The impact of family literacy on the earnings of illiterates:

Evidence from Senegal♣

Leopold R. Sarr

Department of Economics, Cornell University

Cornell Food and Nutrition Policy Program

lrs25@cornell.edu

January 2004

The impact of family literacy on the earnings of illiterates: Evidence from Senegal

Abstract :

This paper investigates the extent to which the sharing of literacy knowledge within the household affects the labor force participation and the earnings of illiterate workers in Senegal. Controlling for selection bias, selective sorting, endogeneity and measurement error in the literacy variables, I find evidence that parental literacy and education do not capture all sources of external literacy benefits and that illiterate members also benefit from other literate members of the household. It also appears that rural workers and male illiterates tend to participate more in the labor market than their urban and female counterparts. The earnings of an illiterate female worker (or an urban worker) are, on average, 44% (or 37%) higher than the ones of another illiterate female (urban) worker whose family's ratio of literate to illiterate members is one point lower. This suggests that policies targeting isolated illiterate households, in both rural and urban zones as well as illiterate women -who appear to be better recipients and generators of external literacy benefits within households-, are likely to mitigate their vulnerability and thus to reduce the incidence of illiteracy and poverty in Senegal.

The impact of family literacy on the earnings of illiterates:

Evidence from Senegal

I- Introduction

Literacy has been so prominent in many studies that it can no longer be overlooked when assessing the welfare of a country, a region or even a household unit[1]. By generating direct benefits such as reduced fertility rates, enhanced productivity or better maternal health and by conferring indirect effects such as lower infant and child mortality, improved children's nutritional status, health and schooling, literacy clearly appears to have the great potential of significantly alleviating poverty and vulnerability[2].

The importance of literacy has been recently demonstrated by Basu and Foster (1998) who provide a new way of measuring literacy that accounts for the externality associated with the presence of a literate member in the household. This new measure is not only superior to the conventional literacy measure which overlooks the benefits that may accrue to proximate illiterates but, it is also relevant to less developed countries in which the literacy rates tend to be low.

In the case of Senegal where over 60% of the adult population is illiterate in 2002[3], a WorldBank report (1996) has revealed that 80% of rural women aged 15-39 are illiterate. The same report shows, for that age cohort, that the female illiteracy rate has ranged as high as between 80 and 90% in the four regions of Diourbel, Tambacounda,

---

[1] At a country level, literacy has come to be one of the major components of the UNDP's Human Development Index used to rank all countries.

[2] There is large evidence of these effects in numerous studies, of which I can cite Thomas, D. et al.(1991), Dreze et al. (1995), Strauss and Thomas (1995), Glick and Sahn (2000), Sahn and Stifel (2002).

[3] The figure comes from a report by UNDP/Ministry of Education in Senegal, 2003. In fact, over the last seven years, the average adult illiteracy rate was about 63.3%. Cf. table 1.

Kolda and Louga while it remained still very high at 65-75% in Kaolack, Fatick, St-Louis and Thies in 1995. Table 1 provides a regional distribution of the illiteracy rate over 1995-2002, and lays out the alarming literacy picture in Senegal.

In such context, assessing the impact of family literacy on the earnings of illiterate workers appears to be of high interest in that there is very little information on what determines their labor force participation and earnings in Senegal. Further, the potential gender differences associated with these determinants remain largely unknown.

As Basu et al. (1999), I attempt to estimate whether the presence of literate members in the household yield externalities for illiterate members[4]. In other words, does having literate members entails higher earnings for the former?

I also will address related questions of whether female illiterates are better recipients of externalities than male illiterates, whether literate women generate more literacy external benefits than men within the household, and whether investment in family literacy yield differing externalities for illiterate workers in the self-employment sector versus the wage sector, or in the rural agricultural sector versus the nonagricultural sector?

From a policy viewpoint, these questions seem critical in the sense that illiterate people -a great share of which is composed of women and poor-, if they do work, tend to be either confined in low paid employment sectors or engaged in self-employment

---

[4] This idea questions the view that parental literacy and education capture all sources of literacy benefits since illiterate members may also benefit from other literate household members. This will lead us to later investigate the extent to which the ratio of literate to illiterate members matters in generating externalities.

activities or simply in low and unskilled jobs[5]. Addressing properly these questions will perhaps be a step toward reducing the incidence of illiteracy in Senegal.

The paper is organized as follows: I will present an overview of the recent literature before sketching the underlying theoretical model of literacy. Section III is devoted to describing the ESAM[6] household data collected in 1994. In section IV, I will estimate the selection model and discuss the policy implications of the results obtained before drawing the concluding remarks.

Table 1: Adult illiteracy rates in Senegal, by region (%)

Regions	1995	1996	1997	1998	1999	2000	2001	2002
Dakar	39.61	43.4	43.4	41.6	39.8	…	…	…
Thiès	64.55	73	73	69.8	66.6	…	…	…
Louga	79.2	81.73	81.73	80.32	78.9	…	…	…
Saint-Louis	71.8	73.5	73.5	70.05	66.6	…	…	…
Tambacounda	78.05	83.2	83.2	79.15	75.1	…	…	…
Kolda	77.86	73	73	73.45	73.9	…	…	…
Ziguinchor	50.26	54.9	54.9	49.95	45	…	…	…
Fatick	73.85	68.5	68.5	71.5	74.5	…	…	…
Kaolack	72.1	53.4	53.4	58.6	63.8	…	…	…
Diourbel	80	84.1	84.1	78	71.9	…	…	…
Sénégal	64.3	67.41	67.41	63.21	59	62.6	61.7	60.7

Source : PNUD/Ministère de l'éducation nationale, Senegal, 2003.

The impact of literacy on labor market outcomes: literature review

In analyzing the role of shared literacy for innovation in the context of Guatemala, Green, Rich and Nesman (1985) provide evidence of the existence of literacy externalities.

Using village data from rural China and Indian, Dreze and Saran (1995) observe that the attitudes and values attached to education as well as the sexual division of labor

---

[5] Some studies in West Africa have shown evidence of this phenomenon, cf. Paul Glewwe (1990) in the case of Ghana, Glick and Sahn (1997) in Guinea and, in the case of Cote d'Ivoire, Wim Vijverberg (1992) and Vijverberg and Van der Gaag (1988).
[6] Enquête Sénégalaise Auprès des Ménages, Direction de la Statistique et de la Prévision, Ministère de l'Economie et des Finances, Sénégal, 1994.

tend to inform the household schooling decisions. While low economic value is often placed on girls'education, Dreze and Saran argue that they are strong incentives to provide women with literacy knowledge, even when they are only involved in domestic chores since other household members are likely to fully capture the external benefits they would generate.

Benham, L. (1974) finds evidence of the husband's earnings being positively correlated with the wife's schooling. But he did not deal with the potential selective sorting that may result from marriage, thus it remains unclear whether that found effect results directly from educational externality or is the outcome of marital sorting.

On the other hand, Basu, Narayan and Ravallion (2002) find that the earnings of illiterate men are on average 15% higher if they come from a household endowed with at least one literate member while women's labor force participation tends to be lower in a literate household. Their Bangladesh sample is however restricted to nonfarm illiterate workers.

Moreover, Basu, K. et al. (2002) find stronger effect of literacy on the earnings of unmarried illiterate women, arguing that the latter has little bargaining power within the household despite the income they bring into the family. Unmarried illiterate women would precisely fully capture the externality since their shared income provides incentives for literate household members to share their literacy knowledge with them to the extent that the sharing does not reverse the household equilibrium balance of power.

It therefore appears that evidence of positive literacy externalities is provided in a number of studies. Meanwhile, there may be instances in which the sharing of family literacy is restricted or given up, especially when it threatens the interests of literate

members within the household. For example, a higher income for an illiterate member will, ceteris paribus, empower her and thus may shift the pattern of household expenditures. Further, if preferences differ enough between illiterate and literate members within a household, a shift in power in favor of illiterate members may trigger the reluctance of literate people from sharing their knowledge. In the case of Senegal where polygamous families are pervasive, such phenomena are likely to occur.

In sum, the existence of literacy external benefits would be an outcome of two opposing forces crystallized in the positive income effect related to potential higher earnings accruing to illiterate members and in the negative effect of the shift in the balance of power between literate and illiterate members.

