ESTIMATING THE RETUNS TO EDUCATION IN CAMEROON INFORMAL SECTOR [(♣)]

NGUETSE TEGOUM Pierre[*]

October 2009

Abstract

This paper discusses the returns to schooling in Cameroon within the informal sector. This is to evaluate to what extent having followed the basic education successfully influence the hourly earnings of workers in the informal sector. Then it analyzes the benefits of the first cycle of secondary education. The methodology used is based on the matching methods and the selection on unobservable models. The data used are those of the survey on employment and the informal sector conducted in 2005 by Cameroon National Institute of Statistics.

The results confirm the positive impact of schooling on the income of informal sector workers. The benefits brought by the completion of basic education are estimated at 20% in the non-agricultural sector and at 28% within the agricultural sector. In addition, if unskilled workers now return to school and obtain the FSLC (or an equivalent certificate) this will increase their income by 22% to 25%. The effects of the possession of the GCE-OL on the income of the workers of the non-agricultural informal sector are estimated to 33%. But in the agricultural sector this certificate may have no effect on earnings. Education also plays a fundamental role in the occupation status of individuals. The probability of entering into the formal sector increases with the level of education.

The study recommends the implementation of measures aiming to increase education supply, especially in rural areas, improving the quality of education. There government should implement a national social education policy in favor of the poor; take measures for the follow up of the informal sector and facilitate the entry of young graduates into the labour market.

JEL Classification: I21, J24, J31

Key words: Returns to education; Informal sector; Selection model; Propensity score matching; Cameroon

---

[(♣)] With this paper I have received the Fist Prize of the Jan Tinbergen 2009 Competition for best papers prepared by young statisticians from developing countries. This competition is organised by the International Statistical Institute (ISI). http://isi.cbs.nl/awards-prizes.htm

This paper has benefited from the technical assistance of Observatoire Economique et Statistique d’Afrique Subsaharienne (AFRISTAT). I would specifically like to thank *Siriki Coulibaly, Aude Vescovo and Christophe Nordman* for their contributions.

[*] Cameroon Ministry of Economy, Planning and Regional Development (MINEPAT, Yaoundé) Email : de_pnguetse@yahoo.fr

1. Introduction

Cameroon has experienced a serious economic slump from 1984 -1993, resulting from, a fall in oil prices and the principal cash products (cocoa, coffee). The financial tensions forced the government to the liquidation and restructuring of many public and parastatals companies and to downsize the civil service. These measures have contributed to deteriorate the labour market and the living conditions of people.

The growth recovered in 1994, following the devaluation of the CFA franc and the initiative for heavily indebted poor countries failed to raise the standard of living of Cameroonians. Indeed, data from the second Cameroon Households survey (ECAM°2) indicate a poverty rate of 40% (INS, 2002). The results of the survey on Employment and the Informal Sector (EESI), in 2005, indicate a situation of under-employment and widespread informal activities (75.8% and 90.4% respectively, INS, 2005a). Because work is the main source of income of Cameroonians, employment should be taken into account in strategies to fight against poverty. That is why the sixth axis of Cameroon Poverty Reduction Strategy Paper (PRSP, 2003) focused on the development and implementation of a national employment policy incorporated to the national poverty strategy.

In this perspective, the question of the benefits of schooling is essential: education significantly influences the income of workers which in turn, determines the state of poverty of households. In fact, education influences the hourly earnings of workers of all the institutional sectors (public, private, formal, non-agricultural informal and agricultural informal). For example, in the agricultural sector, the average hourly income in the main activity increased from 80 CFA francs for non-school workers to 223 CFA francs for workers with a bachelor's degree or higher (INS, 2005a). Furthermore, the educational level of the head of household significantly influences on the likelihood for a household to be poor (INS, 2002). In addition, schooling plays an important role on the mode of absorption of individuals into Cameroon’s labour market. For example, the unemployment rate or the duration of unemployment increases with the level of education (INS, 2005a).

The main objective of this article is to evaluate the benefits of a worker of the informal sector in terms of income, for having completed primary education successfully; that is, obtaining at least the First School Leaving Certificate (FSLC) or an equivalent certificate. It also seeks to estimate the benefits of the first cycle of secondary education (obtaining the General Certificate of Education Ordinary Level (GCE-OL)).

For a long time, as noted by Bennell (1996), many studies on returns to schooling in developing countries were limited to workers of the formal sector while ignoring the informal sector where the benefits of education were supposed to be very low. Then the role of primary education in the informal sector income was recognized with sometimes the reservation that primary education had an impact on informal sector income only if the primary education was completed (3 years of schooling not worth not more than 0 years, what mattered was the completion of this cycle). These findings have prompted international organizations to advocate for the completion of primary education at the expense of advocacy of school attendance. Today, the role of secondary education on informal sector income is also proven (Keupie *et al*, 2008).

In practice, there are two types of methods for assessing the impact: experimental methods and quasi-experimental or non experimental methods. For both methods, there are two groups: the treated group made up of individuals who have received treatment (here, having obtained FSLC) and the non-treated group (control group) comprising individuals who have not received treatment (here, not having obtained the FSLC).

When exposure to treatment is random (case of experimental methods), then, the analysis of the impact of a policy can be measured simply by comparing the average score of the two groups (here, the average income of both groups). If the treatment is not random, as is the case here because the right to obtain the FSLC is not random, this simple comparison is biased. That is why the ordinary least squares method (OLS) proposed by Mincer (1962), naively used, leads to non-convergent estimators because of the endogeneity of education (Heckman *et al*, (1997); Blundell *et al*, (2000), Sianesi Barbara, (2002)). Indeed, there are variables affecting both the fact of obtaining the FSLC and the level of income. It is then necessary to use impact evaluation methods that take into account the differences ex-ante of the variable of interest between individuals of the treated group and individuals of the control group; it is the selection to treatment. These methods are: the matching method, the instrumental variables method, selection models and the double differences method.

The advantage of the matching method is that it is nonparametric, therefore does not conjecture on income and residuals distribution. However, it is based on the assumption that the selection is based on observable variables. This is to say that, what distinguishes a person with the FSLC and an individual not having it can be observed. But this assumption is very strong. There may be unobservable variables affecting the school courses and earnings (for example the intellectual quotient). The verification of this hypothesis requires a lot of variables describing the status of the individual before treatment and predicting the likelihood of obtaining the FSLC.

Accepting this hypothesis obviously solves the problem of endogeneity of the model. Thus, it is not necessary to have an instrument as it is the case in instrumental variables method (IVM)[1], which requires having at least one instrument that affects education, but that does not affect income other than through education. Matching method does not require instrumentals variables. The matching on propensity score summarizes the information contained in a large number of variables explaining exposure treatment into a probability to be treated Rausenbaum *et al*, (1983).

---

[1] See Altonji and Dunn (1996) ; Behrman *et al*. (1996) or Bonjour *et. al*. (2000), for applications of instrumental variables methods.

However, this matching has certain limits. Because obtaining the status of independence that allows the identification of parameters may require the introduction of too many conditional variables; those variables are not always accessible. The relevance of the analysis is also reducing, because the possibilities of matching an individual to another are reduced, when we better explain exposure to treatment. In addition, the matching on observable method is mechanical and based on a purely statistical property, which in practice is difficult to justify from the behaviour of agents Bruno Crépon, (2005). It may be preferable to model jointly the potential earnings of workers and schooling. This yields to the selection on unobservable model which is a parametric model.

The rest of the document is divided as follows. Section 2 presents the data used and the characteristics of informal sector workers. Section 3 presents the models used for estimating the impact of education on the income of informal sector workers. Section 4 presents the results of which the conclusions and recommendations are presented in section 5.

2. Econometric models

2.1 Presentation of the model for the evaluation of the impact of the possession of the FSLC on the earnings of informal sector workers

The causal model of Rubin (1977) is the canonical model for assessing the treatment impact.
Let us note *T* the variable indicating whether the individual is treated or not:

$$T = \begin{cases} 1 & \text{If the worker posses the FSLC (Traeted)} \\ 0 & \text{Otherwise (Non-treated)} \end{cases}$$

The earning of a worker *i* may be expressed as follows:

$$Y_i = T_i.Y_{i,1} + Y_{i,0}(T_i - 1) \qquad (1)$$

If $T_i = 1$, the individual is treated, $Y_i = Y_{i,1}$ .Only $Y_{i,1}$ is observed. The income $Y_{i,0}$ of the individual if he had not been treated is not observed.

If $T_i = 0$ the individual is non-treated, $Y_i = Y_{i,0}$. Only $Y_{i,0}$ is observed. The income $Y_{i,1}$ of the individual if he had been treated is not observed.

$Y_{i,0}$ and $Y_{i,1}$ are the potential results of treatment; but they are never observed simultaneously at the same date for a given individual.

The causal effect of the possession of FSLC on incomes is: $\Delta = Y_{i,1} - Y_{i,0}$. It represents the difference between what would be the situation if the individual was treated and what it would be if he was not.

This effect is unobservable, since only one of the two potential variables is observed for each individual and it is individual; because of this there is a distribution of the causal effect in the population.

