Gender Inequality in Literacy Status and Households Economic Well-Being in Burkina Faso: A Semiparametric Bi-variate Sample Selection Modeling Approach

Ibrahim Niankara

College of Business, Al Ain University of Science and Technology, P.O. Box: 112612, Abu Dhabi, UAE; Tel: +97124444696 Ext: 539 ; Fax: +97124444304 ; E-mail: ibrahim.niankara@aau.ac.ae

Abstract

This paper models the factors explaining households members economic well-being in Burkina Faso, with a focus on the relative influence of gender inequality in literacy status. It does so, using data from the 2014 survey on household living conditions and a semi-parametric bi-variate sample selection modeling approach. This approach compared to the classic Heckman two-step estimator is methodologically innovative because it deals simultaneously with non-random sample selection using conventional systems of two equations, non-linear covariates' effects using spline approach, and the non-normal bivariate distribution using copula functions.

The graphical results from the Lorenz curves combined with the numerical Atkinson and Gini coefficients suggest that inequality in overall per-capita consumption spending among households headed by literates is higher than that of their illiterate counterparts in 2009 and 2014. However, independently of the head of household's literacy status, the level of inequality in total economic well-being decreased from 2009 to 2014. Using the poverty indices of Watts, Sen, Foster ($\alpha = 1$)) we found that poverty among households headed by literates is lower than that of their illiterates counterparts for both years, although overall poverty decreased nationally between 2009 and 2014.

The results also show that although the gender inequality in literacy status does not translate into inequalities in non-food wellness, it does however for food-wellness as female headed households have 38.9% less per-capita food consumption spending than their men counterparts. Combining both food and non-food consumption spending, total economic well-being also seems to exhibit gender inequality as female headed households now have relatively 26.7% less combined per-capita consumption spending.

Keywords: Burkina Faso, Economic Well-Being, Gender inequality, Literacy status, Sample Selection
JEL: C30, D63, I21, I31

1. Introduction

"Imagine not being able to vote in an election, take the right bus, fill out a form at the hospital or an opinion survey because you are not able to read. That's the reality for millions of people today"(Project-Literacy, 2016).

Policymakers and the research community at large have long been interested in the economic well-being of social groups that are particularly vulnerable (Potter, 2014; Benería et al., 2015) - those individuals in society that theoretically cannot improve their own economic well-being, which In addition to children and the disabled, include illiterate adults (Sywelem, 2015), mainly because they have little access to labor markets (Handel, 2003). The Program for the International Assessment of Adult Competencies (PIAAC) as a cyclical, large-scale study of adult skills and life experience focusing on education and employment broadly defines literacy as "understanding, evaluating, using and engaging with written text to participate in society, to achieve one's goals and to develop one's knowledge and potential"(Goodman et al., 2013). In this formulation, improving literacy skills is seen as a key first step to overcoming the obstacles that lock individuals into a cycle of poverty and disadvantage (Heckman and Mosso, 2014; Prete, 2013), and education viewed as the best means of overcoming poverty caused by illiteracy (Cree et al., 2012). This is partly why UNESCO was tasked by the UN General Assembly in September 2015 to play a coordinating and catalyzing role in advancing literacy as part of the Sustainable Development Goals. On the subject, UNESCO Chair in Learning and Literacy at the University of Pennsylvania, Dan Wagner, said:

"Literacy is a key component in achieving the UN's sustainable development goals. Without literacy, each of the 17 goals will be limited by the inability of citizens to be sufficiently informed on key issues, and less empowered to take action. There is a strong argument that tackling illiteracy and low literacy, as a 'foundational' social problem, would pay greater economic dividends than tackling each issue separately."

Illiteracy puts a major burden on the global economy (Suh et al., 2016). In fact, the World Literacy Foundation report (Cree et al., 2012), looking at the cost of illiteracy in emerging and developing countries, as well as the cost of functional illiteracy in the developed world, found that more than 100 million children don't go to school each day, and nearly 800 million people across the globe lack the basic reading and writing skills needed to accomplish simple tasks such as reading a medicine label or filling out a job application, costing the global economy more than $1.19 trillion a year. The annual estimated costs of Illiteracy in terms of GDP loss is 2% for developed nations, and 1.2% for emerging economies. In order to achieve continued economic growth countries need to first achieve an adult literacy rate above 40%. Illiteracy seems to also affect disproportionately the sexes, with women accounting for two thirds of the world's illiterate adults, and In some areas 85% of the female population are unable to read or write, denying them access to learn, earn, vote and thrive. In sub-Saharan Africa, the number of illiterate adults has increased by 37% since

1990. The rate of illiteracy among adults in the region today around 38%, which UNESCO cites as a major barrier to improving child malnutrition and food safety, which in turn costs the global economy nearly$3.5 trillion per year.

As part of a longer-term project to understand and raise awareness of the role illiteracy plays in causing and propagating the biggest social problems such as poverty and inequality, this paper therefore models the gender inequality in literacy status, and look at its effects on households' economic well-being in Burkina Faso using a Bi-variate Sample Selection Modeling Approach. In doing so, this paper aims to contribute to supporting the United Nations' Sustainable Development Goal, "to ensure inclusive and quality education for all and promote lifelong learning", by working towards a clear target: that by 2030, no child will be born at risk of poor literacy in Burkina Faso or elsewhere. As such, the remaining of the paper is organized as follows : Section 2 gives a brief background on Literacy, Economic Well-Being, Inequality and Poverty in Burkina Faso. Section 3 describes the data and analytical strategy. Section 4 presents the Sample Selection Model of Literacy status and Economic Well-Being. Section 5 presents the results, while section 6 concludes the analysis.

2. Literacy, Economic Well-Being, Inequality and Poverty in Burkina Faso

Burkina Faso is a landlocked sahelian country in the loop of the Niger river with an area of 274,200 km2 and an economy based mainly on agriculture, which employs 84% of the population. After more than ten years of adjustment without interruption between 1991 and 2002, the national economy grew at an average rate of 5% per year in real terms, against a population growth of 2.4% per year INSD (2015b). Mass poverty in the country has led the authorities to adopt, since 2000, a strategic framework for the fight against poverty. Some aspects of this framework aim to promote human development and improved social and economic well-being through the eradication of illiteracy. In fact, since 2001 Burkina Faso has developed and implemented the first phase of its ten-year program (programme décennal de développement de l'éducation de base (PDDEB)) to achieve a quantitative and qualitative development of basic education and literacy by "Promoting access to education for the poor."

Despite these measures, to this date the country is still characterized by a low rate of adult literacy. According to a report by the National Institute for Statistics and Demograhy INSD (2015a), four out of five people aged 25 and over have no education. The youth literacy for those aged 15 to 24 has grown at a rate of about 2 percentage points over the past five years. The rate increased from 41% in 2009 to 50% in 2014. In addition, major inequalities persist between urban and rural areas and between the central region and the rest of the twelve regions of the country. The literacy rate is three times higher in urban areas (64.0%) than in rural areas (23.4%). Compare to the national average, the urban literacy rate is almost twice the latter and that of the rural areas is lower of about 10 percentage points.

While poverty, estimated through the incidence, depth and severity is interested in some individuals in the population, namely those living below the poverty line, Inequality on the other hand covers the whole population. Inequality is used to study the relative positions of individuals which, incidentally, are as important as the absolute positions. Inequalities, when pronounced, tend to weaken the social fabric and exacerbate violence. This may explain the social unravel against the former government of Burkina Faso, observed in October of 2014, which as part of the country's history is now a national holiday. This means that even in the absence of poverty, social balance is essential to ensuring economic stability and more importantly the sustainable growth of a nation.

The most common way of measuring economic well-being is using the market value of the goods that are consumed by "economic agents" over a period of time (Edwards, 1999; Fields, 2000). Although it should not be assumed that households do not derive economic well-being from commodities besides goods and services, in this analysis we will focus on household per-capita consumption spending on goods and services, widely accepted as a good indicator for poverty, inequality, and overall well-being (Osberg and Sharpe, 2002; Sumner, 2004). Looking at average measures of economic well-being however does not account for the fact that there are inequalities in economic well-being, that is, wide dispersion in per-capita households' consumption spending, independent of literacy status.

