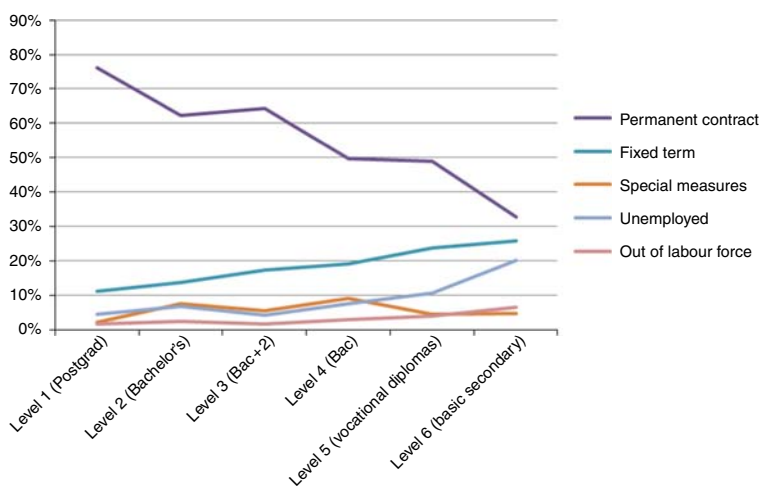
In France, number of years of schooling, or age at the time of finishing education, is not a very useful measure of educational attainment, since around half of the cohort retake one or more years of study in order to attain a certain level or diploma. There are basically four relevant levels of attainment: basic secondary ending at the age of 16 years 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 per cent 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 has a particularly difficult time in the French labour market.

Figures 2–4 provide a more detailed picture of labour market integration in terms of education level. A total of 12 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 per cent), with more than 45 per cent in jobs with standard employment contracts. Between 17 and 22 per cent are in temporary jobs and around 10 per cent are doing national service. At the other end of the education spectrum, for those with only secondary education (14 per cent of the cohort), the rates of unemployment, temporary work and standard employment are almost the same (around 25 per cent in each case).

The main features of this snapshot at 12 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 per cent on fixed term contracts. Less than 6 per cent were unemployed. This contrasts with those at the lower end: a third in permanent employment, 25 per cent on fixed term contracts and 20 per cent unemployed. By 2005, seven years after having left full-time education, integration into “normal” employment was largely complete (80 per cent or more) for high education groups: 4 per cent were unemployed and 6 per cent on temporary contracts. At the extreme, for those with secondary education only, 20 per cent were unemployed and 15 per cent were on fixed term contracts. Only half had permanent jobs. The degree of integration into normal employment for those with a baccalaureat or secondary vocational qualifications was quite high: 8 per cent were unemployed and around 10 per cent in fixed term jobs. Around two thirds were on standard employment contracts.



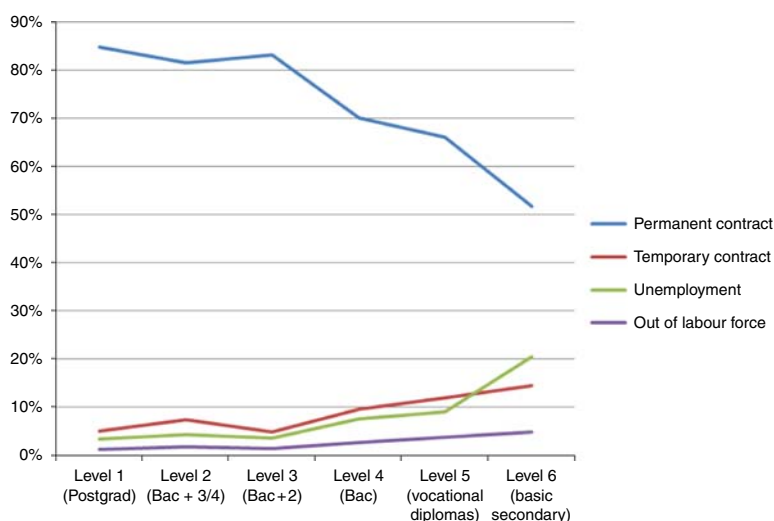
**Figure 2.**  
Labour market status  
of 1998 cohort after  
12 months



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

### 2.3 Transitions between labour market states

These trends mean that there are a large 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) twelve months after leaving education and their status after three years is given in Table I. While there is a certain degree of immobility with a majority having the same status in 2001 as in 1999, 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.