Paola Valenti (2001) has convincingly modeled the distribution of literacy within the household unit by arguing that the literacy externality is an increasing function of the ratio of literate to illiterate members. Unlike Basu and Foster (1998) who argue that proximate illiterate members benefit from externality so long as there is one literate member in the household, she makes the case that additional literate and illiterate members matter in determining the magnitude of the literacy externality. However, it may be the case that some activities require from an illiterate person that she only have access to any literate person, not necessarily someone from the same household.

Finally, as far as I know, not many studies have succeeded in disentangling the literacy effects from the schooling effects when analyzing labor market outcomes. For lack of adequate panel data to account for the full effect of literacy acquisition, a possible avenue to explore is to examine the external effects of family literacy at different levels of education.

In this perspective, Foster and Rosenzweig (1996) show that a farm-household total income depends on the highest education level reached within the household, rather than on the mean or the education level of the household head.

Krishnan (1996) shows how family background factors are key in influencing public sector wages in Ethiopia. She ascribes it to a selection effect since family background, through social networks and connections, determines entry into the public sector. However, family background is found to be unimportant in affecting employment in the private sector. On the other hand, by controlling for family background (parent and/or parents-in-law's schooling) in the labor earnings equation of males in Brazil, Strauss and Thomas (1992) and Lam and Schoeni (1993) find substantial effects. The latter argue that family background variables proxy for unobservable characteristics both in Brazil and in the USA (1993, 2001).

Studies that examine the impact of literacy on outcomes such as infant mortality, child health or schooling generally suggest a stronger and significant effect of mother's literacy. John Gibson (2001) has recently estimated the Basu-Foster literacy externality using a child height production function. Using Papua New Guinean data, he found an externality of 0.76 arguing that any estimated literacy externality based on wage is likely to be lower given that literacy sharing within the household is unlikely to increase an illiterate worker's productivity. However, whether literacy externality is better measured with home-based outcomes is an empirical question since there may be instances in which the sharing of literacy knowledge within the family substantially enhances the illiterate's productivity at work.

II- Intra-household Model of Literacy

Following Basu, Narayan and Ravallion (2002), I develop an intra-household model in which all consumption goods are assumed to be private. Let's represent it by the vector of goods X:

X = (x_1,….., x_k), being a bundle of k goods that the household consumes.

Although literacy is a public good, I will consider private goods that are collectively chosen and consumed by the household[7]. I can then generalize Basu et al.'s model by setting the following maximization problem:

$$Max_x \quad \alpha_j(y_{jA})\sum_{j=i+1}^{M} U_{jA}(x) + \left[1-\alpha_j(y_{iI})\right]\sum_{i=1}^{L} U_{iI}(x) \tag{1}$$

$$s.t. \quad px \le \sum_{i=1}^{L} y_{iI} + \sum_{j=i+1}^{M} y_{jA} \tag{2}$$

where i = 1, …, L correspond to the L literate members in the household ; j = i+1, …, M represent the M illiterate members in the same household; N being the size of the household[8].

$\alpha_j(y_{jA})$ corresponds to the weight attached to illiterate members: it captures in some sense their bargaining power within the household; y_{iI} is the income accruing to the literate member i whereas y_{jA} is the corresponding income for the illiterate member j;

---

[7] Literacy benefits as a public good are non-rival but excludable. They are non-rival because enjoying the benefits of literacy does not reduce the level of benefits that other members can enjoy from having a literate person. However, they are excludable in the sense that a literate person can deliberately refuse to share his knowledge with some individuals. For the sake of simplicity, I restrict the model to private goods.

[8] The capital letter I refers to the French translation for a literate or educated indivudal, i.e., 'Instruit' while the capital subscript A refers to an illiterate member translated by the term 'Analphabete'.

$U_{iA}(x)$ and $U_{iI}(x)$ correspond respectively to the utility of illiterate member j and literate member i within the household. They are assumed to be strictly increasing and concave; P is the price vector of all k goods consumed by the household: P = (p_1,……, p_k).

The optimal bundle is given by:

$$x^* = x\left(\sum y^*_{jA}, \sum y_{iI}, P\right) \qquad (3)$$

where $\sum y_{iI}$ and P are assumed to be constant, so they can be removed from the optimal bundle function. Thus, the optimizing bundle of goods of the household is simply a function of all income brought into the family by illiterate workers, i.e.,

$$x^* = x\left(\sum_{j=i+1}^{M} y^*_{jA}\right) \qquad (4)$$

Whether literate member i shares his knowledge depends on his utility gain, which in turn, depends on the extent to which sharing changes the level and the composition of household expenditures or consumption. In other words, each literate member i will share with each illiterate member j if:

$$U_{iI}\left[x\left(\sum y^*_{jA}\right)\right] \geq U_{iI}\left[x\left(\sum y_{jA}\right)\right] \qquad (5)$$

In fact, literate member i may share with some illiterate j and not with other illiterate j' within the same household if they have differing preferences. Naturally, sharing will occur between literate member r and illiterate member m if $U_{rI}(x) = U_{mA}(x)$.

On the other hand, if all goods consumed by household members are normal, then the utility of literate people will necessarily increase when literacy is shared, i.e.,

$$U'(x) > 0 \text{ and } \alpha'_j\left(y^*_{jA}\right) \succ 0 \qquad (6)$$

9

In a nutshell, the interesting question one would want to ask is what determines the income accruing to the illiterate member, y_{jA}? According to the above theoretical model, it will depend on how much literate members are willing to share their literacy knowledge, which will, in turn, be a function of the characteristics of illiterate members. In other words, the utility functions of both literate and illiterate members will matter in determining the earnings outcome of the latter. Therefore, the following earnings equation can be derived:

$$E_{it} = \gamma + \lambda L_t + \beta X_{it} + \delta Z_t + \varepsilon_{it} \qquad (7)$$

where E_{it} stands for the earnings of illiterate member i in household t; L_t the effective literacy within household t; X_{it}, the individual characteristics; Z_t, the household observable factors and, ε_{it} the standard disturbance term.

III- Data description

The data come from the ESAM, a nationally representative household survey conducted in 1994 by the *Direction de la Prevision et de la Statistique* of the Ministry of Finance in Senegal. About 3300 households were surveyed in 176 clusters called DR (*Districts de Recensement*). The sampling was stratified and the DRs were drawn proportionally to the size of the population.

1-) Literacy picture in Senegal

About 40% of the ESAM's sample are constituted by literate individuals whereas 57% are reported to be illiterate, e.g., cannot read and write in a given language, as defined in the survey. The median age of literate people is 25 while in the illiterate

sample, the median age lies at 33. It appears that illiterate people are on average older than literate persons in the dataset.

When I breakdown the sample by generation with the third generation aging between 15 and 30 years, the second generation ranging between 30 and 60 years and the first generation 60 year or older, the decomposition of the literacy variable reveals that about 63% of the literate sample belongs to the third generation of grand-children while 33% and 4% correspond respectively to the second and first generation of parents and grand-parents.

The total population for which we have information on literacy is composed of 21256 individuals. The share of proximate illiterates in the sample is 38%, which corresponds to about 65% of the overall illiterate sample while the proportion of isolated illiterate individuals is about 21%. Thus, the non-isolated literacy rate in the overall sample is 79%.

Table 2 summarizes different literacy measures by region. It appears that the standard UN's literacy measure ranks Louga as the region with the lowest literacy rate (20.3%) followed by Diourbel (20.9%). Ziguinchor and Dakar are the regions with the highest literacy rates, accounting respectively for 55.9% and 61.1% of their total population while the literacy rates for the remaining six regions range from 22.9% to 36.5%.

Furthermore, it is observable that the Basu-Foster's effective literacy measure is clearly higher across all the regions compared with the standard literacy, simply because of the externality indicator introduced to factor in the presence of proximate illiterates in the households. Table 2 also shows that regions like Diourbel, Louga and Tambacounda

have the greatest share of isolated households with respectively 43.5%, 39% and 38.6%. And even the gender-based isolated illiteracy rate exhibits the same pattern. On the other hand, Dakar and Ziguinchor, because they both have the highest standard literacy rates and Basu-Foster's effective literacy rates, record the lowest share of proximate illiterates in the country[9].