Three parameters are studied:

The average effect of education in the population of educated workers: $\Delta^{TT} = E\left(Y_1 - Y_0 \middle| T = 1\right)$

The average effect of education among the uneducated workers: $\Delta^{TNT} = E\left(Y_1 - Y_0 \middle| T = 0\right)$

The average treatment effect in the population: $\Delta^{ATE} = E\left(Y_1 - Y_0\right)$.

The selection bias in treatment

If the outcome variables are independent of the treatment variable, this is if $\left(Y_0, Y_1\right) \perp T$, then these three parameters of interest are identifiable and equal.

$$\Delta^{TT} = \Delta^{TNT} = \Delta^{ATE} = E\left(Y \middle| T = 1\right) - E\left(Y \middle| T = 0\right) \qquad (2)$$

They may simply be estimated as the difference of the average incomes observed in the group of workers having the FSLC and the group of workers who do not have it. If the independence assumption is no longer satisfied, the difference in average incomes is affected by a selection bias. Indeed,

$$\begin{aligned} E\left(Y \middle| T = 1\right) - E\left(Y \middle| T = 0\right) &= E\left(Y_1 \middle| T = 1\right) - E\left(Y_0 \middle| T = 0\right) \\ &= E\left(Y_1 \middle| T = 1\right) - E\left(Y_0 \middle| T = 1\right) + E\left(Y_0 \middle| T = 1\right) - E\left(Y_0 \middle| T = 0\right) \\ &= \Delta^{TT} + B^{TT} \end{aligned} \qquad (3)$$

Where:

$B^{TT} = E\left(Y_0 \middle| T = 1\right) - E\left(Y_0 \middle| T = 0\right)$. This term is the selection bias. This bias would have been zero if the average income of educated individuals was, in absence of treatment, equal to the one of non-educated workers. In other words, if educated and non-educated workers were similar before treatment. The full independence between the potential outcomes (Y_0, Y_1) and obtaining the FSLC is a highly unlikely scenario.

2.2 Matching on observable characteristics

The alternative to solve the problem of independence is to find a set of observable variables *X* with which the conditional independence between the potential results and obtaining the FSLC is verified; ie we must find the vector *X* such that, $\left(Y_0, Y_1\right) \perp T \middle| X$. This will then make the identification of the parameters of interest be possible.

Each educated individual is associated to a non-educated person called counterfactual, with identical or very close *X* characteristics. The counterfactual individual represents what would have been the situation of the treated individual if it had not been treated. Δ^{TT} can therefore be estimated by the difference between the average income of the group of educated workers and the counterfactual group.

The matching method originally proposed by Rubin (1977) is to match every educated worker *i* to a non-educated person, noted *ĩ(i)*, with the same observable characteristics *X*. For some treated individuals, we can not find an individual having exactly the same characteristics. The estimator proposed by Rubin consists to choose a non-treated individual as close as possible to the treated individual[2].

Matching Methods on the propensity score

The conditional independence property generally requires taking into account a significant number of conditional variables. This problem is solved in part by Rosenbaum and Rubin (1983) showing that if $(Y_0, Y_1) \perp T|X$ then, $(Y_0, Y_1) \perp T|P(X)$ where $P(X)$ is a one-dimension vector summarising the vector *X* of observables. It is therefore sufficient to match individuals on the propensity score, $P(X)$. But once the score $P(X)$ is estimated, it must verify the balance property, that is to say that individuals with the same propensity scores have the same distribution of observable variables irrespective of the status of treatment. We will test the balance property with the algorithm developed by Andrea Ichino and Sascha Becker (2002).

Among the matching methods used, the easiest is the one-to-one matching with replacement. It associated to each treated a non-treated individual having characteristics very close to its own (an individual of the control group can be used more than once). The difference between the logarithms of average hourly earnings of the two groups (of equal size) is then an estimation of the effect of schooling on those who have successfully completed this level education[3]. However, the asymptotic properties (convergence and asymptotic normality) of the estimator Δ^{TT} are unknown.

It is why we will also implement the Epanechnikov kernel matching which Heckman, Ichimura and Todd (1998) have shown its convergence (at a speed of $\sqrt{N}$) and asymptotic normality under certain assumptions of regularity. This method consists to associate an educated individual with a fictional non-educated person, an average person. All non-educated individuals quite close to the educated individual *i* participate in the construction of counterfactual income, with an importance that varies depending on the distance between their score and that of the educated worker *i*. The counterfactual is done with all

---

[2] In practice, the resemblance is measured by Mahalanobis distance. We choose like counterfactual of *i* the individual $\tilde{i}$ verifying: $\tilde{i}(i) = \arg\min_{T_j=0} \|x_i - x_j\|_{\Sigma^{-1}}$ where Σ is the variance-covariance matrix of the characteristics *X* in the population of the treated individuals.

[3] To estimate what the non educated individuals lose because of the lack of education (Δ^{TNT}), one can take back the process while considering like control group the educated individuals.

individuals who are within a given bandwidth *h*. We test the sensitivity of results to multiple values of this parameter.

Observable variables on which the matching will be carried out are first of all variables related to the worker's father when the worker was 15 years old. They are a proxy of the situation of the worker before treatment. It is about the social professional category of the father (High rank officer; Employee; *Independent in reference*), the sector of activity of the father (dummy that takes 1 if the father worked in the formal private sector, 0 otherwise), the branch of activity of the father (Commerce/Industry; Services; *Agriculture/Fisheries/Breeding in reference*) and the level of education (Secondary and higher; Primary; *No education in reference*). On the other hand, will be introduced in the model, individual characteristics which are beyond or independent of his current situation. These variables are: age (and age squared), gender, religion (Christian, Muslim; *Other/No religion in reference*) and the place of birth (Headquarter of province; Headquarter of division/subdivision; *Village in reference*). But finally, we will only keep those of the variables permitting to get a score verifying the balance property.

2.3 Selection on unobservable model

The matching methods are based on the assumption that everything that differentiates educated individuals from non-educated individuals is observable. It is possible that unobservable variables (or variables not available in the database) affect both the likelihood of obtaining the FSLC and the level of income. Thus, we will use the selection on unobservable model, which is another alternative for solving the problem of selectivity. This model has the advantage of modelling simultaneously the potential earnings and the likelihood of obtaining the FSLC. We will implement the following selection model[4]:

$$\begin{cases} Y_i = \beta' Z_i + \Delta T_i + u_i \\ \begin{cases} T_i = 1 & \text{if} \quad \gamma' X_i + e_i > 0 \\ T_i = 0 & otherwise \end{cases} \end{cases} \qquad (4)$$

u_i and e_i follow a bivariate with mean zero and a correlation coefficient.ρ.

We will estimate equation (4) in two stages. We will first estimate the probit model to get a value of the inverse Mills ratio (*Lambda*); then, this variable will be included as an independent in the earnings equation in the second stage.

2.4 Determinants of sectoral allocation and Selection test at the entry into the informal sector

Labour markets in developing countries are segmented, with each having its own specificities as regards to the level of demand, the job quality, the structure and the level of wages (Adams (1991); Schultz

[4] See Heckman (1979).

(2004)). The labour market segmentation in Cameroon can be defined by four sectors: the public sector, the private formal sector, the non-agricultural sector and the agricultural sector. A person of working age may be in one of the following six situations:

0 = be inactive; 1 = active and be unemployed; 2 = work in the public sector; 3 = work in the private formal sector; 4 = work in the informal non-agricultural sector 5 = work in the formal agricultural sector

The determinants of this "choice" can be estimated using a multinomial logistic model. Let L be the variable indicating the situation of working age persons. The utility of being to the institutional sector j is noted U_{ji} and assumed linear, in Q_i, a vector of observable characteristics of the individual i

$$U_{ij} = \phi_{j\,j}^{'} Q_i + \varepsilon_{ij} \qquad (5)$$

The probability that the individual belongs to the sector j_0 is the probability that the utility U_{ij_0} gained from membership in this segment j_0 is higher than the levels of utility that he will reach in the j other sectors, with $j \neq j_0$.

$$\begin{aligned} &\forall\, j_0 = 0,...,5: \quad P\left(L = j_0\right) = P\left(U_{ij_0} > U_{ij}, j \neq j_0, j \in \left[0;5\right]\right) \\ &\forall\, j_0 = 0,...,5: \quad P\left(L = j_0\right) = P\left(\left(\phi_{j_0}' - \phi_j'\right) Q_i > \nu_{ij} - \nu_{ij_0}, j \neq j_0, j \in \left[0;5\right]\right) \end{aligned} \qquad (6)$$

Assuming that the errors terms ε_j are independent and identically distributed according to a Weibull distribution, then the residuals difference follows a logistic distribution and the likelihood of being in the sector j_0 is given by:

$$P\left(L_i = j_0\right) = \exp\left(\phi_{j_0}'.Q_i\right) \Big/ \sum_{j=0}^{5} \phi_j'.Q_i \qquad (7)$$

For the model to be identifiable, ϕ_0 is assumed to be zero.

The effect of a variable q on the probability of belonging to any segment j is given by odds ratios (OR).