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

**Table I.**  
Transition matrix for those employed in both 1999 and 2001 (percentages)

After 12 months	After 3 years			Total
	Permanent	Fixed-term	Special measures	
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

Turning to the unemployed (Table II), 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 per cent 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.

#### 2.4 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, after seven years this process appears to have stabilized and only those with basic secondary education have a non-negligible chance of not being in stable

**Table II.**  
Transitions for those unemployed in 1999 and 2001 (percentages<sup>a</sup>)

Unemployed in	Position in 2001 or 2005				
	Permanent job	Fixed-term contract	Special measures	Unemployed	Out of labour force
1999	34	17	7	29	3
2001	41	14	3	26	7

**Note:** <sup>a</sup>The percentages do not sum to 100 since very small categories such as self-employed, return to education or national service are excluded

employment with a standard labour contract. A further dimension of the process of labour market integration 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. For example, Sicherman and Galor (1990) found that there is greater job mobility of the more highly educated within occupations, and the rate of mobility decreases over time, suggesting that initial job matches are not efficient. Caroleo and Pastore (2015) provided a survey of the extent of over-education in Italy and elsewhere in Europe, and highlighted the wage penalty associated with being in a job requiring fewer skills than those possessed by the individual. There is also recent evidence in the USA that following the Great Recession (after 2008) nearly 50 per cent of recent college graduates were “underemployed” or in jobs not normally requiring a college education (Abel and Deitz, 2016).

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 III. 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 per cent of whom consider their current job unsatisfactory) and this falls to 17 per cent in 2005. In both years, only two-thirds have considered that they are working in a job that corresponds to their qualifications: again, a significant minority (around 30 per cent) consider that they are not in a good job match. In terms of remuneration, only a small proportion (4 per cent) consider that they are well paid. Over half reply that they are quite well paid, and this rises slightly between 2001 and 2005[4].

However, a third of them feel that they are underpaid and 5 per cent say they are very poorly paid. Although 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 SF approach.

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

**Table III.**  
Satisfaction with  
current job  
(percentages)



### 3. The use of stochastic frontiers in labour market analysis

The SF 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 defined in terms of qualifications and labour market experience. 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.

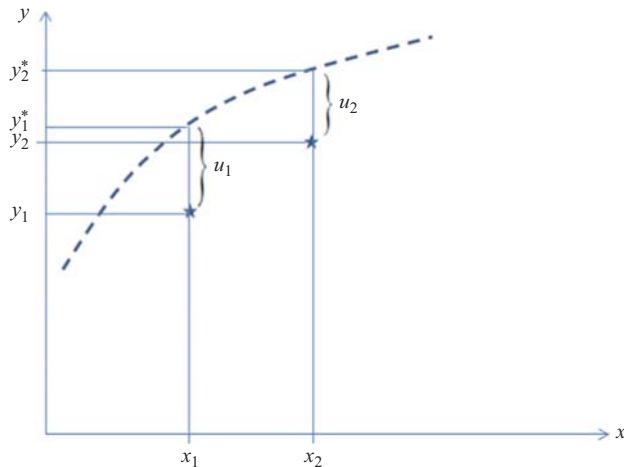
#### 3.1 The stochastic frontier model

The SF model was developed by Aigner *et al.* (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 independently distributed error components:

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

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. This is illustrated in Figure 5. The frontier sets maximum potential output,  $y^*$ , for given inputs  $x$ . The observed value of  $y$  can entail a shortfall,  $u$ , which is, therefore, a measure of inefficiency. Since output for a given combination of inputs cannot be higher than what is technically feasible, as defined by the production function,  $g(x_i; \beta)$ , this 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  $\beta$  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  $u_i \sim |N(\mu_i, \sigma_u^2)|$  (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:

$$f(\varepsilon_i) = \frac{1/\sigma\phi(\varepsilon_i + \mu_i/\sigma) \times \Phi(\mu_i^*/\sigma_*)}{\Phi(\mu_i/\sigma_u)}, \quad (2)$$



