Inequality Decomposition by Regressed-Income Sources in Cameroon

Francis Menjo Baye and Boniface Ngah Epo University of Yaoundé II

Paper Prepared for the Special IARIW-SSA Conference on

Measuring National Income, Wealth, Poverty, and Inequality in African Countries

Cape Town, South Africa, September 28-October 1, 2011

Session 2B: Inequality and Poverty Issues (Parallel)

Time: 2:00 PM-5:30 PM

Inequality Decomposition by Regressed-Income Sources in Cameroon[1]

By

Francis Menjo Baye and Boniface Ngah Epo
Faculty of Economics and Management
University of Yaoundé II, Cameroon
P.O. Box 1365 Yaoundé
E-mail: bayemenjo@yahoo.com and epongahb@yahoo.fr

Abstract
This paper applies the regression-based inequality decomposition approach to explore determinants of income inequality in Cameroon using the 2007 Cameroon household consumption survey. We use a control function approach that tests for potential endogeneity and unobserved heterogeneity of synthetic variables for education and health, while controlling for other correlates of household consumption to generate contributions of regressed-sources to measured income inequality. The synthetic variables for education and health are constructed by the multiple correspondence analysis method to reflect the multidimensional character of health and education. The contribution of each source to measured income inequality is the sum of its weighted marginal contributions in all possible configurations of sources as sanctioned by the Shapley value decomposition rule. Regressed-income sources attributable to education, health, urban residency, household size, fraction of active household members, working in the formal sector and farmland ownership are the main determinants of household income inequality in that order. These results have policy implications for addressing inequalities in the ongoing process of growth, employment and poverty reduction in terms of improving access to labour-intensive employment opportunities, balanced development, family planning and education for all.

Keywords: Regression-based decomposition, Inequality, household Economic well-being and Cameroon.

1. Introduction

Concerns about inequality and its impacts in developing countries have been rekindled in the development agenda of governments and donors as globalisation unfolds. The main observation is that sub-Saharan Africa (SSA) is experiencing rising income inequalities despite efforts made in incorporating a social dimension to the initial structural adjustment programmes. As observed by Wan and Zhou (2005), conventional decompositions by factor components or by population subgroups only provide limited information on the determinants of income inequality. In this context, the regression-based decomposition (RBD) analysis can shed more light to our understanding of factors that determine income inequality (Oyekale et al, 2007). In addition, RBD is more appealing than the standard decomposition exercise because it explores determinants of economic welfare

[1] This paper is an offshoot of research supported financially by the Australian Agency for International Development (AusAID), the Canadian International Development Agency (CIDA) and the International Development Research Centre (IDRC), and technically by the Poverty and Economic Policy (PEP) research Network. We would like to express our appreciation to the PMMA PEP resource persons, Jean -Yves Duclos, Sami Bibi and most especially Abdelkrim Araar for their invaluable mentoring.

econometrically and applies the parameter estimates to compute income sources that explain measured inequality in the distribution of household income (Epo et al, 2010).

While the RBD approach has been routinely applied to investigate factors that explain income inequality (Fields and Yoo, 2000; Morduch and Sicular, 2002; Alayande, 2003; Yuko et al, 2006; Kimhi, 2007), Wan (2002) observed some weaknesses associated with the functional specifications, the treatment of the error term and the exactness of the decomposition procedure. Although Wan (2002; 2004) outlined some orientations on how these shortcomings might be resolved, no study appears to have computed robust estimates that address these issues. The main difficulty is associated with the failure to carry out an exact decomposition of the estimated sources including the error term, putting aside the functional form or the inequality index adopted. This paper attempts to fill some of these lacunas by decomposing measured household income inequality into the different estimated sources and the error term simultaneously using the Shapley value (Shapley, 1953) procedure. In this regard, the marginal contributions of the estimated-income sources and the error term are computed based on the decomposition framework proposed by Shorrocks (1999). This procedure was implemented with the Distributive Analysis Stata Package (DASP 2.1) (Araar and Duclos, 2009) slotted in STATA 10.

Another value addition of this paper is the construction and use of synthetic variables for education and health as regressors in the income generating function that yields factors used in performing the inequality decomposition analysis. This gives a flavour of the multidimensional character of health, education and inequality. In this exercise, education and health carries multiple aspects subsumed as composite variables. They also translate the key role human capital characteristics (Becker, 1967, Grossman. 1972) play regarding household utility and production functions, and therefore household economic well-being.Thus, in keeping with Sen`s (1982) concepts of capabilities and functionings, the synthetic variables capture more policy relevant information on household possession of attributes that may suggest more policy implications.

Understanding how much of total inequality is captured by the regressed-sources is important for targeting the roots of inequality in Cameroon. As observed by Awoyemi and Adekanye (2003), inequality may emanate from: (1) the logical outcome of the market economy, which constitutes pathways for socio-economic segmentation; or (2) the skewed developmental focus in favour of urban dwellers relative to rural populations in terms of access to education, health and other infrastructures. In addition, inequality can be aggravated by poor governance, corruption, poor institutions and administrative inertia. Since these factors affect income redistribution patterns and programs, it is important for governments to be concerned about the underlying factors that perpetuate inequality.

Despite a fall in the incidence of poverty between 1996 and 2001, following the increase in the period 1984-1996, inequality has, at best, stagnated in Cameroon (Araar, 2006; Baye, 2008). Inequality measures show that: (1) while in 1996 the ratio of total expenditures between the poorest quintile and the richest quintile was 1:7, in 2001 this ratio was 1:8; (2) the Gini index marginally increased between 1996 and 2001 (0.406 in 1996 and 0.408 in 2001); (3) while the Gini index for rural areas worsen from 0.345 in 1996 to 0.369 in 2001, that for urban areas ameliorated from 0.449 in 1996 to 0.406 in 2001; and (4) in the cities the gap between the poor and non-poor is more pronounced than in the countryside (INS, 2004;

2005). Overall, between 2001 and 2007 total inequality slightly declined from 0.408 to 0.390, retreating more in cities than rural areas. In terms of inequality decomposition by subgroups, Baye and Fambon (2002) and Baye (2008) found under different dimensions and indicators that within group components overwhelmingly accounted for inequality compared to the between group components. However, the main shortcoming of such analyses is that they fail to identify and quantify the fundamental determinants of either of the two components. Thus we appeal to sources that significantly explain household welfare and its redistribution.

The main objective of this paper is to use the regression-based decomposition approach to explore determinants of income inequality in Cameroon using the 2007 Cameroon household survey. Specifically, it (1) estimates factor-endowments that significantly explain household economic well-being, (2) decomposes the relative importance of estimated-income sources vis-à-vis the residual in accounting for measured income inequality, and (3) derives policy implications on the basis of the analysis. The rest of the paper is organized in five sections. Section 2 gives a review of the literature. Section 3 dwells on the methodology, and the data and variables of interest are explored in Section 4. Section 5 presents the empirical results and Section 6 concludes the paper.

2. Review of literature

Several concepts and measures have been proposed in the literature to characterise inequalities in income distribution (Atkison, 1970; Sen, 1973; Kakwani, 1980; Shorrocks, 1982). An overview of the literature indicates that a lot of research is being done to configure parametric and non-parametric inequality indices into subgroups, income sources, causal factors and other units or characteristics (Heshmati, 2004b). Heshmanti (2004a) pays some attention to the interrelationship between income inequality and the non-income inequality dimensions admitting that while inequality can have many dimensions, economists have long been concerned primarily with the monetary dimension.

Regression-based inequality decomposition measurement can be traced back to Oaxaca (1973) and Blinder (1973). In the early 1990s, Juhn et al. (1993) applied this approach to allow for the decomposition of between-group differences in the full wage distribution rather than the mean of income as in Oaxaca (1973) and Blinder (1973). Bourguignon et al. (2001; 2008) relaxed the requirement of a linear income-generating function of Juhn et al. (1993). Wan (2002) observed that these efforts were devoted to explaining between-group differences in income distributions rather than quantifying contributions of individual determinants to total inequality.

DiNardo et al. (1996) and Deaton (1997) respectively proposed semi-parametric and non-parametric techniques that sought to model and compare the whole distribution of income in terms of density functions. However, as is the case with a lot of semi-parametric or non-parametric methods, the results obtained were rather inconclusive and too aggregated to be of interest to practitioners or policymakers. Fields and Yoo (2000) and Morduch and Sicular (2002) developed a framework for inequality decomposition, which is an extension of Shorrocks (1982; 1984; 1999) approach based wholly and directly on conventional regression equations. This was later extended by Wan (2004) to reveal the enormous flexibility and accommodating characteristics of the RBD approach.

