University of Caen

University of Rennes 1

Working Paper

# Mismatch, On-the-job Training, and Unemployment

**Frédéric Gavrel**

**Jean-Pascal Guironnet**

**Isabelle Lebon**

*University of Caen Basse-Normandie, CREM-CNRS, UMR 6211*

UNICAEN
**université** de **Caen**
Basse-Normandie

cnrs

UNIVERSITÉ DE
**RENNES 1**

# Mismatch, On-the-job Training, and Unemployment

Frédéric Gavrel[*]   Jean-Pascal Guironnet[†]  Isabelle Lebon [‡] [§]

June 2011

### Abstract

In this paper, training, which is seen as a way to reduce the mismatch between workers and jobs, takes place on the job. We show that a general rise in unemployment lowers the probability of on-the-job training by reducing the mismatch. We then close the model by assuming free-entry and study its social efficiency properties. Private educational choices are socially optimal, but job creation is too high under the Hosios condition. Using French data on regional unemployment, we estimate a *probit* model of the training decision and find that on-the-job training is significantly less probable in regions with high unemployment.

**Keywords:** On-the-job training; Mismatch; Equilibrium unemployment; Market efficiency.

**JEL Codes:** H21, J24, J64.

[*]University of Caen, CREM-CNRS, Department of Economics, 19 rue Claude Bloch, BP 5186, 14032 Caen cedex, France; frederic.gavrel@unicaen.fr

[†]University of Caen, CREM-CNRS, Department of Economics, 19 rue Claude Bloch, BP 5186, 14032 Caen cedex, France; jean-pascal.guironnet@unicaen.fr

[‡]University of Caen, CREM-CNRS, Department of Economics, 19 rue Claude Bloch, BP 5186, 14032 Caen cedex, France; isabelle.lebon@unicaen.fr

1

# 1 Introduction

The purpose of this paper is to study the on-the-job training decision. On-the-job training is seen as the means of reducing the mismatch between jobs and workers. This view of on-the-job training has one main implication: As a market tightness increase makes firms have to recruit less suited workers, an unemployment decrease should be associated with an increase in the share of workers who benefit from on-the-job training. We first build a theoretical model to assess the consistency of this argument as well as its welfare implications. Then, we proceed to an empirical test using French data.

Our analytical framework is a matching model with Nash bargaining and free-entry where the mismatch of talents is modelled along the same line as Salop (1979) (see Marimon and Zilibotti (1999)). Workers and jobs lie on a circle, and the distance between two points on this circle measures the mismatch between a job and a worker. Our theoretical contribution has two main innovative features. First, on-the-job training is an optimizing variable. For the sake of simplicity as well as to better fit our data, the educational choice is binary. In line with intuition, we assume that on-the-job training is all the more efficient as the quality of the match is low. As a consequence, there is a mismatch trigger beyond which workers benefit from on-the-job training. Second, in contrast with all circular models (to our knowledge), firms' search is not random. The technology of contacts is described by an extension of the urn-ball model where firms rank their applicants and pick the best one suited. Applicant ranking plays a crucial role in our analysis as it creates a direct relationship between on-the-job training and overall labor market conditions. In the first part of the paper, we state that a market tightness increase, which is associated with a general decrease in unemployment, raises the probability of training by lowering the quality of the matches. Then, we close the model and study market efficiency.

We state two main results. First, private training choices are socially optimal even when wage bargaining is static. This efficiency result comes from the fact that on-the-job training only moves the workers along the circle but does not affect the level of their ability. As a consequence, there is not any holdup phenomenon. Second, assuming that firms internalize the usual congestion effect, job creation appears to be too high. The reason for this inefficiency is that firms do not take into account that in the presence of applicant ranking, an increase in the tightness of the labor market tends to deteriorate the quality of the matches, raising then the training expenditures. In the last part of the paper, we confront our positive results with empirical evidence. Using fine French data, we estimate a probit model of the on-the-job training decision. The unemployment rate is used as a proxy for market tightness. In accordance with the prediction of the theoretical model, all other things being equal, the probability of being trained is significantly lower in the regions where the unemployment rate is

high.

To our knowledge, our view of on-the-job training is new. So the relations between our contribution and labor theory are essentially methodological. Many papers use the circular model to account for two-sided horizontal heterogeneity; but in those papers, search is random. Conversely, two authors introduce an urn-ball process with applicant ranking in a matching model of the labor market (Moen (1999), Gavrel (2009)); but in those papers, firms are identical and workers are vertically differentiated[1]. However, our work joins this recent development of labor theory which reexamines the issue of training choices in the presence of market imperfections. Search frictions create two types of inefficiencies. Pissarides' seminal papers focused on the first one which can be referred to as the matching problem (Pissarides (2000)). The matching problem generates rents which prompt the firms to provide their employees with training, whether the acquired skills are specific or general. In other words, search frictions allow one to go beyond Becker as Acemoglu and Pischke (1999 a) put it (see also Lechthaler and Snower (2008)). During the last decade, labor theory emphasized the second implication of search frictions which can be referred to as the assignment problem. In the presence of market frictions, jobs are not necessarily held by the right man; Teulings and Gautier (2004) provide a brief survey of this literature. Depicting the technology of contacts with an extension of the urn-ball model allows us to better account of these two distinct market imperfections. For that purpose, we assume that workers' search is partially oriented. Unemployed workers pick a job in the neighborhood of their skill type. The size of this search area determines the extent of the mismatch, hence the need for on-the-job training. Conversely, it is worth noting that in our setting the probability of filling a vacancy does not depend on this search area. As usual, this probability only depends on the tightness of the labor market. In the extreme case where the workers perfectly orient their search, the matching problem remains but the assignment problem disappears and on-the-job training is not needed any longer.

Partially oriented search has another implication. In usual two-sided matching models, workers do not not orient their search at all. As a consequence, they are likely to encounter jobs with which a match would yield a negative surplus because the fit is too bad. In such a case, workers will prefer to wait for a better-suited job. With partially oriented search and on-the-job training (which raises the surplus) such an event becomes improbable. In the model we develop all matches are assumed to be viable. In other words there is no job rejection in the sense of Pissarides (1984).

To conclude our discussion about the position of our contribution in labor theory, let us return to the origins and compare our view of training with Becker's seminal work.

---

[1] Gavrel and Lebon (2009) use an urn-ball with applicant ranking in a model with horizontally differentiated workers. However, in this paper, workers' differentiation is binary.

Becker (1964) drew the very useful distinction between general and specific training. In a competitive market where workers receive their marginal product, firms could never recoup their investment in general skills, so they will never pay for general training (Acemoglu and Pischke (1999 a)). This prediction seems to contradict the empirical evidence. Acemoglu (1997) goes beyond Becker, by showing that search frictions give the firms an incentive to sponsor general training by alleviating the holdup threat (Pigou (1912)). Along the same line of research, the presence of information asymmetries can do the same job, for instance see Chang and Wang (1996) and Acemoglu and Pischke (1998, 1999 b).

In our setting, training is rather specific than general. However, the reason why on-the-job training exists is not because it is specific. Without market imperfections, workers would go to the right job and training would be useless. In other words, our contribution highlights the fact that market imperfections can generate a need for training in the workplace, which would not exist in a competitive economy; whereas, the previous branch of literature shows how market imperfections make employer-sponsored training possible.