**Figure 5.**  
The concept of a stochastic earnings frontier and earnings inefficiency

where  $\sigma^2 = \sigma_v^2 + \sigma_u^2$ ,  $\mu_i^* = 1/\sigma^2(\mu_i\sigma_v^2 + \varepsilon_i\sigma_u^2)$  and  $\sigma_* = \sigma_v^2\sigma_u^2/\sigma^2$ .  $\phi(\cdot)$  and  $\Phi(\cdot)$  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,  $\varepsilon_i$ , will be negatively skewed. An additional feature of the SF 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  $y_i - g(x_i; \hat{\beta})$ . 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:

$$E(u_i|\varepsilon_i) = \mu_i^* + \sigma_* \frac{\phi(\mu_i^*/\sigma_*)}{\Phi(\mu_i^*/\sigma_*)}. \quad (3)$$

As earnings are expressed in logarithms, the term ( $u_i$ ) specified in Equation (1) is the difference between the log of frontier earnings and the log of observed (actual) earnings. Thus,  $\exp(-u)$  corresponds to Earnings Efficiency (EFF) which is the ratio of observed earnings to frontier earnings. A worker  $i$  manages to capture (100 per cent  $\times$  EFF $_i$ ) of the maximum attainable earnings of her demographic group, given her human capital endowment. Efficiency can be estimated using:

$$\text{EFF}_i = E(\exp(-u_i)|\varepsilon_i) = \exp\left(-\mu_i^* + \frac{1}{2}\sigma_*^2\right) \frac{\phi(\mu_i^*/\sigma_* - \sigma_*)}{\Phi(\mu_i^*/\sigma_* - \sigma_*)}.$$

The parameters are estimated by maximum likelihood. The parameters in the vector  $\beta$  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:

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

(see Kumbhakar *et al.*, 2015).

### 3.2 Stochastic earnings frontiers

To the best of our knowledge, the first published[5] application of the method of SFs 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 SF method in relation to job search. Herzog *et al.* (1985) used 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.*, 2007; Diaz and 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 *et al.*, 2010).

Daneshvary *et al.* (1992) provided an analysis of immigrant assimilation in the USA that could 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 USA, 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 Polachek (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.

#### 4. A stochastic earnings frontier model for young persons in France

##### 4.1 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 terms of education, apprenticeships, internships and actual labour market experience. This is the specification that is most widely adopted in the literature as the notion of frontier is concerned with whether the individual attains the potential earnings consistent with what (s)he brings to the labour market. Education is defined as a dummy variable for the highest diploma obtained as years of education are not relevant in the French context as explained above. There are five educational 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[6]. We remove the outliers by excluding workers whose monthly wage is in the first and 99th percentiles of the distribution. Descriptive statistics are presented in Tables AI and AII.

As the distribution of the non-negative error term will be positively skewed, the composite error will be negatively skewed. Before proceeding to estimate the SF, it is customary to first undertake a skewness test. Schmidt and Lin (1984) and Coelli (1995) have proposed tests based on ordinary least squares residuals obtained for the equation for the dependent variable as if there was a single error term. We also use the generalised likelihood ratio (LR) test of inefficiency, suggested by Kumbhakar *et al.* (2015), as an additional check.

We estimate stochastic earnings frontier models using the truncated normal distribution ( $u_i \sim |N(\mu_i, \sigma_u^2)|, \mu_i = \delta'z_i$ ), 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.

##### 4.2 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. Greene (2010) and Lai (2015) have presented different approaches to

estimating an SF model with sample selection. Lai (2015) proposed a model which uses the closed skew-normal distribution to derive a *closed* form of the likelihood function of the SF model with sample selection. The selection model estimated on the entire sample (working and not working) is:

$$d_i = 1(w_i^T \gamma + e_i > 0).$$

The above model 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 that 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 normalised error term (with variance equal to 1).

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

In practice, Lai (2015) used 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  $\gamma$  are estimated (the results used here are given in Table AIII). In the second step, the remaining parameters are estimated conditional on the estimated  $\gamma$ . Because of the two-step procedure, the Murphy-Topel variance estimator is used to adjust the asymptotic variance of the estimates from the SF model with sample selection. Since the estimates are obtained from SF models that are highly nonlinear (Greene, 2010) and estimated using maximum likelihood methods, frontier earnings identification can be achieved by functional form (see Yen, 2005). Nevertheless, our specification of the model is such that there are fewer regressors included in the frontier equation compared to the selection equation.