For a dummy variable: $OR\left(q, j\right) = P\left(L = j \middle| q = 1\right) \Big/ P\left(L = j \middle| q = 0\right)$

For a continuous variable: $OR\left(q, j\right) = P\left(T = j \middle| q = n+1\right) \Big/ P\left(T = j \middle| q = n\right)$

Paid workers of the informal sector are not chosen randomly in the working age population. The restriction of earnings equation on these workers is therefore potentially biased by a selection at the entry into the informal sector. In this case where the selection variable has several modalities, Lee's model (1983)[5] which is an extension of the Heckman method helps to estimate the earnings equations while testing the hypothesis on selection of segment.

For the implementation of this model will not consider the treatment endogeneity bias induced by unobservable variables. It is not easy to control both treatment endogeneity and selection to the entry into institutional sectors. In addition, the modality *inactive* and the modality *active unemployed* will be grouped together. Moreover, it is also necessary to assimilate unpaid workers (mostly family-aids and apprentices) to inactive or unemployed persons, since they will not belong to the estimation of the earnings equations (since the dependent variable is the logarithm of the hourly income of the main job)[6].

Thus, in the multinomial logit model at the first stage we have the five following modalities:
0 = unpaid (inactive, unemployed and unpaid worker); 1 = paid worker in the public sector; 2 = paid worker in the formal private sector; 3 = paid worker in the private sector and non-agricultural paid workers 4 = paid worker in informal private sector agriculture.

The earnings equation is then written:

$$Y_{ij} = \alpha_j + \beta_j' Z_i + \Delta_j T_i + u_{ij} \quad \text{j=1,2,3,4} \qquad (8)$$

Y_{ij} occurs only if the sector *j* is chosen by the individual *i*.

But, there is a bias because, the residuals u_{ij} are correlated to the residuals (ε_j) of the sectoral allocation equation. We will therefore estimate the following equation:

$$Y_{ij} = \alpha_j + \beta_j' Z_i + \Delta_j T_i + \lambda_{ij} + \kappa_{ij} \quad \text{j=1,2,3,4} \qquad (9)$$

λ_j corrects the selection bias created by the fact that belonging to the sector rather than another may potentially be the result of the action of unobservable variables. In equation (9), residuals κ_j are now independent of error terms ε_j of sectoral allocation equation. We will implement this equation with Bourguignon et al. (2004)'s Stata program[7].

In the regressions, we will make use of exploratory data analysis (EDA) techniques in order to obtain results that are stable, insensitive to possible outliers that could bias the estimates. (Tukey (1997); Bienias *et al* (1994)).

[5] This model doesn't pose a problem even when Assumption IIA (Independence of Irrelevant Alternatives) is not verified.

[6] See Kuepie *et al*. (2008).

[7] This program is available at the following address: http\\:www.pse.ens.fr\senior\gurgand\selmlog13.htm

3. The data and some descriptive statistics of the labour market

3.1 Presentation of the survey

The data used are those of the Survey on Employment and the Informal Sector (EESI) conducted in 2005 by the National Institute of Statistics of Cameroon. This is a nation wide operation with two phases. During the first phase, was collected the socio demographic and employment data. The second phase is a survey on non-agricultural informal production units identified during the first phase. The methodology of the survey EESI is actually that of phases 1 and 2 of a 1-2-3 survey; meaning that phase 3 on household consumption was not done. Only data from the first phase are used here.

The sampling data set used for the survey is the result of the mapping of the third General Population and Housing Census conducted in 2005. A sample of 8°540 households had been drawn following a two degree stratified survey design (stratification is done according to the ten provinces and the area of residence; that is: urban, semi-urban or rural).

The working age population is, in accordance with international recommendations, all individuals aged 15 years and above. The concept of informal sector chosen for the EESI survey is the one adopted by the 1993 System of National Accounts (set of international standards to establish a framework for the production of national accounts statistics). The distinction between sectors is made at the enterprise level, on the basis of administrative record and the fact of keeping formal accounts. The informal enterprises (or informal production units (IPU)) are those that do no have a taxpayer number and / or do not keep formal accounts. Informal sector workers are persons exercising their main job in informal establishments.

The informal sector can be divided into two segments: the agricultural sector and non-agricultural sector. The agricultural informal sector includes workers of informal production units whose main activities are: agriculture, livestock (including poultry) and the manufacture of products of animal origin, hunting, fishery and pisiculture. The non-agricultural sector comprises workers engaged in non-agricultural IPU (industry, commerce, services).

The variable of income used in the estimations is the logarithm of hourly income based on the declared monthly income and the number of hours worked. Income includes salary, end of the year bonuses, profit sharing, paid leaves, benefits in kind. For self-employed, it refers to the profit or the mixed income of their production unit. And for dependent employees (apprentices and family-aids), their earnings is the sum of bonuses in cash or in kind they received if these elements have a regular character.

3.2 Descriptive Statistics

The Cameroonian informal sector employs 89.4% of Cameroonian workers aged of 15 and above. Workers of this sector are younger than those working in the formal sector. The average age is 32.6 years in the non-agricultural sector and 37.2 years in the agricultural sector against 37.8 years in the formal sector. Women constitute the main workforce of the informal sector. They represent half the workforce of non-agricultural informal enterprises and 53.9% of the workforce of the traditional primary sector. Conversely, in the formal sector, only one worker out of four is a woman (24.4%).

Workers of the formal sector are more educated and more skilled than those of the informal sector. However, the workforce of the non-agricultural sector is relatively qualified as 56% of the workers of this informal sector have completed primary education and 4, 4% of them have obtained the GCE AL, degree or a higher education certificate. But in the agricultural sector, workers are generally less educated because almost three quarters of them have not obtained the FSLC.

The number of hours worked per week is higher in the non-agricultural sector than in the agricultural sector, due to the fact that agricultural activities are more constrained by the length of the day. Moreover, the effect of the possession of FSLC is different in the two sectors since it increases the weekly working time in non-agricultural activities and decreases it in agricultural activities. The formal sector workers work in average more than those of the informal sector.

The average income (monthly or hourly) in the non-agricultural sector is more than twice that of agricultural sector. Possession of the FSLC increases the average hourly income of 38% in the non-agricultural sector. This increment is about 79% in the informal agricultural sector; a worker of the informal agricultural sector earns more if he has graduated, but he also works longer. Incomes in the formal sector are very high compared to those of the informal sector, in fact, a worker of the formal sector averagely earns 3.9 times more than one exercising in non-agricultural sector and 8.8 times more than a worker of the rural sector.

Table 1 : Workers' characteristics and jobs' characteristics according to the institutional sector (15 years old and above)

Variables		Institutional sector		
		Formal	Non-agricultural informal	Agricultural informal
Percentage of persons working in the sector		10.7	37.1	52.2
Average age		37.8	32.6	37.2
Proportion of women		24.4	49.4	53.9
Average number of school years completed		10.9	5.9	3.4
Level of education	No education	1.1	19.4	35.8
	Primary	17.8	40.0	46.8
	Secondary 1er cycle	23.9	28.0	14.4
	Secondary 2nd cycle and +	57.3	12.6	2.9
	Total	100.0	100.0	100.0
Highest certificate	No certificate	6.9	43.8	73.8
	FSLC	29.3	40.4	22.5
	GCE-OL/PROBATOIRE	25.0	11.4	2.9
	GCE-AL and +	38.9	4.4	0.8
	Total	100.0	100.0	100.0
Number of hours worked per week	No certificate	50.9	41.7	39.1
	FSLC and more	44.2	45.0	32.1
	Total	44.6	43.6	37.1
Average monthly income (Average hourly income) (in CFAF)	No certificate	51°659 (249)	22 902 (162)	11 485 (86)
	FSLC and more	118°433 (713)	32 150 (224)	15 942 (154)
	Total	113°847 (682)	28 263 (198)	12 771 (105)

Source : EESI (2005), Phase1. Our calculations; weighted data.

4. Results

The results are based on the sample of informal sector paid workers, whose fathers were alive when they were 15 years and have briefed the question of the level of the father.

4.1 Naive estimate of the effect of FSLC on earnings in the informal sector

The OLS regressions carried out indicate that whatever the segment, the model is globally significant at 1%. The variable FSLC is also significant at 1%. But all the variables are not significant in both models. Indeed, the variable *potential experience* is not significant in the agricultural informal segment and variable *religion* is not in the non-agricultural segment. We note that several other factors (such as the characteristics of production units) may also explain the income of workers in the informal sector in Cameroon, since the two models explain less than 13% of income dispersion.

The results show that the effect of FSLC on workers of the informal sector is quite important, particularly in the agricultural segment where it reached 38% against 30% in the non-agricultural segment. But these estimates are biased because the possession of FSLC is not randomly distributed, it depends on certain factors that can be observed or not observed. We will therefore proceed to a selectivity bias correction on observable variables then a correction on unobservable variables.