Regarding empirical literature, papers about on-the-job training are numerous (see Leuven (2004) for a survey). This literature mostly focuses on the effect of on-the-job training on wages. The reason for this is that the effect on wages is presumed to reflect the effect on productivity. In that respect, we would like to mention that explaining the training decision also provides a more indirect, but valid information about the productivity effect of on-the job training. In addition, our theoretical model clearly highlights one serious problem when attempting to evaluate the productivity enhancing effect of on-the-job training by estimating a wage equation. Trained workers do not necessarily receive higher wages. On the contrary, training may reveal a bad match being then correlated with low wages. In other words, one should control for the extent of the mismatch, quite a difficult task. Some papers estimate an on-the-job training equation in order to palliate the selection bias following the method of Heckman (1979). The usual regressors are characteristics of the workers (educational levels, gender, experience...) or of the jobs they hold (sector of activity, size of the firm...). Most papers omit the unemployment rate of workers' employment area. This may be because the required data were not available. One exception is Arulampalam and Booth (2001) who use UK data.

The paper is organized as follows. Section 2 presents the analytical framework. In Section 3, we study the effect of an increase in the tightness of the labor market on the probability of on-the-job training. Next, we close the model and we study its comparative statics as well as its welfare properties (Section 4). The comparative statics put the emphasis on the effects of an increase in the size of the workers' search area. Section 5 gives an assessment of the empirical relevance of our main positive result. The last section summarizes the paper and offers some final comments.

4

# 2 Analytical framework

## 2.1 Market structure

Our analytical framework is an extension of the usual search matching model with static Nash bargaining and free-entry. On-the-job search is ruled out.

Relative to Pissarides (2000), we introduce *ex ante* heterogeneity in the same way as Marimon and Zilibotti (1999) (see also Gavrel and Lebon (2008), Gautier *et al.* (2010)). With *ex ante* heterogeneity, the output of a job depends on the quality of the match with the worker who holds it. In contrast to Marimon and Zilibotti (1999), the relation between the output and the quality of a match is not exogenous. Here, it depends on on-the-job training which is a choice variable. In addition, search is not random. Firms rank their applicants and pick the best one.

The economy is comprised of two sets of risk-neutral agents: workers and firms. Each active firm offers a single job. All agents have the same discount rate $r$. Firms are infinitely-lived[2] whereas workers have a finite life expectancy of $1/m$. Time is continuous and the Poisson rate $m$ measures workers' labor market exit rate. Each worker who leaves the market is replaced with a newcomer. The measure of the total labor force is constant and normalized to one.

In order to describe the differentiation of workers and jobs, we use the tool of analysis of Salop (1979). Workers and firms are uniformly distributed along a circle of unit length (see Figure 1). The location of a worker on this circle represents the "**type**" of his/her skill. Likewise, the position of a firm on the circle represents its "**type**", that is the skill which perfectly suits its needs.

Let us consider two points $i$ and $j$ on the circle of the skills. Let $x$ be the distance between both points ($0 \leq x \leq \frac{1}{2}$). This distance measures the mismatch between the "type" of a firm located in $i$ and the "type" of a worker located in $j$. Consequently, the output of job $i$ when matched with worker $j$ is a decreasing function in the distance $x$.

In what follows, for the sake of simplicity, we will restrict ourselves to a stationary equilibrium where $u$ unemployed workers and $v$ vacant jobs are uniformly distributed along the circle of skills[3]. Thus, as the length of the circle is normalized to one, $u$ ($v$) also represents the density of unemployed workers (vacant jobs) in any point of the circumference. The ratio $\theta = v/u$ is referred to as the tightness of the labor market. Another consequence of the uniformity of both distributions is that the asset value

---

[2]Exogenous job destruction is thus ruled out.

[3]In that respect, one could point out that the circular model is built so as to generate such an equilibrium.

Figure 1: The circle of skills
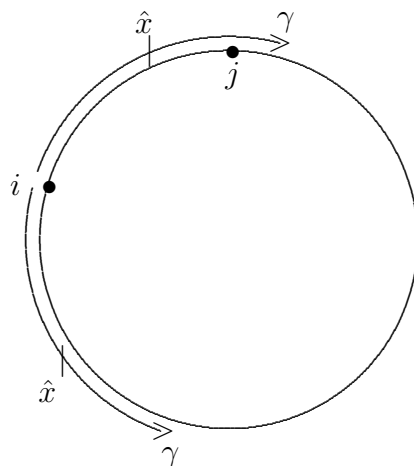


of a vacancy, denoted by $V$, does not depend on its location on the circle.

Concerning job creation, we adopt the usual assumption of free entry. Firms freely enter the labor market as long as profit opportunities exist. In equilibrium, all profit opportunities from new jobs are exploited, thus driving the value of vacant jobs to zero. We then have:

$$V = 0 \tag{1}$$

Our setting has two main innovative features. First, concerning the hiring process, firms rank their applicants and pick the best suited one. Second, aiming at reducing the effect of the mismatch on the output, workers can be trained on the job.

Let us provide a detailed exposition of these two main assumptions.

## 2.2 Hiring process

In most matching models of the labor market, search is random. In Marimon and Zilibotti (1999), for example, firms hire the first worker they meet insofar as the mismatch is low enough as to generate a positive (private) surplus. Two exceptions are Moen (1999) and Gavrel (2009). Gavrel uses an urn-ball with applicant ranking to study the effect of a general rise in unemployment on skill requirements and Moen applies the same tool of analysis in a study of the efficiency of educational choices when workers use human capital as a means to compete for jobs. In these papers, firms are homogenous while workers are vertically differentiated. On the other hand, in our setting, workers and jobs are horizontally differentiated, but the reasoning is similar.

Workers' search is partially oriented. Specifically, we admit that the job which a worker draws belongs to the subset of the vacancies which types are not further than some positive distance $\gamma$ from her/his own type. For $\gamma = 0$, workers' search is perfectly oriented; the labor market is divided into as many sub-markets as worker (job) types (similarly to the model of vertical differentiation of Mortensen and Pissarides (1999)). Conversely, for $\gamma = 1/2$, workers' search is random (as usual in circular models). In general, firms will have several applicants of different types. Firms are assumed to have full knowledge of the sample of their applicants. They then pick the best suited one.

Following Moen (1999), we obtain a continuous-time matching process by assuming that during a time interval $[t, t + dt]$, firms advertise their vacancies with probability $bdt$ while (unemployed) workers randomly send an application to one of the (advertised) vacancies with probability $adt$. In other words, vacancies are advertised at Poisson rate $b$ while applications are sent at Poisson rate $a$. The ratio $a/b$ is denoted by $\lambda$.

During the delay $dt$, $(a\ u\ dt)$ unemployed workers send an application to one of the $(b\ v\ dt)$ advertised vacancies at random. Let $q(x,.)dxdt$ denote the probability for an advertised vacancy to be filled with a worker whose mismatch stays on the range $[x, x + dx]$ during the delay $dt$. In other words, $bq(x,.)$ is the rate at which firms with a vacancy hire workers whose mismatch (with the offered job) $x$ belongs to the subset $[x, x + dx]$.

Concerning the hiring process, we state the following lemma:

**Lemma 1**. *In the circular matching model with applicant ranking and continuous time, the density $q(x, \theta)$ is given by*:

$$q(x, \theta) = \frac{\lambda}{\gamma\theta} \exp(-\frac{\lambda x}{\gamma\theta}) \tag{2}$$

**Proof**. In order to compute the probability of hiring a worker of mismatch $x$, let us first consider the probability for an advertised vacancy not to be drawn by any applicant of mismatch lower than $x$ $(x \leq \gamma)$[4]. It is given by: $[1 - \frac{1}{2\gamma bv}]^{2xau}$. Assuming that vacancies are numerous, this probability tends to: $[\exp(-\frac{\lambda x}{\gamma \theta})]$.