Compared with the unconditional approach that amounts to an accounting exercise, the regression-based methods, depending on the modelling strategy, provide possibilities to quantify the conditional roles of various attributes in a multivariate context that allows for heterogeneity in responses. A range of different applications of the regression-based income inequality decomposition literature exist. Among these, note can be made of the extension by Morduch and Sicular (2002) where the composition of income by the different sources is observed, and the case where different income sources are accounted for by different income regimes in Farm-household income (Bardham and Boucher, 1998; Yuko et al., 2006; and Kimhi, 2007).

Wan (2002; 2004), however, noted that most regression-based income inequality decompositions usually ignored or incorrectly treated the constant and the residual terms. While encountering a constant as a source of income inequality in empirical analysis of income distribution is possible but rare, the presence of a constant is a rule rather than an exception in regression equations. Wan (2004) correctly questions the pertinence of discarding the residual term in standard regression-based decompositions. Although this term or its estimated counterpart is a white noise by definition, which means it does not affect the mean of the dependent variable in the estimated regression equation, nor does it affect the shape of the empirical Lorenz curve, its presence or absence does result in different income density functions and thus influences income distribution and measured inequality. The value added of including this term in decomposition analysis is that it indicates the proportion of the contribution of sources which are not captured by the income generating function when explaining inequality. Consequently, the potential and real advantage of this approach will be undermined and further advances in this area might be hampered, if this term is not appropriately treated.

Research on regression-based income inequality decomposition analysis is just beginning to gain prominence in SAA. Among these efforts, one can cite Alayande (2003) and Oyekale et al. (2007) who applied this analysis to Nigerian data. In the case of Cameroon, only the attempt by Tabi (2009) has been made to the best of our knowledge. In spite of these attempts, including synthetic variables, controlling for potential endogeneity, unobserved heterogeneity and using survey-based linear models, as well as computing the share of regressed-income sources simultaneously with the predicted residual to measured income inequality via marginal contributions as sanctioned by the Shapley value decomposition rule are the value additions of this paper.

3. Methodology

We briefly exposed the control function econometric model we intend to estimate before exploring the regression-based inequality decomposition framework.

3.1. Regression model

To generate reliable parameter estimates needed for the inequality decomposition exercise, we have to assume that both health and education are jointly and simultaneously determined with household welfare, thus we present health and education separately in the household income generating function that follows:

$$Y = w_1\delta_y + \sum_{k=1}^{2}\eta_k HC_k + \varepsilon_1 \qquad 1$$

where, Y and B_k, are household economic wellbeing and endogenous determinants of wellbeing such as health and education; w_1 is a vector of exogenous covariates such as individual, household, and community characteristics; δ_y is a vector of parameters including the constant term and those of exogenous explanatory variables that correlate with the income generating function to be estimated; η_k are parameters of the potential endogenous explanatory variables (health and education) in the economic wellbeing function; and ε_1 is the error term.

The estimation of the parameters η_k would show the effects of health and education on household economic wellbeing, but these estimates are likely to be bias. In this context, potential instruments are needed in order to consistently estimate effects of health and education on household welfare. The reduced form equation of household demand for health and education that accommodates such instrumental variables takes the form:

$$HC_k = w\delta_{hck} + \varepsilon_{2k} \qquad 2$$

where, HC_k is household health/education; w is a vector of exogenous variables, comprising of w_1 covariates that belong to the outcome equation and a vector of instrumental variables, w_2, that affect each of the endogenous inputs B_k (k=1, 2), but have no direct influence on household economic wellbeing generating function, Y; δ_{hck} are vectors of parameters of exogenous explanatory variables in the reduced form health/education equation to be estimated and ε_{2k} are error terms.

To account for the potential endogeneity and heterogeneity of responses of unobservable variables, Equation 1 can be augmented to Equation 2, which is the control function model of interest.

$$Y = w_1\delta_y + \sum_{k=1}^{2}\eta_k HC_k + \sum_{k=1}^{2}\alpha_k\hat{\varepsilon}_{2k} + \lambda(\hat{\varepsilon}_{2k} * HC_k) + \varepsilon \qquad 3$$

where, $\hat{\varepsilon}_{2k}$ is fitted residual of an endogenous input, derived from the reduced form model (Equation 2). The predicted residual, $\hat{\varepsilon}_{2k}$, serves as the control for unobservable variables that correlate with HC_k, thus allowing these endogenous inputs to be treated as if they were exogenous covariates during estimation; $(\hat{\varepsilon}_{2k} * HC_k)$ is the interaction of the predicted residuals with the actual values of each of the potential endogenous variables; ε is a composite error term comprising ε_1 and the unpredicted part of ε_2, and δ, η, α *and* λ are parameters to be estimated.

3.2. Regression-based inequality decomposition framework

Shorrocks (1982) established a measure of inequality[2] written as a weighted sum of income:

$$I(y) = \sum_i a_i(y) y_i \tag{4}$$

where a_i is the income share of household *i*, y_i is the income of household *i*, *y* is total income, *I(y)* is the weighted sum of total household income, corresponding to an inequality measure, and $a_i(y)$ is the ethical weight attributed to individual *i* based on the vector of income *y*. Since household income may be observed as the sum of income from M sources or endowments, $y_i = \sum_{m=1}^{M} y_{i,m}$, the inequality measure can now be written as the sum-specific component S_m:

$$I(y) = \sum_i a_i(y) \sum_m y_{i,m} = \sum_m \left(\sum_i a_i(y) y_{i,m} \right) = \sum_m S_m \tag{5}$$

The proportional contribution of each income source may be obtained by dividing the sum-specific component by $I(y)$. Thus, the relative contribution of income source *m*, S_m, can be written as:

$$S_m = \frac{\sum_i a_i(y) y_{i,m}}{I(y)} \tag{6}$$

Shorrocks (1982) observed that the weight $a_i(y)$ may be chosen in an arbitrary manner leading to an infinite number of possibilities and proposed a unique decomposition rule that satisfy the following: (a) if a new distribution is obtained by multiplying all incomes by a constant, measured inequality should be the same under both distribution and; (b) if total income is divided into two components whose factor distributions are permutations of each other, their contributions to total inequality are equal.

Without loss of generality, we can express our estimating econometric model in Equation 3 in the form:

$$y = XB + \varepsilon \tag{7}$$

Using such a form Morduch and Sicular (2002) extended the decomposition rule (6) to a regression-based decomposition by determinants of household income. Where *X* is a vector of explanatory variables with the first column made of a vector of constants $\alpha = (1,1,.....,1)$, β is a vector of parameters to be estimated and ε is the residual term.

Given the vector of consistently estimated parameters ($\hat{\beta}$), income (Equation 7) can be expressed as a sum of the estimated-income source flows and the predicted error term ($\hat{\varepsilon}$) as in equation (8):

$$y = X\hat{\beta} + \hat{\varepsilon} \tag{8}$$

Since the econometric results yield estimates of the income source flows attributed to household variables, they allow us to make use of decomposition by income sources (or factor endowments). By construction, total income is the sum of these estimated income source flows (plus the predicted regression residual):

---

[2] Early attempts at decomposing inequality by subgroups include Theil (1972), Bourguiognon (1979), Shorrocks (1980, 1984) and Foster and Shneyerov (1997).

$$y_i = \sum_{m=0}^{M+1} \ddot{\hat{y}}_{i,m} \tag{9}$$

where $\ddot{\hat{y}}_{i,m} = \hat{\beta}_m x_{i,m}$ for m=0,1, 2, …, M and $\ddot{\hat{y}}_{i,m} = \hat{\varepsilon}_i$ for m=M+1.

Substituting equation (9) into equation (6), we obtain the share of inequality attributable to the income source, $\ddot{\hat{y}}_{i,m}$ as:

$$S_m = \frac{\hat{\beta}_m \sum_i a_i(y) x_{i,m}}{I(y)} \tag{10}$$

$\hat{\beta}_m$ is estimated coefficient associated with income source m, $x_{i,m}$ is the income source m attributable to household *i*, $\sum_i a_i(y)$ is the sum of weights attributable to households and, *I(y)* is the total income inequality index.