Table 2: Naive estimation of returns to education in the informal sector: OLS

Variables	**Non-agricultural**		**Agricultural**	
	Coefficient	Standard error	Coefficient	Standard error
FSLC	0.30***	0.04	0.38***	0.05
Potential experience	0.01***	0.01	0.01	0.01
(Potential experience)^2 /100	-0.01	0.01	-0.01	0.01
More than 32 years old	0.18***	0.04	0.21***	0.06
Female	-0.30***	0.03	-0.45***	0.05
Union (Married or Free union)	0.06**	0.03	0.08*	0.05
Christian	0.05	0.07	0.20**	0.08
Moslem	0.01	0.07	0.11	0.10
Migrant	0.07**	0.03	0.08**	0.05
Urban milieu	0.12***	0.03	0.27***	0.06
Constant	4.52***	0.08	3.92***	0.10
Statistics of the model				
Adjusted R2 (%)	11.3		12.8	
Observations	2°391		1°571	

Source : EESI (2005), Phase1. Our calculations.
: significant at 10 %; **: significant at 5 %; *: significant at 1 %.*

4.2 Correction of treatment selectivity based on observable variables: matching

Estimating the probability of obtaining FSLC

Table 3 summarizes the results on the estimation of the determinants of possession of FSLC. Whatever the segment, statistics on the quality of the model are satisfactory. In the non-agricultural sector, 44% of the variability is explained by the model against 27% in the rural sector. In addition, the model is able to allocate at least 83% of individuals in their observed categories.

The main variable on worker's father that explains the school course of a worker of the informal sector in Cameroon is the level of education. Indeed, a child whose father had a level of primary education had four times more chances to obtain the FSLC compared to a child whose father had not been to school. This odd ratio is more than seven if this individual is compared to a worker whose father had reached the secondary education. For workers of the non-agricultural informal sector, the institutional sector and the branch of activity of the father have also influenced their school attendance. Because their parents were mostly working in the informal sector, they were less paid than those working in the formal sector. Therefore, they did not have enough financial resources to enrol their children at school.

Individual characteristics such as age, sex, religion and place of birth have also influenced the likelihood of obtaining FSLC. Whatever the segment, the effect of age is concave, with a pick around 32 years. This means that a child born around 1971 was more likely to get the FSLC than a child born before or after. Persons born in rural areas were less likely to get the FSLC than those born in urban areas (province, division or subdivision headquarters); they were generally used as labour force in farm activities. This report helps to show the negative impact of child labour on child education. Finally, men have got at least two times more chances to get the FSLC than women, because of gender discrimination and certain traditions/customs that still hamper the education of Cameroonian young girls.

<u>Table 3:</u> Estimated propensity for workers of the informal sector to obtain the FSLC

		Non-agricultural		**Agricultural**	
		Odds-ratio	SE	Odds-ratio	SE
Variables related to the father					
CSP (ref. : Self-employed worked)	*High rank officer*	1.38	0.38	0.74	0.27
	Employee	1.00	0.26	0.53**	0.17
Sector of activity (ref: formal)	*Informal*	1.50	0.38	2.16***	0.70
Branch of activity (ref. : Agriculture, Fishery, Hearing)	*Commerce/Industry*	1.85***	0.32	1.46*	0.34
	Services	1.61***	0.31	1.54*	0.41
Level of education (ref. : No education)	*Primary*	4.91***	0.71	4.27***	0.63
	Secondary and +	11.34***	2.86	7.52***	2.56
Variables related to the worker					
Age		1.34***	0.04	1.14***	0.03
Age squared		1.00***	0.00	1.00***	0.00
Female		0.31***	0.04	0.41***	0.05
Religion (ref. : Other/No religion)	*Christian*	2.37***	0.60	2.53***	0.65
	Moslem	0.15***	0.04	0.18***	0.07
Lieu de naissance (ref. : village)	*Headquarter of Division/Subdivision*	1.90***	0.26	2.00***	0.30
	Headquarter of province	1.98***	0.37	0.62	0.22
Statistics of the model					
Pseudo R^2 (in %)		44.0		27.0	
Area Under Roc Curve (AURC) en %		90.7		83.1	
Observations		2°382		1°581	

<u>Source</u> : EESI (2005), Phase1. Our calculations.
: significant at 10 %; **: significant at 5 %; *: significant at 1 %.*

Distribution of the propensity score and analysis of the common support

With the variables used to model the likelihood to obtain FSLC we have implement the algorithm of Ichino and Becker (2002) to identify variables permitting to have a score verifying the balance propriety. Results show that, whatever the segment all the initially selected variables are balanced at the threshold 0.1%, thus these variables were retained in the computation of the propensity score.

The propensity score is simply the predicted probability of obtaining the FSLC derived from the logit equation modelling the likelihood of obtaining the FSLC. Individuals are matched with respect to the segment to which they belong: non-agricultural workers on one side and agricultural workers on the other side. Before matching, it is necessary to analyze the spectrum of the score distributions in both groups (treated and non-treated) to identify individuals who fall within the common support.

The common support has been determined using the rule of min-max. This rule compares the minimum and maximum score in both groups (treated and non-treated). Individuals who are on the common support are those whose score is equal or greater than the maximum of the minimum values and less or equal to the minimum of maximum values. The application of this rule shows that whatever the institutional sector, more than 95% of individuals are on the common support. Therefore, individuals from both groups have close characteristics looking to observable variables.

Table 4: Statistics on the score and the common support

	Non-agricultural informal		Agricultural informal	
Score	Treated	Non-treated	Treated	Non-treated
Minimum	0.0623	0.0001	0.0474	0.0002
Maximum	0.994	0.988	0.986	0.956
common Support				
Observations	938	1°471	602	997
Percentage	95.1		96.4	

Source : EESI (2005), Phase1. Our calculations.

Matching

We tested two matching methods: one-to-one matching with replacement and Epanechnikov kernel matching. Both methods have been restricted to the support common because the inclusion of individuals who are out of this region biases estimates. The two techniques effectively permit to reduce the differences between the average characteristics of the treatment and the control groups. But kernel methods are more efficient. As shown in the tables A4 and A5 (in appendix), they better bring closer the two groups in terms of average characteristics. For example, considering the agricultural segment, we can note that with one-to-one matching the average characteristics of the treatment group are significantly different to those of the control group with regard to the dummy variables *Christian*, *High rank officer* and *Commerce/Industry*. Whereas, with kernel Epanechnikov matching, no average characteristic is significantly different between the two groups. Moreover, whatever the segment, there is no significant difference of the average effect of education on the treated when the bandwidth varies between 0.04 and 0.08. We finally adopted the bandwidth $h = 0.06$; this value was also used by Blundell and Sianesi Barbara (2001).

Table 4 presents the average treatment effects after matching. It shows that the returns to primary education for workers of the informal sector in Cameroon are considerable and significant at the threshold 1%. These benefits are lower than those obtained with the OLS method which is biased.

In the agricultural segment, the returns of FSLC on the hourly earnings of workers who have this certificate are estimated to 20%. In other words, if these workers had not successfully completed primary education, their incomes would have been 20% less than what they have now. Furthermore, if workers of this segment not having the FSLC had gotten it, their income would improve by 23%. So, if workers not having the FSLC return to school and got the certificate, the impact on their income would be at least equal to the initial training received by workers currently graduated, assuming that the age at which the certificate is obtained does not affect the treatment returns. The average benefit of basic education on the workers of the non-agricultural informal sector is an increase of their earnings by 21%.

In the agricultural segment, the returns to schooling are even greater. Indeed, returns of primary education on the income of workers holding the FSLC are about 28%. While, non-graduated workers would have earned 25% more if they had graduated. The average benefits of FSLC on the income of agricultural workers are estimated to 26%.

In summary, basic education plays an important role on the income of Cameroon informal sector workers. This proves the importance of human capital on income levels, poverty reduction and economic growth.

Table 5: Returns to basic education in the informal sector: matching method

Returns of FSLC	**Non agricultural**		**Agricultural**	
	Estimation	SE	Estimation	SE
On the income of workers holding the FSLC : Δ^{TT}	20.0***	4.9	27.6***	6.1
On the income of workers not having the FSLC : Δ^{TNT}	22.7***	7.2	25.1***	7.4
On the income of workers : Δ^{ATE}	21.0***	5.5	26.0***	5.9

Source : EESI (2005), Phase1. Our calculations
Bootstrapped standard errors (200 replications)
: significant at 10 %; **: significant at 5 %; *: significant at 1 %.*

4.3 Correction of treatment selectivity based on unobservable variables

The model reveals two important results. First, there is a bias induced by unobservable; because, whatever the segment, the variable *Lambda* that captures the action of unobservable variables is significant at 1%. But its negative sign indicates a negative influence of these unobservable on the income of workers. Secondly, having followed the basic education significantly influences (at the threshold 1%) the income of workers of the informal sector. The average treatment effect of the FSLC is estimated at 22% in the non-agricultural segment and it is around 28% in the agricultural segment. These effects are similar to those obtained with the matching model but are significantly lower than the values obtained with the naive model which overstates parameters.