In fact, as shown in the report INSD (2015b), high inequalities exist between 2009 and 2014, as 20% of the wealthiest individuals share 44.1% of total consumption expenditure while the remaining 80% of the population share the remaining 56% of consumer spending. In general, inequality can be measured using the Gini Coefficient[1], which compares not just two households, but all households in describing inequality. Again, a larger Gini coefficient means greater inequality in per-capita consumption spending. One of the benefits of the Gini coefficient here is that it standardizes inequality to the absolute amount of the per-capita consumption variable. The analysis of inequality complements that of poverty and provides a better way of characterizing economic well-being in Burkina Faso.

3. Data and Analytical Strategy

The empirical analysis is based upon data from the National Survey on Household Living Conditions (EICVM, 2009 and 2014) administered by the National Institute for Statistics and demography (INSD) in Burkina Faso. The survey uses a two-level stratified random sampling with weights that produce nationally representative estimates for households' annual per-capita consumption spending, and a wide range of demographic and socioeconomic characteristics for the

---

[1]The Gini coefficient is calculated by taking the differences between the per-capita consumptions of every household (this would be n times (n-1) differences for a sample of n households), averaging them, and dividing by two times the average household consumption.

civilian, non institutionalized population in Burkina Faso. Primary sampling units were selected with probability proportional to their size, and the secondary sampling units or households selected with equal probability within those primary sampling units. The EICVM 2009 and 2014 were collected during a period of twelve months each. Our analysis included 8,404 from the 2009 survey and 10411 from the 2014 survey after accounting for variables selection and missing data constraints. The two data sets were combined into a panel of cross-sections over the two periods (2009 and 2014) with a total sample size of 18815 households for our descriptive analysis of inequality and poverty.

3.1. Economic Well-Being Outcome Measures

As previously stated we use households' per-capita consumption spending on Food and Non-food items as our measures of economic well-being. This allows us to distinguish three levels of economic well-being: Food-wellness, Non-Food-wellness, and Total Economic Well-Being.

3.1.1. Food-wellness

This measure of households' economic well-being is obtained by dividing total household spending on food consumption during the year by the size of the household. That is $\frac{dalim}{HHsize}$, which is households' per-capita food consumption spending.

3.1.2. Non-Food-wellness

This measure of households' economic well-being is obtained by dividing the total household spending on non-food consumption during the year by the size of the household. That is $\frac{danlim}{HHsize}$, which is households' per-capita spending on non-food consumption.

3.1.3. Total Economic Well-Being

This measure of well-being is obtained by dividing the total household spending on food and non-food consumption during the year by the size of the household. That is $\frac{deptotnd}{HHsize}$, which represents the total households' per-capita spending on food and non-food consumption.

3.1.4. Poverty Status

This is a binary measure constructed using the total households' per-capita spending on food and non-food consumption, and the poverty line. It is assumed that households as rational optimizers are looking to maximize their utility from the consumption of goods and services they purchase annually. In this quest to satisfy their needs, households choose to spend yearly on consumption, an amount either greater than the poverty line (in which case the household is viewed

as "non-poor") or less than the poverty line (in which case the household is viewed as "poor")[2]. This suggests that households per-capita annual spending is an indicator of household poverty status and in this way, being characterized as "non-poor" or "poor" is a deliberate choice of the household to spend annually per-capita, an amount "greater than" or "less than" the poverty line in the satisfaction of their needs. The binary poverty status variable is therefore given by :

$$PoverStatus = \begin{cases} Nonpoor = 1 \ \ \forall \ \ \frac{deptotnd}{HHsize} > Poverty \ line, \\ Poor = 0 \ \ \forall \ \ \frac{deptotnd}{HHsize} \leq Poverty \ line \end{cases} \quad (1)$$

3.1.5. Independents variables

Like any scientific study using evidence from observational data, our interests here centers on a postulated causal influence from the attributes and environments of households' members in Burkina Faso to their responses, or per-capita annual consumption spending choices. It is assumed that these spending choices reflect households' poverty status as well. So in choosing the covariates to be included in the model, the question that needs to be addressed is: In addition to literacy status, what other factors affect households economic well-being in Burkina Faso?

Keeping in mind that the primary goal of this empirical analysis is not to model the determinants of households' economic well-being in Burkina Faso, but to measure the impact that heads of household's literacy status has on economic well-being, as measured by per-capita household consumption spending; therefore, the primary independent variable is the literacy status of the head of household. In order to achieve the study goal hoverer we need to account for the effects of the other covariates affecting this relationship such as Age, Sex, Marital status, and Education of the head of household, Residency status, and Household Size. Table (1) provides definitions and summary statistics for the variables used in the analysis.

4. The Sample Selection Model of Literacy status and Economic Well-Being

In its prevailing and traditional definition, literacy is regarded as central to helping people obtain and retain employment, which is the key to moving them from dependency toward greater self-sufficiency. This functionalist definition, espoused by many policymakers, and employers, is based on the assumption that there are jobs for the financially poor who are able to improve their literacy skills (Hornbeck and Salamon, 1991). Following this view and the Human capital theory framework, we can predict relatively higher wages and greater economic prospects for the more

---

[2]This definition of poverty status, is given by the National Institute for Statistics and demography (INSD) in Burkina faso which was the agency in charge of the EICVM 2009 and 2014 surveys. This poverty line was 130,375 CFA Franc for 2009, and 153,530 CFA Franc for 2014

Table 1: Summary Description of the Variables used in the Econometric Modeling

	10411	Mean	SD
LnCapSpendg	log(annual household Overall per-capita spending)	12.220	0.689
LnCapSpendgF	log(annual household per-capita spending on Food)	11.539	0.680
LnCapSpendgNF	log(annual household per-capita spending on Non-Food)	11.440	0.782
HHsize	the number of people living in the household	7.479	4.969
Age	the age in years of the head of household	46.567	15.517
AgeSqr	the age squared in years of the head of household	2409.263	1596.270
Poverty Status			
Poor (ref)	= 1 if household is considered "poor"		
NonPoor	= 1 if household is considered "non-poor"		
Literacy Status			
Illiterate (ref)	= 1 if household head is "Illiterate"		
Literate	= 1 if household head is "Literate"		
Education Level			
None (ref)	= 1 if head of household has no education		
Primary	= 1 if head of household has only a primary education		
Secondary	= 1 if head of household has only a secondary education		
Higher	= 1 if head of household has some higher education		
Gender			
Male (ref)	= 1 if head of household is Male		
Female	= 1 if head of household is Female		
Marital Status			
Single (ref)	= 1 if head of household is single		
Married	= 1 if head of household is married		
Widow	= 1 if head of household is a widow		
Residency Status			
Urban (ref)	= 1 if Household lives in Urban area		
Rural	= 1 if Household lives in Rural area		

Source: 2014 EICVM Data Set by the National Institute for statistics and Demography in Burkina Faso.

(ref) represents the reference category for the nominal variable with more than one level.

For each numeric variable, the mean and standard deviation are provided.

literate workers because of greater productivity. The higher wages in turn translate into higher per-capita consumption spending, and thus higher economic well-being.