Wan (2004) extends the linear regression model by proposing the general form:

$$y = F(X) + \varepsilon \tag{11}$$

where F(X) allows for any form for the income generating function, which could be linear with the presence of the constant term or highly non-linear with the absence of this term. In terms of estimated income sources, equation 11 can take the form:

$$y = \hat{y} + \hat{\varepsilon} = \hat{\alpha} + \tilde{y} + \hat{\varepsilon} \tag{12}$$

where $\hat{y}$ is the deterministic part of the estimated income generating function, $\hat{\alpha}$ is the estimated constant, $\tilde{y}$ is the vector of estimated income sources excluding the estimated constant term and $\hat{\varepsilon}$ is the predicted residual. As noted by Wan (2004) and reviewed by Araar and Duclos (2009), relative inequality indices are not defined when the average of the variable of interest equals zero (the case of the residual). In addition, inequality indices equal zero when the variable of interest is a constant (the case of the estimated constant). To deal with these two problems, Wan (2004) proposed some basic rules. According to Araar and Duclos (2009), let total income be expressed in terms of regressed-income source flows and the predicted residual as:

$y = y_0 + y_1 + y_2 + ... + y_m + ... + y_M + \hat{\varepsilon}$,

$\hat{y} = y_0 + y_1 + y_2 + ... + y_m + ... + y_M$ and

$\tilde{y} = y_1 + y_2 + ... + y_m + ... + y_M$, where $y_m = \hat{\beta}_m X_m$.

Using $I(\cdot)$ as an inequality measure, then overall income inequality can be decomposed into the contribution of the constant term $I(y_0)$, the contribution of the estimated income sources $I(\tilde{y})$ and the contribution of the predicted residual $I(\hat{\varepsilon})$ as follows:

$$I(y) = I(y_0) + I(\tilde{y}) + I(\hat{\varepsilon}) \tag{13}$$

The contribution of the estimated constant to measured income inequality: $I(\hat{\varepsilon}) = I(\hat{y}) - I(\tilde{y})$ and the contribution of the predicted residual to measured income inequality: $I(\hat{\varepsilon}) = I(y) - I(\hat{y})$.

The difference between $I(y)$ and $I(\hat{y})$ is important. The ranking by y and $\hat{y}$ differs and would be equivalent if and only if there is a perfect fit of the income-generating function.

Looking at it from this perspective, the decomposition makes intuitive, as well as theoretical sense. Decomposing equation 11 entails that the disturbance term is irrelevant and does not affect income inequality. This is not true because in addition to earlier discussions, one should note that $I(y) \neq I(\hat{y})$ unless all $\varepsilon = 0$. One way to treat the residual term is to discard it altogether because the residuals are not explainable by the structural-income generating function. If this is the case, one could focus on $\hat{y}$ and obtained further decomposition results. This, however, is not recommended. The residual term can be viewed as representing factors or determinants other than those included in the regression model. Ignoring, ε, is certainly unwise as it does contain useful information and its contribution once identified, can inform policymakers on how much included factors can explain the overall inequality.

In general, there are two main approaches for the decomposition of total inequality by income sources: the analytical approach, which is based on algebraic developments to define an inequality index by a formula that shows the contribution of income sources, and the Shapley value approach, which is based on the expected marginal contribution of income sources to measured inequality. We use the two procedures in this paper for comparative purposes. In terms of inequality indices, we use the Gini, the coefficient of variation and the generalized class of entropy measures.

Using the Shapley value to generate the expected components of the different income sources that account for inequality in terms of marginal contributions, the basic idea hinges on the Shapley value concept as developed by Shorrocks (1999). According to the established rule, the entry of an extra factor in a set of factors permits the factor to benefit a marginal gain or loss commensurate to what it brings into the set. In this paper, the different estimated sources act as factors that explain inequality (Araar and Duclos, 2009).

When we are dealing with the semi-log linear functional form, the income generating function takes the form: $\log(y) = y_0 + y_1 + y_2 + ... + y_m + ... + y_M + \hat{\varepsilon}$ and with this specification, we have:

$$y = Exp(y_0 + y_1 + y_2 + ... + y_m + ... + y_M + \hat{\varepsilon}) = Exp(y_0) * \prod_{m=1}^{M} Exp(y_m) * Exp(\hat{\varepsilon}) \qquad (14)$$

This implies that adding a constant will have no impact on measured inequality since we are simply multiplying by a scalar.

4. Data and Variables of Interest

The data used in this study is the 2007 Cameroon household consumption survey (CHCS III). The CHCS III survey was collected by the National Institute of Statistics in the period May to July 2007. The targeted sample consisted of 12, 000 households of which 11391 were effectively visited. Its aim was to update knowledge on poverty and welfare status in Cameroon by providing indicators that capture the living standards of the local population. This survey covers the national territory. The two principal cities, Yaoundé and Douala, were considered as two urban strata. In addition, each of the 10 regions was divided into three strata- urban (large towns with at least 50, 000 inhabitants); semi-urban (small towns with at least 10, 000 inhabitants and at most 50, 000 inhabitants) and rural strata (settlements with less than 10, 000 inhabitants). In all, 32 strata were established for this survey. This comprises 12 urban strata (Yaoundé, Douala and the urban strata for the 10 regions that make up Cameroon), 10 semi-urban strata and 10 rural strata.

Two types of sampling designs were undertaken depending on the zone of residence. In the main cities of Yaoundé and Douala, a two-stage sampling frame was adopted. For other areas, a three-stage random sampling frame was adopted following the sequence city-primary sampling unit-household. 12 households were survey in each primary sampling unit in Yaoundé and Douala and 18 households in each primary sampling unit in other areas. The primary sampling units were chosen on the basis of the number of people residing in a particular area. Primary sampling units for urban areas were numbered 001 to 699. For rural areas, the numbering was from 700 to 900.

Data used for this analysis comprises both observed and synthetic variables. Based on the observed data obtained from the CHCS III household survey, the following variables were selected. The dependent variable considered as a proxy for income or production or well-being was household expenditure per capita. This variable is derived by dividing the total household expenditure by the number of individuals living in the household. The assumption with this variable is that there are no economies of scale in the household. The following independent variables were considered. Household size indicated the number of people living in a particular household at a given point in time. Age of household head indicates the age of the household head at the time of the survey. Fraction of active household members was generated as the proportion of active and working adults living in the household. The variable working in the formal sector was constructed to indicate that the household head is employed in the formal sector. The variable owning farmland indicates households in which the household head owns exploitable farmland and most farmland is inherited or owned communally. In terms of geography, pure urban areas were chosen, excluding pure rural areas and semi-urban areas to avoid perfect collinearity.

Variables instrumenting for education and health were related to access to information technology and housing quality - ownership of television, radio, and number of sleeping rooms. These values are captured at cluster level and expressed as cluster percentage shares. The idea here is that a given household cannot influence a societal variable (community variable), thus considering the cluster percentage shares in each primary sampling unit reduces potential endogeneity (Baye and Epo, 2009; Mwabu et al., 2000). The choice of the first two variables indicate the key role of communication in affecting education (Bailey, 2009; Fedotova, 2008) and health (Jackson et al., 1998; International Institute of Communication and Development (IICD) health sector report, 2008). The number of rooms reflects the role of adequate housing on health (WHO Regional Office for Europe’s Health Evidence Network, 2005; Douglas et al., 2003) and educational outcomes (Cheshire and Sheppard, 2002). Concerning the first two variables, one could argue that access to information channeled by radios and televisions owned by households will positively impact on education and health. The last variable shows the important role of housing quality on education and health. The main idea vehicle here is inspired from Becker (1962).

We constructed synthetic variables for education and health by the multiple correspondence analysis (MCA) method that captures the multidimensional notion of health and education. Moreover, as noted by Thomas (2001), it is widely recognized that health is multidimensional - reflecting the combination of an array of factors that include physical, mental and social well-being, genotype and phenotype influences, as well as expectations and information. Education is also multidimensional and includes amount of time spent in school, nature of the curriculum, quality of schooling at each stage, extent of learning in school, post-schooling training and skill acquisition. Modalities used to construct each of these synthetic variables included a wide range of questions that capture their multidimensional character and translate more public policy relevant information. (See, Appendix 1). The ordering of the various scores were generated and normalized to treat for the presence of negative values which may cloud the classification of observations and interpretation of results. Variables selected for our empirical work alongside their sample means and standard deviations are hosted in Table 1.