Let us now consider the probability for a sample of candidates to contain at least one worker whose mismatch remains on the range $[x, x + dx]$. This probability tends to $(\frac{\lambda dx}{\gamma \theta})$ when $dx$ tends to zero.

Finally, an advertised vacancy will be filled with a worker of mismatch $x$ only if the sample of its applicants does not contain any better-suited worker. This proves Lemma 1.
**Q.E.D.**

Integrating $q(x, \theta)$ on the range $[0, \gamma]$ gives the probability of filling an advertised vacancy, called $Q(\theta)$. We get:

$$Q(\theta) = 1 - \exp(-\frac{\lambda}{\theta}) \qquad (3)$$

Consequently, the number of job-worker matches per period, $M$, is given by the constant-returns matching function: $M = bvQ(\theta)$. Notice that the derivative $Q'(\theta)$ is negative. We also deduce the probability for an applicant of finding a job: $P(= (\theta/\lambda)Q(\theta))$. One can show that $P'(\theta) > 0$. The probability $P$ can also be obtained from the following integral:

$$P = \int_0^\gamma \exp(-\frac{\lambda x}{\gamma \theta})\frac{1}{\gamma}dx$$

In this expression, the term $[\exp(-\frac{\lambda x}{\gamma \theta})\frac{1}{\gamma}dx]$ represents the probability for an applicant of finding a job with which the mismatch would lie on the range $[x, x + dx]$.

It is worth noting that probabilities $Q$ and $P$ do not depend on $\gamma$. As mentioned in the introduction, this extension of the urn-ball model perfectly dissociates the matching issue from the assignment issue.

We denote the density of the mismatch $x$ among employed workers (*i.e.* the set of occupied jobs) as $\rho(x, \theta)$. We have:

$$\rho(x, \theta) = \frac{q(x, \theta)}{Q(\theta)} \qquad (4)$$

The properties of the function $\rho(x, \theta)$ play a crucial in the following.

---

[4]For simplistic reasons, we consider countable sets. However, the results extend to continuums.

## 2.3 Training, mismatch and productivity

When a new match occurs, the agents decide on whether to provide the worker with on-the-job training or not. If yes, training takes place immediately[5] and the cost $F$ must be paid.

On-the-job training is then assumed to be a discrete variable $I$. $I$ is equal to one when the worker is trained and, if not, it is equal to zero. So the output of the job depends on two variables: the match quality, measured by the distance $x$ ($0 \leq x \leq \gamma$), the dummy variable $I$. The output is then a function: $y(x, I)$.

This function, referred to as the productivity function, is assumed to verify the assumptions below.

First, the output $y(.)$ is a decreasing function of the mismatch $x$:

$$\frac{\partial y(x,I)}{\partial x} < 0, \text{ for } I = 0, 1 \text{ and } 0 < x \leq 1/2 \tag{PF1}$$

Second, on-the-job training raises the output if and only if the mismatch is positive. We then have:

$$y(0, 1) = y(0, 0) \tag{PF2a}$$

and:

$$y(x, 1) - y(x, 0) > 0, \text{ for } 0 < x \leq 1/2 \tag{PF2b}$$

From assumptions (PF2a, b), we deduce that:

$$y(x, 0) < y(x, 1) < y(0, 0), \text{ for } 0 < x \leq 1/2$$

The previous inequalities mean that training does not improve the general skills of workers. In other words, on-the-job training only moves workers along the circle of skills.

Third, the return to on-the-job training is an increasing function of the mismatch $x$:

$$\frac{\partial(y(x,1) - y(x,0))}{\partial x} > 0, \text{ for } 0 < x \leq 1/2 \tag{PF3}$$

Assumption (PF3) signifies that in line with intuition, training in the workplace is all the more efficient as the mismatch is high.

---

[5]Assuming that this process lasts $n$ periods is unlikely to affect the results.

# 3 On-the-job training as a response to unemployment

## 3.1 On-the-job training decision

In fact, the presence of on-the-job training raises two different issues. The first one is: Why training takes place once the worker is matched with a job and not before? The second one is: Why do firms pay (part of) the training costs?

In Acemoglu and Pischke (1999 b, section 5), training improves the general skills of the workers, *i.e.* the level of their ability. As a consequence, the authors need to explain why workers are not able to invest in general skills before they encounter a firm. Their favorite answer to this question is that credit constraints impede the workers to finance any investment, whereas firms can raise the funds they need. As search frictions enable the firms to recoup the training costs, the credit problems faced by workers resolve the two puzzles of on-the-job training.

As Lechthaler and Snower (2008) point out, relating on-the job training to credit market failures in a model where all the agents are risk-neutral sounds very *ad hoc.*

In our setup, it is obvious that training cannot take place before a match occurs. Unemployed workers do not know which type of job they will find. So, we do not need any credit market imperfections, and we have no reason to assume that firms bear the whole training investment (contrarily to Acemoglu and Pischke (idem)). As a consequence, on-the-job training will take place as soon as it raises the (private) surplus of a match.

For obvious reasons, the training decision will depend on the extent of the mismatch between the type of the worker and the type of the job. If the mismatch is low, the training cost $F$ will be higher than the return to on-the-job training. More precisely, we shall show that there exists a training trigger, denoted by $\hat{x}$, below which the agents choose not to pay the training cost ($I = 0$).

Assuming Nash bargaining, the on-the-job training decision will derive from the maximization of the value of the matches. So let us consider the *ex ante* surplus of a match of quality $x$ (net of the training cost F). This asset value is given by:

$$-IF + IS(x, 1) + (1 - I)S(x, 0) \tag{5}$$

with $S(x, I)$ being the *ex post* private surplus of a match (gross of the training costs).[6] The surplus $S(x, I)$ is deduced from the *ex post* value of a filled job and the lifetime utilities of a worker. The *ex post* value of employing a worker, denoted by $J(x, I)$, is a function of the mismatch and of the training dummy variable.[7] This asset value

---

[6]For $I = 0$, the *ex post* asset values are obviously equal to the *ex ante* values.

[7]idem.

satisfies:

$$(r + m)[J(x, I) - V] = y(x, I) - w(x, I) - rV \tag{6}$$

where $w(x, I)$ is the wage which depends on the mismatch and the training choice.

On the other hand, the *ex post* lifetime utility of an employed worker, called $W(x, I)$ verifies:

$$(r + m)[W(x, I) - U] = w(x, I) - (r + m)U \tag{7}$$

where $U$ is the lifetime utility of an unemployed worker.
Like $V$ the value $U$ does not depend on the location of the worker on the skills circle.
Combining equations (6) and (7) yields the *ex post* surplus $S(x, I)$. We obtain:

$$(r + m)S(x, I) = y(x, I) - (r + m)UrV \tag{8}$$

As mentioned in the introduction, partially oriented search coupled with on-the-job training make all matches be viable. Formally, it is assumed that:

$$-F + S(x, 1) \geq 0, \forall x \leq \gamma$$

Under this assumption, the mismatch trigger $\hat{x}$ is obtained by equalizing the value of a match with or without on-the-job training. It results that:

$$y(\hat{x}, 1) - y(\hat{x}, 0) = (r + m)F \tag{9}$$

Because on-the-job training is all the more efficient as the mismatch is high (assumption (PF3)), workers with a higher mismatch than $\hat{x}$ will benefit from on-the-job training.
Intuitively, an increase in the cost $F$ raises the training trigger $\hat{x}$. In other words, the share of the workers who benefit from on-the-job training decreases when the cost $F$ rises.
In our setting, on-the-job training critically depends on workers' search. If workers orient their search in the close neighborhood of their skill type ($\gamma \leq \hat{x}$) the match quality would never be that bad so as to motivate training. In the following, we will restrict ourselves to the case where workers' search orientation is bad ($\gamma > \hat{x}$). Consequently, employed workers whose mismatch is low ($x \leq \hat{x}$) will not be trained; whereas, employed workers whose mismatch is high ($x > \hat{x}$) will benefit from training.