This ranking of regions by the standard literacy rates is reversed when one uses the effective literacy rate. In particular, four of the ten regions have moved either one step forward or backward in comparison to the classification set by the standard literacy rate. Finally, women appear to be the group that suffers the most from illiteracy, regardless of the literacy measure considered. Although the gender bias in literacy is not that acute, the gap is smaller for the Basu-Foster measure relatively to the standard literacy rate.

The low figures in most of the regions somewhat mirror the nationwide literacy's picture. A comparison between tables 1 and 2 corroborates this observation.

Table 2: Literacy measures in Senegal, by region (%)

REGION	R	Rm	Rf	P	I	Pm	Pf	Im	If	1-I	BasuFoster*	T
Louga	20.3	12.7	7.6	40.7	39	13.8	26.9	16.9	22.1	61	30	1263
Diourbel	20.9	14.2	6.7	35.5	43.5	10.4	25.1	17.4	26.1	56.5	29.8	1267
Kolda	22.9	16.8	6.1	43.4	29.4	15.5	29.9	14.4	15	70.6	33.8	1209
Tambacounda	23.4	15.9	7.5	37.9	38.6	12.9	25	17.7	20.9	61.4	33	875
Thies	29.2	11.7	17.5	43.4	27.3	16.6	26.8	12.4	15	72.7	40	2619
St-Louis	32.5	18.8	13.7	39.2	28.3	12.9	26.3	11	17.3	71.7	42.3	2090
Kaolack	33.7	21.9	11.8	44.9	21.4	15.3	29.6	9.5	11.9	78.6	45	2357
Fatick	36.5	22.6	13.9	40.7	22.8	13.4	27.3	10.8	12	77.2	46.7	1184
Ziguinchor	55.9	35.3	20.6	35.2	8.9	11.4	23.8	4.2	4.6	91.1	65	1270
Dakar	61.1	35.1	26	32.2	6.7	9.6	22.6	2.8	3.8	93.3	69	7122
Overall	41.2	24.8	16.4	38	20.8	12.5	25.5	9.1	11.7	79.2	51	21256

Source: Author's calculations from ESAM I, 1994.

---

[9] The literacy situation in Diourbel can be largely explained by religious reasons whereas Louga and Tambacounda seem to share the characteristics of either backwardness or remoteness from the center, Dakar where opportunities to learn are more widespread.

Moreover, according to UNESCO, the overall Senegalese population literacy rate in 1995 was about 33.1% with 43% of adult male and 23.2% of adult females being able to read and write a simple sentence about their everyday life and understand its meaning[10].

Legend

R: the sample standard literacy rate	I: the sample isolated illiteracy rate.
Rm: the sample male literacy rate	Pm: the male proximate illiteracy rate.
Rf: the sample female literacy rate	Pf : the female proximate illiteracy rate.
P: the sample proximate illiteracy rate	Im: the male isolated illiteracy rate.
(1-I): the no-isolated literacy rate	If: the female isolated illiteracy rate.
BF: the Basu-Foster effective literacy rate	T: the sample overall adult population.

*: To compute the Basu-Foster's effective literacy measures, I have assumed an externality indicator of 0.25, meaning that a proximate illiterate member is equivalent to a quarter of a literate member when one wants to measure effective literacy.

2-) Employment patterns of the illiterate labor force

Tables 3 and 4 display the distribution of the illiterate labor force in Senegal in 1994[11]. It appears that independent or self-employed workers constitute the largest portion of the illiterate labor force population, accounting for almost 67% of the overall illiterate sample.

---

[10] The literacy definition and the statistics are taken from the UNESCO World Education Report: Teachers and Teaching in a Changing World, 1998.

[11] The lower bound age of 15 is generally considered to be the starting age for entering the labor force. However, in the survey, the definition adopted was 10 years or older as the entry age into the labor force.

Among these self-employed, 39.1% are working in the agriculture sector and 27.8% in non-agriculture branches of the Economy. Further, among self-employed, I find nine workers in the agriculture sector and twenty-four workers in the nonfarm sector who received no income from their principal activity: they probably earned their annual income from being engaged in a secondary activity. Indeed, this occurs to 43 illiterate workers across all socio-professional categories.

Wage employees and apprentice represent respectively 8.0% and 4.8% of the illiterate population while the share of manual or unskilled workers amounts to 26.9%. Among wage employees, I find that 7.9% of the overall working population is recorded as unskilled workers and the remaining 0.7% as simply employees[12].

Finally, the unemployment rate among illiterate workers seems to be low (4.1%), perhaps due to the potential unreports of employment status from number of illiterate members. More specifically, the female unemployment rate is 3.3 points lower than the male unemployment rate (3.7%). However, given that many illiterate operate in the informal sector or perform multiple small tasks over the year, they may not consider themselves as regular workers, especially in urban areas.

The dataset shows that about 96% of the illiterate labor force is employed whereas only 4% do not have an occupation. By examining both table no.3 and 4, it becomes clear that most illiterate workers are self-employed, whether they are engaged in the agricultural or the non-agricultural sector. They form about 70% of workers with occupation. The next largest socio-professional category being the manual workers with a share of 28.1% of the total number of employed workers.

---

[12] Note that all individuals reported as holding internship are literate, as one would expect.

When I decompose the data by professional status, employees working for family businesses add up to 15.8% while wage employees account for 8.4% of the total working illiterate population.

For both types of decomposition, the share of female workers happens to dominate across all self-employed, wage employees and family member employees. On the other hand, not surprisingly, male workers are more likely to be found in manual work and apprenticeship.

Labor force participation of illiterate workers

Now, if we consider the overall adult population (15 years and above, including the literate sample), the labor force participation rate of the illiterate individuals is around 36.4%, which corresponds to a male and female labor force participation rates of 16.5% and 19.9% respectively. Illiterate women appear then to participate more in the labor market than illiterate men: 7.2% of females are reported as non-agricultural independent workers whereas only 2.9% males are operating in the same sector. This is also true for agricultural independent workers, wage employees, family member employees. Further, a smaller number of illiterate women (0.2%) are unemployed compared to illiterate men (1.4%). The same employment patterns are observed when I restrict the adult population to individuals aged 15-65.

Table 3: illiterate workers: labor force, participation and employment rates,
by socio-professional categories (%)

Socio-professional categories	illiterate labor force			Labor particip. rates			Employment distribution		
	All	Male	Female	All	Male	Female	All	Male	Female
Non-agricultural independent workers	27.8	8	19.8	10.1	2.9	7.2	29	8.3	20.7
Agricultural independent workers	39.1	18.8	20.3	14.2	6.8	7.4	40.8	19.6	21.2
Manual workers	26.9	14.1	12.8	9.8	5.1	4.7	28.1	14.7	13.4
Employees	2	0.8	1.2	0.7	0.3	0.4	2.1	0.8	1.3
Unemployed workers	4.1	3.7	0.4	1.5	1.4	0.2	-	-	-
Total	-	45.4	54.5	36.4	16.5	19.9	-	43.4	56.6

Source: Author's calculations from ESAM I, 1994.

Table 4: illiterate workers: labor force, participation and employment rates,
by professional status (%)

Professional status	illiterate labor force			Labor particip. rates			Employment distribution		
	All	Male	Female	All	Male	Female	All	Male	Female
Jobber/journeyman	1.6	1	0.6	0.6	0.4	0.2	1.7	1	0.7
Independent worker	64.8	25.4	39.4	23.6	9.3	14.3	67.7	26.6	41.1
Wage employee	8	3.3	4.8	2.9	1.2	1.7	8.4	3.4	5
Household member employee	15.2	6.6	8.6	5.5	2.4	3.1	15.8	6.8	9
Apprentice	4.8	4.3	0.4	1.7	1.6	0.1	4.9	4.5	0.4
Other	1.4	1.1	0.3	0.5	0.4	0.1	1.4	1.1	0.3
Unemployed worker	4.1	3.7	0.4	1.5	1.4	0.2	-	-	-
Total	-	45.4	54.5	36.4	16.5	19.9	-	43.4	56.5

Source: Author's calculations from ESAM I, 1994.