Table 6: **Returns to basic education in the informal sector: selection on unobservable model**

Variables	Non-agricultural		Agricultural	
	Coefficient	Standard error	Coefficient	Standard error
FSLC	0.22***	0.04	0.28***	0.05
Potential experience	0.01 **	0.01	0.00	0.01
(Potential experience)^2 /100	0.01	0.01	0.01	0.01
More than 32 years old	0.22***	0.04	0.29***	0.06
Female	-0.26***	0.03	-0.36***	0.05
Union (Married or Free union)	0.08**	0.03	0.11**	0.05
Christian	-0.02	0.07	0.05	0.08
Moslem	0.18**	0.07	0.21**	0.10
Migrant	0.08**	0.03	0.08*	0.05
Urban milieu	0.08***	0.03	0.23***	0.06
Lambda	-0.29***	0.04	-0.40***	0.06
Constant	4.75***	0.08	4.39***	0.12
Statistics of the model				
Adjusted R2 (%)	13.0		15.1	
Observations	2°532		1°659	

Source : EESI (2005), Phase1. Our calculations.
: significant at 10 %; **: significant at 5 %; *: significant at 1 %.*

4.4 Returns to the first cycle of secondary education on informal sector workers

With the same methodology, we now analyze the effect of obtaining the GCE-OL (General Certificate of Education, Ordinary Level) on informal sector workers. Here the treatment group is made of workers exercising in the informal sector who have the GCE-OL or a certificate that is superior and the control group regroups workers who have just had the FSLC. This allows us to assess the net impact of the first cycle of secondary education on the incomes of workers. We apply the kernel Epanechnikov matching (taking the bandwidth $h = 0.06$) and we also implement the selection on unobservable model.

Selection on observable

Observable variables that we consider in the computation of the propensity score are the same we have used in estimating the probability of obtaining FSLC. This score assesses the likelihood of obtaining the GCE-OL conditionally to the possession of the FSLC. The examination of spectra of scores shows that in the non-agricultural segment 99.6% of observations fall within the common support against 96.5% in the agricultural segment.

The matching shows that the returns to GCE-OL on the income of workers of the non-agricultural sector who have this certificate are estimated to 33%. On the other hand, if workers who have FSLC return to school and obtain the GCE-OL, this will increase their income by 30% assuming that the age at which the certificate is obtained does not affect the benefits it provides. The average effect of the first cycle of secondary education on workers of the non-agricultural sector having the FSLC is estimated at 31%. In contrary, in the agricultural segment the average effect of GCE-OL on the income of workers may be very low, quite zero; in fact, none of the three parameters is significant.

Table 7: Returns to the secondary education first cycle in the informal sector: matching method

Returns to GCE-OL	**Non agricultural**		**Agricultural**	
	Estimation	SE	Estimation	SE
On the income of workers holding the GCE-OL : Δ^{TT}	33.0***	4.9	12.9	12.3
On the income of workers not having the GCE-OL but holders of the FSLC : Δ^{TNT}	29.9***	5.4	21.6	13.9
On the income of workers having the FSLC : Δ^{ATE}	31.0***	5.1	20.2	12.7

Source : EESI (2005), Phase1. Our calculations
Bootstrapped standard errors (200 replications)
: significant at 10 %; **: significant at 5 %; *: significant at 1 %.*

Selection on unobservable

The model selection on unobservable is justified in both segment, since the inverse Mills ratio (*Lambda*) is significant at 1% threshold, which confirms the existence of unobservable variables affecting both the possession of GCE-OL and the income. Several control variables related to the potential experience, religion and marital status are not significant. But excluding these variables does not significantly affect the other coefficients.

In the non-agricultural sector, the variable GCE-OL reflecting the possession or not of the GCE-OL is significant at threshold of 1% and it indicates that the average benefits of this certificate on the income of workers of the non-agricultural informal sector are around 31%. In opposite, in the rural sector this variable is not significant. So the possession of GCE-OL would have no impact on the income of workers exercising in agricultural activities. The results obtained with the selection on unobservable model thus converge with those of the matching method.

Table 8: Returns to the secondary education first cycle in the informal sector: selection on unobservable model

Variables	Non-agricultural		Agricultural	
	Coefficient	Standard error	Coefficient	Standard error
GCE-OL	0.31***	0.04	0.11	0.11
Potential experience	0.02*	0.01	0.02	0.01
(Potential experience)^2 /100	-0.04	0.05	-0.03	0.03
More than 32 years old	0.09*	0.05	0.27***	0.08
Female	-0.18***	0.04	0.41**	0.18
Union (Married or Free union)	0.06	0.04	-0.03	0.08
Christian	-0.09	0.09	0.22	0.18
Moslem	0.00	0.10	-0.01	0.23
Migrant	0.14***	0.04	-0.04	0.07
Urban milieu	0.05	0.04	0.06	0.09
Lambda	-0.35***	0.06	-1.02***	0.23
Constant	5.19***	0.13	5.67***	0.31
Statistics of the model				
Adjusted R2 (%)	14.9		11.1	
Observations	1°471		574	

Source : EESI (2005), Phase1. Our calculations.
: significant at 10 %; **: significant at 5 %; *: significant at 1 %.*

4.5 Selection at the entry into the informal sector and determinants of sectoral allocation

We are now testing the selection at the entry into the informal sector; since the previous earnings equations have been restricted, ignoring the existence of the formal sector; they can therefore be biased. Then, we will search for the determinants of sectoral allocation. The sample is the set of potentially active persons (aged of 15 years and above) interviewed during the survey EESI.

Selection test at the entry into the informal sector

The results of the test (see Table 9 below) show that the inverse Mills ratio (*Lambda*) is significant and positive in the equations of formal sector segments (public and private) and it is negative in the informal sector equations. But in the non-agricultural sector, this variable is not significant, even at the threshold 10%. So in formal sectors (public or private), the unobserved characteristics affecting the sectoral "choice" of an individual also affect his wage once he gets in this sector. In the rural segment these characteristics play harmfully on the potential earnings of workers and in the non-agricultural segment they have no impact. Thus, Cameroonian workers exercising in the informal sector have not made their choice so as to maximize their potential earnings as it should have been the case in a competitive market. They therefore found themselves there against their wish; because they were unable to enter into the formal sector.

<u>Table 9:</u> Selectivity test at the entry into the labour market segments

Variables	Public	Private formal	Non-agricultural Informal	Agricultural informal
FSLC	0.45***	0.22**	0.32***	0.36***
Potential experience	0.03***	0.03***	0.02***	0.00
(Potential experience)^2 /100	-0.05*	-0.04	-0.05***	0.00
More than 32 years old	0.18***	0.27***	0.20***	0.24***
Female	-0.13***	0.37***	-0.34***	-0.36***
Union (Married or Free union)	0.00	0.15***	0.09***	0.04
Christian	0.05	0.02	0.19*	0.28***
Moslem	-0.12	0.00	0.10***	0.30***
Migrant	0.14***	0.06	0.07***	0.08**
Urban milieu	0.28***	0.17***	0.17***	0.22***
Constant	6.10***	6.09***	4.31***	3.74***
Lambda *(Selectivity test)*	0.51***	0.71***	-0.04	-0.15***
Statistics of the model				
Adjusted R2 (%)	30.4	29.2	10.5	9.5
Observations	1°204	1°160	6°572	3°540

<u>Source</u> *: EESI (2005), Phase1. Our calculations.*
: significant at 10 %; **: significant at 5 %; *: significant at 1 %.*

Determinants of sectoral allocation

The sectoral allocation model presented in appendix (table A10) shows that Hausman-McFadden (1984) specification test which indicates the probability to wrongly reject the null hypothesis (of Independence of Irrelevant Alternatives) is significant at 1%. This is to say that, sectoral allocation is a random process. A working age person first chooses to participate into the labour market or to be inactive. Then, active persons are then divided between unemployed persons and workers of the different segments.

We note that variables related to family environment play a major role in the decision of individuals to participate or not into the labour market. Family responsibilities apprehended by the number of little children and the fact of being the household head persuade individuals to seek for a job. As against, the presence of inactive persons in the household negatively influences the chances of its members to enter the job market. The involvement of individuals in the informal sector is determined by variables related to family environment. The effect of these variables on the probability for persons to participate in the labour market was also highlighted by El Aynaoui, (1998) with Morocco data.

Education plays a fundamental role on the occupational status of people. The probability of being unemployed increases with the level of education; qualified people preferring to remain unemployed rather than engaging themselves into the informal sector characterized by precarious jobs and low incomes. Moreover, the probability of entering formal segments increases sharply with the level of education. But it is the opposite effect in the informal segments where the probability of entry is decreasing with the level of education.

5. Conclusion

The study envisaged to analyze the returns to schooling on workers of the Cameroonian informal sector. We have implemented matching on observables methods and selection on unobservables models to assess the effects of basic education on the income of people working in the informal sector (non-agricultural and agricultural). The study has also analyzed the benefits of the first cycle of secondary education on these workers.

The results obtained with both methods are converging and confirm the positive impact of education on the incomes of informal sector workers. The benefits brought by the completion of basic education (possession of the FSLC) are estimated to 20% in the non-agricultural sector and to 28% in the rural sector. But, if non-educated workers now return to school and obtain the FSLC (or an equivalent certificate), this will increase their income by 22% to 25%, assuming that the age at which the certificate is obtained does not affect the potential benefits it provides.