5. Empirical results

5.1. Descriptive statistics

Weighted descriptive statistics for the CHCS III survey indicated that 17.8 million people lived in Cameroon in 2007 (Table 1). The statistics identify that 55% of the total population live in pure rural areas and 35% in pure urban areas. The average age of household head was 44 years. Descriptive statistics indicate that 79 percent of the household interviewed were male. Sixty percent of households interviewed own farmland. In rural areas 20 percent of the households interviewed were headed by women, and 78% of these household owned or exploited farmland. In urban areas, 23% of the total populations interviewed are women. Averagely, households had six members. On average, one-fifth of household members were active and working. Regarding the formal sector, 15% of household heads worked in the formal sector. 52.54% of households own a radio, while about 33% of household own a television in the general population. The cluster percentage share of owning a radio and television were 0.13 and 0.09, respectively. For the number of rooms this value was 0.3489.

The descriptive statistics of the different modalities used to construct the synthetic variables were also computed. For instance, for the composite variable health, the average time to get to the nearest health district is 35mins. The average distance to the nearest health district is 2.8 kilometers. Over 56% of households chose to consult traditional doctors compared to 8% that visit health districts when they are sick. As for education, average distance to the nearest public school is less than a kilometer. For the nearest private school the distance is between 1 and 2 kilometers. The average time to get to the nearest public school is 25 minutes. To get to the nearest private schools, needs, on average, 35 minutes. 77% of household heads have at least gone to school. 71 % of household heads can read and write (NIS, 2007, 2008).

Table 1: Weighted Descriptive Statistics

Variable	Mean	SD	Min	Max
Outcome Variables				
Log Total Expenditure Per Head	12.427	0.6914	11.1852	16.244
Education *	1.0251	0.3762	0.04123	1.5352
Health*	0.6790	0.3878	0	1.4839
Household Size	6.4763	3.9868	1	43[3]
Age	44.395	14.279	11	99
Gender (1=male and 0=otherwise)	0.7907	0.4067	0	1
Fraction of Active Household Members	0.2090	0.1865	0	1
Formal Sector (1= yes and 0=otherwise)	0.1481	0. 3552	0	1
Own Farmland (1= yes and 0=otherwise)	0.6075	0.4883	0	1
Regions				
Pure Urban	0.3531	0.4779	0	1
Pure Semi-Urban	0.0973	0.2965	0	1
Pure Rural	0.5593	0. 4964	0	1
Instruments for composite variables for education and health				
Household own Radio (Cluster percentage Share)	0.1274	0.0512	0	0.2973
Household own Television(Cluster percentage Share)	0.0915	0.0942	0	0.4506
Number of rooms (Cluster percentage Share)	0.3489	0.1765	0.0615	1.3256
Control for Unobservable variables				
Education residual	$-1.16{*}e^{-10}$	0.2754	-1.2369	0.7949
Health residual	$2.75{*}e^{-10}$	0.3731	-0.8285	0.9455
Education times its residual	0.0758	0.2673	-0.7702	1.1751
Health times its residual	0.1392	0.3155	-0.2649	1.3063

Source: Computed by Authors using CHCS III (2007) and STATA 10. Notes: Variables with stars are synthetic variables obtained from the MCA.

5.2. Regression Results

Table 2 hosts the OLS (Column 1), the two stage least square (Column 2), the parsimonious control function estimates (Column 3) and the control function estimate with the interaction terms (Column 4), that identify correlates of household welfare. The models were globally significant with R-squares about 0.50. Findings may suggest that the IV 2SLS and Control function approach produce more robust results than the Ordinary Least Square approach because they account for the potential endogeneity bias. For instance, the estimated parameters for education and health by the 2SLS (Table 2; columns 2) increased almost three fold, compared to the OLS estimates in Table 2: Column 1. The estimates for education and health obtained from the CFA (Table 2; column 4) were higher than the 2SLS estimates. These results may be revealing the value added of the CFA over the other estimation techniques. This observation indicates the importance of properly estimating the structural parameters to correctly attribute effects for policy guidance. Furthermore, the fitted residual of the composite variables for education and health in the Control Function Approach

[3] Traditional fondoms in rural areas in Cameroon have very large household sizes

estimates significantly reduces expenditure. This entails an endogenous relation that negatively affects expenditure patterns of household members. In addition, controlling for non-linear interactions between education and unobservables, the interaction term was significant for education. Since the interaction term for health was not significant, we elect to use the parsimonious control function estimates, which are equivalent to the 2SLS estimates in the analysis that follows.

We also test for the relevance, strength and exogeneity of instruments (Table 2; column 2). According to Shea (1997), the first-stage F statistic and the partial R^2 convey vital information as to the validity and relevance of instruments in the case of a single endogenous variable. The first-stage F statistic on excluded instruments are 986 and 91.5, respectively (p-value=0.000) for the synthetic variables for education and health. The Cragg–Donald statistic is needed to assess the strength of excluded instruments (Stock and Yogo, 2004). This value was 39.97, greater than the Stock-Yogo weak ID test critical values: 10% maximal IV relative bias of 13.34. Tests at the bottom of Table 2 also show the education and health are indeed endogenous (Durbin-Wu-Hausman Chi-square Statistic = 519, p-value=0.00) which indicates that the OLS estimates are not reliable for inference, implying that the IV estimates are preferred.

Table 2, Column 4 reveals that education and health associate positively with household welfare. Access to better education enhances knowledge and choices made in the face of employment opportunities, production and labour market exigencies, which improve household income. Education reflects an opportunity for employment, thus the potentials to generate income. This finding corroborates the result obtained by Awoyemi (2003) for Nigeria; Morduch and Sicular (2002) for China and Maria and Jose (2008) for Cape Verde. The value added, however, is that it is based on a synthetic variable for education. In terms of health, the ability to access a district health center, short distances to these centers and quality services imply that these variables are likely to be positively associated with better handling of ill-health that might prevent individuals from undertaking income generating activities. In addition, economies of scale are generated from good health in terms of more labour market participation because health implies fewer sick days per annum.

Table 2 also hosts non-synthetic variables that correlate positively with household economic welfare. These variables are age of the household head, fraction of active household members, working in the formal sector and being a male household head. Working in the formal sector implies having a steady source of income, as well as other advantages like being able to borrow money and to have an adequate insurance policy. These tend to positively impact on household economic well-being.

Table 2: Determinants of Household Economic Well-being - Dependent variable is log of household expenditure per head

Variable	Ordinary Least Square (1)	Two-Stage Least Squared (2)	Control function excluding interaction term (3)	Control function including interaction term (4)
Endogenous variables				
Education	0.3133***	0.9991***	0.9991***	1.0281***
	(7.73)	(14.32)	(17.08)	(17.48)
Health	0.2000***	0.7044***	0.7044***	0.7101***
	(10.01)	(5.08)	(6.05)	(6.10)
Included Exogenous variables				
Household Size	-0.0267***	-0.0243***	-0.0243***	-0.0241***
	(-5.81)	(-12.08)	(-14.40)	(-14.31)
Age	0.0011**	0.0022***	0.0022***	0.0022***
	(2.01)	(2.98)	(3.55)	(3.56)
Gender (Male=1 and 0=otherwise)	0.0358**	0.1153***	0.1153***	0.1159***
	(2.37)	(5.09)	(6.08)	(6.12)
Fraction of Active Household members	0.9192***	0.9772***	0.9772***	0.9819***
	(17.70)	(26.90)	(32.09)	(32.26)
Formal Sector (1= yes and 0=otherwise)	0.3436***	0.1855***	0.1855***	0.1858***
	(14.42)	(10.13)	(12.08)	(12.11)
Household own farmland (1= yes and 0=otherwise)	-0.1289***	-0.0520***	-0.0520***	-0.0519***
	(-5.68)	(-3.60)	(4.29)	(-4.28)
Pure Urban area	0.4432***	0.1770***	0.1770***	0.1759***
	(15.34)	(7.87)	(9.39)	(9.34)
Constant	11.7432***	10.6289***	10.6289***	10.5663***
	(172.04)	(155.83)	(185.86)	(180.39)
Control function variables				
Predicted residual Education			-0. 822***	-1.0277***
			(-13.52)	(-13.67)
Predicted residual for Health			-0. 5278***	-0.6010***
			(-4.51)	(-4.85)
Education times its predicted residual				0.1999***
				(4.65)
Health times its predicted residual				0.0807
				(1.42)
R-Squared	0.4929		0.5161	0.5171
Centred/Adjusted R-squared		0.9979	0.5156	0.5165
Fisher Test [p-value]	329; [0.000]	918.8; [0.0]	1103; [0.00]	936.9; [0.000]
Partial R-Squared for Education		0.1380		
Test of excluded instruments: F-stat[p-value]		607; [0.00]		
Partial R-Squared for Health		0.0216		
Test of excluded instruments: F-stat[p-value]		83.7; [0.00]		
Test of Joint Significance of Identifying Variables/Cragg-Donald weak Identification test				
F-Stat [10 % Relative Bias]		39.97 [13.34]		
Underidentification tests (Aderson canon corr. LR statistics)				
Chi-Sq [p-value]		119; [0.00]		
Sargan statistics (Overidentification test of all instruments)				
Chi-Sq (1) [p-value]		40; [0.00]		
Endogeneity test of endogenous regressors				
Chi-Sq (2) [p-value]		519; [0.00]		
Number of Observation	11391-total population: 17.9 million			

Source: Computed by Authors using STATA 10. Notes: ***, ** and * are 1, 5 and 10 percent significance levels, respectively. Variables in parenthesis are t-student values. Sampling weights are used and the standard errors are adjusted for survey design.