# 4   On-the-job training and market tightness

From the density $\rho(x, \theta)$, we deduce the share of the (employed) workers who benefit from on-the-job training, called $\Phi$:

$$\Phi \equiv \int_{\hat{x}}^{\gamma} \rho(x, \theta) dx \tag{10}$$

Notice that $\Phi$ also represents the probability that an employed worker (drawn at random) benefits from on-the-job training. As the training trigger $\hat{x}$ is an increasing function of $F$, an increase in the training costs lowers $\Phi$.

Concerning the effect of an increase in market tightness on on-the-job training, we state the following Proposition:

**Proposition 1**. *An increase in market tightness ($\theta$) raises the probability of training ($\Phi$). The same holds for an increase in the workers' search area ($\gamma$).*

**Proof** Let us first consider the derivative of the density $\rho(x, \theta)$ with respect to $\theta$. This derivative has the same sign as:

$$G(x, \theta) \equiv \frac{x}{\gamma} - \frac{\theta^2}{\lambda} \frac{P'(\theta)}{P(\theta)}$$

Let us denote by $\tilde{x}$ the value of the distance $x$ such that $G(x, \theta) = 0$. One can check that: $0 < \tilde{x} < \gamma$. As $G(.)$ grows with $x$, we obtain:

$$\frac{\partial \rho(x, \theta)}{\partial \theta} < 0, \text{ for } 0 < x < \tilde{x}$$

$$\frac{\partial \rho(x, \theta)}{\partial \theta} > 0, \text{ for } \tilde{x} < x < \gamma$$

This means that regarding the sign of the derivative of $\Phi(\theta, .)$ with respect to $\theta$, two cases must *a priori* be distinguished. If $\tilde{x} < \hat{x} < \gamma$, we deduce:

$$\frac{\partial \Phi(\theta, .)}{\partial \theta} = \int_{\hat{x}}^{\gamma} \frac{\partial \rho(x, \theta)}{\partial \theta} dx > 0$$

On the contrary, if $\hat{x} < \tilde{x} < \gamma$, we obtain:

$$\frac{\partial \Phi(\theta, .)}{\partial \theta} = \int_{\hat{x}}^{\gamma} \frac{\partial \rho(x, \theta)}{\partial \theta} dx = -\int_{0}^{\hat{x}} \frac{\partial \rho(x, \theta)}{\partial \theta} dx > 0$$

This proves the first statement of Proposition 1. Now the probability of training can be rewritten as:

$$\Phi = \frac{1}{Q} \int_{\hat{x}}^{\gamma} q(\theta, x) dx = \frac{\exp^{-\lambda \hat{x}/\gamma \theta} - e^{-\lambda/\theta}}{Q}$$

This shows that $\Phi$ grows with $\gamma$ and completes the proof of Proposition 1.
**Q.E.D.**

As an increase in market tightness lowers the unemployment rate, Proposition 1 also means that an increase in overall unemployment is associated with a decrease in probability $\Phi$. More unemployment reduces the mismatch, hence the need for on-the-job training.

The reason for this result is the following: given the fact that firms pick the best-suited worker they meet, the match quality is an order statistic which is the maximum of the queue of applicants. As the expected maximum of a sample from a given distribution rises with the sample size, a decrease in the number of applicants per firm (*i.e.* an increase in market tightness) lowers the expected quality of the best application. As a consequence, the probability that the two parties reduce the mismatch by using on-the-job training rises. In other words, an increase in the share of the workers who benefit from on-the-job training can be seen as a response to an unemployment decrease. That the probability $\Phi$ grows with the distance $\gamma$ is not surprising. A well oriented workers' search improves the assignment of workers to jobs, reducing then the need for training in the workplace. Remember that when the search distance $\gamma$ becomes lower than the training trigger $\hat{x}$, the mismatch is not high enough so as to induce an investment in skills any longer.

# 5 Equilibrium and comparative statics

## 5.1 Equilibrium

So far, the tightness of the labor market was taken as an exogenous variable. We now close the model by using the free-entry assumption and define an equilibrium of the labor market. We first have to compute the expected (private) surplus of a match. From the definition of probability $\Phi$, we deduce that the lifetime utility of an unemployed worker verifies:

$$(r + m)U = d + aP\beta(-\Phi F + \bar{S}) \tag{11}$$

where $d$ is the utility of leisure and $\bar{S}$ is the average *ex post* private surplus. Formally, $\bar{S}$ is defined as follows:

$$\bar{S} = \int_0^{\hat{x}} \rho(\theta, x) S(x, 0) dx + \int_{\hat{x}}^{\gamma} \rho(\theta, x) S(x, 1) dx$$

Under the assumption of free entry (equation (1)), combining equations (8) and (11) yields:

$$(r + m + \beta aP)\bar{S} = \bar{y} - d + \beta aP\Phi F \tag{12}$$

with $\bar{y}$ being the average output. Variable $\bar{y}$ is a function of $\theta$ and $\hat{x}$ which is given by:

$$\bar{y} = \bar{y}(\theta, \hat{x}) = \int_0^{\hat{x}} \rho(\theta, x) y(x, 0) dx + \int_{\hat{x}}^{\gamma} \rho(\theta, x) y(x, 1) dx \tag{13}$$

On the other hand, under the assumption of free entry, the value of a vacancy, $V$, satisfies:

$$rV = -c + bQ(1 - \beta)[-\Phi F + \bar{S}] = 0 \tag{14}$$

with $c$ being the cost to keep a vacancy open.

Substitution of (12) into (14) yields the following equilibrium equation:

$$-c + bQ(\theta)(1 - \beta)\frac{\bar{y}(\theta, \hat{x}) - d - (r + m)\Phi(\theta, \hat{x})F}{r + m + \beta aP(\theta)} = 0 \tag{15}$$

In sum, we can define an equilibrium of the labor market as follows:

**Definition 1**. *An equilibrium of the labor market is a pair of variables $(\theta, \hat{x})$ which jointly satisfy equations (9) and (15).*

From the variables $\theta$ and $\hat{x}$, we deduce the equilibrium values of the probability $\Phi$, the average output $\bar{y}$, and the unemployment rate $u$. The unemployment rate is derived from flow equilibrium:

$$u = \frac{m}{m + aP}$$

In the comparative statics as well as in the efficiency study, we will need to know how the market tightness affects the average surplus of a match, hence the following quantity:

$$H \equiv \bar{y} - d - (r + m)\Phi F \tag{16}$$

Using Proposition 1, we state the following lemma:

**Lemma 2**. *H is a decreasing function of the market tightness $(\theta)$.*
**Proof** See Appendix A.

As already noted, an increase in market tightness raises the expected mismatch by compelling the firms to hire ill-suited workers. Despite the effect of on-the-job training, this causes a decrease in the expected surplus $\bar{S}$ (for a given value of $P$).

## 5.2 Workers' search, unemployment and training

In our setting, on-the-job training results from the mismatch between workers and jobs. As the efficiency of the technology of contacts critically depends on the degree of orientation of workers' search, we naturally put the emphasis on the effects of an increase in the distance $\gamma$. We state the following proposition.

**Proposition 2**. *An increase in the size of workers' search area ($\gamma$): -(i) reduces job creation ($\theta$) -(ii) raises the probability of on-the-job training ($\Phi$).*
**Proof** See Appendix B.