The effects of the completion of the first cycle of secondary education on the incomes of the workers of non-agricultural sector are even more important. The possession of GCE-OL helps to increase by 33% the income of those who have this certificate, while the loss of workers who have stopped at the FSLC is about 30%. The average treatment effect of GCE-OL on the income of the non-agricultural sector workers is estimated to 31%. But, in the agricultural segment the returns of this qualification would be quite zero. However, this result should be confirmed by other studies.

In addition, the selection test at the entry into the informal sector has revealed the existence of a selectivity bias affecting the results of the agricultural sector. However, the results of the non-agricultural sector are not affected by this bias. This test also showed that the Cameroonian workers exercising in the informal sector have not made their choice so as to maximize their potential earnings as it should have been the case in a competitive market. They therefore found themselves there because they were unable to enter into the formal sector.

Education plays a fundamental role in the occupation status of people. The probability of being unemployed and entering into the formal sector increases with the level of education; on the contrary the probability of entering the informal sector declines with schooling. The entering in this sector is mainly determined by the family environment.

In summary, the study puts the spotlight on the role of basic education and the first cycle of secondary education in Cameroon informal sector; which yields uncertain individual returns. Lessons learned will therefore appeal for greater accessibility to education at least until the first cycle of secondary education. The Cameroonian government should intervene to improve the access to education and its quality. Because even if primary education was declared free of charge in Cameroon since 2000, we must acknowledge that the results of this policy are not very suitable. Indeed, education supply of is still very low since there is a lack of school infrastructure in many rural areas. There is also an overcrowding of pupils, a lack of teachers, a lack of equipment etc. The consequences are high repetition rates and high dropout rates.

The Cameroonian government should recruit more teachers, build and equip schools, strengthen vocational training and also develop and implement a national social education policy to help poor parents who have no means to buy school materials for their children. In a medium term, free schooling could be extended to the first cycle of secondary education. Furthermore, the government could regulate the informal sector actors; introduce incentive programmes to encourage young graduates to enter the informal sector (for example, granting of credits, tax exemption for a number of years, etc.). The central administration may also organize free of charge vocational training (teach rural farmers new agricultural technologies with high efficiency) and implement policies for the follow-up of efficient informal production units in order to facilitate their transition into the formal sector.

REFERENCES

Adams, J. (1991), « The Rural Labour Market in Zimbabwe », Development and Change, 22 (2), pp. 297-320.

Altonji, J. and T. Dunn (1996), « Using Siblings to Estimate the Effect of Schooling Quality on Wages », Review of Economics and Statistics, 78,665-671.

Bennell, P. (1996), « Rates of Return on Education: Does the Conventional Pattern Prevail in Sub-Saharan Africa? », World Development, 24 (1), pp. 183-199.

Blundell and Sianesi Barbara, (2001), « Estimating the Returns to Education: Models, Methods and Results », Economic Journal.

Bienias Julia; Lassman David; Scheleur Scott and Hogan Howard; (1994), "Improving Outlier Detection in Two Established Surveys," American Statistical Association.

Bourguignon, F., Fournier M. and Gurgand M. (2004), « Selection Bias Corrections Based on the Multinomial Logit Model: Monte-Carlo Comparisons », DELTA working paper, September

Bruno Crépon, (2005) « Econométrie Linéaire », INSEE France.

EL Aynaoui Jean-Philippe (1998), « Participation, choix occupationnel et gains sur un marche du travail segmente : une analyse appliquée au cas du Maroc », Centre d'économie du développement, Université Montesqieu-Bordeau IV-France

Hausman J., McFadden D. (1984), « Specification Tests for the Multinomial Logit Model », *Econometrica*, Vol. 52, n° 5

Heckman J., (1979), « Sample selection bias as a specification error, *Econometrica* », Vol. 47, N°1, p. 153

Heckman J., Ichimura Hidehiko and Todd Petra. (1998), « Matching as an Econometric Evaluation Estimator ». Review of Economic Studies 65:261-294.

Ichino Andrea et Becker Sascha (2002), « Estimation of average treatment effects based on propensity score », Laboratorio R. Revelli Centre for Employment Studies.

Institut National de la Statistique (2002), « Conditions de vie des populations et profil de pauvreté au Cameroun en 2001 », Deuxième Enquête Camerounaise auprès des ménages, INS Cameroun.

Institut National de la Statistique (2005a), « Phase 1 : Enquête Emploi », Enquête sur l'Emploi et le Secteur Informel, INS Cameroun

Institut National de la Statistique (2005b), « Document de méthodologie », Enquête sur l'Emploi et le Secteur Informel, INS Cameroun.

Keupie Mathias, Nordman Christophe and Roubeau François (2008), « Education and Earnings in Urban West Africa», Journal of Comparative Economics, YJCEC 2111.

Lee, L.-F. (1983), « Generalized Econometric Models with Selectivity », Econometrica, 51(2), pp.507-512.

Mincer, Jacob (1962) « On-the-job training: costs, returns, and some implications », Journal of political economy, Supplement 1962, Vol. 70

République du Cameroun (2003) : Document de stratégie de réduction de la pauvreté (DSRP)

Rosenbaum Paul R. and Donald B. Rubin. (1983), « Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score », The American Statistician 39 (1): 33–38.

Rubin Donald B. (1977), « Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies », Journal of Educational Psychology 66 (5): 688–701.

Schultz, T. P. (2004), "Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education", Journal of African Economies, 13, AERC Supplement, pp. ii95-ii148

Sianesi Barbara (2002), « Estimating the Returns to Education », IFAU-Institute For Labor Market Policy Evaluation.

Tukey, John W., (1977), *EDA: Exploratory Data Analysis*, Addison-Wesley, MA

APPENDIX

Table A1: Distribution of workers in the informal sector by category

Informal sector	Employer/ High rank officer	Own account worker	Employee/ Labourer	Apprentice/ Family aid	Total
Non-agricultural	5.1	59.7	21.6	13.5	100.0
Agricultural	2.2	63.3	2.0	32.6	100.0
Together	3.4	61.8	10.2	24.7	100.0

Source : EESI (2005), Phase1. Our calculations. Weighted data

Table A2: Distribution of the sample of informal sector workers considered in estimating the returns to education

Survey region	Non agricultural	Agricultural	Total
Douala	422	10	432
Yaounde	359	9	368
Adamaoua	164	104	268
Centre-yde	178	163	341
East	123	179	302
Far north	273	92	365
Littoral-dla	113	134	247
North	234	237	471
North west	201	243	444
West	238	298	536
South	77	52	129
South west	150	138	288
Cameroon	2532	1659	4191

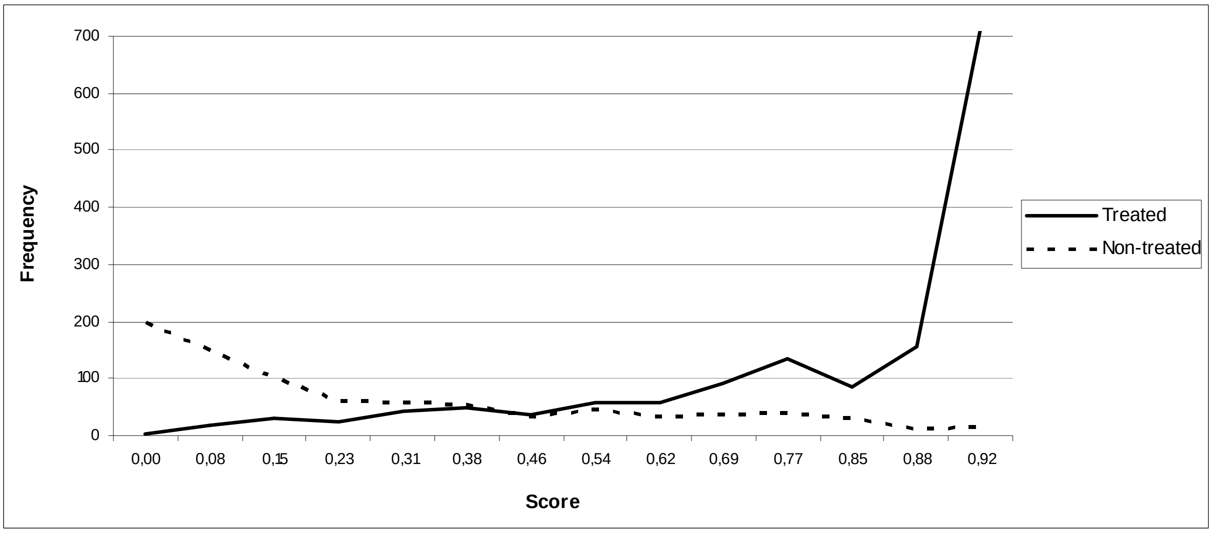
Source : EESI (2005), Phase1. Our calculations.