The Bi-variate Sample Selection Model of Literacy status and Economic Well-Being then specifies a binary participation equation defining whether or not the head of household is literate, and an outcome equation capturing household economic well-being as measured by annual per-capita consumption spending. More specifically, if y_1^* is the latent utility defining whether or not the head of household is literate, and y_2^* the level of annual per-capita consumption spending by the household, then the Bivariate sample selection model comprises the literacy status equation:

$$LiterStatus = \begin{cases} 1 & \text{if } y_1^\star > 0 \\ 0 & \text{if } y_1^\star \leq 0 \end{cases} \tag{2}$$

and the well-being equation:

$$wellBeing = \begin{cases} y_2^* > Poverty\ line, & \text{if } y_1^\star > 0 \\ y_2^* \leq Poverty\ line & \text{if } y_1^\star \leq 0 \end{cases} \tag{3}$$

The standard model specifies a linear model with additive errors for the latent variables, so

$$y_1^\star = x_1^* \beta_1 + \epsilon_1, \tag{4}$$

$$y_2^\star = x_2^* \beta_2 + \epsilon_2 \tag{5}$$

The challenge arises in estimating β_2 when ϵ_1 and ϵ_2 are correlated. Following (Cameron and Trivedi, 2005, 547-548) estimation by Maximum likelihood is straightforward if we make the additional assumption that the correlated errors are jointly normally distributed and homosckedastic, with

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right] \tag{6}$$

Where the normalization $\sigma_1^2 = 1$ is used for identification purposes. The conditional truncated mean of the model is obtained with:

$$E[wellBeing | \mathbf{x}, y_1^\star > 0] = E[\mathbf{x}_2' \beta_2 + \epsilon_2 | \mathbf{x}_1' \beta_1 + \epsilon_1 > 0] \tag{7}$$

$$= \mathbf{x}_2' \beta_2 + E[\epsilon_2 | \epsilon_1 > -\mathbf{x}_1' \beta_1] \tag{8}$$

where $\mathbf{x}$ denotes the union of x_1 and x_2. If the errors ϵ_1 and ϵ_2 are independent then the last term simplifies to $E[\epsilon_2] = 0$, and ordinary least squares regression of the Well-being outcome on $\mathbf{x}_2$

will give a consistent estimate of β_2. Any correlation between the two errors however means that the truncated mean is no longer $\mathbf{x}_2'\beta_2$ and we need to account for selection. The standard in the literature when correlation exists is to use the two step estimator by Heckman (1979) to obtain $E[\epsilon_2|\epsilon_1 > -\mathbf{x}_1'\beta_1]$.

In the current analysis, we will instead rely on semiparametric estimation of sample selection models as presented in (Cameron and Trivedi, 2005, 565-566) . This estimation is more challenging but allows us to relax the assumption that the errors (ϵ_1, ϵ_2) are jointly normally distributed. It uses as starting point the expression in equation (7) and gives:

$$E[wellBeing_i|\mathbf{x}_i, y_{1i}^{\star} > 0] = \mathbf{x}_{2i}'\beta_2 + E[\epsilon_2|\epsilon_1 > -\mathbf{x}_{1i}'\beta_1] \tag{9}$$

$$= \mathbf{x}_{2i}'\beta_2 + g(\mathbf{x}_{1i}'\beta_1) \tag{10}$$

where the second equality assumes that $\epsilon_{2i}|\mathbf{x}_i$, ϵ_{1i} has distribution that depends on $\mathbf{x}_{1i}$. The distribution of (ϵ_1, ϵ_2) is left unspecified so the function $g(.)$ is unknown, leading to a semiparametric estimation. since it is possible that $g(\mathbf{x}_1'\beta_1) = \mathbf{x}_1'\beta_1$, identification in this model with $g(.)$ unspecified requires an exclusion restriction that at least one component of $\mathbf{x}_1$ does not appear in $\mathbf{x}_2$. The more uncorrelated $\mathbf{x}_1'\beta_1$ is with $\mathbf{x}_2$ the better β_2 and $g(.)$ can be distinguished. The model in equation (9) is called partially linear, and can be estimated using semi-parametric methods as presented in Wojtys et al. (2016).

5. Results

The results are presented in two parts, the first part gives a description of the distribution of Economic well-being in Burkina Faso, using Households per-capita consumption spending on food and non-food items as measures respectively of Food-wellness, and Non-food wellness. The sum of the two per-capita consumption spending is taken as household total economic well-being. Using the economic well-being measures we study the distribution of inequality and poverty in Burkina Faso in the next subsection. The second part of the results addresses the implemented econometric estimations in the analysis. All the results presented in this section are produced using functions provided by the R Statistical Software (R Core Team, 2015).

5.1. Descriptive results and Interpretation

Using our total economic well-being measures for 2009 and 2014, obtained from the EICVM Data Sets we compute the summary statistics and welfare functions presented in table (2) using tools from the R library Zeileis (2014)

We use the Atkinson welfare function (see Zeileis (2014)) with $\epsilon = 1$, which means a medium concern for inequality. Table(2) shows that welfare has risen by 08.13% with this measure, mainly

	2009	2014
N	8404	10411
μ	259622.7	272978.9
Median	167323.8	181489.6
Welfare	12.10956	12.21986

Table 2: Summary Statistics and welfare function

because the mean per-capita consumption spending has increased by 13356.22 between 2009 and 2014. For both years, the median is lower than the mean, due to the asymmetry of the distribution of total per-capita consumption spending.
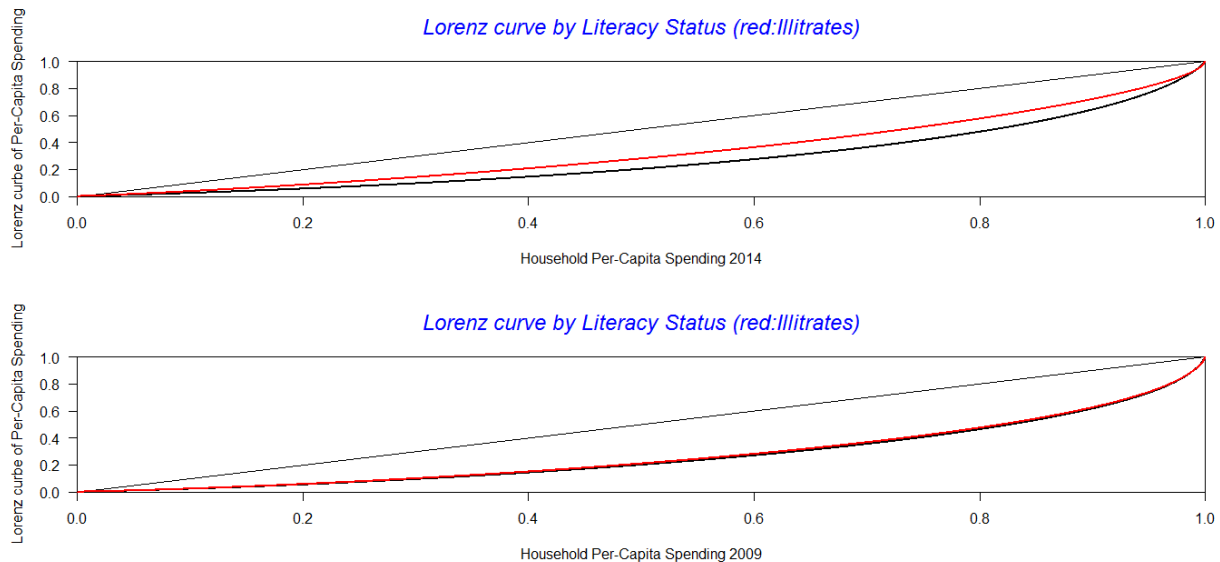
Figure 1: Inequalities in the distribution of economic well-being by literacy status for 2009 and 2014

The figure (1) above presents the Lorenz curves showing the inequalities in the distribution of economic well-being by literacy status (Literate, Iliterate) in 2009 and 2014. The lorenz curves are computed using the "Lc" function in the R library Zeileis (2014). They suggest for the year 2009 that inequality in overall per-capita consumption spending (economic well-being) among households with literate heads was fairly similar to that among households with illiterate heads. However this figure changed in 2014 where inequality among illiterates was relatively smaller than that observed among literates. These graphical results by the lorenz curves are supported by the Inequality measures in table (3). In fact both the Atkinson and Gini coefficients for the households

with illiterate heads are relatively smaller than those of the literate in 2009 and 2014. This suggest that the inequalities in total economic well-being among illiterate headed households are lower than that of their literate counterparts for both years.