Table of summary statistics

The variables used in the estimations are summarized in table 5. The mean age of the illiterate sample is about 23 years and 41% are male whereas 45% live in urban areas. About 84% of them are proximate illiterate in the sense that they live in a household where at least one person is literate. Further, only 8% of male proximate illiterates live in a household where no female is literate while about 37% of female proximate illiterates live in a household where at least one female is literate. About 38% of the illiterate population live in households in which primary education is the highest schooling level reached, whereas only 23% of them belong to households in which secondary schooling was the highest level. About 26% of illiterates live in a house where the head is educated, but only one household over 5 is headed by a woman. The mean numbers of children less than 2 and 5 year are respectively about 1.0 and 2.4.

When I breakdown the sample into labor market participants and nonparticipants, about 64% of illiterates are in the former category. I also observe that, in the overall nonparticipants's sample, about 64% are literate and only around 33% are illiterate.

Table 5: Summary of statistics

Variable	Overall illiterate sample		illiterate labor force participants		illiterate labor force nonparticipants	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
literacy	0.844	0.363	0.840	0.367	0.853	0.355
ratio_loi	0.458	1.017	0.377	0.896	0.606	1.193
ratiosq	1.244	8.386	0.945	6.600	1.789	10.884
attendschool	0.129	0.335	0.080	0.271	0.218	0.413
age	22.941	20.407	21.808	19.622	24.997	21.611
assetindex	-0.714	1.605	-0.964	1.410	-0.261	1.824
hoh_edu	0.261	0.439	0.218	0.413	0.339	0.473
hoh_fem	0.199	0.399	0.155	0.362	0.280	0.449
highedup_hhd	0.382	0.486	0.377	0.485	0.391	0.488
highedus_hhd	0.231	0.422	0.190	0.393	0.305	0.460
propchildren	34.680	14.510	35.768	14.405	32.705	14.492
nbchild2	0.968	0.985	0.979	0.987	0.948	0.982
nbchild5	2.398	1.857	2.455	1.878	2.295	1.813
wolof	0.405	0.491	0.401	0.490	0.414	0.493
pular	0.259	0.438	0.244	0.430	0.286	0.452
sereer	0.162	0.369	0.168	0.374	0.152	0.359
diola	0.031	0.174	0.033	0.179	0.028	0.164
urbain	0.449	0.497	0.377	0.485	0.580	0.494
gender	0.415	0.493	0.493	0.500	0.273	0.446
fproanal	0.370	0.483	0.322	0.467	0.455	0.498
mproanal	0.083	0.276	0.098	0.298	0.055	0.228
N	18950		12219		6731	

Definition of variables:

literacy: a dummy taking 1 if any household member is literate and zero otherwise;

attendschool: being 1 if the illiterate member ever attended school, zero otherwise;

ratio_loi: the ratio of literate to illiterate members in the household;

assetindex: an index of household asset generated from principal component analysis

that proxies for household socio-economic status;

hoh_edu: a dummy taking 1 if the head of household is educated and zero otherwise;

hoh_fem: dummy taking 1 if the head of household is female and zero otherwise;

highedup_hhd & *highedus_hhd*: two variables capturing the highest level of education

reached within the household: 1 if it is primary/secondary and zero otherwise;

propchildren: captures the share of children less than 10 within the household;

nbchild2: the number of children less than two year old in the household;

nbchild5: the number of children less than five year old in the household;

urbain: a dummy taking 1 if one lives in an urban zone, zero otherwise;

gender: a dummy taking 1 if male and zero otherwise;

fproanal: a dummy taking 1 if a proximate illiterate female lives with at least one

        literate female member and zero otherwise;

mproanal: a dummy taking 1 if a proximate illiterate male lives with no female literate

        member and zero otherwise[13].

Dependent variable

The logarithm of annual earnings used as the dependent variable corresponds, in the data, to the gross pay received by the worker: it includes the basic salary plus taxes and subscriptions to the National pension Fund. Allowances, bonuses and other benefits related to overtime work are included in the salary so long as they have the same periodicity as the basic salary. Some workers such as the *sourgha*[14] receive partially or totally a salary in kind. Therefore, their gross annual earnings include both monetary and in-kind compensations.

If many workers earn most of their income from their principal activity, there are number of workers, especially self-employed illiterates who tend to be engaged in more than one activity and thus they might earn more income from their secondary activity. For instance, the *tâcherons* are a special category of workers who do not hold a single activity but, depending on the available opportunities, perform different tasks underlay by formal or informal contracts. In computing their gross earnings, not only do I consider all

---

[13] The number of observations N is lower for the last variable.

[14] These workers are generally agriculture workers who work in farms and lands owned by their employers.

compensations received from their main activity but also all income earned from other tasks performed over the year. Some of these payments are in kind and thus were added to their regular annual earnings[15].

IV- Econometric specifications, results and discussion

1- Specification of an illiterate worker's earnings equation

Specifying an earnings equation traditionally requires accounting for the fact some people in the sample are not working or they chose not to work because of some unobserved characteristics. If these unobserved factors affecting both sample selection and labor market outcomes are correlated, then we are in presence of selection bias. To properly account for the biases induced by the selection process into the labor force, I will consider Heckman selection bias model[16]. The general framework for these selectivity models can be derived as follows:

$$Y_t^* = X_t'\beta + \varepsilon_t \qquad \text{(outcome equation)} \qquad (8)$$

$$d_t^* = Z_t'\delta + \eta_t \qquad \text{(selection equation)} \qquad (9)$$

Y_t^* is observed if $d_t = 1$ ($d_t^* \geq 0$), i.e., the earnings is observed whenever an illiterate worker participates in the labor market; otherwise Y_t^* is not observed and $d_t = 0$ iff $d_t^* < 0$. Further, there is selection bias iff $E(\varepsilon_t, \eta_t) \neq 0$.

---

[15] The profit earned by running small scale businesses were generally missing in the data or filled as zero, therefore I could not include them in their annual earnings

[16] However, I will back up the estimated parameters with two alternative selectivity bias models so as to provide them further robustness check.

Note that η_t follows respectively a normal, a logistic or a uniform distribution if one estimates a Heckman's model, a Lee's selectivity model or an Olsen's selection bias model.

In estimating the impact of family literacy on the labor force participation and the earnings of illiterate members, I will deal with three major econometric issues: identification of parameters, sample selection bias, and the endogeneity of some RHS variables, in particular the literacy variables, in both the selection and the outcome equations. Further, I will test the existence of literacy externalities and whether women are better recipient or generators of externalities than men. I also will check whether the ratio of literate to illiterate members matters in generating literacy externalities. Finally, the issue of measurement error in the literacy variables, which may bias the estimators, deserves some attention.

Kernel density estimations provide a first step in making some distributional assumptions on the data and in deriving the likely functional forms of our model's specifications[17]. Therefore, model I is specified in the following form:

Selection equation $\qquad P_{ijk} = \gamma + \alpha L_{jk} + \beta X_{ijk} + \delta Z_{jk} + \varepsilon_{ijk}$ (8)

Earnings equation $\qquad W_{ijk} = \lambda + \mu L_{jk} + \nu X_{ijk} + \xi Z_{jk} + \eta_{ijk}$ (9)

where P_{ijk} and W_{ijk} correspond respectively to the labor force participation and the earnings of illiterate member i in household j in community k; L_{jk} captures the effective

---

[17] The graphs are available upon request. They basically show that the effective literacy variable enters linearly the earnings'equation whereas both variables RLI and age a nonlinear shape. The kernel mean estimates were also computed for other RHS variables. Furthermore, kernel density functions for illiterates' earnings tend to mimic the normal density, reinforcing the argument in favor of using the Heckman selection bias model.

literacy within household j in community k; X_{ijk}, the illiterate i's characteristics; Z_{jk}, the household j observable factors (including ethnicity) and, ε_{ijk} and η_{ijk} the classical disturbance terms.