#### 4.3 The endogeneity of education

The possible endogeneity of education in earnings equations is a key issue in the labour economics literature. The econometric issue is that bias may result from correlation between the error term in the earnings equation – unobserved individual characteristics that give rise to higher earnings – and the individual’s choice of education level. The usual approach to taking on board the possibility that education and unobserved individual factors are correlated is to use instrumental variable methods. The econometric approach adopted here gives rise to a nonlinear specification, which makes the use of the standard IV approach more complicated, although approaches to dealing with endogeneity are beginning to be developed for continuous explanatory variables (see e.g. Amsler *et al.*, 2016 and Karakaplan and Kutlu, 2017). However, it is the ‘multiple treatment’ nature of the specification here that poses the biggest problem since the education variable is a mixture of an ordered and unordered multinomial variable (composed of 24 dummy variables) and would be difficult to handle even in a standard linear or binary probit model using approaches such as those suggested in Heckman *et al.* (2006) and Chesher and Rosen (2017). The composite two error specification in the SF model renders these approaches intractable in our application. Unobserved factors, however, are taken into account in the correction for sample selection.

## 5. Results

Using data on those who are in employment, we first undertake skewness tests since if the composite error term is not negatively skewed the SF is no different from the standard linear single error specification. The tests indicate that the residual is negatively skewed in 2001: (Table IV). The Schmidt and Lin statistic is  $-0.249$  (critical value  $-0.052$ ) and the Coelli test statistic is  $-19.61$  (critical value  $-1.96$ ). Therefore, the earnings frontier with two error

	2001		2005	
	Without specialisation	With specialisation	Without specialisation	With specialisation
<i>Skewness tests</i>				
Schmidt and Lin (1984)	-0.249 (-0.280) <sup>a</sup>	-0.184 (-0.237)	0.055**	0.090
Coelli (1995)	-19.608 (-11.451)	-14.501 (-9.667)	2.252**	3.695
Type of skewness	Negative (Negative)	Negative (Negative)	Positive	Positive
<i>Likelihood ratio tests of inefficiency<sup>b</sup></i>				
No correction for sample selection	3,227.989 (116.064)	2,549.656 (90.122)	0	0
With correction	2,179.206 (246.187)	2,142.113 (350.007)	0	0
<b>Notes:</b> <sup>a</sup> Negative skewness and inefficiency tests for 2001 using only those individuals who are still present in 2005 are given in parentheses; <sup>b</sup> The null hypothesis ( $H_0$ ) is that the variance for the inefficiency component is zero in a half-normal specification (one degree of freedom). Critical value 5.412 at the 1 per cent significance level. All estimates are significant at 1 per cent except those marked ** (significant at 5 per cent), * (significant at 10 per cent) or ns (not significant at 10 per cent)				

**Table IV.**  
Negative skewness  
and inefficiency tests

components is applicable in 2001. This is not the case for 2005, where the tests indicate that the residual is positively skewed and this is incompatible with the usual SF specification[7]. Finally, the LR tests presented in the lower half of Table IV confirm the presence of inefficiency in 2001. The test also indicates that there is no earnings inefficiency in 2005. These conclusions hold even when there is a correction for sample selection. We, thus, concentrate on results for the stochastic earnings frontier model estimated for 2001 only. The parameters are estimated by maximum likelihood and are presented in Table V.

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 only 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 per cent[8] 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 per cent, respectively. Someone leaving school with only basic education will earn 9 per cent less than an individual who has the baccalaureat on the frontier, around 4 per cent 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 per cent, corresponding to an increase of 6.2 per cent a year for fully occupied individuals at this early stage in their careers.

Using the formula above (Equation (3)), the average estimated efficiency of earnings is 81.7 per cent (the median is 83.4 per cent). 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. A positive coefficient means that earnings are further from the frontier level. 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 VI). This suggests that young females face more difficulties in attaining their frontier earnings that can be interpreted as

Dependent variable: log earnings	Correction for sample selection	No correction for sample selection	
Level 1 – post graduate	0.606	0.605	
Level 2 – bachelor's	0.353	0.353	
Level 3 – two year	0.280	0.280	
Level 4 – Baccalaureat – high school diploma	0.098	0.098	
Level 5 – vocational diploma	0.026	0.026	
Level 6 – secondary only		Reference	
Experience	0.005	0.005	
Internship	0.045	0.045	
Apprenticeship	0.044	0.044	
Constant	6.893	6.895	
<i>Determinants of inefficiency</i>			
Stability	0.058	0.058	
Female	0.249	0.249	
Female responsible for child under 5	0.051	0.051	
Couple	-0.013**	-0.013**	
Immigrant origins	-0.010 ns	-0.010 ns	
Work in urban area	-0.065	-0.065	
Workers immobility	0.076	0.076	
Delay in school	-0.004**	-0.004**	
Education in Ile-de-France	-0.178	-0.178	
Constant	0.023 ns	0.023 ns	
sigma_u	0.167	0.167	
sigma_v	0.212	0.212	
$\rho$	0.011 ns	–	
Average efficiency (median)	81.79 (83.51)	81.36 (83.03)	
Log $L$	-3,758.325	-3,758.344	
Observations		37,087	