	2009		2014	
N	8404		10411	
	Literate	**Illiterate**	**Literate**	**Illiterate**
Atkinson	0.184	0.174	0.166	0.096
Gini	0.472	0.457	0.455	0.336

Table 3: Inequality measures by literacy status for 2009 and 2014
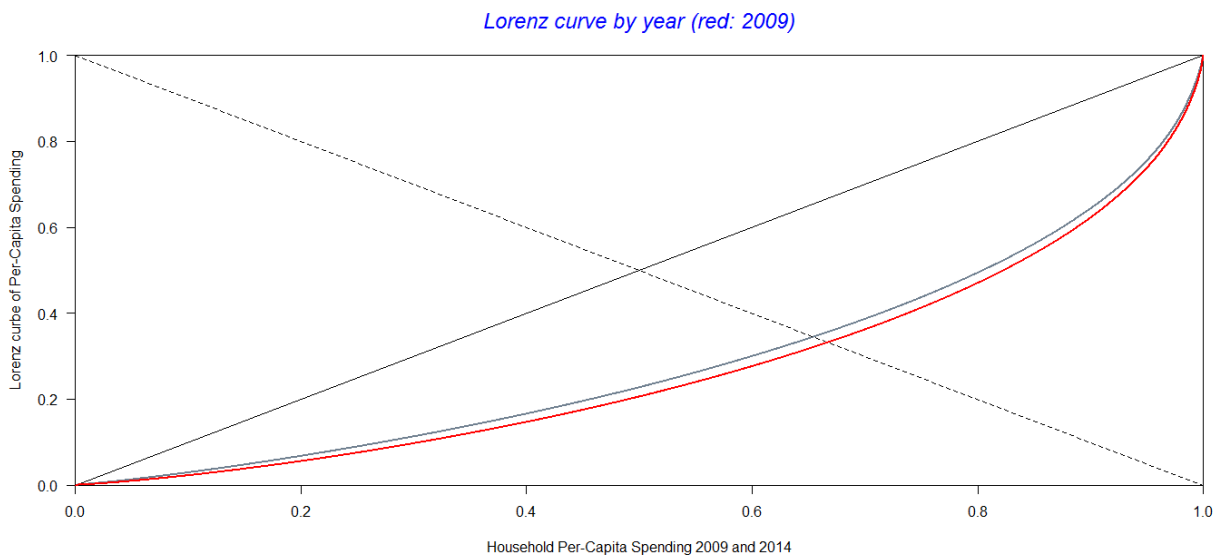
Figure 2: Inequalities in the distribution of economic well-being 2009 and 2014

Looking at the lorenz curves in figure (2), as shown by the red (2009) and gray (2014) curves, and independently of the head of household's literacy status, we see that inequality in total household per-capita consumption spending (economic well-being) among households, has decreased between 2009 and 2014. These graphical results by the lorenz curves are also supported by the overall Inequality measures in table (4). In fact both the Atkinson and Gini coefficients measures decreased from 2009 to 2014. These suggest a reduction in the level of inequality in total economic well-being among households in Burkina Faso.

	2009	2014
n	8404	10411
Adkinson	0.178	0.152
Gini	0.462	0.427

Table 4: Overall Inequality Measures for 2009 and 2014

Now looking at the distribution of Poverty by literacy status (Literate, Illiterate) as shown in Table(5), for 2009 and 2014, All three poverty indices (ie. Watts, Sen, Foster($\alpha = 1$)), computed using the "pov" function in the R library "ineq" (see Zeileis (2014)), paint the same picture. Their values are relatively higher for the households with illiterate heads compared to their literate counterparts, for 2009 and 2014. These results suggest that for both years, the incidence of poverty among illiterate headed households is higher than that among their literate counterparts.

	2009		**2014**	
Poverty Line	**130350**		**153530**	
	Literate	**Illiterate**	**Literate**	**Illiterate**
Watts	0.036	0.101	0.066	0.169
Sen	0.041	0.112	0.073	0.178
Foster ($\alpha = 1$)	0.146	0.355	0.227	0.480

Table 5: Poverty Measures by Literacy Status for 2009 and 2014

The distribution of overall poverty nationally without distinction of literacy status for the years 2009 and 2014 as shown in table(6) suggests that overall poverty has decreased nationally, as shown by the decrease in the coefficients values for the indices of Watts, Sen and the head count measure of Foster ($\alpha = 0$).

	2009	2014
Poverty Line	130350	153530
Watts	0.148	0.134
Sen	0.149	0.142
Foster ($\alpha = 0$)	8.785	5.742
Foster ($\alpha = 1$)	0.357	0.392

Table 6: Overall Poverty Measures for 2009 and 2014

5.2. Econometric results and Interpretation

This section presents the results from three estimated specifications corresponding respectively to the bivariate models of : 1) literacy status and Food-wellness , 2) Literacy status and Non-Food Wellness, 3) Literacy status and Total Economic Well-being. For each of the three model specifications, we present two estimation results :One from the classic Bivariate Estimation (based on Heckman (1979) sample selection model using the two-step and maximum likelihood approaches), and the other from the Semi-parametric Bivariate Estimation (based on the penalized maximum likelihood framework proposed by Marra and Radice (2013) for the bivariate normal case) which deals simultaneously with non-random sample selection using conventional systems of two equations, with non-linear covariate effects using spline approach, and also copes with non-normal bivariate distributions (see equation(6)). In doing so, we use a novel approach for relaxing the bivariate normality assumption in selection models using non-linear copula functions following McGovern et al. (2015).

The outcome equations in the Bivariate Economic well-being models are log-linear, therefore the literal interpretation of the estimated $\hat{\beta}$ coefficients is that a one-unit increase in the explanatory variable will produce an expected increase in log(well-being) of $\hat{\beta}$ units. In terms of the level of well-being itself, this means that the expected value of well-being is multiplied by $e^{(\hat{\beta})}$ So in terms of effects of changes in the explanatory variable X on Y(unlogged):

1. Each 1-unit increase in X multiplies the expected value of Y by $e^{(\hat{\beta})}$
2. For small values of $\hat{\beta}$, approximately $e^{(\hat{\beta})} = 1 + \hat{\beta}$. We can use this for the following approximation for a quick interpretation of the coefficients: $100 *(\hat{\beta})$ is the expected percentage change in Y for a unit increase in X. For instance for $\hat{\beta} = 0.06$, $e^{(0.06)} = 1.06$, so a 1-unit change in X corresponds to (approximately) an expected increase in Y of 6%.
3. In sum, $\hat{\beta}$ is an elasticity measure capturing the responsiveness of the outcome variable to a change in one explanatory variable. Therefore, for every one unit change in the independent variable X, the dependent variable Y changes by a $\hat{\beta}\%$.

As shown by the descriptive results in figure(1) and table (3) Economic well-being is distributed unequally based on heads of household literacy status. This indicates the possibility of non-random selection in our study sample as the sub-sample consisting of households headed by literate heads may differ in important characteristics from the sub-sample of households headed by illiterates. As such estimates based on models that ignore such a non-random selection, such as single equation modeling may be biased and inconsistent Heckman (1979). If the link between literacy status and Economic well-being is through observables alone, then selection bias can be avoided by accounting for these variables. However, if the link is through unobservables as well then inconsistent parameter estimates are obtained when using a classic single equation modeling approach.

Moreover, two other aspects may complicate modeling the relationship between the set of covariates and Economic Well-being: The first of which relates to variables such as age and education, likely to have a non-linear relationship with both literacy status and Economic well-being outcomes. This is because these variables embody productivity and life-cycle effects that are likely to have non-linear influences on the outcomes(Wojtys et al., 2016). Imposing a priori linear relationship (or non-linear by simply using for example quadratic polynomials) could mean failing to capture the true and more complex relationships. Finally, the often criticized assumption of bivariate normality as shown in equation (6) between the selection equation and the outcome equation may be too restrictive for applied work and is typically made for mathematical convenience. Therefore in our Semi-parametric estimation of the models, the effect of the numeric "age" covariate is flexibly modeled using thin plate regression splines with basis dimensions equal to 10 and penalties based on second order derivatives, as opposed to the quadratic polynomial specification in the classic estimation. The results based on the classicHeckman (1979) two-step estimator with maximum likelihood, and the semi-parametric copula regression spline with penalized likelihood (Wojtys et al., 2016, Section 2.2), are presented in tables (7), (8), and (9) respectively for the three estimated specifications (1), (2), and (3) discussed above.