Subsequent specifications will include a variable capturing the ratio of literates to illiterates within the household (R) and its quadratic terms plus additional covariates capturing community and regional effects (C). In that respect, the kernel estimates presented above suggest specifying the following model IV:

Selection equation $P_{ijk} = \gamma + \alpha L_{jk} + \acute{\alpha} R_{jk} + \theta R^2_{jk} + \beta X_{ijk} + \delta Z_{jk} + \kappa C_k + \varepsilon_{ijk}$ (8)

Earnings equation $W_{ijk} = \lambda + \mu L_{jk} + \acute{\upsilon} R_{jk} + \zeta R^2_{jk} + \nu X_{ijk} + \xi Z_{jk} + \rho C_k + \eta_{ijk}$ (9)

For identification purposes, Z_{jk} does not contain, in the selection equation, the same variables as in the earnings'equation[18].

2- Estimation Results

All outcome equations are estimated using as dependent variable the log of annual earnings of an illiterate worker. Unlike in Basu et al., our dependent variable comprises income earned by an illiterate worker from both her principal and secondary activity[19]. The selection equations are estimated with the dependent variable capturing whether an illiterate person participates or not in the labor market.

---

[18] This condition is not required for Heckman selection bias model. The fact that both selection and outcome equations can be identified even when they hold the same RHS variables results from the nonlinearity of the inverse Mills' ratio which in turn arises from the assumption of normality in the probit model. However, note that, in the Olsen's selection bias model for instance, that identifying condition is necessary.

[19] In the case of Senegal, many workers, especially self-employed, are involved in more than one activity, and in some cases, they earn more income from their secondary activity. Therefore, not accounting for that would probably provide a truncated picture of the reality. Since that information is available in the dataset, I include individual earnings from all activities.

Our estimation strategy is to first run an OLS and then compare its coefficients with the ones of a selectivity bias model. In particular, for outcome equations that exhibit a non significant coefficient for the selectivity-correction term (lambda), I test whether the OLS coefficients are significantly different from the selectivity model's coefficients. To do so, I decompose the illiterate sample into urban/rural samples and male/female groups[20]. Finally, the tests permit to decompose the data by sector of employment and thus to distinguish between self-employed illiterate workers and wage and other illiterate workers[21].

2-1) Estimations from OLS and Heckman selection bias model

Standard OLS regressions suggest that, having at least one literate member in the household raises an illiterate's earnings in rural areas across all four models whereas in the gender decomposition, the literacy coefficient, despite being non-significant, is higher for female illiterate workers than for male workers. On the other hand, when the variable capturing the ratio of literate to illiterate members (RLI henceforth) is included, the literacy coefficient drops, at least in the male, female and urban samples. The coefficient associated with RLI is very significant in both the female and the urban sample (cf. Panels A1 to A4).

However, it is well-known that the restriction of the sample to illiterate earners may remove a nontrivial and nonrandom portion of the sample illiterate population. Thus, to correct for these potential biases associated with standard OLS regressions, I run estimations based on three different selection bias models but I only present the results

---

[20] This breakdown is based on statistical tests run on the different model specifications: they show, on the one hand, that female illiterate workers behave differently from illiterate male workers and, on the other, that rural workers display, in the labor market, a distinct behavior from that of urban workers.
[21] These tests are available upon request.

based on the Heckman selection bias model[22]. Estimations from the selection equation show that illiterate male workers are more likely to participate in the labor market than female workers. Likewise, urban illiterate workers from households in which at least one member is literate are more likely participate in the labor market. However, both illiterate female workers and rural workers whose family's RLI are higher are more likely to work. In the earnings'equation, the returns to effective literacy for rural workers range between 0.44 and 0.50 across the four models. As for the RLI variable in model IV, an illiterate female worker will earn, on average, a wage that is 26% higher than that of another female worker whose family's RLI is one point lower[23].

Panel A1	Model I (Full illiterate sample)											
	OLS Earnings equation				Heckman Earnings equ.				Participation equation			
	Fem.	Male	Rural	Urban	Fem.	Male	Rural	Urban	Fem.	Male	Rural	Urban
literacy	0.185	0.18	0.331	0.118	0.185	0.163	0.362	0.117	0.045	0.102	0.015	0.093
	(1.55)	(1.40)	(1.66)	(1.16)	(1.53)	(1.21)	(1.80)	(1.12)	(1.25)	(2.19)	(0.39)	(2.26)
lambda@	-		-		4.087	1.085	-2.48	3.526				
					(2.99)	(0.70)	(-0.82)	(3.07)				
R^2	0.22	0.25	0.19	0.28	0.24	0.25	0.19	0.29				
N	433	415	186	662	433	415	186	662	11083	7867	10439	8511

Panel A2	Model II (Full illiterate sample)											
	OLS Earnings equation				Heckman Earnings equ.				Participation equation			
	Fem.	Male	Rural	Urban	Fem.	Male	Rural	Urban	Fem.	Male	Rural	Urban
literacy	0.107	0.113	0.364	0.038	0.11	0.108	0.404	0.061	0.042	0.102	0.003	0.101
	(0.85)	(0.84)	(1.67)	(0.36)	(0.88)	(0.78)	(1.84)	(0.57)	(1.15)	(2.15)	(0.07)	(2.42)
RLI	0.177	0.03	0.176	0.098	0.169	0.019	0.176	0.064	0.034	0.012	0.063	-0
	(2.95)	(0.37)	(1.20)	(2.00)	(2.87)	(0.23)	(1.21)	(1.31)	(2.83)	(0.51)	(2.13)	(-0.02)
lambda@	-		-		2.419	0.758	-2.55	2.71	-		-	
					(1.84)	(0.46)	(-0.82)	(2.35)				
R^2	0.25	0.26	0.20	0.29	0.26	0.26	0.20	0.29				
N	433	415	186	662	433	415	186	662	11083	7867	10439	8511

---

[22] The estimations from the Lee and the Olsen's selection bias models are available upon request. In a nutshell, the results obtained from these two models are similar to the Heckman's model, in particular the associated coefficients and the R-squared are very close.

[23] The high significativity of the inverse Mills'ratio in the female sample across all four models, points to some correlation between the error terms of the selection and outcome equations. In cases in which the selectivity term is non significant, I test whether the OLS coefficient is statistically different from the selectivity coefficient.

Panel A3	**Model III (Full illiterate sample)**											
	OLS Earnings equation				Heckman Earnings equ.				Participation equation			
	Fem.	Male	Rural	Urban	Fem.	Male	Rural	Urban	Fem.	Male	Rural	Urban
literacy	0.111	0.109	0.328	0.046	0.115	0.101	0.366	0.072	0.043	0.101	0.011	0.103
	(0.89)	(0.80)	(1.57)	(0.43)	(0.93)	(0.72)	(1.73)	(0.68)	(1.17)	(2.14)	(0.27)	(2.47)
RLI	0.166	0.026	0.095	0.096	0.141	0.011	0.108	0.06	0.029	0.013	0.064	-0.01
	(2.70)	(0.30)	(0.62)	(1.89)	(2.35)	(0.12)	(0.70)	(1.15)	(2.40)	(0.52)	(2.11)	(-0.55)
lambda@	-		-		3.184	1.1	-2.72	2.594	-		-	
					(2.53)	(0.67)	(-0.94)	(2.16)				
R^2	0.27	0.27	0.25	0.29	0.28	0.27	0.25	0.30				
N	433	415	186	662	433	415	186	662	11083	7867	10439	8511

Panel A4	**Model IV (Full illiterate sample)**											
	OLS Earnings equation				Heckman Earnings equ.				Participation equation			
	Fem.	Male	Rural	Urban	Fem.	Male	Rural	Urban	Fem.	Male	Rural	Urban
literacy	0.037	0.029	0.344	-0.05	0.047	0.029	0.379	-0.02	0.04	0.102	0.013	0.108
	(0.30)	(0.21)	(1.60)	(-0.44)	(0.38)	(0.21)	(1.75)	(-0.23)	(1.08)	(2.14)	(0.34)	(2.57)
RLI	0.497	0.343	-0.08	0.426	0.447	0.333	-0.03	0.375	0.043	0.01	0.029	-0.03
	(5.69)	(2.20)	(-0.18)	(4.99)	(5.03)	(1.98)	(-0.07)	(4.12)	(1.93)	(0.23)	(0.54)	(-1.14)
RLI^2	-0.08	-0.07	0.067	-0.07	-0.07	-0.07	0.054	-0.07	-0	4E-04	0.01	0.002
	(-5.89)	(-2.05)	(0.50)	(-4.34)	(-5.11)	(-1.95)	(0.41)	(-3.87)	(-0.72)	(0.08)	(0.75)	(1.01)
lambda@	-		-		2.24	0.273	-2.72	1.524	-		-	
					(1.79)	(0.16)	(-0.93)	(1.24)				
R^2	0.30	0.28	0.25	0.31	0.30	0.28	0.25	0.31				
N	433	415	186	662	433	415	186	662	11083	7867	10439	8511

2-2) Estimations by age group and sector of employment

Estimations run by age group show that, the magnitude of the literacy effect rises as the worker gets older. This is true for both the selection and the outcome equations

(results not shown). In the participation equations, the effect is very significant for the age group 46-55 and less significant for the age groups 26-35 and 36-45 but the coefficient associated with the variable RLI is generally insignificant. On the other hand, in the outcome equations, all coefficients related to the literacy variable are insignificant except for workers of less than 15 years old for whom the sign is negative. However, RLI appears significantly different from zero across age groups 15-25, 26-35, 36-45 and 46-55 with a rate of return ranging between 0.15 and 0.61. The results also show the existence of diminishing returns to RLI as workers get older.