Table A3: Distribution of the sample of the study by the highest Certificate

Certificate	Non agricultural		Agricultural	
	Observations	Frequency	Observations	Frequency
No certificate	972	38.4	1054	63.5
FSLC	991	39.1	514	31.0
GCE-OL and +	569	22.5	91	5.5
Together	2532	100.0	1659	100.0

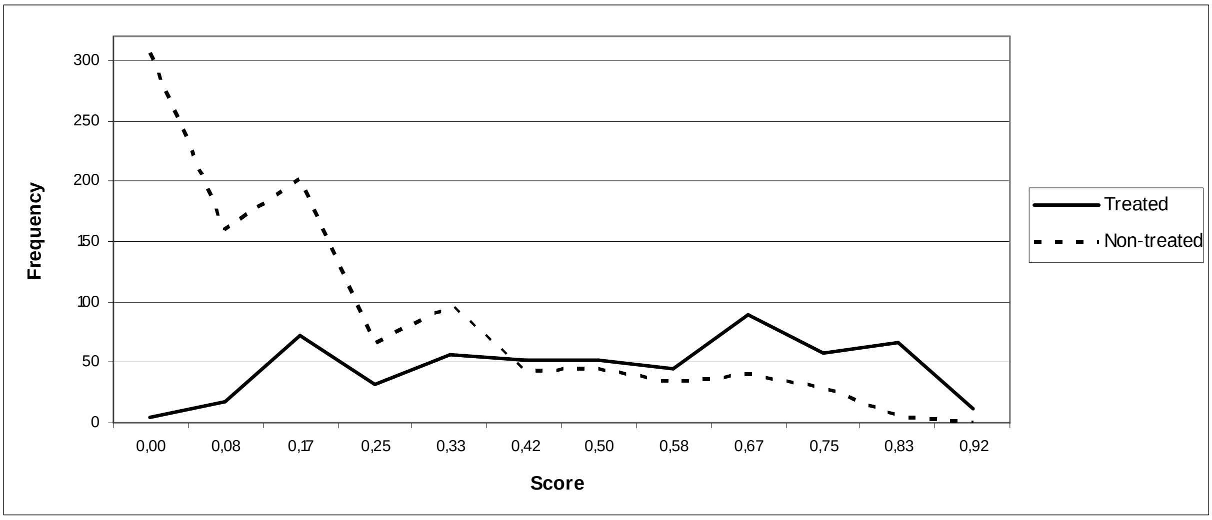
Source : EESI (2005), Phase1. Our calculations.

<u>Figure A1:</u> Distribution of the score of non-agricultural sector workers (FSLC / No FSLC)



<u>Source</u> : EESI (2005), Phase1. Our calculations

<u>Figure A2:</u> Distribution of the score of agricultural sector workers (FSLC / No FSLC)



<u>Source</u> : EESI (2005), Phase1. Our calculations

Table A4: Average characteristics of workers of the non-agricultural sector before and after matching methods

Characteristics	Before matching		After matching				
				One-to-one	bandwidth *h=0.04*	bandwidth *h=0.06*	bandwidth *h=0.08*
	TG	Difference (TG-CG)	TG	Difference (TG-CG)	Difference (TG-CG)	Difference (TG-CG)	Difference (TG-CG)
Log of the hourly income	*4.723*	*0.305****	***5.007***	***0.167****	***0.199****	***0.200****	***0.201****
Individual characteristics							
Age	36.084	-5.704***	30.407	-0.257	0.318	0.332	0.322
Female	0.468	-0.061***	0.432	0.067***	0.059***	0.057***	0.053***
Christian	0.436	0.388***	0.814	-0.035**	-0.008	-0.008	-0.008
Moslem	0.514	-0.389***	0.133	0.025**	-0.001	-0.001	-0.002
Province	0.141	0.146***	0.272	0.051***	0.053***	0.056***	0.059***
Division	0.427	0.040*	0.467	-0.018	-0.023	-0.025	-0.023
Father's characteristics							
Primary	0.215	0.232***	0.475	-0.061***	-0.046**	-0.047**	-0.047**
Secondary and +	0.063	0.239***	0.260	0.073***	0.043***	0.045***	0.046***
High rank officer	0.065	0.117***	0.154	0.004	0.017	0.018	0.018
Employee	0.148	0.172***	0.319	-0.016	-0.027	-0.026	-0.025
Commerce/ Industry	0.179	0.029*	0.211	0.021	-0.003	-0.002	-0.001
Services	0.199	0.225***	0.398	0.015	0.007	0.008	0.011
Informal sector	0.146	0.283***	0.396	-0.003	0.008	0.010	0.014

Source : EESI (2005), Phase1. Our calculations.
: significant at 10 %; **: significant at 5 %; *: significant at 1 %.*
***TG**= Treated Group* ***CG**= Control Group*

<u>Table A5:</u> Average characteristics of workers of the agricultural sector before and after matching methods

Characteristics	Before matching		After matching				
				One-to-one	bandwidth *h=0.04*	bandwidth *h=0.06*	bandwidth *h=0.08*
	TG	Difference (TG-CG)	TG	Difference (TG-CG)	Difference (TG-CG)	Difference (TG-CG)	Difference (TG-CG)
Log of the hourly income	***4.598***	***0.371****	***4.594***	***0.226****	***0.272****	***0.276****	***0.279****
Individual characteristics							
Age	36.760	-7.883***	36.784	0.339	0.389	0.301	0.252
Female	0.451	-0.103***	0.453	-0.005	0.004	-0.001	-0.002
Christian	0.894	0.213***	0.894	-0.032*	-0.020	-0.017	-0.016
Moslem	0.055	-0.166***	0.055	0.008	0.016	0.015	0.015
Province	0.043	0.014*	0.043	0.010	0.001	0.000	-0.001
Division	0.349	0.155***	0.346	0.017	0.013	0.010	0.012
Father's characteristics							
Primary	0.474	0.305***	0.477	-0.013	0.029	0.032	0.032
Secondary and +	0.099	0.078***	0.095	0.000	-0.003	-0.006	-0.004
High rank officer	0.063	0.027***	0.058	0.030**	0.014	0.014	0.014
Employee	0.197	0.116***	0.198	-0.033	-0.011	-0.006	-0.001
Commerce/ Industry	0.109	0.036*	0.110	0.038**	0.017	0.015	0.015
Services	0.203	0.126***	0.199	-0.002	0.005	0.012	0.018
Informal sector	0.213	0.149***	0.209	-0.008	0.000	0.004	0.009

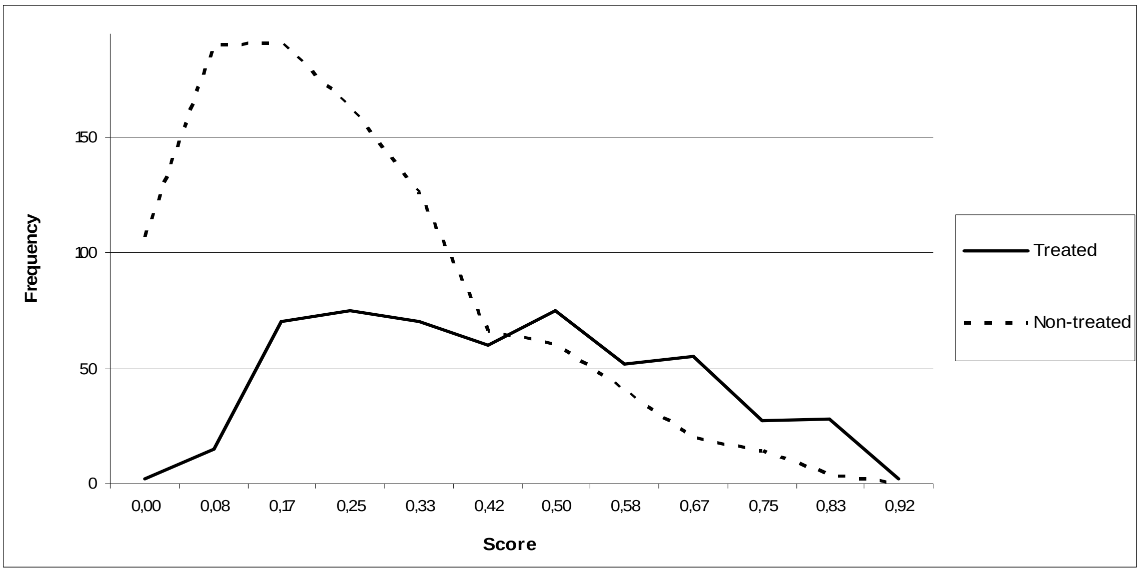
<u>Source</u> : EESI (2005), Phase1. Our calculations.
: significant at 10 %; **: significant at 5 %; *: significant at 1 %.*
***TG**= Treated Group* *__CG__= Control Group*

Table A6: Estimation of the propensity of having the GCE-OL among informal sector workers holders of the FSLC