5.2.1. Literacy and Food-wellness

Starting with the results of the bivariate model of Literacy status and Food-Wellness, figure(3) shows that the errors in the latent variable models (equations (4) and (6)) for LiterStatus and LnCapSpendgF have a positive high correlation $\hat{\theta} = 0.91$ that is highly statistically significant as shown by the 95% confidence interval (0.889, 0.924) in table (7). $\hat{\sigma} = 0.835$ is also statistically significant as shown by its 95% confidence interval (0.804, 0.865). These are indications that the unobserved factors affecting the head of household literacy status, also affect the household's food-wellness as measured by per-capita food consumption spending. The estimated overall average from the food wellness outcome in the model is 10.8 which is also statistically significant as shown by the 95% confidence interval of (10.2, 11.3).
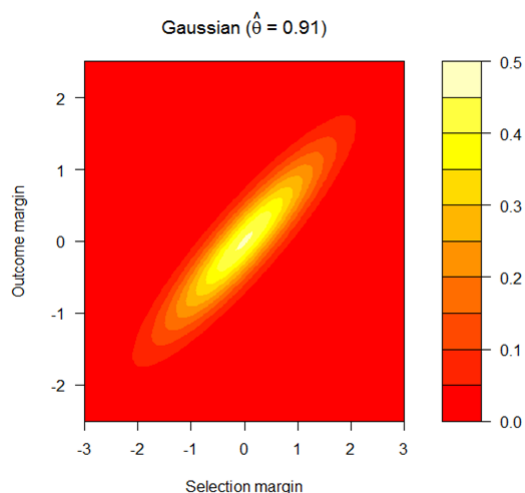
Figure 3: Bivariate distribution of Literacy status and Food wellness

As for the literacy status (selection) equation of this model, the results presented in the third column of table (7) show that all the variables (Gender, Marital Status, and Place of Residency) which enter the model parametrically, are statistically significant at the 1% level. In fact, a gender inequality in literacy status shows up with female heads of households being 51.7% less likely than males to have a literate status. Also compared to singles, married and widowed heads of household have respectively 51.8% and 74% less chances of being literate. Finally, heads of households living in rural areas have 80.6% less chances of being literate compared to their urban counterparts. For the smoothed "age" term, calculated as discussed in section 2.3 of (Wojtys et al., 2016), the p-value ($< 2e - 16$) and estimated degrees of freedom (edf $= 7.356$) indicates that age does have a significant impact on the head of household literacy status, and as shown by the smooth function estimate and 95% confidence interval in figure(4) this effect is negative.
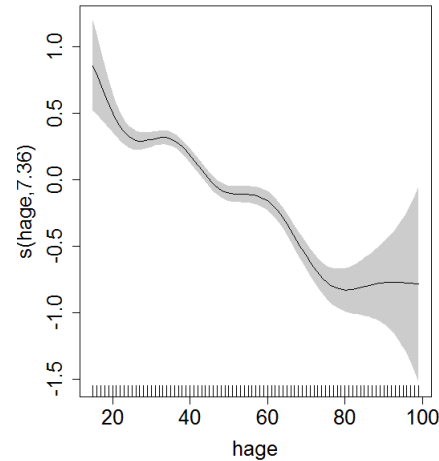
Figure 4: Smooth function estimates and 95% confidence bands from the age variable in the literacy status equation of the food-wellness model

Regarding the outcome equation in the food-wellness model, the fourth column of table (7) shows that all the variables (Gender, Marital Status, and Place of Residency, Secondary and Higher Education) which enter the model parametrically, are statistically significant at the 1% level, except for Primary education which is significant at the 5% level. In fact, the results suggest that the gender inequality previously observed in literacy status translates into food-wellness since households headed by females show a 38.9% less per-capita Food consumption spending compared to those headed by males. Also, compared to households headed by singles, those headed by married and widowed heads have respectively 64.9% and 76.9% less per-capita Food consumption spending, which means relatively lower food-wellness. Moreover, households living in rural areas have 70.7% less per-capita Food consumption sending compared to their urban counterparts.

Looking at the effects of education, compared to households where the head has no education, those where the head has a primary education have 5.4% less per-capita food consumption spending, while those with secondary and higher educations have respectively 13.2%, and 48.1% more per-capita food consumption spending. These effects of education suggest that simply completing a primary degree does not give an edge on food-wellness over the uneducated, as a matter of fact, the results seems to suggest a comparative disadvantage. However, once a head of household passes primary education and acquires a secondary or a higher degree, then the secured per-capita food-wellness over the uneducated household is both statistically and economically significant. This further supports the idea that furthering education beyond the primary level provides for better food security in Burkina Faso. Finally looking at the effect of household size, the coefficient estimate in the fourth column of table (7) suggest that for every one unit increase in household

size, household's food wellness as measured by per-capita food consumption spending decreases by 5.5%. This might be explained by the fact that there is now more mouth to feed with increased household size, decreasing then the amount of food received by each member.

Now turning to the smoothed "age" term in the food-wellness outcome equation, the p-value ($< 2e - 16$) and estimated degrees of freedom (edf $= 6.644$) indicates that the age of the head of household does have a significant impact on household food-wellness. As shown by the smooth function estimate and 95% confidence interval in figure(5) this effect is quadratic, with a pick around the 30s from which it starts decreasing.
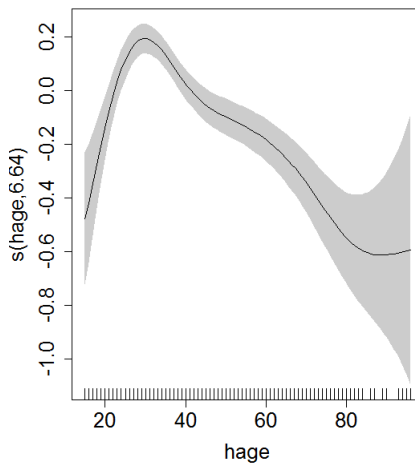

Figure 5: Smooth function estimates and 95% confidence bands from the age variable in the outcome equation of the food-wellness model

5.2.2. Literacy and Non-Food Wellness

Looking at the results of the bivariate model of Literacy and Non-Food-Wellness, figure(6) shows that the errors in the latent variable models (equations (4) and (6)) for LiterStatus and LnCapSpendgNF have a negative correlation $\hat{\theta} = -0.37$ that is statistically significant as shown by the 95% confidence interval (-0.648,-0.011) in table (8). $\hat{\sigma} = 0.616$ is also statistically significant as shown by its 95% confidence interval (0.564,0.673). These indicate that the unobserved factors affecting the head of household literacy status, also affect the household's Non-food-wellness as measured by household per-capita consumption spending on Non-Food items. The estimated overall average from the Non-food-wellness outcome in the model is 11.81 which is statistically significant as shown by the 95% confidence interval of (08.85, 14.77).