**Note:** All estimates are significant at 1 per cent except those marked \*(significant at 5 per cent), \*(significant at 10 per cent) or ns (not significant at 10 per cent)

**Table V.**  
Stochastic frontier estimates – truncated-normal (2001) – sample selection correction

Determinants	No controls for education specialisation	With controls for education specialisation	
Stability	0.034	0.032	
Female	0.146	0.138	
Female responsible for child under 5	0.030	0.030	
Couple	-0.007	-0.007	
Work in urban area	-0.038	-0.034	
Workers immobility	0.045	0.042	
Delay in school	-0.002	-0.005	
Education in Ile-de-France	-0.104	-0.109	

**Table VI.**  
Average marginal effects for inefficiency determinants (only statistically significant influences are presented)

evidence of discrimination as argued by Diaz and Sanchez (2011). Those having pursued their studies outside the Paris region or living outside of the main urban areas have larger 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 and being geographically immobile tend to earn slightly less than their potential earnings.

A possible source of misspecification concerns the nature of the diploma obtained. In the estimated model various forms of qualification are aggregated by level, but within each 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 there is earnings inefficiency in 2001 but not 2005 (see Table IV). Furthermore, although there is some weak evidence of selectivity, the difference between the estimated coefficients with and without correction for selectivity is negligible (see Table VII). Average efficiency is slightly higher at 83 per cent (a median of 85 per cent).

It is clear that a non-specialist postgraduate diploma corresponds to lower frontier earnings (63 per cent more than secondary education only) than say master's degree in engineering (a premium of 100 per cent) or one obtained at a business school (a premium of 108 per cent). 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 per cent more on the frontier than someone with secondary only and this premium is 74 per cent 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 V) 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.

As pointed out above, the Generation Survey involves telephone interviews of a cohort of young persons leaving the French educational system. For obvious reasons, it is difficult to keep track of all of those interviewed in the first wave. One major issue encountered with this study is the large attrition rate between 2001 and 2005. If for those who are still present in 2005 there is no evidence of inefficiency in 2001 (i.e. they already had earnings on the frontier in 2001), then any observed improvement in efficiency between 2001 and 2005 would be spurious. To check the robustness of the finding that there has been a reduction in inefficiency between 2001 and 2005, the null hypothesis of that there is no earnings inefficiency in 2001 is tested using data solely for those who are still present in 2005 wave. Both the skewness and the LR tests reject the hypothesis of the absence of inefficiency in 2001 at the 1 per cent significance level (Table IV). Thus, considering only those who are still present in 2005, it is still evident that inefficiency is present in 2001 and has disappeared by 2005.

## **6. Conclusions**

The integration of young persons in the French labour market is a relatively long process which is inversely related to the education level. The institutional framework for employment contracts