The fraction of active household members (the ratio of active household members to the household size) contributes positively to household income through the reasoning that an increase in the number of individuals in a given household undertaking income generating activities entails greater income generation with positive effects on household economic welfare. This result is similar to that obtained by Yuko et al. (2006) for farm households in Korea. Age correlates positively with household welfare at the 1% level. This finding is similar to the results obtained by Babatude et al. (2008) in studying determinants of poverty in South-Western Nigeria. Along gender lines, households headed by men endowed with higher economic welfare because of the likelihood of male heads obtaining jobs more easily than their female counterparts or the discrimination in the job market in favour of men.

Variables that downgrade household welfare are household size and ownership of farmland. Other things being equal, farmland ownership is expected to impact positively on household economic welfare. The negative and significant sign of farmland ownership may be indicating that households might not be operating their farm holdings profitably, but since formal safety-nets like insurance, unemployment benefits and old age pension facilities are not accessible to informal sector operators in Cameroon, they might sensibly continue to operate production units even if such units are economically unprofitable. Thus farm ownership might as well impact household economic well-being negatively. Moreover, the mean opportunity cost of rural labour typically approaches zero. We verified this atypical behavior by looking for the correlation between farmland ownership and the dependent variable. This correlation was indeed negative. Moreover, the bulk of the rural population (about 85%) has a household member who has access to farmland, whereas the rest of the rural population operates mainly in the formal sector, which is an important income generating factor. The negative on farmland may be a mechanical outcome of this observation.

The relationship between household size and household income was confirmed to be negative by the correlation matrix. This indicates that a higher number of "dependents" or individual residing in a particular household will tended to exert a lot of pressure on the meager household income and consequently an overall deterioration in well-being. The findings on farmland and household size corroborate those by Oyakele et al. (2007) in their study of urban and rural poverty in Nigeria.

Urban residency tends to increase household productivity and income generation, while rural residency instead reduces household economic welfare. Generally, households living in urban areas are exposed to many opportunities which are incomes generating than rural dwellers and that may explain why poverty levels appear lower in urban regions. This finding is in tandem with those by Alemayehu et al. (2005) for Nigeria and Mwabu et al. (2000) for Kenya.

5.3. *Regression-based inequality decomposition Results*

To decompose measured income inequality by regressed sources, we compute contributions of the various estimated factors using the analytical and the Shapley value-based approaches (Table 3). The difference between the analytical approach and the Shapley value approach is that, while the former computes inequality by the Gini index as the product of the income shares and the coefficients of concentration, the latter is based on a set of axioms (Shorrocks,

1999) and have the merit of computing the weighted marginal contributions of an estimated income source in various coalitions of income sources. These weighted contributions exactly sum up to the considered inequality measure. In Column 1 of Table 3, putting aside the constant term, the estimated income sources for education, health, age, fraction of active household members and urban residency had the highest income shares. The income sources: household size, owning farmland, the predicted residual for health and education registered negative income shares.

Column 2 of Table 3 shows results of the Gini decomposition generated by the analytical approach. The absolute contributions to inequality of education and residing in urban areas are highest. Other sources that contributed in explaining inequality were working in the formal sector, fraction of active household members, health, household size and owning farmland and. The estimated income source for education reveals the key role education plays over time in enhancing well-being and exacerbating inequality. This result is similar to the findings by Oyakale et al. (2007). Differences in educational achievements imply differences in the ability to earn income and consequently disparities in expenditure. Consequently, disparities in access to school infrastructure and knowledge acquisition as indicated by the synthetic variable for education affect household expenditure. This is reflected in the gaps in well-being between those households endowed with this attribute and those that do not have this attribute.

Urban residency engendered a higher relative contribution to income inequality. This implies that disparities in urban-rural contributions worsened overall income inequality. The ratio of active household members to household size had the fourth highest contribution in explaining inequality in the distribution of household well-being after those in formal sector employment. This implies that a larger number of active household members will improve household chances of labour market participation and appears as an important source of inequality in the distribution of living standards. Formal sector workers fared better in terms of well-being than informal sector employees and consequently contribute positively to measured income inequality.

The estimated sources for age, gender and the predicted residual for education and health inputs contributed in reducing inequality. The residual term contributed in increasing inequality. Variables not captured by our income generating function tend to account for up to about 24 % of measured income inequality (Column 2). This share is scaled down when one amalgamates it with the residuals for education and health. The residual term informs policymakers and others as to how much included-factors can explain the overall inequality. The overall Gini index accounted for was 0.3864.

Column 3, of Table 3 also host inequality decomposition of the Gini index based on the Shapley value. Sources that largely explain inequality were education and health. The relative contributions of these factors sum up to 33%. Other sources that contributed in explaining inequality were the fraction of active household members, household size, age of household head, working in the formal sector, owning farmland, the predicted residuals for education and health, and urban residency. The relative contributions of these regressed sources sum up to 26%.

Table 3: Decomposition of total inequality by estimated income sources

Income Sources	Analytical Approach		Shapley value Approach
	Income Shares (1)	Gini Index (2)	Gini Index (3)
Composite Variables			
Education*	0.8452	0.1524	0.0895
		(0.3944)	(0.2317)
Health*	0.1605	0.0166	0.0366
		(0.0431)	(0.0948)
Observed Variables			
Household Size	-0.0113	0.0021	0.0207
		(0.0055)	(0.0537)
Age of Household head	0.2149	-0.0045	0.0010
		(-0.0118)	(0.0025)
Gender(1=male and 0=otherwise)	0.0530	-0.0021	0.0008
		(-0.0055)	(0.0020)
Fraction of Active Household Members	0.2736	0.0526	0.0167
		(0.1362)	(0.0433)
Formal Sector (1=working in the formal sector and 0=otherwise)	0.0661	0.0585	0.0132
		(0.1515)	(0.0342)
Household own farmland (1=Own farmland and 0=otherwise)	-0.0397	0.0115	0.0051
		(0.0298)	(0.0132)
Pure Urban Area	0.1080	0.0790	0.0216
		(0.2046)	(0.0559)
Complementary Sources for education and health			
Predicted residual for Educational	-0.7678	-0.0122	0.0196
		(-0.0315)	(0.0508)
Predicted residual for Health	-0.0712	-0.0026	0.0077
		(-0.0069)	(0.0199)
Residual	0.0000	0.0942	0.1538
		(0.2439)	(0.3981)
Constant term	0.1686	-0.0592	
		(-0.1533)	
Total value	1.000	0.3864	0.3864
		(1.000)	(1.000)

Source: Computed by authors using STATA 10 and the DASP 2.1 Software developed by Araar and Duclos (2009).
Notes: Income sources with stars are synthetic variables obtained from the MCA approach. Values in brackets are the relative contributions.

Although health had a positive contribution, its magnitude is relatively small compared to education. The reason for the small contribution of health in measured income inequality is attributable to the modalities used to construct this composite variable. These modalities are fixed in nature, comprising durable public investments such as type of health structure constructed and appreciation of health services, which are quasi-accessible to both poor and rich households, and slow to vary overtime. However, the composite health indicator captures inequality relative to the dimensions outlined in Appendix 3.1. In terms of location, urban residency contributed about 5% in accounting for measured income inequality. This result indicates that, while poverty is lower because urban dwellers are exposed to more opportunities than rural residents, inequality within the urban dwellers is higher. In contrast, rural areas may tend to host many poor households and disparities among them are low. This result has implications for policies that curb push-factors of rural-urban migration.