The first statement of Proposition 2 is not very surprising as an increase in workers' search area deteriorates the assignment of workers to jobs, lowering then the value of filled jobs. The increase in the mismatch also explains why trained workers become more numerous. However, with applicant ranking, the market tightness decrease tends to reduce the average mismatch. Due to this indirect effect, the proof of statement (ii) is not straightforward.

# 6 Market efficiency

One important issue is the efficiency of such a decentralized equilibrium. We show that, when viewed as a mean of reducing the mismatch between workers and jobs, private training choices are socially optimal. On the other hand, firms tend to create too many jobs.
Along the same line as Hosios (1990) and Pissarides (2000), let us consider a social planner who is only subject to the search frictions and can redistribute income among agents at no cost. Assuming that the interest rate is equal to zero[8], this planner maximizes the social surplus flow (per head), called $\Sigma$. We have:

$$\Sigma \equiv (1-u)\overline{y} + ud - \theta uc - Pu\Phi(\theta, \hat{x})F \tag{17}$$

Notice that in the expression $\Sigma$, the quantity $(Pu\Phi)$ measures the flow of the workers who benefit from on-the-job training per period. Due to flow equilibrium, this quantity is equal to $(m(1-u)\Phi)$.

---

[8]The results extend to a positive interest rate. Proof is available upon request from the authors.

It is well known that in the basic matching model, job creation is efficient if and only if the Hosios condition holds. The elasticity of the matching function with respect to unemployment, $\eta$, must be equal to the bargaining strength of workers, $\beta$. As we want to put the emphasis on the implications of the ranking of applicants, we shall assume that firms internalize the so-called congestion effect ($\eta = \beta$).
The following result can be established:

**Proposition 3**. *Under the Hosios condition, a decentralized equilibrium of the labor market is -(i) efficient in terms of on-the-job training but -(ii) inefficient in terms of job creation; the market tightness ($\theta$) is too high.*

**Proof** Let us first study the efficiency of the training trigger. The derivative of the social surplus $\Sigma$ with respect to $\hat{x}$ has the same sign as:

$$-[y(\hat{x}, 1) - y(\hat{x}, 0)] + mF$$

For $r = 0$, the previous expression is nil in a decentralized equilibrium, see equation (9). This proves statement (i).
Let us now turn to job creation. Under the Hosios' condition, the derivative of the social surplus with respect to $\theta$ reduces to (for $r = 0$):

$$(1 - u)[\frac{\partial \bar{y}}{\partial \theta} - m\frac{\partial \Phi}{\partial \theta}]F$$

From Lemma 2, we deduce that this derivative is negative. This proves Proposition 3.
***Q.E.D.***

The efficiency of on-the-job training[9] choices may look counterintuitive, as one might expect that Nash bargaining gives rise to a holdup phenomenon (see Grout (1985)). The reason why this phenomenon does not occur in our setting (whether the wages are renegotiated or not) is that on-the-job training does not improve the outside opportunities of the workers (the lifetime utility of unemployment)[10]. Training only moves the workers along the circle. Their skill type changes but their ability level remains the same. It is worth noting that on-the-job search is unlikely to affect this result.

---

[9]Notice that the mismatch trigger is partially efficient. In other words it is not a social optimum but efficient for a given value of market tightness.

[10]One could object that a trained worker is endowed with two skill types. This objection is relevant but taking it in account is unlikely to substantially affect the efficiency result. We surmise that the training trigger would become a little too high.

The inefficiency of job creation is not surprising. From Proposition 1, we know that an increase in the tightness of the labor market lowers the quality of the best candidates. As firms do not internalize this effect, they create too many vacancies[11] .

# 7    Empirical evidence

Our theoretical study has one main positive prediction: an increase in unemployment should lower the probability of on-the-job training. In order to get an assessment of the empirical relevance of our view of training, we are now going to estimate an econometric model of the training decision.
We will first present our data, then discuss our results.

## 7.1    Data

Our data set is based on the latest cohort of the Education-Training-Occupation survey (*Formation Qualification Profession*, referred to as FQP hereafter), conducted by the French National Statistical Agency (INSEE) in 2003.

Using a complex sampling design, the survey covers all men and women in metropolitan France. In this survey, a cohort of individuals is investigated over the period 2002-2003. FQP provides many information on return to education: the efficiency of the educational system, the impact of social origins on academic and professional performances, as well as the impact of vocational training on careers in terms of earnings and mobility. The questionnaire contains five parts: worker mobility, academic education, vocational training, social origins, and earnings.
The reference population is constituted of individuals aged between 18 and 65 who live in France in main home. The final sample contains about 40,000 individuals. The survey is made with face-to-face interviews. After the description of the household, the questionnaire takes about 30 minutes per person.
For the purposes of this study, we have restricted the sample to the workers of the private sector over the period 2002-2003.This period of study is specially attractive, as until 2004, the on-the-job training particulary depends on firm decisions. After this date, a reform in France allows the workers to personally ask for on-the-job training, even to obtain another job. We also excluded individuals whose training was financed by public funds as well as entrepreneurs, farmers and self-employed workers. Furthermore, individuals who went into retirement and those who had been unemployed over 2002-2003 have been removed from the sample. All in all, our sample contains 9,760 workers, 22.5 percent out of them were involved in at least

---

[11]Gavrel (2009) obtains a similar result in a model where workers are vertically differentiated.

one on-the-job training event. Appendix C provides a description of the variables we used.

Concerning unemployment, metropolitan France is divided into 22 local areas (*Régions*). In this study, the Local Area Rate of Unemployment (referred to as *LARU* hereafter) measures the unemployment rate of these 22 French areas over 2002-2003. The *LARU* variable is based on the "Labour Force Survey" (referred to as LFS, 2003, INSEE).

## 7.2 Econometric model and results

Workers' probability of being trained on-the-job is assumed to be determined by the following probit model:

$$I_i = \begin{cases} 1 & if \ \ I_i^* > 0 \\ 0 & if \ \ I_i^* \leq 0 \end{cases} \quad with \ \ I_i^* = \alpha \Psi_i + \epsilon_i$$

$I_i^*$ is a latent variable explaining the probability of being trained on-the-job and $\Psi_i$ is a set of independent variables (see Table 1), $\alpha$ is the vector of parameters and $\epsilon_i$ is the error term, assumed to be normally distributed.

As Box and Tidwell (1962), we used a Box-Cox transformation to the independent variables (Box and Cox, 1964). This method has two advantages: (i) first, it transforms continuous variables into an almost normal distribution, (ii) second, it accounts for a possible nonlinear relation between dependent and independent variables.

According to this method, a nonlinear relation is found for *LARU* variable whereas a linear relation yields better results for training and experience variables. We thus applied this transformation to only one continuous independent variable, *LARU* (see nonlinear model in Table 1).

The variable *LARU* is then transformed as follows:

$$LARUT_i = \begin{cases} ((LARU_i + k)^\pi - 1)/(\pi g^{\pi-1}) & if \ \ \pi \neq 0 \\ g[\ln(LARU_i + k)] & if \ \ \pi = 0 \end{cases}$$

This family of transformations of *LARU* variable is controlled by the parameter $\pi$[12], the parameter $k$ is used to rescale *LARU* so that it is strictly positive[13], the parameter $g$ is used to scale the resulting values, $g$ is taken as the geometric mean of the $LARU_i$

---

[12]Estimated in maximizing the likelihood function by testing different values of parameter $\pi$.

[13]In the study, *LARU* is a variable strictly positive. Then $k = 0$ in our case.

observations. Linear, square root, inverse, cubic, and so on are all special cases of Box-Cox transformations.

Estimating the linear model leads to the expected signs for most variables (Table 1).