		Non-agricultural		Agricultural	
		Odds-ratio	SE	Odds-ratio	SE
Variables related to the father					
CSP (ref. : Self-employed worked)	*High rank officer*	1.58**	0.40	0.45	0.51
	Employee	1.07	0.26	0.14*	0.16
Sector of activity (ref: formal)	*Informal*	1.43*	0.31	2.69	2.98
Branch of activity (ref. : Agriculture, Fishery, Hearing)	*Commerce/Industry*	1.12	0.21	1.33	0.72
	Services	1.23	0.24	0.60	0.38
Level of education (ref. : No education)	*Primary*	1.51**	0.26	1.77*	0.59
	Secondary and +	3.63***	0.76	1.04	0.67
Variables related to the worker					
Age		1.39***	0.06	1.05	0.06
Age squared		1.00***	0.00	1.00	0.00
Female		0.38***	0.05	0.01***	0.01
Religion (ref. : Other/No religion)	*Christian*	1.31	0.37	0.50	0.29
	Moslem	0.89	0.29	0.29	0.25
Place of birth (ref. : village)	*Headquarter of Division/Subdivision*	2.66***	0.48	2.21**	0.71
	Headquarter of province	1.58***	0.26	3.82*	2.67
Statistics of the model					
Pseudo R^2 (in %)		15.6		22.8	
Area Under Roc Curve (AURC) en %		75.7		82.9	
Observations		1°513		571	

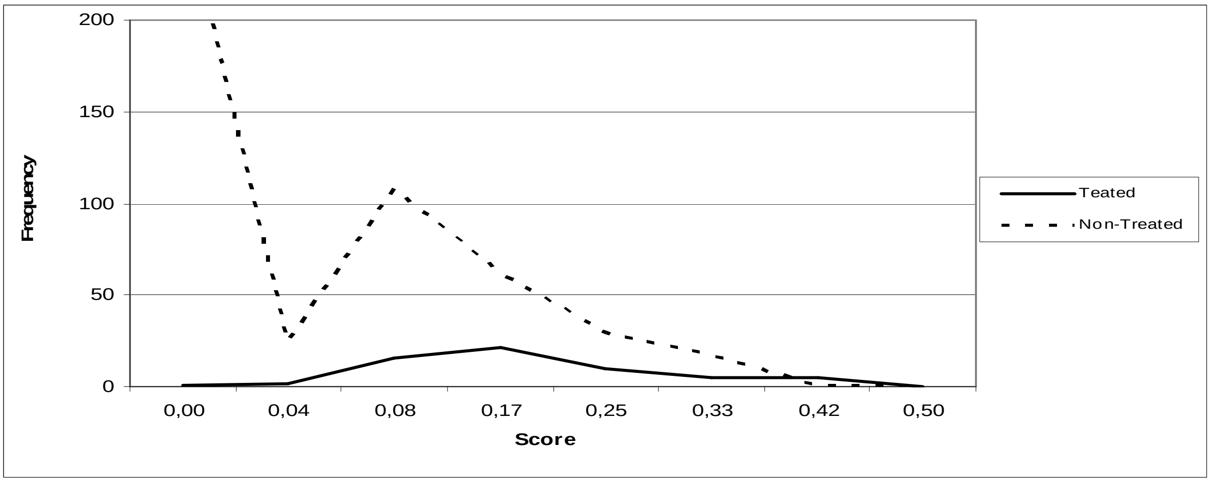
Source : EESI (2005), Phase1. Our calculations.
: significant at 10 %; **: significant at 5 %; *: significant at 1 %.*

Figure A3: Distribution of the score among the non-agricultural sector workers having the FSLC



Source : EESI (2005), Phase1. Our calculations.

<u>Figure A4:</u> Distribution of the score among the agricultural sector workers having the FSLC



<u>Source</u> : EESI (2005), Phase1. Our calculations.

<u>Table A7:</u> Distribution of the sample of people of 15 years old and above according to the activity situation

Situation of the person	Observations	Frequency
Inactive	5848	26.2
Unemployed	1681	7.5
Public sector	1226	5.5
Private formal sector	1199	5.4
Non-agricultural informal sector	7659	34.3
Agricultural informal sector	4739	21.2
Total	22352	100.0

<u>Source</u> : EESI (2005), Phase1. Our calculations.

<u>Table A8:</u> Summary characteristics of the sample of persons of 15 years old and above

Variable	Mean	Std Dev
Milieu of residence		
Yaoundé/Douala	0.325	0.003
Other towns	0.280	0.003
Rural	0.395	0.003
Individual characteristics		
No education	0.146	0.002
Primary	0.317	0.003
Secondary 1er cycle	0.303	0.003
Secondary 2nd cycle and +	0.234	0.003
Age	30.715	0.086
Female	0.510	0.003
Family environment		
Household head	0.363	0.003
Be in union(married or free union)	0.409	0.003
Proportion of children of 0-4 years in the household	0.122	0.001
Proportion of children of 5-9 years in the household	0.122	0.001
Proportion of jobless persons in the household	0.637	0.002
Number of hours devoted to domestic works per week	16.208	0.106
Number of hours devoted to schooling per week	4.237	0.084

<u>Source</u> : EESI (2005), Phase1. Our calculations

<u>Table A9:</u> Multinomial selection Model (odds ratios)

		Public	Private formal	Non-agricultural informal	Agricultural informal
Milieu of residence (ref.: *rural milieu*)	*Yaoundé/Douala*	0.48***	1.39***	1.14**	0.03***
	Other Towns	1.08	1.41***	1.49***	0.30***
Level of education (ref. *: No education*)	*Primary*	10.18***	4.94***	1.34***	2.00***
	Sec 1er cycle	26.05***	8.13***	1.28***	1.27***
	Sec 2nd cycle and +	121.34***	15.17***	0,83**	0.55***
Age		2.02***	1.83***	1.44***	1.45***
Age squared		0.99***	0.99***	1.00***	1.00***
Female		0.76**	0.33***	0.71***	0.66***
Household head		6.91***	6.08***	5.35***	4.77***
Be in union		1.41***	0.96	0.86***	1.20***
% children of 0-4 years old in the household		8.63***	14.96***	30.14***	74.88***
% children of 5-9 years old in the household		13.87***	25.46***	41.00***	91.28***
% working age persons in the household who have no income		0.00***	0.00***	0.00***	0.00***
Number of hours devoted to domestic works per week		0.97***	0.96***	0.99***	1.00
Number of hours devoted to schooling per week		0.96***	0.94***	0.96***	0.95***
Statistics of the model					
Pseudo R^2 (%)		40.9			
LR $\chi^2(60)$		24181.11			
Observations		22°352			

<u>Source</u> : EESI (2005), Phase1. Our calculations.
: significant at 10 %; **: significant at 5 %; *: significant at 1 %.*

Table A10: **Determinants of the sectoral allocation**

Variables		Unemployed (widely speaking))		Public		Private formal		Non-agricultural informal		Agricultural informal	
		OR	***SE***	***OR***	***SE***	***OR***	***SE***	***OR***	***SE***	***OR***	***SE***
Milieu of residence (ref.: *rural milieu*)	*Yaoundé/Douala*	1.25***	0.10	0.32	0.03	0.93	0.09	0.71***	0.04	0.02***	0.00
	Other Towns	1.49***	0.12	0.86	0.09	1.11	0.12	1.14**	0.07	0.23***	0.01
Level of education (ref. *: No education*)	*Primary*	4.87***	0.69	16.48	5.53	8.63***	1.92	2.35***	0.17	2.78***	0.20
	Sec 1^er^ cycle	4.78***	0.68	33.66	11.22	10.67***	2.39	1.70***	0.13	1.43***	0.12
	Sec 2^nd^ cycle and +	4.81***	0.71	146.57	48.58	18.34***	4.12	0.97***	0.08	0.63***	0.07
Age		1,40***	0.02	2,05	0.05	1.81***	0.04	1.41***	0.01	1.34***	0.01
Age squared		1,00***	0.00	0,99	0.00	0.99***	0.00	1.00***	0.00	1.00***	0.00
Female		0,63***	0.05	0,67	0.07	0.30***	0.03	0.64***	0.04	0.57***	0.04
Household head		2,08***	0.19	5,81	0.65	5.34***	0.58	4.41***	0.32	2.97***	0.24
Be in union		0,87*	0.07	1,11	0.11	0.76***	0.07	0.63***	0.04	0.79***	0.05
% children of 0-4 years old in the household		0,76	0.17	2,90	0.88	4.45***	1.34	7.55***	1.30	9.34***	1.82
% children of 5-9 years old in the household		0,51***	0.12	5,53	1.68	9.62***	2.93	14.05***	2.46	19.42***	3.82
% working age persons in the household who have no income		0,83	0.15	0,00	0.00	0.00***	0.00	0.00***	0.00	0.00***	0.00
Number of hours devoted to domestic works per week		1,01***	0.00	0,98	0.00	0.97***	0.00	0.99***	0.00	1.00	0.00
Number of hours devoted to schooling per week		0,97***	0.00	0,95	0.01	0.93***	0.01	0.95***	0.00	0.95***	0.00
Statistics of the model											
Pseudo R^2 (%)		32.0									
LR $\chi^2(75)$		22260. 31									
Hausman test ($\chi^2(60)$)		345. 37									
Observations		22°352									

Source : EESI (2005), Phase1. Our calculations.
: significant at 10 %; **: significant at 5 %; *: significant at 1 %.*