Total inequality computed by the Gini index was 0.3864 (Column 3). This value is similar to the Gini index of 0.390 obtained by the National Institute of Statistic using total expenditures per adult equivalent computed from the same survey data (INS, 2008). The contribution of the predicted residual term to income inequality in this case is 39%. As indicated earlier, the residual term informs the political entrepreneurs as to how much regressed-sources can explain the overall measured inequality. In this case, included variables accounted for over 60% of total inequality. This indicates that policy makers may choose to design policies accordingly to deal with inequality based on included variables with some confidence. However, more investigations are needed to increase the margin of confidence in addressing the problem of inequality.

The marginal contributions of the estimated income sources using the Gini approach are illustrated next. The Gini index is deemed as the inequality index that behaves the best in reporting our results because it is good for decomposition by sources (Araar, 2006). These marginal contributions are based on the notion of the Shapley value concept developed by Shorrocks (1999), where a regressed-income source joins a league of sources and the marginal contributions are calculated. Thus the Shapley value-based component of each regressed-income source to measured income inequality is the weighted mean of the marginal contributions of the source in all configurations of sources including the residual. These contributions are generated by the DASP 2.1 software package. The level of entry indicates the position in which a regressed source is introduced to a set of already existing sources. The introduction of each source into a coalition of sources can be envisaged as a policy-mix.

In Appendix 3, Table A hosts marginal contributions of included and excluded regressed-income sources to measured income inequality along different configurations of sources. For instance, of the weighted mean of marginal contribution of the composite variable education of about 0.0895 to measured income inequality of 0.3864, about 0.015 is realised at level 1, that is, in the absence of other regressed-income sources and the predicted residual (see, Table 3 and Table A). As the effect of other regressed-income sources are progressively taken into consideration from level 2 through level 12, the sum of the remaining weighted marginal contributions of education is 0.074 (Table A). Whereas the source education at all levels of entry registered no negative, the source predicted residual of education subsequently registered an inequality equalizing trend from the eight level of entry. The implication here is that promoting only education for all would be equity augmenting, but promoting it alongside policies that curb inequality in other income sources would enhance the effectiveness of the education for all policy.

The second estimated income source with the highest marginal contribution is the composite variable for health. Its marginal contribution in explaining inequality at level one is 0.012. This makes up about 32 percent of the total share of this source (0.0366) in accounting for observed inequality. Progressively including other estimated sources increases the impact of this source in explaining inequality. Its impact decreases progressively. This finding consolidates the observation made earlier as concerns the source complementary health inputs. What can be drawn is that the combine results of this first two sources show that health constitutes a key factor in human capital development because it contribute to household utility and productivity. Consequently, targeting modalities used in constructing the synthetic-variable for health for policy formulation will help dissipate inequality. For instance, ameliorating the working conditions of health workers will ameliorate personnel public relations in welcoming and following-up patients. The CHCS III survey questioned individuals on their reasons for dissatisfaction with public health facilities. One of the main findings was the poor reception of health personnel. This modality was captured in the composite health variable. This indicates the important role health might play in perpetuating or reducing inequality. This reveals that policies that try to reduce inequality in access to health facilities are important. However, other policies that target other dimensions of well-being should be consolidated as well.

For household size, working in the formal sector, urban residency and owning farmland, we witnessed at certain levels of entry positive and negative values (Table A). The variable household size when considered alone (level 1) has a weighted marginal impact of 0.0047. This amounts to about one-quarter of the total impact of this source in explaining observed inequality. At the seventh level, the weighted marginal contribution of this source becomes constant.

A key result that can be identified from this reading is the role of spatial inequality, as made explicit by the estimated source area of residence, in explaining observed inequality is the source urban residency. For urban residency, of the weighted mean of marginal contributions of about 6% of measured income inequality, about 14% is realised at level 1, that is, in the absence of other income sources. As the effect of other income sources is progressively considered from level 2 through level 12, the remaining 86% of the weighted marginal contributions of urban residency is captured. Policies that encourage rural development would be inequality reducing, and would tend to be more effective if additional policy instruments are used to target other sources of measured income inequality. The indication of our analysis is that packaging policy instruments to address the problem of inequality in the distribution of living standards would be more effective than implementing policies in solo.

6. Conclusions and Policy Recommendation

This paper aimed at investigating regressed-sources that account for measured income inequality in Cameroon using the 2007 Cameroon household consumption survey. In this endeavor, efforts were made to resolve the shortcomings identified in the literature concerning studies that apply the regression-based inequality decomposition approach. Use was also made of synthetic variables for education and health constructed by the multiple correspondent analysis method. Both the STATA 10 and the DASP 2.1 packages were used to generate results. The Control Function modelling adopted in this paper has the advantage over other modelling approaches because it can be used to purge structural parameter estimates of potential biases due to endogeneity and heterogeneity of endogenous variables with unobservable variables. Both the analytical and Shapley value decomposition procedures were applied to compute the contributions of the estimated income sources in explaining measured inequality, while the preferred Shapley value approach was used to illustrate the weighted marginal contributions of the estimated-income sources.

The composite variables for education and health -human capital characteristics, were positively and significantly associated with household economic welfare. Non-synthetic variables that also associated positively with household economic well-being were fraction of active household members, working in the formal sector, age of household head, living in urban areas and being a male headed household. Household size and owning farmland related negatively with the income generating function.

Estimated-income sources such as education, health, fraction of active household members and working in the formal sector were prominent in accounting for measured income inequality. Urban residency also largely contributed to measured income inequality. Assuming that there is no guidance as to the correct measure of income inequality to based policy advice, included variables explain 60-75% of total inequality, meaning the residual takes about 25-40%, and policy makers may choose to design policies accordingly to deal with inequality and ignore other factors with some margin of confidence. In this study, we elected to base policy implications on the Gini coefficient because of its popularity and desirable properties, and the Shapley value-based contributions as heralded in the literature. The joint contribution of education and health in accounting for total inequality was 33%, indicating the key role human capital characteristics play in explaining observed inequality in the redistribution of household income.

The component of each regressed-source to measured income inequality was the sum of the weighted marginal contributions of that source in all configurations of sources as sanctioned by the Shapley value approach. In the case of the synthetic variable – education, of the weighted mean of the marginal contributions of about 23.2% of measured total income inequality of 39%, about 4% is realised in the absence of other regressed-sources. As its effect in leagues of other regressed-sources was progressively taken into consideration, the weighted marginal contributions of education reduced progressively, while accounting for the remaining 19.4% of measured income inequality. The implication was that promoting only education for all would be equity augmenting, but promoting it alongside considerations that target other regressed-sources of inequality would enhance the effectiveness of the education for all policy. Thus there seems to be more wisdom in packaging policy instruments when addressing problems of inequality than implementing policies unaccompanied.

The following policy suggestions to curb income inequality can be distilled from this study: (a) education for all would be inequality reducing. This implies particularly targeting rural areas and women; (b) pro-poor health spending by government such as the recent commitment to treat uncomplicated malaria free-of-charge for children below five years would be inequality reducing; (c) Labour-intensive economic activities if encouraged can enhance broad-based employability of the citizenry; (d) Increasing access to family planning can be welfare and equity enhancing. This may ensure adequate human capital development in terms of education and health, which act as engines for the acquisition of future standards of living; and (e) anchoring on broad-based infrastructural developments and agricultural modernization in rural areas would bridge rural-urban disparities and check rural exodus.

Bibliography

Adams, R. H.Jr. (2001). "Non-Farm Income, Inequality and Land in Rural Egypt." *Economic Development and Cultural Changes*, 50: 339-363.

Alayande, B. (2003). 'Development of Inequality Reconsidered: some Evidence from Nigeria.' Paper presented at the UNU/WIDER conference on Inequality, Poverty and Human Wellbeing in Helsinki, Finland between 29th and 31st of May, 2003.

Alemayehu, G., Neik de Jong, M.S. Kimenyi and, G., Mwabu. (2005). 'Determinants of Poverty in Kenya: A Household Level Analysis'. Working Paper 2005-44, Department of Economics, University of Connecticut.

Araar, A and Duclos, J. Y. (2009). Distributive Analysis Stata Package (DASP) 2.1 software. University of Laval, CIRPEE and the Poverty Economic and Policy Research Network.

Araar, A. (2006). "On the decomposition of the Gini Coeffiecient: an Exact Approach, with Illustration Using Cameroonian Data", Cahiers de recherché, 0602, CIRPEE.

Araar, A. (2009). "The Hybrid Multidimensional Index of Inequality." Cahier de recherche/Working Paper 09-45. Centre Interuniversitaire sur le Risque, les Politiques Economique et sur L'emploi (CIRPEE). October 2009

Arrow, K., J. (1951). *Social Choice and Individual Values*. New York: Wiley, 138 p.