Overall using the AIC and BIC criterion presented in table (8), the semi-parametric model with (AIC $= 18080.78$, and BIC $= 18302.09$) is preferred over the classical model with (AIC $= 18174.14$,

Table 7: Food Wellness: Estimates from the Parametric and Semi-parametric Sample Selection Models

	Classic Bivariate	Sample Selection	Semi-Parametric Bivariate	Sample Selection
	LiterStatus	LnCapSpendgF	LiterStatus	LnCapSpendgF
CONST	1.535***	12.510***	0.608***	12.333***
	(0.067)†	(0.090)	(0.060)	(0.052)
FEMALE	-0.517***	-0.406***	-0.517***	-0.389***
	(0.054)	(0.049)	(0.054)	(0.048)
AGE	-0.020***	-0.001	(edf = 7.356)	(edf = 6.644)
	(0.001)	(0.004)	p-val =$(< 2e - 16)$	p-val =$(< 2e - 16)$
MARRIED	-0.512***	-0.550***	-0.518***	-0.649***
	(0.060)	(0.046)	(0.061)	(0.047)
WIDOW	-0.743***	-0.675***	-0.740***	-0.769***
	(0.092)	(0.083)	(0.091)	(0.083)
RURAL	-0.809***	-0.718***	-0.806 ***	-0.707***
	(0.028)	(0.029)	(0.027)	(0.028)
AGESq		-0.0001**		—
		(0.00004)		—
PRIMARYED		-0.055*		-0.054*
		(0.023)		(0.022)
SECONDARYED		0.121***		0.132***
		(0.026)		(0.025)
HIGHERED		0.494***		0.481***
		(0.041)		(0.041)
HHSIZE		-0.057***		-0.055***
		(0.002)		(0.002)
$\hat{\sigma}$		0.843		0.835
		(0.811, 0.876)††		(0.804, 0.865)
$\hat{\theta}$		0.907		0.908
		(0.888, 0.922)		(0.889, 0.924)
AIC		17734.890		17629.75
BIC		17872.650		17847.27
AVER		10.800		10.8
		(10.20, 11.30)		(10.2, 11.3)

† standard deviation of the parameters in parentheses

†† The 95% confidence intervals for sigma and theta

*** is 0.01% level significance; ** is 1% level significance ; * is 5% level significance

and BIC =18311.9). Once again, showing that relaxing the classical assumptions provides for a better model. In the semi-parametric estimation, the effect of the "age" covariate is flexibly modeled using thin plate regression splines, with convergence diagnostics suggesting satisfactory convergence of the algorithm as shown by the largest absolute gradient value of (7.668202e-06)

which is close to zero, and the positive definite observed information matrix.
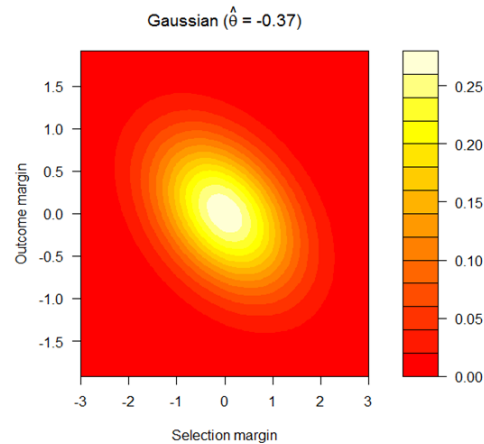

Figure 6: Bivariate distribution of Literacy status and Non-Food wellness

The results from the literacy status (selection) equation as presented in the third column of table (8) suggest that all the variables (Gender, Marital Status, and Place of Residency) which enter the model parametrically, are statistically significant at the 5% level. However, a gender inequality exists with respect to literacy status as female heads of household are 50.1% less likely than males to be literate.

With regards to marital status, compared to singles, married and widowed heads of household have respectively 37.1% and 63.2% less chances of being literate. Finally, head of households living in rural areas have 82.4% less chances of being literate compared to their counterparts in urban areas. For the smoothed "age" term, calculated as discussed in section 2.3 of (Wojtys et al., 2016), the p-value $(< 2e - 16)$ and estimated degrees of freedom (edf = 7.133) indicate that age does have a significant negative impact on the head of household literacy status, as confirmed by the smooth function estimate and 95% confidence interval in figure(7).
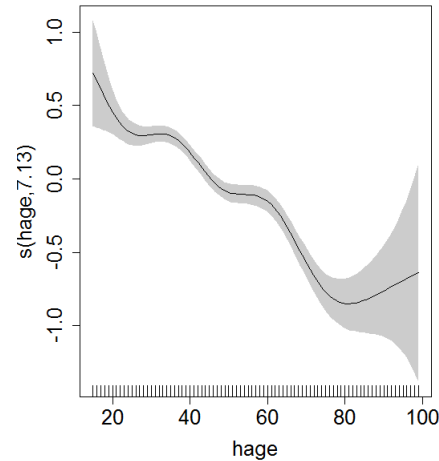
Figure 7: Smooth function estimates and 95% confidence bands from the age variable in the literacy status equation of the Non-food-wellness model

Regarding the outcome equation in the Non-food-wellness model, the fourth column of table (8) shows that except for the Widow marital status which is significant at the 5% level, all the remaining variables (Married marital status, Rural residency status, Primary, Secondary and Higher Education levels, and Household size) which enter the model parametrically, are all statistically significant at the 1% level. However, the results suggest that the gender inequality previously observed in literacy status and food-wellness does not translate into non-food-wellness since households headed by females seem to have 9.7% more (not less) per-capita non-food consumption spending compared to those headed by males, although this effect only reaches significance at the 10% level, but failing at the 5% level. Also, compared to households headed by singles, those headed by married and widowed heads have respectively 21.5% and 19.4% less per-capita non-food consumption spending, meaning relatively lower non-food-wellness. In addition, households living in rural areas have 30.8% less per-capita non-food consumption spending compared to their urban counterparts.

Looking at the effect of education, compared to households where the head has no education, those where the head has a primary education have 18.2% more per-capita non-food consumption spending, while those with secondary and higher educations have respectively 60.2%, and over 100% more per-capita non-food consumption spending. These observed effects of education on non-food-wellness suggest that simply completing a primary degree does give an edge on non-food-wellness over the uneducated, although this was not the case for food wellness. Furthermore, once a head of household passes primary school to acquire a secondary or higher education, then the secured per-capita non-food-wellness over the uneducated household head is statistically and

economically significant. Finally looking at the effect of household size, the coefficient estimate in the fourth column of table (8) suggest that for every one unit increase in household size, household's non-food-wellness as measured by per-capita consumption spending on non-food items decreases by 6.1%.

Turning now to the smoothed "age" term in the non-food-wellness outcome equation, calculated as in (Wojtys et al., 2016, section 2.3), the p-value ($< 2e - 16$) and estimated degrees of freedom (edf = 7.39) indicates that the age of the head of household does have a significant impact on household non-food-wellness. However, as opposed to the food-wellness case where this effect is quadratic, for the non-food-wellness case the effect is purely increasing as shown by the smooth function estimate and 95% confidence interval in figure(8).
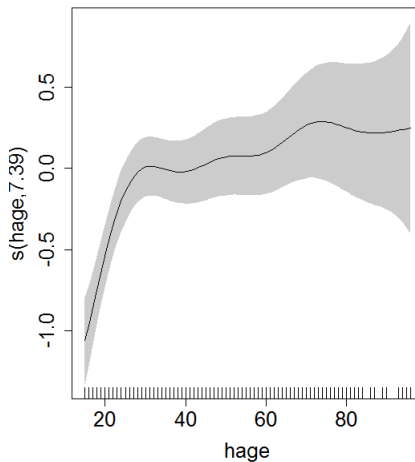

Figure 8: Smooth function estimates and 95% confidence bands from the age variable in the outcome equation of the non-food-wellness model

5.2.3. Literacy and Total Economic Well-being

Moving now to the results of the bivariate model of Literacy and total economic Well-Being, figure(9) shows that the errors in the latent variable models (equations (4) and (6)) for LiterStatus and LnCapSpendg have a positive high correlation $\hat{\theta} = 0.855$ that is statistically significant as shown by the 95% confidence interval (0.824, 0.885) in table (9). $\hat{\sigma} = 0.732$ is also statistically significant as shown by its 95% confidence interval (0.699, 0.767). These are indications that the unobserved factors affecting the head of household literacy status, also affect household's total economic well-Being as measured by per-capita household consumption spending on combined food and non-Food items. The estimated overall average from the total economic Well-Being outcome

Table 8: Non-Food Wellness: Estimates from the Parametric and Semi-parametric Sample Selection Models