The first interesting result is that, contrary to most Mincerian wage equations, we do not observe gender discrimination in training choices. Gender differences are likely to occur in the earlier stages of professional insertion (see Barros et al. (2010)). In contrast, on-the-job training appears to be influenced by the individual origin: workers with French parents have a significantly higher probability of being trained. In line with our intuitions, it seems that training choices are sensible to the potential duration of matches: workers with permanent labor contracts benefit from training more frequently (see "Permanent" variable). This effect is highly significant contrary to the "Full-Time" variable which seems to have no influence. The need to use a computer on the job also raises the probability of being trained. Globally the probability of on-the-job training is positively correlated with the occupation hierarchy (see the "Occupations" variables): other things equal, Unskilled Workers are the least trained whereas Executives are often trained on the job. As one could expected, Technicians - an occupation requiring very specific skills - benefit from the highest probability of being trained on the job. Not surprisingly, on-the-job training is more frequent in large compagnies where the opportunity cost of training (the output loss) is likely to be lower.

Let us now come to our main variable of interest ("$LARU$"). In accordance with the predictions of the theoretical model, the on-the-job training probability is a decreasing function of the unemployment rate. Less unemployment lowers the number of applicants per vacant jobs. Firms then have to hire ill-suited workers. As it improves the quality of matches, on-the-job training becomes more frequent[14].

## 8   Conclusion

In this paper, we presented a novel explanation of on-the-job training. In an imperfect labor market, training in the workplace can be used to reduce the mismatch between workers and jobs. As workers have no idea of the type of the job they will hold, training necessarily takes place once the meeting occurred, hence on the job. Using a circular matching model with applicant ranking and bargained wages, we first stated

---

[14]Furthermore, the nonlinear model suggests that the (positive) response of the training probability to an unemployment cut is all the stronger as the labor market is tighter. An inverse transformation of the $LARU$ variable - close to $(-x^{-2})$ - slightly improves the goodness of fit of our model.

that, in line with intuition, an increase in unemployment lowers the probability of on-the job training by improving the expected quality of the best application. Using French fine data about regional unemployment, we have shown that our main positive prediction is supported by empirical evidence.

We also showed that private training decisions are efficient whereas job creation tends to be too high as long as firms internalize the usual congestion effect. This means that in our setting subsidies to on-the-job training are useless.

To conclude, we would like to add two comments.

First, we are aware that the empirical correlation between unemployment and training in the workplace can be interpreted in a different way. In particular, one could point out that on-the-job training for an already-employed worker might be a substitute for hiring a new one when a shock imposes a reorganization of the production process. As hiring is more costly when the applicants are fewer, an increase in unemployment should make more firms prefer hiring than training, reducing then the probability of on-the-job training. In fact, this argument is not fundamentally different from ours as in both cases on-the job training is used to improve the quality of the match between a worker and a job.

According to our view, training in the workplace can be seen as a substitute for an improvement in the assignment of workers to jobs. This means that the training costs might be reduced with no loss in market efficiency by enhancing the performances of the technology of contacts. To address this issue, one should make workers' search orientation endogenous. This is a line for further research.

# References

Acemoglu, D., 1997. Training and Innovation in an Imperfect Labor Market. *Review of Economic Studies* 64, 445-464.

Acemoglu, D., Pischke, J-S., 1998. Why do Firms Train? Theory and Evidence. *Quarterly Journal of Economics* 113, 79-119.

Acemoglu, D., Pischke, J-S., 1999a. Beyond Becker: Training in Imperfect Labor Markets. *Economic Journal* 109, 112-142.

Acemoglu, D., Pischke, J-S., 1999b. The Structure of Wages and Investment in General Training. *Journal of Political Economy* 107, 539-572.

Albrecht, J., Gautier, P., Vroman, S., 2003. Matching with Multiple Applications. *Economics Letters* 78, 67-70.

Arulampalam, W., Booth, A., 2001. Learning and Earning: Do Multiple Training

Events Pay? A Decade of Evidence from a Cohort of Young British Men. *Economica* 68, 379-400.

Barros, C.P., Guironnet, J-P, Peypoch, N., 2010. How to Quickly Get a Job? The Transition from Higher Education to French Labor Larket by a Survival Model. *Applied Economics*, DOI:10.1080/00036840903166251.

Becker, G.,1964. *Human Capital.* The University of Chicago Press, Chicago.

Blanchard, O., Fisher, S., 1989. *Lectures on Macroeconomics.* MIT Press, Cambridge, MA.

Box, G.E.P., Cox, D.R., 1964. An Analysis of Transformations (with Discussion). *Journal of the royal statistical society* Series B 26, 211-252.

Box, G.E.P., Tidwell, P.W., 1962. Transformation of the Independent Variables. *Technometrics* 4, 531-550.

Chang, C., Wang, Y., 1996. Human Capital Investment under asymmetric information: The Pigovian Conjecture Revisited. *Journal of Labor Economics* 14, 505-519.

Gautier, P.A., Teulings, C.N., Van Vuuren, A., 2010. On-the-Job Search, Mismatch and Efficiency. *Review of Economic Studies* 77, 245-272.

Gavrel, F., Lebon, I., 2008. Minimum Wage, Unemployment Benefits and Labour Market Efficiency. *Louvain Economic Review* 74, 53-75 (in French).

Gavrel, F., Lebon, I, 2009. Endogenous Job Specialization. *Economic Modelling*, 26, 328-334.

Gavrel, F., 2009. Technical Skill Bias as a Response of Firms to Unemployment. A Matching Model with Applicant Ranking and Endogenous Skill Requirements. *Labour Economics* 16, 304-310.

Grout, P., 1984. Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach. *Econometrica* 52, 449-460.

Heckman J., 1979. Sample Selection Bias as a Specification Error. *Econometrica* 47, 153-161.

Hosios A.J., 1990. On the Efficiency of Matching and Related Models of Search and Unemployment. *Review of Economic Studies* 57, 279-298.

Lechthaler, W., Snower, D.J., 2008. Minimum Wages and Training. *Labour Economics* 15, 1223-1237.

Leuven, L., 2004. A Review of the Wage Returns to Private Sector Training. Mimeo, University of Amsterdam and CREST-INSEE.

Marimon, R., Zilibotti, F., 1999. Unemployment vs. Mismatch of Talents: Reconsidering Unemployment Benefits. *Economic Journal* 109, 266-291.

Moen, E.R., 1999. Education, Ranking and Competition for Jobs. *Journal of Labor Economics* 17, 694-723.

Mortensen, D.T., Pissarides, C.A., 1999. Unemployment Responses to "Skill-Biased" Technology Shocks: The Role of Labour Market Policy. Economic Journal 109, 242-265.

Pigou, A.C., 1912. *Wealth and Welfare.* MacMillan, London.

Pissarides, C.A., 1984. Efficient Job Rejection. *Economic Journal* 96, 97-108.
Pissarides, C.A., 2000. *Equilibrium Unemployment Theory.* $2^{nd}$ edition, MIT Press, Cambridge, MA.

Salop, S., 1979. Monopolistic Competition with Outside Goods. *Bell Journal of Economics* 10, 223-236.

Teulings, C., Gautier, P., 2004. The Right Man for the Job. *Review of Economic Studies* 71, 553-580.