Furthermore, I run the same model IV considering now the sector of employment (results not shown). It appears that, illiterate members are less likely to participate in the labor market either as self-employed, the higher RLI is or when they live in a literate household. In the earnings equations, only the sample of rural wage workers and other workers than self-employed display significant coefficients for the RLI variable.

2-3) Endogeneity of the literacy variables and other covariates

One important issue I haven't dealt with so far but which can produce biased and inconsistent estimates is the potential endogeneity of the literacy variables. In fact, if some unobserved factors such as the quest for more bargaining power within the household or a standing tradition of knowledge sharing that prevails in the household, are correlated with effective literacy or the ratio of literate to illiterate members, then not accounting for that will inevitably bias the parameter estimates of these variables. Correcting for the endogeneity bias requires identifying exogenous sources of variation in literacy and RLI that are uncorrelated with unobserved characteristics and do not affect the labor force participation and earnings of a given illiterate worker.

Our instrumental variable strategy will thus consist of running Models IV using a linear combination of the community averages that serve as instruments for both the literacy and RLI variables[24]. I will simultaneously address the endogeneity of other covariates such as the household asset index, the number of children less than two or five and the head of household's level of education.

Labor participation equation

The results from estimating the participation equation in model IV show that the effective literacy variable is significant for both the male and the urban groups and their respective coefficient suggests that male illiterate workers and urban earners are more likely to participate in the labor market than their female or rural counterparts (Panel A4). On the other hand, RLI appears very significant in the female sample and in both the urban and rural groups. But, in the latter category, the sign is negative and the magnitude of the coefficient much larger in absolute value than that in the urban sample. This suggests that rural workers tend to reduce their labor market participation as the ratio of literate to illiterate members increases within the household. Further, women whose family's RLI is the same as that of men, are more likely to participate in the labor market.

Outcome equation: what determines illiterates' earnings?

The estimations based on the outcome equations show that the selectivity correction term is significant for only the female sample[25]. The coefficient associated with effective literacy is significant in the male sample, providing evidence that an

---

[24] More specifically, I regress the endogenous variable on some instruments and then plug the fitted value into the original selection or outcome equation.

[25] I rely, for identification purposes, on household characteristics such as the household asset index, the proportion of children in the household, a dummy for whether the head is female and the highest level of education reached in the family, as exclusion restrictions. The likelihood ratio tests performed on these identifying variables confirm that they matter in the selection equation since they are jointly significant at the 0.1% significance level.

illiterate male worker living in a literate household is expected to earn more than another illiterate male worker who belongs to a household in which all members are illiterate. Whereas the coefficient associated with RLI suggests that illiterate women are better recipients of literacy externalities than men. Further, an illiterate female worker will earn, on average a wage that is 44% higher than that of another illiterate woman whose family's RLI is one point lower[26]. At the margin, it appears that providing literacy knowledge to an extra female household member could reduce, in no small part, the vulnerability facing a family.

Wald tests performed on the OLS coefficients versus the selectivity coefficients - whenever the selectivity correction term is insignificant- show that the OLS coefficients are not jointly statistically different from those of the selection estimations at the 1% significance level, for the male, urban and rural samples.

The Wu-Hausman test comparing the OLS estimator with the 2SLS estimator shows that OLS is adequate in the rural sample, thus suggesting a literacy gain of about 0.41 for a worker living in a literate household[27]. On the other hand, the 2SLS estimator appears more efficient than the OLS estimator of effective literacy for the male sample or the OLS estimator of RLI in the urban group.

Factors capturing whether literate men or women generate more literacy externalities were generally non significant in most regressions although illiterate members tend to participate less in the labor market when they live with at least one

---

[26] Note that the returns to literacy may be overstated because I could not account for inter-household or extra-household externalities. The data do not permit it.

[27] Rather than applying the selectivity coefficient, I compare the standard OLS coefficient with that of the 2SLS since the inverse Mills'ratio is insignificant. J. Jinardo and J. Dinardo (1997) have a nice presentation of the test, pp.338-42.

female literate person. Further, the gains from literacy resulting from women suggest that they generate higher externalities than men.

When I add regional dummies to model IV, the results show that the returns from being in Dakar -the capital city- range between 0.65 and 0.81 across the female, male and urban samples[28].

Finally, given that the matrix of RHS variables is full rank and that the order condition is satisfied, the identification condition for the estimated parameters is therefore met[29]. Moreover, both the participation and the earnings equations are exactly identified in so far as the number of instruments corresponds to the number of endogenous regressors.

Panel B1	Model IV corrected for endogeneity (Full illiterate sample)											
	OLS Earnings equation				Heckman Earnings equ.				Participation equation			
	Fem.	Male	Rural	Urban	Fem.	Male	Rural	Urban	Fem.	Male	Rural	Urban
literacy	-0.05	1.771	1.496	0.92	0.047	1.694	1.589	0.822	0.153	0.662	-0.06	0.341
	(-0.07)	(2.26)	(1.07)	(1.65)	(0.07)	(2.15)	(1.15)	(1.40)	(0.94)	(3.03)	(-0.26)	(2.08)
RLI	0.417	0.097	0.156	0.254	0.403	0.116	0.188	0.254	0.079	0.031	-0.28	0.117
	(3.78)	(0.70)	(0.42)	(2.63)	(3.65)	(0.83)	(0.49)	(2.62)	(2.19)	(0.62)	(-4.64)	(3.33)
RLI^2	-0.02	-0	-0.01	-0.01	-0.018	-0.004	-0.007	-0.010	-0.003	-0.003	0.012	-0.007
	(-3.11)	(-0.49)	(-0.23)	(-2.07)	(-3.2)	(-0.53)	(-0.25)	(-1.93)	(-1.86)	(-0.96)	(2.95)	(-3.94)
lambda@	-		-		2.836	-1.806	-1.753	-1.256	-		-	
					(1.71)	(-1.05)	(-1.15)	(-0.84)				
R^2	0.30	0.27	0.25	0.32	0.31	0.28	0.25	0.32				
N	433	415	186	662	433	415	186	662	11083	7867	10439	8511

The number in parenthesis corresponds to the z-statistics in the selection equations. In the earnings equations, it represents the t-statistics obtained from computing the robust estimates that allow for clustering at the household level.

@ : corresponds to the inverse Mills' ratio and the number in parenthesis is the t-statistics associated with it.

RLI: ratio of literate to illiterate household members

Participation/Earnings Equations

Model IV: Model III + ratiosq

---

[28] The inclusion of ten regional dummies turn out to be uninformative, so only the Dakar dummy were kept and it appears very significantly different from zero and positive, suggesting, as expected, that workers living in Dakar are more likely to participate in the labor market and to earn higher income than those from the rest of the country. However, since most employment opportunities are concentrated in Dakar and for that reason, given that many individuals may choose to live in the capital, it is likely that some unobservable factors will be correlated with the Dakar dummy in the disturbance term, at least in the urban sample.

[29] In both the selection and the outcome equations, the number of excluded exogenous variables is at least as great as the number of included endogenous variables.