	Classic Bivariate	Sample Selection	Semi-Parametric Bivariate	Sample Selection
	LiterStatus	LnCapSpendgNF	LiterStatus	LnCapSpendgNF
CONST	1.379***	11.830***	0.482***	12.487***
	(0.067)†	(0.096)	(0.061)	(0.068)
FEMALE	-0.492***	0.102*	-0.501***	0.097
	(0.054)	(0.049)	(0.055)	(0.057)
AGE	-0.020***	0.022***	(edf = 7.133)	(edf = 7.39)
	(0.001)	(0.004)	p-val =($< 2e - 16$)	p-val =($< 2e - 16$)
MARRIED	-0.345***	-0.132**	-0.371***	-0.215***
	(0.060)	(0.041)	(0.063)	(0.045)
WIDOW	-0.615***	-0.092	-0.632***	-0.194*
	(0.092)	(0.082)	(0.093)	(0.089)
RURAL	-0.827***	-0.267***	-0.824***	-0.308***
	(0.028)	(0.047)	(0.028)	(0.066)
AGESq		-0.0001**		—
		(0.00004)		—
PRIMARYED		0.181***		0.182***
		(0.026)		(0.026)
SECONDARYED		0.593***		0.602***
		(0.029)		(0.029)
HIGHERED		1.212***		1.192***
		(0.044)		(0.044)
HHSIZE		-0.062***		-0.061***
		(0.002)		(0.002)
$\hat{\sigma}$		0.645		0.616
		(0.598, 0.692)††		(0.564,0.673)
$\hat{\theta}$		-0.485		-0.375
		(-0.65, -0.271)		(-0.648,-0.011)
AIC		18174.14		18080.78
BIC		18311.9		18302.09
AVER		11.900		11.810
		(09.93, 13.87)		(08.85, 14.77)

† standard deviation of the parameters in parentheses

†† The 95% confidence intervals for sigma and theta

*** is 0.01% level significance; ** is 1% level significance ; * is 5% level significance

in the model is 11.6 which is statistically significant as shown by the 95% confidence interval of (10.9 , 12.3).
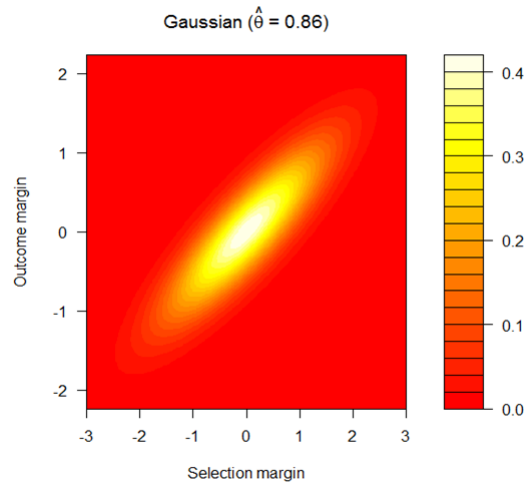
Figure 9: Bivariate distribution of Literacy status and Total Economic well-Being

Overall using the AIC and BIC criterion presented in table (9) for the two models, the semi-parametric model with (AIC = 17238.2 , and BIC = 17455.37) is preferred over the classical model with (AIC = 17452.3 , and BIC = 17590.06). This once again shows that relaxing the classical assumptions provides for a better model. In the semi-parametric estimation with the effect of the "age" covariate flexibly modeled, convergence diagnostics suggest satisfactory convergence of the algorithm as shown by the largest absolute gradient value of (9.097074e-06) which is close to zero, and the positive definite observed information matrix.

In relation to the literacy status (selection) equation of this model, the results presented in the third column of table (9) show that all the variables (Gender, Marital Status, and Place of Residency) which enter the model parametrically, are statistically significant at the 1% level. In fact, the results exhibit an overall gender inequality with respect to literacy status since female heads of households are 49.9% less likely than males to be literate.

Also compared to singles, married and widowed heads of household have respectively 45.5% and 69.8% less chances of being literate. Finally, heads of households living in rural areas have 81.3% less chances of being literate compared to their counterparts in urban areas. For the smoothed "age" term, the p-value ($< 2e - 16$) and estimated degrees of freedom (edf = 7.255) indicate that age does have a significant impact on the head of household literacy status, as confirmed by the smooth function estimate and 95% confidence interval in figure(10).
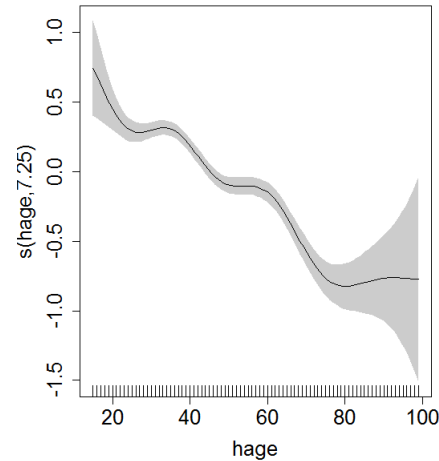
Figure 10: Smooth function estimates and 95% confidence bands from the age variable in the literacy status equation of the total economic well-Being model

Regarding the outcome equation in the total economic well-Being model, the fourth column of table (9) shows that except for Primary education which is significant at the 5% level, all the other variables (Gender, Marital Status, and Place of Residency, Secondary and Higher Education) which enter the model parametrically, are statistically significant at the 1% level. In fact, the results show that the gender inequality previously observed in literacy status, and food-wellness also translates into total economic well-being since households headed by females have 26.7% less per-capita combined food and non-food consumption spending compared to those headed by males. Also, compared to households headed by singles, those headed by married and widowed heads have respectively 51.9% and 64% less per-capita combined food and non-food consumption spending, meaning a relatively lower total economic well-Being. In addition, households living in rural areas have 71.2% less per-capita combined food and non-food consumption spending compared to their urban counterparts. Looking at the effects of education, compared to households where the head has no education, those where the head has a primary, secondary or higher education have respectively 5% , 35.9% , and 85.4% more per-capita combined food and non-food consumption spending. Note that the magnitudes of the effects of education are all positives and increasing with increasing levels of education. Finally, looking at the effect of household size, the coefficient estimate in the fourth column of table (9) suggests that for every one unit increase in household size, household's total economic Well-being as measured by combined per-capita food and non-food consumption spending decreases by 5.8%.

Turning to the smoothed "age" term in the total economic well-being outcome equation, the p-value ($< 2e - 16$) and estimated degrees of freedom (edf = 6.697) indicate that the age of the

24

head of household does have a significant impact on household total economic well-being. However, as opposed to the non-food-wellness case where this effect is increasing, here this effect on total economic well-being is quadratic as shown by the smooth function estimate and 95% confidence interval in figure(11).
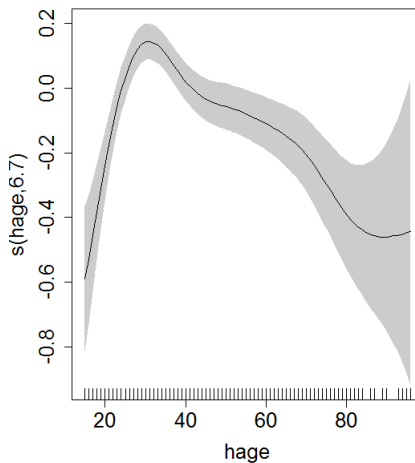

Figure 11: Smooth function estimates and 95% confidence bands from the age variable in the outcome equation of the total economic well-being model

6. Conclusion

The motivation for this empirical analysis was the desire to understand and raise awareness of the role that illiteracy plays in causing and propagating some of the biggest social problems such as poverty and inequality. To this end, the paper modeled the gender inequality in literacy status, and its effects on households economic well-being in Burkina Faso using a semi-parametric bivariate sample selection modeling approach. In this formulation, economic well-being was captured by households' per-capita consumption spending on food and non-food items from the National Survey on Household Living Conditions(EICVM, 2014) administered by the National Institute for Statistics and demography (INSD) in Burkina faso.