Dependent variable: log earnings	Correction for sample selection	No correction for sample selection	
Post graduate	0.489	0.485	
Humanities and social sciences	0.049 ns	0.049 ns	
Business school	0.239	0.240	
Sciences and technology	0.137	0.137	
Teaching	-0.050 ns	-0.050 ns	
Engineering	0.206	0.205	
Bachelor's	0.288	0.285	
Humanities and social sciences	0.032 ns	0.032 ns	
Business school	0.266	0.266	
Sciences and technology	0.115	0.114	
Teaching	0.089	0.088	
Tertiary	0.002 ns	0.003 ns	
Two year diploma	0.266	0.263	
Agriculture	-0.128	-0.128	
Industry	0.004 ns	0.003 ns	
Tertiary	0.037	0.037	
Baccalaureat – High School diploma	0.099	0.097	
Agriculture	-0.068	-0.070	
Industry	0.015 ns	0.014 ns	
Tertiary	-0.007 ns	-0.007 ns	
Vocational diploma	0.096	0.093	
Agriculture	-0.118	-0.118	
Industry	-0.048 ns	-0.047 ns	
Tertiary	-0.094	-0.093	
Secondary only			Reference
No particular specialisation			Reference
Experience	0.005	0.005	
Internship	0.038	0.038	
Apprenticeship	0.038	0.037	
Constant	6.874	6.890	
<i>Determinants of inefficiency</i>			
Desire for stability	0.058	0.058	
Female	0.254	0.252	
Female responsible for child under 5	0.055	0.054	
Couple	-0.013**	-0.013**	
Immigrant origins	-0.006 ns	-0.007 ns	
Urban area	-0.062	-0.062	
Immobility	0.077	0.076	
Delay in school	-0.009	-0.009	
Education in Ile-de-France	-0.199	-0.199	
Constant	-0.003 ns	-0.001 ns	
sigma_u	0.166	0.166	
sigma_v	0.212	0.212	
rho	0.090*	-	
Average Efficiency (Median)	83.05 (84.90)	82.65 (84.44)	
Log L	-3,306.910	-3,308.283	
Observations		37,087	

**Note:** All estimates are significant at 1 per cent except those marked \*\* (significant at 5 per cent), \* (significant at 10 per cent) or ns (not significant at 10 per cent)

**Table VII.**  
Stochastic frontier estimates in 2001 for truncated-normal specification with and without sample selection correction – inclusion of educational specialisation

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 that 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.

SF 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 per cent three years 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.

Seven years after the cohort had left full-time education, these factors are no longer sources of inefficiency and earnings differences reflect not only productivity related to human capital but also compensating earnings differentials related to the career and personal choices made. Over time this situation tends to be rectified as job mobility leads to improved matching and less inefficiency. The contribution of this paper is to show that the stochastic earnings frontier would, therefore, seem to be a useful and appropriate tool for modelling the process of labour market integration young persons in the same way that it has been found for analysing the same kind of process for other groups such as migrants and the long-term unemployed. Inefficient initial job matches are made and these tend to be corrected subsequently through job mobility.

## Notes

1. The unemployment rate for under 25 s has risen from 10 per cent in 1980 to 25 per cent in 2015.
2. For the cohort that left full-time education in 1998, for example, 23 per cent of those employed were on fixed-term contracts three years later in 2001, and 14 per cent 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".
4. The extent of each form of satisfaction is increasing with education level. It increases between 2001 and 2005 for all educational categories except for the highest (postgraduate qualifications) where proportion satisfied is either stable or falls slightly.
5. An earlier version was presented at a conference in 1982.
6. In all of the models presented a squared experience term was added and found to have very small coefficient and in only one case was significant at a 5 per cent significance level.
7. Polachek and Yoon (1996) have considered the case where there may be two-sided inefficiency.
8. This and the other returns are calculated as  $\exp(\beta)-1$  where  $\beta$  is the estimated coefficient.

## References

- Abel, J. and Deitz, R. (2016), "Underemployment in the early careers of college graduates following the great recession", NBER Working Papers No. 22654, National Bureau of Economic Research, Cambridge, MA.
- Aigner, D., Knox Lovell, C. and Schmidt, P. (1977), "Formulation and estimation of stochastic frontier production models", *Journal of Econometrics*, Vol. 6 No. 1, pp. 21-37.