2-4) Behavior of other covariates

In all estimations, the gender and zone variables suggest that male or urban illiterate workers tend to earn more income than their female or rural counterparts whereas rural or male illiterates tend to participate more in the labor market. This could point to the existence of some gender discrimination in the labor market.

Although some of the RHS variables in the selection equations are potentially endogenous, the results reveal that the wealthier the family the more unlikely an illiterate member is to participate in the labor market. The variable capturing the head of household's level of education appears to be positive and very significant for the male, female and rural samples. However, having a female headed-household tends systematically to reduce the probability that an illiterate member enters the labor market, across all groups. This might be due to the fact that the labor participation of a female head of household serves as a substitute for other household members. Thus, the family's burden falls on that single household member, increasing its exposure to vulnerability[30].

On the one hand, there is a negative correlation between the labor force participation of illiterate members and the fact that the highest level of education reached within the household is primary or secondary, the effect of the latter dummy being larger. This suggests that illiterate members tend to rely on educated members with potentially better chances of finding a job[31]. Further, illiterate members who ever attended school tend to participate less in the labor market.

---

[30] Even the negative sign associated with the variable *fproanal* seems to corroborate this explanation.
[31] The variable capturing the effect of having a family whose highest level of education is university was systematically insignificant in all estimations and was thus removed. It would be interesting to break down the effect at different levels of schooling attainment within each of the three education levels but unfortunately such information is not available in the dataset.

Finally, the higher the share of children aged less than 10 in the household, the more an illiterate member will participate in the labor market. Further breakdown of the children cohort shows that illiterate members tend to reduce their labor market participation when the family has more children under two years old in the female, male and rural samples. This finding is consistent with Gronau's story (1973)[32]. However, when I add a variable capturing the number of children under five, the sign is reversed suggesting that as children grow, their illiterate parents or illiterate elder siblings might return or enter the labor market.

The variable age is significant in both the participation and the earnings equations but it has a negative sign in the former for the female sample, suggesting that the younger the female illiterate member, the more likely she is to participate in the labor market. This might result from young girls working as domestics, for lack of employment opportunities.

Models III and IV include ethnic covariates, based on the assumption that ethnicity potentially influences entry into the labor market and earnings. I consider a dummy for each of the four major ethnic groups. The results show that being *wolof, pular* or *sereer* tend to be associated with lower participation rates in the labor market, unlike illiterates who are *diola*. These findings seem to be somewhat puzzling since one would expect the *wolof* and perhaps the *pular* and the *sereer* to be more connected and more likely to get a job. Are they unobserved factors correlated with being *wolof, pular* or *sereer* that influence labor force participation and earnings? A possible explanation is that the *wolof, pular* and *sereer* migrate domestically or internationally more than the *diola*

---

[32] Gronau argues that fewer women who have small children participate in the labor market. They have a higher reservation wage since their time at home is more valuable.

do, on average, and those who remain in the village or in the country tend to rely more on migrants or immigrants'remittances, inducing them to reduce their labor market participation.

2-5) Omitted variable bias

The endogeneity issue arising from the variables included in the estimations inclines to exclude number of RHS variables so as to avoid serious biases in the estimates. However, this is done at the cost of introducing another type of bias –the omitted variable bias-, which may turn out to be even worse. For that reason, I run the Ramsey RESET tests to check for potential omitted variable biases. The tests performed on Model IV show that there is no bias arising from the omission of some important RHS variables for both the female and the rural groups of illiterates[33]. However, for both the male and urban samples, I correct the bias by introducing additional RHS variables controlling for their potential endogeneity[34]. The results show that an urban worker is expected to earn about 37% more than another urban worker whose family's RLI is one point lower.

2-6) Selective sorting through family formation

As far as sorting -via family formation- is concerned, unobservable characteristics related to marriage may also affect the labor market productivity of illiterate workers. For instance, an educated working husband may help his illiterate wife find a job through his connections or provide her with enough capital to start a business or be self-employed.

---

[33] The null hypothesis that Model IV has no omitted variables cannot be rejected even at the 25% significance level.

[34] For robustness check, I regress both literacy variables on some household's characteristics. I then take the predicted values from these regressions and interact them with the literacy variables in the original model IV. The interaction terms being insignificant provides further evidence that there is no bias arising from unobserved factors correlated with the instruments. Moreover, the likelihood ratio tests show that the identifying variables matter in the participation equations.

On the other hand, a wealthy illiterate husband may influence the level of his family literacy by marrying an educated woman. Thus, to deal with this issue of selective sorting, I will compare the full illiterate sample with that of unmarried illiterate members.

When the sample is restricted to unmarried illiterate members, the effective literacy variable in the selection equation is positive and significant in both male and urban sample, as in the full illiterate sample. As for the RLI variable, both the female and urban samples exhibit significant and positive coefficients suggesting that women and urban workers are more likely to enter the labor market the higher their family's RLI. However, the sign is reversed in the rural sample, as in the full illiterate sample.

In the earnings equations, the effective literacy coefficient is insignificant but RLI is significant in both the unmarried urban group and the sample of single women[35]. The effect of family literacy is larger for unmarried women workers (0.47) relatively to the full sample (0.44). Likewise, unmarried urban workers enjoy a higher literacy gain compared to the full sample (0.40 vs 0.37). This might point to some selective sorting arising from family formation among illiterate workers. For instance, an urban married worker exhibits a rate of return to literacy that is 6% lower than that of a single urban worker. However, when I consider the sample of urban married workers who are polygamous or whose husband is polygamous, their literacy gain exceeds that of single urban workers by 21%[36].

---

[35] As far as the effective literacy variable is concerned, the fact that earnings gains are much higher for male workers might reflect the fact that men have generally more bargaining power than women within the household. Even though unmarried male workers are on average younger than married male workers and earn much less than them, they still tend to benefit more from literacy externalities than women.

[36] I got similar results for the sample of married women versus single women.

2-7) Measurement error in the literacy variables

An important issue in estimating literacy externalities is how the literacy variable is measured. Undoubtedly, self-report of literacy competences is subject to measurement error. For instance, the illiteracy rate of individuals with incomplete primary education is likely to be understated. In the presence of measurement error, it is well-known that OLS estimators are inconsistent and biased and that using instrumental variables eliminates attenuation bias due to random-noise measurement error-in-variables. However, the IV approach may not eliminate the bias resulting from unobserved characteristics correlated with instruments.

Moreover, two opposing effects seem to play in the estimated returns to literacy. The first effect is related to the fact that, as you add correlated regressors, the estimated effect decreases. On the other hand, as you control for variables correlated with literacy, you identify the literacy parameters from increasingly noisy information. In other words, the higher the correlation between the literacy variables and the other covariates, the larger the increase in the magnitude of the measurement error bias. In the dataset, the correlations between the literacy variables and other regressors are quite weak, limiting thereby the share of the change in estimated literacy gains that is due to increased measurement error bias. Consider the following probability limit of the estimated literacy gain:

$$p\lim \hat{\beta}_l = \beta_l - \beta_l \frac{\lambda}{1 - R_l^2} \qquad (10)$$

where β_l is the true literacy coefficient; λ, the proportion of the measurement error in the observed literacy variable or more technically the noise-to-signal ratio in measured

literacy and R_l^2 is the R-squared obtained from regressing the literacy variable on all

other regressors[37]. From equation (10), I can easily derive the measurement error bias:

$$m = -\frac{\lambda \beta_l}{1 - R_l^2} \tag{11}$$

Assuming that the proportion of measurement error in the observed literacy is 20% and that the true estimated literacy gain is 30%, I can compute the measurement error bias resulting from the different models'regressions[38]. Panel B2 shows that only 3.1% of the observed change in the returns to effective literacy is due to measurement error bias arising from passing from the regression of model I to that of model IV in the both the female and the male samples whereas in the urban and rural samples, it is respectively 6.0% and 27.1%. For the RLI variable, when I go from estimating model II to model IV, the share of observed change in the returns to RLI that is due to measurement error bias, ranges from 0.2% to 1.4% across the four samples.