Atkinson, A. (1970). "On the Measurement of Inequality." *Journal of Economic Theory*, n°2, p. 244-263.

Awoyemi and Adekanye. (2003). 'Gender Analysis of Economics Efficiency in Cassava-Based Farm holdings in South-Western Nigeria.' *Journal of Economic and Rural Development*. Department of Agricultural Economics, University of Ibadan,pp. 45-59.

Babatunde, R.O., Olorunsanya, E.O. and A.D., Adejola. (2008). "Assessment of Rural Household Poverty: Evidence from South-western Nigeria." *American-Eurasian J. Agric. & Environ. Sci.*, 3 (6): 900-905, 2008.

Bailey, H. (2009). "Issues Affecting the Social Sustainability of Telecentres in Developing Contexts: A Field Study of Sixteen Telecentres in Jamaica." *The Electronic Journal on Information Systems in Developing Countries*, 36(4).

Barham, B., and S., Boucher. (1998). "Mitigation, Remittances and inequality: Estimating the Net effects of Migration on Income Distribution." *Journal of Economics*. 55: 307-331.

Baye M., F and Epo, N, B. (2009). "Explaining Inter-Household Gender Inequality in Cameroon: An Oaxaca-Blinder Approach". Paper presented at the 65th annual congress of the International Institute of Public Finance (IIPF), in Cape Town, South Africa, from the 13th to 16th August 2009.

Baye, F. M. (2006). "Growth, Redistribution and Poverty changes in Cameroon: A Shapley Decomposition Analysis". *Journal of African Economies, 15, No.4, pp. 543-570.*

Baye, F. M. (2008). "Exact Configuration of Poverty, Inequality and Polarisation trends in the Distribution of Well-being in Cameroon". Final report submitted to AERC research workshop, Nairobi, Kenya.

Baye, F. M. and Fambon, S. (2002). "Decomposition of Inequality in the Distribution of Living Standards in Cameroon", African Journal of Economic Policy, vol. 9, N°2, pp 51-75.

Baye, M. F and S, Fambon. (2010). "Parental Literacy and Child Health Production in Cameroon." *African Journal of Economic Policy*, vol. 17, No. 2. ISSN 1116-4875.

Baye, M., F and S., Fambon. (2002). “Decomposition of Inequality in the Distribution of Living standard in Cameroon.” *African Journal of Economic Policy*, 9, No. 2, pp.51-75.

Becker, G., S. (1962). “Investment in Human Capital: A Theoretical Analysis.” *The Journal of Political Economy*, Vol. 70, No. 5, Part 2: Investment in Human Beings, pp. 9-49.

Becker, G., S. (1967). *Human Capital and the Personal Distribution of Income*, Woytinski Lecture, Ann Arbor: University of Michigan Press.

Blinder, A.S. (1973). ‘Wage Discrimination: Reduced Form and Structural Estimates.’ *Journal of Human Resources, 8:436-55.*

Bourguignon, F, Ferreira Francisco H. G. and L Phillippe. (2008). “Beyond Oaxaca-Blinder: Accounting for Differences in Household Income Distributions”. *Journal of Economic Inequality*, Vol. 6, No.2: 117-148.

Bourguignon, F. (1979). “Decomposing Income Inequality Measures.” *Econometrica*, 47, 901-920.

Bourguignon, F., M. Fournier and M. Gurgand. (2001). ‘Fast Development with a stable Income Distribution: Taiwan, 1979-1994’. *Review of Income and Wealth, 47(2):139-63.*

Cancian, M. and D. Reed. (1998). “Assessing the Effects of Wives Earning on Family Income Inequality.” *Review of Economics and Statistics, 80: 73-79.*

Cheshire, P and S, Stephen. (2002). “The Welfare Economics of Land Use Planning.” Journal of Urban Economics, 52(2), pp. 242-69.

Deaton, A.(1997). *The Analysis of Household Surveys.* John Hopkins University Press: Baltimore.

DiNardo, J., N.M. Fortin and T. Lemiux. (1996). ‘Labour Market Institution and the Distribution of Wages, 1973-1992: A semi-parametric Approach.’ *Econometrica 64(5):1001-44.*

Douglas M, Thomson H, and Gaughan M. (2003). Health Impact Assessment of Housing Improvements: A Guide, Public Health Institute of Scotland, Glasgow.

Epo, B., N, Baye, F., M and N., A., T, Manga. (2010). “Explaining Spatial and Inter-temporal sources of Poverty, Inequality and Gender Disparities in Cameroon: A Regression-Based Decomposition Analysis”. Poverty and Economic Policy (PEP) Research Network. Research project No. 11321. June 2010. www.pep-net.org .

Epo, B., N. (2010). “Moderate and Persistant Determinants of Poverty in Cameroon”. Working paper series publications. Social Science Research Network (SSRN), *Urban Economics and Regional Studies eJournal*. (June 2010).

Fedotova, E. (2008). E-Skills: Catalyst to Opportunity. *Baltic IT&T Review*, 2008(3).

Fields, G.S. and G. Yoo. (2000). ‘Falling Labour Income Inequality in Korea’s Economic Growth: Patterns and Underlying Causes.’ *Review of Income and Wealth, 46(2): 139-59.*

Foster, J.E. and A., A., Shneyerov. (1997). “A General Class of Additively Decomposable Inequality indices.” Discussion Paper No. 97-W10, Department of Economics, Vanderbilt University.

Garen, J. (1984). “The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable”. *Econometrica*, Vol.52, No. 5, pp. 1199–1218.

Grossman, M. (1972). “On the Concept of Health Capital and the Demand for Health”. *Journal of Political Economy*, 80, 223-255.

Heshmati A. (2004a). “The World Distribution of Income and Income Inequality: A Review of the Economics Literature.” *Journal of World Systems Research*, XII, I, july 2006, 61-107.

Heshmati A. (2004b). “A review of Decomposition Income Inequality”. IZA Discussion Paper No. 1221. July 2004.

Institue Nationale des Statistiques. (2008). *Tendance, Profil et Déterminant de la Pauvreté au Cameroun en 2007.* Disponible sur *www.statistic-Cameroun.org.*

Institue Nationale des Statistiques.(2002a). « ECAM II : Document de Méthodologie », Yaoundé.

International Institute of Communication and Development, (IICD), (2008). Health Sector Report. July, 2008. www.iidc.org.

Jackson, L.D and Duffy, B., K., eds. Health Communication. Westport, CT: Greenwood.

Juhn, C., K. Murphy and B. Pierce. (1993). ‘Wage Inequality and the Rise in Returns to Skill.’ Journal of Political Economy.

Kakwani, N. (1980). “On a Class of Poverty Measures.” *Econometrica*, n°48, p. 437-446.

Kimhi, A. (2007). “Regression-Based Inequality Decomposition: A Critical review and Application to Farm-Household Income Data.” Agricultural Economics Department, The Hebrew University, Israel.

Maria, F., O., M and M., F., Jose. (2008). “ The Determinants of Poverty in Cape Verde. An Econometric Approach”. *Bulletin of Statistic & amp; Economics*.

Mincer, J. (1958). “Investment in Human Capital and the Personal Income Distribution”. *Journal of Political Economy*, Vol. 66; 4, pp 281-302.

Morduch, J. and T, Sicular. (2002). ‘Rethinking Inequality Decomposition with Evidence from Rural China’. *Economic Journal, 112(January), 93-106*.

Mwabu, G. (2009). “The Production of Child Health in Kenya: A Structural Model of Birth Weight”. *Journal of African Economies*, Vol. 18, No. 2, pp. 212-260, Advance Access published July 22, 2008.

Mwabu, G., M. Wafula., R.Gesami, J. Kirimi, G., Ndenge’e, T., kiriti, F., Minene, M., Chemngich and, J., Mariara. (2000). ‘Poverty in Kenya: Profile and Determinants’. Nairobi: University of Nairobi and Ministry of Finance and Planning.

Mwabu, G., M. Wafula., R.Gesami, J. Kirimi, G., Ndenge’e, T., kiriti, F., Minene, M., Chemngich and, J., Mariara. (2000). ‘Poverty in Kenya: Profile and Determinants’. Nairobi: University of Nairobi and Ministry of Finance and Planning.

National Institute of Statistics. (2007). *Enquête Camerounaise Auprès des Ménages (ECAM III).* National Institute of Statistics, 2007.