- Amsler, C., Prokhorov, V. and Schmidt, P. (2016), "Endogeneity in stochastic frontier models", *Journal of Econometrics*, Vol. 190 No. 2, pp. 280-288.
- Azzalini, A. (2005), "The skew-normal distribution and related multivariate families", *Scandinavian Journal of Statistics*, Vol. 32 No. 2, pp. 159-188.
- Béduwé, C., Dupray, A. and El Akremi, A. (2018), "Perceived job insecurity in early careers and human capital accumulation", CEREQ Working Paper No. 6, Marseille.
- 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*, Vol. 18 No. 6, pp. 934-945.
- Caroleo, F. and Pastore, F. (2015), "Overeducation: a disease of the school-to-work transition", IZA Discussion Paper No. 9049, Bonn.
- Chesher, A. and Rosen, A. (2017), "Generalised instrumental variable models", *Econometrica*, Vol. 85 No. 3, pp. 959-989.
- Coelli, T. (1995), "Estimators and hypothesis tests for a stochastic frontier function: a Monte Carlo analysis", *Journal of Productivity Analysis*, Vol. 6 No. 3, pp. 247-268.
- Couppié, T., Gasquet, C. and Lopez, A. (2006), "Les sept premières années de vie active de la Génération 98: entre insertion et débuts de carrière", *CEREQ Bref*, No. 234, p. 9.
- Daneshvary, N., Herzog, H.W., Hofler, R.A. and Schlottmann, A. (1992), "Job search and immigrant assimilation: an earnings frontier approach", *Review of Economics and Statistics*, Vol. 74 No. 3, pp. 482-492.
- Diaz, M.A. and Sanchez, R. (2011), "Gender and potential wage in Europe: a stochastic frontier approach", *International Journal of Manpower*, Vol. 32 No. 4, pp. 410-425.
- Dupray, A. (2005), "Les mobilités en début de vie professionnelle. Externes ou internes aux entreprises, des évolutions aux effets très différents", *CEREQ Bref*, No. 216, p. 4.
- Garcia-Prieto, C. and Gomez-Costilla, P. (2017), "Gender wag gap and education: a stochastic frontier approach", *International Journal of Manpower*, Vol. 38 No. 3, pp. 504-516.
- Granier, P., Joseph, O. and Joutard, X. (2011), "Military service and the integration of young men into the labor market based on their level of education", *Revue Economique*, Vol. 62 No. 4, pp. 651-686.
- Greene, W. (2010), "A stochastic frontier model with sample selection", *Journal of Productivity Analysis*, Vol. 34 No. 1, pp. 15-24.
- Heckman, J.J., Urzua, S. and Vytlacil, E. (2006), "Understanding instrumental variables in models with essential heterogeneity", *Review of Economics and Statistics*, Vol. 88 No. 3, pp. 389-432.
- Herzog, H., Hofler, R.A. and Schlottmann, A. (1985), "Life on the frontier: migrant information, earnings and past mobility", *Review of Economics and Statistics*, Vol. 67 No. 3, pp. 373-382.
- Hofler, R.A. and Murphy, K. (1992), "Underpaid and overworked: measuring the effect of imperfect information on wages", *Economic Inquiry*, Vol. 30 No. 3, pp. 511-529.
- Hofler, R.A. and Polachek, S. (1985), "A new approach to measuring wage ignorance in labor markets", *Journal of Economics and Business*, Vol. 77 No. 3, pp. 267-276.
- Jensen, U., Gartner, H. and Rassler, S. (2010), "Estimating German overqualification with stochastic earnings frontiers", *Advances in Statistical Analysis*, Vol. 94 No. 1, pp. 33-51.
- Karakaplan, A. and Kutlu, L. (2017), "Handling endogeneity in stochastic frontier analysis", *Economics Bulletin*, Vol. 37 No. 2, pp. 889-901.
- Kumbhakar, S.C., Wang, H. and Horncastle, A. (2015), *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata*, Cambridge University Press, Cambridge.
- Lai, H.P. (2015), "Maximum likelihood estimation of the stochastic frontier model with endogenous switching or sample selection", *Journal of Productivity Analysis*, Vol. 43 No. 1, pp. 105-117.
- Lang, G. (2005), "The difference between wages and wage potentials: earnings disadvantages of immigrants in Germany", *Journal of Economic Inequality*, Vol. 3 No. 1, pp. 21-42.
- Lopez, A. and Moncel, N. (2006), "Les enquetes 'Generation' du CEREQ", *NetDoc*, No. 25, p. 4, available at: [www.cereq.fr/cereq/netdoc25.pdf](http://www.cereq.fr/cereq/netdoc25.pdf)

- Mazari, Z. and Recotillet, I. (2013), "Generation 2004: des débuts de trajectoire durablement marqués par la crise?", *Bref du CEREQ*, No. 311, p. 4.
- Polachek, S.W. and Yoon, B.B.J. (1996), "Panel estimates of a two-tiered earnings frontier", *Journal of Applied Econometrics*, Vol. 11 No. 2, pp. 169-178.
- Robinson, M.D. and Wunnava, P.V. (1989), "Measuring direct discrimination in labor markets using a frontier approach: evidence from CPS female earnings data", *Southern Economic Journal*, Vol. 56 No. 1, pp. 212-218.
- Schmidt, P. and Lin, T.F. (1984), "Simple tests of alternative specifications in stochastic frontier models", *Journal of Econometrics*, Vol. 24 No. 3, pp. 349-361.
- Sicherman, M. and Galor, O. (1990), "A theory of career mobility", *Journal of Political Economy*, Vol. 98 No. 1, pp. 169-192.
- Yen, S. (2005), "A multivariate sample-selection model: estimating cigarette and Alcohol demands with Zero observations", *American Journal of Agricultural Economics*, Vol. 87 No. 2, pp. 453-466.