In relation to the descriptive analysis, the graphical results by the lorenz curve combined with the Inequality measures (Atkinson and Gini coefficients) suggested that the inequalities in overall per-capita consumption spending among literates headed households were higher than that of their illiterates counterparts for the years 2009 and 2014. Also, independently of the head of household's literacy status, it was found that the level of inequality in total economic well-being

Table 9: Total Economic Well-Being: Estimates from the Parametric and Semi-parametric Sample Selection Models

	Classic Bivariate	Sample Selection	Semi-Parametric Bivariate	Sample Selection
	LiterStatus	LnCapSpendg	LiterStatus	LnCapSpendg
CONST	1.397***	12.930***	0.554***	12.991***
	(0.068)†	(0.087)	(0.061)	(0.048)
FEMALE	-0.489***	0.013	-0.499***	-0.267***
	(0.054)	(0.044)	(0.054)	(0.044)
AGE	-0.020***	0.017***	(edf = 7.255)	(edf = 6.697)
	(0.001)	(0.004)	p-val =($< 2e-16$)	p-val =($< 2e-16$)
MARRIED	-0.360***	-0.296***	-0.455***	-0.519***
	(0.061)	(0.037)	(0.062)	(0.043)
WIDOW	-0.633***	-0.206**	-0.698***	-0.640***
	(0.092)	(0.072)	(0.092)	(0.076)
RURAL	-0.827***	-0.184***	-0.813***	-0.712***
	(0.028)	(0.035)	(0.027)	(0.028)
AGESq		-0.0001**		—
		(0.00004)		—
PRIMARYED		0.059*		0.050*
		(0.234)		(0.022)
SECONDARYED		0.377***		0.359***
		(0.026)		(0.025)
HIGHERED		0.876***		0.854***
		(0.040)		(0.039)
HHSIZE		-0.061***		-0.058***
		(0.002)		(0.002)
$\hat{\sigma}$		0.593		0.732
		(0.558, 0.63)††		(0.7, 0.767)
$\hat{\theta}$		-0.522		0.855
		(-0.649,-0.368)		(0.824,0.885)
AIC		17452.3		17238.2
BIC		17590.06		17455.37
AVER		12.700		11.600
		(11.20, 14.10)		(10.9, 12.3)

† standard deviation of the parameters in parentheses

†† The 95% confidence intervals for sigma and theta

*** is 0.01% level significance; ** is 1% level significance ; * is 5% level significance

among households in Burkina Faso decreased from 2009 to 2014. On the other hand the analysis of the distribution of poverty by literacy status (Literate, Illiterate) as shown for the years 2009 and 2014, using the three poverty indices (Watts, Sen, Foster($\alpha = 1$)), suggested that poverty among literates headed households was lower than that of their illiterates counterparts for both periods.

In addition, the distribution of overall poverty nationally without distinction of literacy status for the years 2009 and 2014 suggested that overall poverty had decreased nationally between the two periods, as shown by the decrease in the coefficients values for the indices of Watts, Sen and the Head count measure of Foster ($\alpha = 0$).

In regards to the econometric analysis, the classic sample selection modeling based on Heckman (1979) two-step estimator, and the semi-parametric bivariate sample selection modeling approach (Wojtys et al., 2016) were applied to the modeling of food-wellness, non-food wellness, and total economic well-being respectively. In all three cases the AIC and BIC criterion showed the semi-parametric sample selection modeling framework to perform better than its fully parametric counterpart. As described in the previous sections the results do support the existence of gender inequality in literacy status in all three cases, with women heads of households having less chances of showing a literate status than their men counterparts. In relation to the outcome equations, the results showed that although this observed gender inequality in literacy status did not show up in non-food-wellness, it did however translate into inequality in not only food-wellness with 38.9% less per-capita food consumption spending for female headed households, but also into inequality in overall wellness as female headed households showed 26.7% less per-capita combined consumption spending than male headed households.

Together, these results confirm the previously mentioned conclusion by UNESCO that "illiteracy is a major barrier to improving child malnutrition and food safety", and even more so when the illiteracy is affecting female heads of households. Overall the study consolidates the idea that reducing gender inequality in literacy skills is indeed an important step to overcoming the obstacles that lock individuals into a cycle of poverty and disadvantage, and education the best mean to overcoming poverty caused by illiteracy as shown by the global commitments through the 2030 agenda for sustainable Development, and the UN 2030 education for all pledge.

References

Benería, L., Berik, G., Floro, M., 2015. Gender, development and globalization: economics as if all people mattered. Routledge.

Cameron, A. C., Trivedi, P. K., 2005. Microeconometrics: methods and applications. Cambridge university press.

Cree, A., Kay, A., Steward, J., 2012. The economic and social cost of illiteracy: A snapshot of illiteracy in a global context. World Literacy Foundation.

Edwards, A. N., 1999. Dynamics of economic well-being: Poverty, 2009–2011. Journal of Human Resources 34, 557–88.

Fields, G. S., 2000. The dynamics of poverty, inequality and economic well-being: African economic growth in comparative perspective.

Goodman, M., Finnegan, R., Mohadjer, L., Krenzke, T., Hogan, J., 2013. Literacy, numeracy, and problem solving in technology-rich environments among us adults: Results from the program for the international assessment of adult competencies 2012. first look. nces 2014-008. National Center for Education Statistics.

Handel, M. J., 2003. Skills mismatch in the labor market. Annual Review of Sociology, 135–165.

Heckman, J. J., 1979. Sample selection bias as specification error. Econometrica 47 (1), 153–161.

Heckman, J. J., Mosso, S., 2014. The economics of human development and social mobility. Annual Review of Economics 6 (1), 689–733.
URL https://doi.org/10.1146/annurev-economics-080213-040753

Hornbeck, D. W., Salamon, L. M., 1991. Human capital and America's future: an economic strategy for the'90s. Johns Hopkins Univ Pr.

INSD, 2015a. Rapport enquete multisectorielle continue (emc) 2014: Alphabétisation et scolarisation. Tech. rep., Institut national de la statistique et de la démographie (INSD), http://www.insd.bf/n/index.php/publications?id=151.

INSD, 2015b. Rapport enquete multisectorielle continue (emc) 2014: Profil de pauvreté et d'inégalités. Tech. rep., Institut national de la statistique et de la démographie (INSD), http://www.insd.bf/n/index.php/publications?id=151.

Marra, G., Radice, R., 2013. Estimation of a regression spline sample selection model. Computational Statistics & Data Analysis 61, 158–173.

McGovern, M. E., Bärnighausen, T., Marra, G., Radice, R., 2015. On the assumption of bivariate normality in selection models: a copula approach applied to estimating hiv prevalence. Epidemiology (Cambridge, Mass.) 26 (2), 229.

Osberg, L., Sharpe, A., 2002. An index of economic well–being for selected oecd countries. Review of Income and Wealth 48 (3), 291–316.

Potter, D., 2014. Extending literacy: Addressing the needs of marginalised groups in papua new guinea and australia. This article was originally published as: Potter, D.(2014). Extending literacy: Addressing the needs of marginalised groups in Papua New Guinea and Australia. PNG Journal of Literacy, 1 (1), 1-14. ISSN: 2313-2507.

Prete, A. L., 2013. Economic literacy, inequality, and financial development. Economics Letters 118 (1), 74–76.

Project-Literacy, 2016. A call-to-action for education and literacy. Project Literacy Online Platform, http://buyimpossible.com/collections/project-literacy/this.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

Suh, S. W., Han, J. W., Park, J. Y., Hong, J. W., Kim, K., Kim, T., Lee, K. H., Han, G., Jeong, H., Seo, J., et al., 2016. Impacts of illiteracy on the risk of dementia: a global health perspective. Journal of Alzheimer's Disease 53 (2), 731–741.

Sumner, A., 2004. Economic well-being and non-economic well-being.

Sywelem, M. M. G., 2015. Literacy and adult education in egypt: Achievements and challenges. American Journal of Educational Research 3 (7), 793–799.
URL http://pubs.sciepub.com/education/3/7/1

Wojtys, M., Marra, G., Radice, R., 2016. Copula regression spline sample selection models: the r package semiparsamplesel. Journal of Statistical Software 71 (6).

Zeileis, A., 2014. ineq: Measuring Inequality, Concentration, and Poverty. R package version 0.2-13.
URL https://CRAN.R-project.org/package=ineq