National Institute of Statistics. (2008). *Tendance, Profil et Déterminant de la Pauvreté au Cameroun en 2007.* Disponible sur *www.statistic-Cameroun.org*. May 2009.

National Institute of Statistics. (August, 2005). *Les Comptes Nationaux Du Cameroun 1993-2004*. Cameroun.

National Institute of Statistics. (October, 2004). *Annuaire Statistique du Cameroun*. Edition 2004. Tome III.

Oaxaca, R. (1973). ‘Male-Female Wage Differences in Urban Labour Markets.’ *International Economic Review 14(3): 693-709*.

Oyekale, A.S, A.I., Adeoti and T.O., Oyekale. (2007). “Regression-Based Approach to Income Inequality Decomposition in Rural and Urban Nigeria.” *International Journal of Economic Perspective*, vol. *1*, Issue, 2, pp. 45-54.

Schultz, T., W. (1961). “Investment in Human Capital: Reply.” *The American Economic Review*, Vol. 51, No. 5, pp. 1035-1039.

Sen, A., K. (1981). “Poverty and Famine: an Essay on Entitlement and Deprivation”. Oxford: Claredon Press.

Sen, A.K. (1973). “On Economic Inequality”. Oxford University Press, Oxford.

Shapley, L. (1953). ‘A Value for n-Person Games’. In: H. W. Kuhn and A. W. Tucker, eds., *Contributions to the Theory of Games*, Vol. 2, Princeton University Press.

Shea, J. (1997) ‘Instrument Relevance in Multivariate Linear Models: A Simple Measure’, *Review of Economics and Statistics,* 49 (2): 348–52.Shorrocks, A., F. (1980). “The Class of Additively Decomposable inequality Measures.” *Econometrica*, 48, 613-625.

Shorrocks, A.F. (1982). “Inequality Decomposition by Factor Components.” *Econometrica,65,* pp.*177-207*.

Shorrocks, A.F. (1984). “Inequality Decomposition by Population Subgroups.” *Econometrica, 52, 1396-1385*.

Shorrocks, A.F. (1999). “Decomposition procedure for Distribution Analysis: A unified Framework Based on Shapely value”. University of Essex and Institute for Fiscal studies, Colchester, England.

Stock, J., H. and M. Yogo (2004) ‘Testing for Weak Instruments in Linear IV Regression’, Mimeo, Department of Economics, Harvard University.

Tabi, A., J. (2009). *Components, Sources and Determinants of Aggregate Inequality Trends in the Distribution of Household Economic Well-being in Cameroon, 1996-2001*. Undefended and Unpublished PhD Thesis. University of Yaoundé II, Soa, Cameroon.

Theil, H. (1972). *Statistical Decomposition Analysis.* (North Holland, Amsterdam).

Wan, G.H. (2002). ‘Regression-Based Inequality Decomposition: Pitfalls and a Solution Procedure.’ UNU-WIDER Discussion Paper No. 2002/101. Helsinki.

Wan, G.H. (2004). ‘Accounting For Income Inequality in Rural China: A Regression-Based Approach.’ *Journal of Comparative Economics*. 32(2), 348-63.

Wan, G.H. and Z.Y. Zhou. (2005). ‘Income Inequality in Rural China: Regression-Based Decomposition Using Household Data.’ *Review of Development Economics, 9, 107-20.*

WHO Regional Office for Europe’s Health Evidence Network. (2005). Is housing improvement a potential health improvement strategy? February 2005.

Yuko,A., J.M. Kim. and A., Kimhi. (2006). ‘Determinants of Income Inequality among Korean Farm Households’. Economic Research Centre, Discussion paper, No. 161, School of Economics, NAGOYA University.

Appendix 1: Ingredients of the synthetic variables for education and health

Dimension 1: ***Education and related basic infrastructures***
Knowing how to read and write
Already attended schools
First reason for dissatisfaction regarding the closest public primary school
First reason for dissatisfaction regarding the closest private primary school
Distance to go to the nearest public primary school (0,1,2,3,4,5 or 6km and more.)
Distance to go to the nearest private primary school (0,1,2,3,4,5 or 6km and more.)
Required Time to go the nearest primary public school
(0-5min/6-15min/16-25min/26-35min/36-45min/ 46min or more)
Required Time to go the nearest private public school
(0-5min/6-15min/16-25min/26-35min/36-45min/ 46min or more)

Dimension 2: ***Health and related basic infrastructures***
Sector of consultation
Type of sanitary centre
Appreciation of health status
First reason for dissatisfaction regarding the closest sanitary centre
Distance to go to the nearest sanitary centre (0,1,2,3,4,5 or 6km and more.)
Required Time to go the nearest sanitary centre
(0-5min/6-15min/16-25min/26-35min/36-45min/ 46min or more)

Appendix 2: Weighted Reduced Form Estimates for Education and Health

Variable	Education (1)	Health (2)
Included Exogenous variables		
Household Size	-0.0058***	0.0045***
	(-7.09)	(4.09)
Age of Household head	-0.0039***	0.0029***
	(-20.64)	(11.06)
Gender (Male=1 and 0 otherwise)	0.0009	-0.1315***
	(0.13)	(-14.83)
Fraction of Active Household members	-0.0705***	0.0218
	(-4.17)	(0.95)
Formal Sector (1=working in the formal sector and 0=otherwise)	0.1051***	0.0733***
	(13.48)	(6.94)
Household own farmland (1=Own farmland and 0=otherwise)	-0.0368***	-0.0135
	(-5.46)	(-1.48)
Pure Urban area	0.1226***	0.0236**
	(14.42)	(2.05)
Constant	1.0947***	0.4948***
	(74.55)	(24.87)
Excluded Exogenous variables		
Household Own Radio (Cluster percentage shares)	0.2556***	0.7790***
	(5.36)	(12.05)
Household Own Television (Cluster percentage shares)	1.6520***	0.3182***
	(37.14)	(5.28)
Number of Room (Cluster percentage shares)	-0.0003***	-0.0001
	(-11.64)	(-1.59)
R-Squared	0.4643	0.0744
Adjusted R-Squared	0.4638	0.0736
Fisher Test [p-value]	986.2; [0.00]	91.53; [0.00]
Number of Observation	11391	11391
Total population	17.8 million	17.8 million

Source: Computed by Authors using STATA 10. Notes: ***, ** and * are 1, 5 and 10 percent significance levels, respectively. Variables in parenthesis are t-student values.

Appendix 3: Table A: Marginal contributions of the various estimated income sources based on the Shapley value Approach for 2007

Estimated income Sources	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9	Level 10	Level 11	Level 12	Level 13
Education*	0.0155	0.0115	0.0092	0.0078	0.0068	0.0061	0.0056	0.0052	0.0049	0.0046	0.0043	0.0041	0.0040
Health*	0.0119	0.0077	0.0052	0.0037	0.0026	0.0019	0.0014	0.0010	0.0007	0.0004	0.0002	0.0000	-0.0001
Household Size	0.0047	0.0028	0.0019	0.0015	0.0013	0.0012	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011
Age Cohorts	0.0013	0.0004	0.0001	0.0000	-0.0000	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
Sex(1=male & 0=otherwise)	0.0012	0.0004	0.0001	0.0000	-0.0000	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
Fraction of Active Household Members	0.0040	0.0023	0.0015	0.0012	0.0010	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008
Formal Sector (1=working in the formal sector and 0=otherwise)	0.0019	0.0012	0.0009	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0010	0.0010	0.0011	0.0011
Household own farmland (1=Own farmland and 0=otherwise)	0.0009	0.0004	0.0003	0.0002	0.0003	0.0003	0.0003	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004
Pure Urban Area	0.0031	0.0019	0.0014	0.0013	0.0013	0.0013	0.0014	0.0015	0.0015	0.0016	0.0017	0.0018	0.0018
Complementary Educational Input	0.0100	0.0061	0.0038	0.0024	0.0015	0.0008	0.0003	-0.0002	-0.0005	-0.0008	-0.0011	-0.0012	-0.0014
Complementary Health Input	0.0083	0.0045	0.0024	0.0011	0.0003	-0.0002	-0.0006	-0.0009	-0.0011	-0.0013	-0.0014	-0.0016	-0.0017
Residual	0.0214	0.0172	0.0146	0.0130	0.0118	0.0110	0.0103	0.0098	0.0094	0.0090	0.0086	0.0083	0.0081

Source: Computed by Authors using DASP 2.1 distributive software slotted in STATA 10.

Notes: Levels indicate the point of entry of an estimated source into a coalition of sources. Results are reported in four decimal places.