### Further reading

- Greene, W. (2008), "The econometric approach to efficiency analysis", in Fried, H., Lovell, K. and Schmidt, S. (Eds), *The Measurement of Productive Efficiency*, Oxford University Press, Oxford.
- Polachek, S.W. and Xiang, J. (2006), "The effects of incomplete employee wage information: a cross-country analysis", *Research in Labor Economics*, Vol. 24, pp. 35-75.

### Appendix

Variables	2001		2005	
	Mean	SD	Mean	SD
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
Specialisation 1 – Agriculture	0.036	0.187	0.048	0.215
Specialisation 2 – Industry	0.280	0.449	0.288	0.453
Specialisation 3 – Tertiary	0.339	0.474	0.361	0.480
Specialisation 4 – Humanities and social sciences	0.113	0.317	0.088	0.284
Specialisation 5 – Business schools	0.013	0.113	0.008	0.090
Specialisation 6 – Sciences and technology	0.053	0.225	0.039	0.194
Specialisation 7 – Engineering	0.023	0.150	0.024	0.152
Specialisation 8 – Teaching	0.032	0.176	0.032	0.177
Specialisation 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 AI.**  
Descriptive statistics –  
whole sample

Variables	2001		2005	
	Mean	SD	Mean	SD
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
Specialisation 1 – Agriculture	0.036	0.186	0.048	0.215
Specialisation 2 – Industry	0.284	0.451	0.289	0.453
Specialisation 3 – Tertiary	0.334	0.472	0.358	0.479
Specialisation 4 – Humanities and social sciences	0.115	0.319	0.093	0.290
Specialisation 5 – Business schools	0.014	0.117	0.008	0.089
Specialisation 6 – Sciences and technology	0.056	0.231	0.040	0.197
Specialisation 7 – Engineering	0.025	0.157	0.025	0.155
Specialisation 8 – Teaching	0.035	0.184	0.034	0.182
Specialisation 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 earnings	1,252.387	430.863	1,540.256	527.2447
Observations	31,087		10,028	

**Table AII.**  
Descriptive statistics –  
for employed  
individuals only

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 ns
Secondary only	Reference
Experience	0.080
Internship	0.011 ns
Apprenticeship	0.095*
Agriculture – two year diploma	0.284 ns
Agriculture – Baccalaureat	0.284*
Agriculture – Vocational diploma	–0.095 ns
Industry – two year diploma	0.195*
Industry – Baccalaureat	0.278
Industry – Vocational diploma	–0.067 ns
Tertiary – bachelor's	–0.265 ns
Tertiary – two year diploma	0.167*
Tertiary – Baccalaureat	0.156**
Tertiary Vocational diploma	–0.218 ns
Humanities and social sciences – post graduate	–0.103 ns
Humanities and social sciences – bachelor's	–0.191 ns
Business schools – post graduate	–0.161 ns
Business schools – bachelor's	–0.157 ns
Sciences and technology – post graduate	–0.139 ns
Sciences and technology – bachelor's	0.086 ns
Engineering – postgraduate	0.156 ns
Teaching – post graduate	–0.157 ns
Teaching – bachelor's	0.316 ns
General	Reference
Desire for stability	0.095
Female	–0.266
Responsible for child under 5	–0.210
Couple	0.036 ns
Immigrant origins	–0.101**
Urban area	–0.015 ns
Immobility	0.120
Delay in school	–0.023**
Education in Ile-de-France	0.018 ns
Constant	–0.867
Log L	–8,978.783
Observations	41,597

**Note:** All coefficients significant at 1 per cent except, \*\* (significant at 5 per cent), \* (significant at 10 per cent) and ns (not significant)

**Table AIII.**  
Probit estimates – (for  
sample selection  
correction)

### Corresponding author

Stephen Bazen can be contacted at: Stephen.BAZEN@univ-amu.fr