Measurement error bias and change in observed returns to literacy variables

Panel J4	Female		Male		Rural		Urban	
	literacy	RLI	literacy	RLI	literacy	RLI	literacy	RLI
m_b (Model I)	0.00627	-	0.00611	-	0.00614	-	0.00627	-
m_b (Model II)	0.00630	0.00763	0.00614	0.00720	0.00617	0.00686	0.00628	0.00704
m_b (Model IV)	0.00632	0.00767	0.00615	0.00721	0.00618	0.00688	0.00618	0.00708
Δm_b	4.4E-05	3.5E-05	4.1E-05	7.8E-06	4.6E-05	2.5E-05	-8.4E-05	4.3E-05
$\Delta \beta$	-0.1388	0.2783	-0.1340	0.3136	0.0171	-0.2033	-0.1405	0.3108
$|\Delta m_b|/|\Delta \beta|$	3.1%	1.3%	3.1%	0.2%	27.1%	1.2%	6.0%	1.4%

m_b :Measurement error bias computed from the formula with the assumption of λ=20% and β_1=30% .
$\Delta \beta$:Change in observed returns to literacy/RLI from model I or II to model IV.

---

[37] Assuming that literacy is measured with error, $L_i^* = L_i + \mu_i$, where μ_i corresponds to pure measurement error uncorrelated with L, the noise-to-signal λ will be defined as: $\lambda = \frac{V(\mu)}{V(L^*)}$.

[38] The measurement error 0.2 appears to be a reasonable value for λ given that the literacy information in the data mirrors the nationwide picture. On the other hand, since most of the literacy gains reported tend to exceed 30%, I take it as a lower bound given the unaccounted potential inter-household externalities.

3-) Policy discussion

From the above discussion, it appears that policies targeting isolated illiterate households, in both rural and urban zones as well as illiterate women -who appear to be better recipients and better generators of external literacy benefits within households-, are likely to mitigate their vulnerability and thus to reduce the incidence of illiteracy. Unlike general adult literacy campaigns, such targeted policies are potentially cost-effective.

The data have also shown that a region such as Diourbel is populated with an important share of isolated illiterate households, mainly because the religion has been impervious to formal schooling. In such cases, policies enabling coranic schools to impart literacy and numeracy skills to both boys and girls could prove very effective in reducing the vulnerability of isolated households[39]. In the same respect, regions remotely located from the capital city and/or those lacking employment opportunities such as Tambacounda, Louga and Kolda could be potential targets for adult literacy campaigns. However, further microeconomic analysis of the characteristics of households located in these regions would be needed to better understand their behaviors and the underlying causes of their sustained vulnerability.

---

[39] It would be preferable that the literacy knowledge be taught in Latin rather than Arabic alphabet since many jobs would require such basic skills. This will probably induce additional teacher training costs.

V- CONCLUSION

While the sociological literature has long studied social interactions within the family and how individuals are influenced by the behavior of some reference group, there is a growing interest among economists to formalize social interactions and interdependent preferences.

In that respect, Kaushik Basu and James Foster (1998) have developed a theoretical intra-household model of literacy that accounts for the externality generated by the presence of a literate member in the household.

In my paper, I empirically estimate the extent to which the sharing of literacy knowledge within the family influences the labor force participation and earnings of illiterate members. Controlling for selection bias, selective family sorting, endogeneity bias and measurement error in the literacy variables, I find evidence that parental literacy and education do not capture all sources of literacy externalities, suggesting therefore that illiterate members do benefit from other literate household members.

It also appears that rural and male workers tend to participate more in the labor market than their urban or female counterparts. Further, the Senegalese household data permit to distinguish between rural and urban illiterate workers and to include earnings resulting not only from primary but also from secondary activity. Unlike previous empirical studies, I account for the distribution of literacy within the household by including a variable capturing the ratio of literate to illiterate members. The results show that an illiterate female worker (or an urban worker) is expected to earn a wage that is about 44% (or 37%) higher than that of another female (urban) worker whose family's ratio of literate to illiterate individuals is lower by one point.

These findings suggest that policies targeting isolated illiterate households, in both rural and urban zones as well as illiterate women -who appear to be better recipients and better generators of external literacy benefits within households-, are likely to mitigate their vulnerability and thus to reduce the incidence of illiteracy and poverty in Senegal.

The limited information provided by the ESAM I cross-sectional data underscores the need to collect, for instance, detailed microeconomic data on illiterate 'entrepreneurs' operating both in the formal and informal sectors in Senegal, so as to better address the concerns raised by estimating literacy externalities, in particular how literacy and education could translate into higher productivity[40]. Furthermore, the existence of panel data could potentially help disentangle the earnings'effect of changes in both the family's effective literacy and the literacy distribution in households over time.

---

[40] In Senegal, there is a nontrivial network of informal 'entrepreneurs' with religious connections that extend beyond the country, spanning Europe, the USA and even now South East Asia. However, many of these 'entrepreneurs' are generally able to read and write in Arabic.

References

Journals

Basu, K., and J. Foster, 1998, On Measuring Literacy, *Economic Journal* 108, 1733-49.

Basu, K., A. Narayan, M. Ravallion, 2002, Is literacy shared within households? Theory and Evidence for Bangladesh, *Labour Economics* 8, pp.649-665.

Benham, L., 1974, Benefits of Women's Education Within Marriage, *Journal of Political Economy* 82(1), S57-71.

Dreze, J., and M. Saran, 1995, Primary Education and Economic Development in China and India: Overview and Two Case Studies, in: K. Basu, P. K. Pattanaik and K. Suzumura, Eds, *Choice, Welfare and Development* (Clarendon Press, Oxford).

Foster, A.D. and M.R. Rosenzweig, 1996, Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution, *American Economic Review* 86, 931-53.

Gibson, J., 2001, Literacy and Intrahousehold Externalities, World Development 29: 155-166.

Glick, P. and D.E. Sahn, 2000, 'Schooling of girls and boys in a West African country: The effects of parental education, income, and household structure', *Economics of Education Review*, vol.19 (1), pp.63-87

Green, S. E., T.A. Rich and E.G. Nesman, 1985, Beyond Individual Literacy: The Role of Shared Literacy for Innovation in Guatemala, Human Organization 44, 313-21.

Heckman, James J., 1990, Varieties of Selection Bias, *American Economic Review*, vol. 8(2)

Krishnan, P., 1966, Family Background, Education and Employment in Urban Ethiopia, *Oxford Bulletin of Economic and Statistics* 58, 167-183.

Lam, D, and R.F. Schoeni, 1993, Effects of Family Background on Earnings and Returns to Schooling: Evidence for Brazil, *Journal of Political Economy* 101(4), 710-740.

Lam, D, and R.F. Schoeni, 2001, Family Ties and Labor Markets in the United States and Brazil, *Journal of Human Resources,* vol. 29(4).

Lee, Lung-Fei, 1982, Some Approaches to the Correction of Selectivity Bias, *Review of Economic Studies*, vol.49(3).

Olsen, R. J., 1980, A Least Squares Correction for Selectivity Bias, *Econometrica*, vol. 48(7).

Sahn, D. E and Stifel, D. C., 2002, Parental preferences for nutrition of boys and girls: evidence from Africa, *Journal of Development Studies*, vol.39(1), Oct.

Strauss, J., and D. Thomas, 1995, Human Resources: Empirical Modeling of Household and Family decisions, in: J. Behrman and T.N. Srinivasan, Eds, *Handbook of Development Economics*, volume 3 (North-Holland, Amsterdam).

Thomas, D., J. Strauss and M. Henrique, 1991, How Does Mother's Education Affect Child Height, *Journal of Human Resources*, vol. 26(2).

Vijverberg, Wim P. M., 1993, Educational Investments and returns for Women and Men in Cote d'Ivoire, *Journal of Human Resources*, vol. 28(4).

Unpublished manuscripts

Valenti, Paola M., 2001, Should we be concerned about the distribution of literacy across households? An Axiomatic Investigation, unpublished manuscript, Cornell University, Ithaca (NY), USA.

Statistical data sources

Direction de la Prévision et de la Statistique, Ministère de l'Economie et des Finances du Sénégal, Dakar: République du Sénégal, 1995.

Ministère de l'Education Nationale du Sénégal, Dakar: République du Sénégal, 2003.