# Appendix A. Proof of Lemma 2

Two cases must be distinguished.
In the first case , $\hat{x} < \tilde{x}$. As the output $y$ is a decreasing function of $x$, we have (see the proof of Proposition 1):

$$\frac{\partial \rho}{\partial \theta} y(x, 0) < \frac{\partial \rho}{\partial \theta} y(\hat{x}, 0)$$

for $0 < x < \hat{x}$.

$$\frac{\partial \rho}{\partial \theta} y(x, 1) < \frac{\partial \rho}{\partial \theta} y(\tilde{x}, 1)$$

for $\hat{x} < x < \tilde{x}$.
And,

$$\frac{\partial \rho}{\partial \theta} y(x, 1) < \frac{\partial \rho}{\partial \theta} y(\tilde{x}, 1)$$

for $\tilde{x} < x < \gamma$.

The derivative of $H$ with respect to $\theta$ then satisfies:

$$\frac{\partial H}{\partial \theta} < -\frac{\partial \Phi}{\partial \theta}y(\hat{x}, 0) + \frac{\partial \Phi}{\partial \theta}y(\tilde{x}, 1) - (r+m)\frac{\partial \Phi}{\partial \theta}F$$

As $\frac{\partial \Phi}{\partial \theta} > 0$ and $y(\hat{x}, 1) > y(\tilde{x}, 1)$, equation (9) implies:

$$\frac{\partial H}{\partial \theta} < \frac{\partial \Phi}{\partial \theta}[y(\hat{x}, 1) - y(\hat{x}, 0) - (r+m)F] = 0$$

This proves Lemma 2 in the first case. In the second case, a similar reasoning shows that:

$$\frac{\partial H}{\partial \theta} < \frac{\partial \Phi}{\partial \theta}[y(\hat{x}, 1) - y(\tilde{x}, 0) - (r+m)F]$$

As $y(\hat{x}, 0) < y(\tilde{x}, 0)$, equation (9) implies that the derivative of $H$ is also negative for $\hat{x} > \tilde{x}$.

**Q.E.D.**

# Appendix B. Proof of Proposition 2

Statement (i).

To prove the first statement, we need to study the effect of $\gamma$ on $H$. We have:

$$\frac{\partial H}{\partial \gamma} = \frac{\partial \bar{y}}{\partial \gamma} - (r+m)\frac{\partial \Phi}{\partial \gamma}F$$

Let us first consider the partial derivative:

$$\frac{\partial \bar{y}}{\partial \gamma} = \int_0^\gamma \frac{\partial \rho}{\partial \gamma}y(x, I(x))dx + \rho(\gamma, \theta)y(\gamma, 1)$$

Let $\xi$ denote the product $(\gamma\theta/\lambda)$. One can show that:

$$\frac{\partial \rho}{\partial \gamma} < (>)0$$

for:

$$x < (>)\xi$$

Three cases must be distinguished. Let us first assume that $\hat{x} < \xi < \gamma$.
Using the fact that:

$$\int_0^{\hat{x}} \frac{\partial \rho}{\partial \gamma} dx = -\frac{\partial \Phi}{\partial \gamma}$$

we obtain the following inequality:

$$\frac{\partial \bar{y}}{\partial \gamma} < [\frac{\partial \Phi}{\partial \gamma} - \rho(\gamma, \theta)]y(\xi, 1) + \rho(\gamma, \theta)y(\gamma, 1) - \frac{\partial \Phi}{\partial \gamma}y(\hat{x}, 0)$$

It results that:

$$\frac{\partial H}{\partial \gamma} < \frac{\partial \Phi}{\partial \gamma}y(\xi, 1) - \rho(\gamma, \theta)[y(\xi, 1) - y(\gamma, 1)] - \frac{\partial \Phi}{\partial \gamma}y(\hat{x}, 0) - (r + m)\frac{\partial \Phi}{\partial \gamma}F$$

As the output is a decreasing function of the mismatch, this implies that:

$$\frac{\partial H}{\partial \gamma} < \frac{\partial \Phi}{\partial \gamma}[y(\xi, 1) - y(\hat{x}, 0) - (r + m)F]$$

We know that the derivative $\frac{\partial \Phi}{\partial \gamma}$ is positive (Proposition 1). Therefore, equation (9) implies that the derivative of $H$ has the same sign as:

$$y(\hat{x}, 1) - y(\hat{x}, 0) - (r + m)F - [y(\hat{x}, 1) - y(\xi, 1)] = -[y(\hat{x}, 1) - y(\xi, 1)] < 0$$

In the opposite case ($\xi < \hat{x}$), we obtain:

$$\frac{\partial \bar{y}}{\partial \gamma} < [\frac{\partial \Phi}{\partial \gamma} - \rho(\gamma, \theta)]y(\hat{x}, 1) + \rho(\gamma, \theta)y(\gamma, 1) - \frac{\partial \Phi}{\partial \gamma}y(\xi, 0)$$

It results that:

$$\frac{\partial H}{\partial \gamma} < \frac{\partial \Phi}{\partial \gamma}y(\hat{x}, 1) - \rho(\gamma, \theta)[y(\hat{x}, 1) - y(\gamma, 1)] - \frac{\partial \Phi}{\partial \gamma}y(\xi, 0) - (r + m)\frac{\partial \Phi}{\partial \gamma}F$$

As the output is a decreasing function of the mismatch, the previous inequality implies that:

$$\frac{\partial H}{\partial \gamma} < \frac{\partial \Phi}{\partial \gamma}[y(\hat{x}, 1) - y(\xi, 0) - (r + m)F]$$

From equation (9), we then deduce that the derivative of $H$ has the same sign as:

$$y(\hat{x}, 1) - y(\hat{x}, 0) - (r + m)F - [y(\xi, 0) - y(\hat{x}, 0)] = -[y(\xi, 0) - y(\hat{x}, 0)] < 0$$

Finally, if $\xi > \gamma$, we have:

$$\frac{\partial \bar{y}}{\partial \gamma} < \frac{\partial \Phi}{\partial \gamma}[y(\gamma, 1) - y(\hat{x}, 0)]$$

The derivative of $H$ then satisfies:

$$\frac{\partial H}{\partial \gamma} < \frac{\partial \Phi}{\partial \gamma}[y(\gamma, 1) - y(\hat{x}, 0)] - (r + m)\frac{\partial \Phi}{\partial \gamma}F$$

As the output is a decreasing function of the mismatch, this implies that:

$$\frac{\partial H}{\partial \gamma} < \frac{\partial \Phi}{\partial \gamma}[y(\hat{x}, 1) - y(\hat{x}, 0) - (r + m)F - (y(\hat{x}, 1) - y(\gamma, 1))] < 0$$

To sum up, holding $\theta$ as a constant, the derivative of $H$ with respect to $\gamma$ is always negative. Let us now consider the left hand side of equation (15). As $Q'(\theta) < 0$, $P'(\theta) > 0$ and $\frac{\partial H}{\partial \theta} < 0$ (Lemma 2), its derivative with respect to $\theta$ is negative. Therefore, an increase in $\gamma$ necessarily decreases the equilibrium value of $\theta$. Let $\theta(\gamma)$ be this equilibrium value. We then have:

$$\theta'(\gamma) < 0$$

Statement (ii).
Let us now turn to the second statement of Proposition 2. To prove that an increase in $\gamma$ raises the probability $\Phi$ despite the decrease in $\theta$, we first state that the equilibrium value of $\xi$ ($= \gamma\theta/\lambda$) is an increasing function in $\gamma$.
Step 1. Let us consider the product $(QH)$. From equation (15), as $P'(\theta) > 0$, we deduce that an increase in $\gamma$ necessarily decreases the equilibrium value of $(QH)$, that is:

$$(QH)'(\gamma) < 0$$

On the other hand, the product $(QH)$ can be seen as a function of variables $\gamma$ and $\xi$:

$$QH = (QH)(\gamma, \xi)$$

The partial derivative of this function with respect to $\gamma$ is given by:

$$\frac{\partial(QH)(\gamma, \xi)}{\partial \gamma} = q(\gamma, \xi)[y(\gamma, 1) - (r + m)F]$$

Using equation (9), this derivative can be rewritten as:

$$\frac{\partial (QH)(\gamma, \xi)}{\partial \gamma} = q(\gamma, \xi)[(y(\gamma, 1) - y(\gamma, 0)) - (y(\xi, 1) - y(\xi, 0)) + y(\gamma, 0)]$$

Under the assumption (PF3), we obtain:

$$\frac{\partial (QH)(\gamma, \xi)}{\partial \gamma} = q(\gamma, \xi)y(\gamma, 0)] > 0$$

We established above that, holding $\theta$ as a constant, the derivative of $H$ with respect to $\gamma$ is negative. This latter derivative can be written as follows:

$$\frac{1}{Q}[\frac{\partial (QH)(\gamma, \xi)}{\partial \gamma} + \frac{\partial (QH)(\gamma, \xi)}{\partial \xi}\frac{\theta}{\lambda}] < 0$$

Combining the two previous inequalities yields:

$$\frac{\partial (QH)(\gamma, \xi)}{\partial \xi} < 0$$

Let $\xi(\gamma)$ be the equilibrium value of $\xi$. The derivative $(QH)'(\gamma)$ can be rewritten as follows:

$$\frac{\partial (QH)(\gamma, \xi)}{\partial \gamma} + \frac{\partial (QH)(\gamma, \xi)}{\partial \xi}\xi'(\gamma) < 0$$

As the derivative of $(QH)(\gamma, \xi)$ with respect to $\gamma$ (with respect to $\xi$) is positive (negative), the latter inequality implies that:

$$\xi'(\gamma) > 0$$

In words, an increase in $\gamma$ raises the product $(\gamma \theta)$ despite the decrease of $\theta$.

Step 2. Let $\Phi(\gamma)$ be the equilibrium value of the probability $\Phi$. One can check that the derivative $\Phi'(\gamma)$ has the same sign as:

$$Qe^{-\frac{\hat{x}}{\xi}}\frac{\hat{x}}{\xi^2}\xi'(\gamma) + (1 - e^{-\frac{\hat{x}}{\xi}})Q'(\theta)\theta'(\gamma)$$

As $\xi'(\gamma) > 0$, $Q'(\theta) < 0$, and $\theta'(\gamma) < 0$, this proves that $\Phi'(\gamma) > 0$.
**Q.E.D.**

# Appendix C. Data description

The sample retained for our study is described in table 2.

Table 1: Determinants of the Probability of Being Trained on-the-job

| Parameter | Linear Model | Std. Error | Nonlinear Model | Std error |
|---|---|---|---|---|
| **Maximum Likelihood Estimation** | | | | |
| Intercept | -1.039*** | 0.149 | 31.965*** | 4.482 |
| Sex | -0.001 | 0.037 | -0.004 | 0.037 |
| French | 0.292*** | 0.062 | 0.287*** | 0.062 |
| Computer | 0.434*** | 0.039 | 0.436*** | 0.039 |
| Full-time | 0.032 | 0.052 | 0.037 | 0.052 |
| Change | 0.310*** | 0.036 | 0.310*** | 0.036 |
| Tenure | -0.001*** | 0.0001 | -0.001*** | 0.0001 |
| Experience | -0.001*** | 0.0001 | -0.0005*** | 0.0001 |
| LARU(T) | -0.077*** | 0.011 | -0.098*** | 0.013 |
| Permanent | 0.565*** | 0.089 | 0.563*** | 0.089 |
| *Occupations (Ref: Employee)* | | | | |
| Unskilled Worker | -0.340*** | 0.075 | -0.344*** | 0.076 |
| Skilled Worker | 0.090* | 0.053 | 0.090* | 0.053 |
| Manager | 0.237*** | 0.063 | 0.237*** | 0.063 |
| Technician | 0.360*** | 0.056 | 0.361*** | 0.056 |
| Executive | 0.265*** | 0.047 | 0.272*** | 0.047 |
| *Industries (Ref: Services)* | | | | |
| Manufacturing | -0.123*** | 0.037 | -0.126*** | 0.038 |
| Building | -0.096 | 0.065 | -0.097 | 0.065 |
| *Plant Size (Ref: > 500)* | | | | |
| <10 | -0.672*** | 0.053 | -0.677*** | 0.053 |
| 10-49 | -0.496*** | 0.045 | -0.496*** | 0.045 |
| 50-199 | -0.280*** | 0.061 | -0.280*** | 0.061 |
| 100-199 | -0.281*** | 0.059 | -0.283*** | 0.059 |
| 200-499 | -0.261*** | 0.058 | -0.264*** | 0.058 |
| Akaike Information Criterion | 9014.74 | | 9004.08 | |
| Concordant Percentage | 75.3 | | 75.4 | |

Note: *** significant to 1%, ** significant to 5% and * significant to 10%

$\pi$=-2.19 for LARUT variable in the nonlinear model.

Source: FQP 2003, INSEE.

Table 2: Data description

| Variable | Description | Mean | Std. error | Min. | Max. |
|---|---|---|---|---|---|
| I (dependent variable) | Coded 1 if the individual have be trained, 0 otherwise | 0.225 | 0.418 | 0 | 1 |
| Sex | Coded 1 for men, 0 otherwise | 0.584 | 0.493 | 0 | 1 |
| French | Coded 1 for individual of French descent, 0 otherwise | 0.908 | 0.290 | 0 | 1 |
| UW | Unskilled workers (binary code) | 0.100 | 0.300 | 0 | 1 |
| SW | Skilled workers (binary code) | 0.232 | 0.422 | 0 | 1 |
| Employee | Employee (binary code) | 0.348 | 0.476 | 0 | 1 |
| Manager | Manager (binary code) | 0.064 | 0.245 | 0 | 1 |
| Technician | Technician (binary code) | 0.232 | 0.422 | 0 | 1 |
| Executive | Executive (binary code) | 0.162 | 0.369 | 0 | 1 |
| Computer | Use of computer on the job (binary code) | 0.553 | 0.497 | 0 | 1 |
| Full-time | Full-time workers (binary code) | 0.877 | 0.329 | 0 | 1 |
| Manufacturing | Manufacturing industries (binary code) | 0.299 | 0.458 | 0 | 1 |
| Building | Building industries (binary code) | 0.095 | 0.293 | 0 | 1 |
| < 10 | Plant of less than 10 workers (binary code) | 0.175 | 0.380 | 0 | 1 |
| 10-49 | Plant with 10 to 49 workers (binary code) | 0.196 | 0.397 | 0 | 1 |
| 50-99 | Plant with 50 to 99 workers (binary code) | 0.073 | 0.260 | 0 | 1 |
| 100-199 | Plant with 100 to 199 workers (binary code) | 0.077 | 0.266 | 0 | 1 |
| 200-499 | Plant with 200 to 499 workers (binary code) | 0.078 | 0.269 | 0 | 1 |
| > 500 | Plant with more than 500 workers (binary code) | 0.401 | 0.490 | 0 | 1 |
| Change | Job change within the firm (binary code) | 0.305 | 0.460 | 0 | 1 |
| Tenure | Worker's tenure within the firm | 130.288 | 119.361 | 0 | 552.000 |
| Experience | Worker's experience in others firms | 106.085 | 116.757 | 0 | 576.000 |
| LARU | Local Area Rate of Unemployment between 2002-2003 | 8.109 | 1.394 | 6.100 | 12.200 |
| Permanent | Permanent contract (binary code) | 0.941 | 0.236 | 0 | 1 |