Discussion Papers In Economics And Business

An Update of the Returns to Education in Kenya: Accounting both endogeneity and sample selection biases

Kentaro Shimada, Zeba Khan, Suguru Mizunoya and Ayako Wakano

Discussion Paper 16-18

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

An Update of the Returns to Education in Kenya:
Accounting both endogeneity and sample selection biases

Kentaro Shimada, Zeba Khan, Suguru Mizunoya and Ayako Wakano

Discussion Paper 16-18

May 2016

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

An Update of the Returns to Education in Kenya:

Accounting both endogeneity and sample selection biases

Kentaro Shimada[a], Zeba Khan[b], Suguru Mizunoya[c] and Ayako Wakano[d]

Abstract: The study is latest to estimate returns to education after the introduction of free primary education in 2003 in Kenya, simultaneously addressing two sources of biases due to endogeneity of schooling and earnings, and sample selection. Using the 2005-2006 Kenya Integrated Household and Budget Survey, the paper finds that (a) returns to additional year of schooling are 14.9% for males and 13.5% for females with a continuous education variable, but the returns to females are consistently higher than males when returns are estimated by level of education, (b) returns to education increases for higher levels of education i.e., the classical pattern of diminishing return to schooling does not hold true for both males and females in Kenya, and (c) the use of joint IV-Heckman method adjust the endogeneity and sample selection biases introduced by OLS and IV.

JEL code: I26, I25, O55

Key Words: Returns to Education; Education Policy; Human Capital; Gender; Kenya

a- Ph. D candidate, Graduate School of International Cooperation Studies, Kobe University
Email address: shimaken629@gmail.com

b- Research Assistant, Faculty of Social Science, Chinese University of Hong Kong
Email address: khanzeba@hotmail.com

c- Assistant Professor, Faculty of Social Science (Global Studies), Chinese University of Hong Kong
Email address: suguru777@gmail.com

d- Ph. D candidate, Graduate School of Economics, Osaka University
Email address: ayakowakano@gmail.com

1. Introduction

The positive impact of additional schooling on earnings, captured by the "private" rate of return on education, is widely accepted as a measure of educational productivity, and incentive for households to invest in human capital. The rate of return on education is useful in (i) explaining patterns of educational demand, (ii) guiding how best to prioritize the distribution of public resources in education, and (iii) identifying areas which need incentives and promotion for private investment. The compilation of work by Psacharopoulos (1973,1981,1985,1994) and Psacharopoulos and Patrinos (2002) have found rates of return to education to be consistently higher in developing regions (Sub-Saharan Africa, Latin America and the Caribbean) compared to the world average, and the returns tend to decline with increasing levels of education. As these returns to education, especially at the primary level, exceed returns obtained from investments in physical capital, Psacharopoulos's inference led to a paradigm shift in global focus on expanding access to primary education in developing countries, resulting in considerable flow of donor money in these nations.

As one of five such countries included in Psacharopoulos' initial international comparative account of returns to education (1973) – Kenya has a comparatively longer history of return to education studies. Private and social returns to education in Kenya have been estimated to additional years of schooling in various ways, and also according to the discrete cycles of primary, secondary, and tertiary education.

Kenyan education has witnessed significant expansion and change after gaining independence in 1963. In the 1960s and 70s, apart from overseeing a rapid increase in primary enrollment, Kenya's first President Kenyatta also called for a grassroots-driven *harambee* movement, which led to a parallel enlargement of secondary education. Later in 1985 a shift from the previously British colonial-style 7-4-2-3 educational system to a 8-4-4 structure under President Moi extended primary schooling by one year in order to implement a more practical, vocational curriculum (Buchmann, 1999). The overhauling of the education system in Kenya with a purpose to achieve "Education for All", led to the introduction of free primary education in January 2003, which resulted in gross primary enrollment to touch approximately 107%in 2003, in comparison to 92% in 2002 (World Bank, 2015). According to Tooley, Dixon and Stanfield (2008), the enrolment in primary education in 2003, increased to 7.2 million from 5.9 million. Following the establishment of free primary education framework, the Kenyan government in 2008, implemented the free day secondary education programme, which aimed to make secondary education more accessible. Predictably, therefore, returns to education in Kenya should vary over time (Appleton, Bigsten, and Manda, 1996) due to structural changes arising from policy interventions. Updating these estimates based on comparatively recent data thus should be of policymaking interest.

Furthermore, in recent years, evidence has emerged indicating the classic pattern of diminishing rates of return to education by level suggested by Psacharopoulous and Patrinos (2002) may no longer hold true for the majority of developing countries. Rather, analysis of data from the 1990s and early 2000s suggests that the slope of the earnings function increases with education level (Colclough, Kingdon, and Patrinos, 2010).

It is noteworthy to state that existing study on return of education in Kenya suffer from data and methodological deficiencies. They include lack of gender disaggregation, and limited sample coverage (e.g., only workers in Nairobi were covered). For example, Patrinos and Psacharopoulos (2002) exhibit an extensive list of returns of education from various sources for different countries, but the returns in Kenya are not disaggregated by sex; Manda, Mwabu, and Kimenyi (2002) did not account for endogeneity issue; and Johnson (1972), Knight and Sabot (1987), and Armitage and Sabot (1987) used samples only from Nairobi. As the large amount of public budget is spent for the education sector and the education is one of the key national strategies, estimating the returns of education using nationally representative data for both sex is critical to evaluate and plan effective education and labour market policies.

Returns to an additional year of schooling at both, overall and discrete levels of education have been estimated for many developing countries using the semi-logarithmic earnings function (Mincer, 1974), which has become ubiquitous for its undemanding data requirements. However, the scarcity of quality data introduces biases to the estimated returns to education with the use of ordinary least squares (OLS) methods. None of the previously published studies on return to education in Kenya addressed the issue of endogeneity arising from the correlation between education and earnings, and non-representativeness of the wage-working sample compared to the overall population. This study simultaneously addresses these two problems by adopting IV and Heckman method for computation of returns to additional years of schooling.

Thus this study aims to update the returns of education in Kenya whose latest estimates were made using the Welfare and monitoring Survey 1994 through the lenses of gender. The paper addresses both endogeneity and selectivity biases, by employing instrumental variable with Heckman's two-step procedure based on the suggestion of Wooldridge (2002). There is paucity of studies on return of education in Kenya undertaken after the Millennium Development Goals was envisaged by the United Nations in 2000, and after the implementation of key education policies in Kenya i.e., the free primary education policy. The study is, to the best of our knowledge, the first time that the approach which addresses both biases has been used to estimate returns to education for both sexes in a developing African country; noteworthy to mention that previous authors have limited its use to married women in Kazahkstan (Arabsheibani and Mussurov, 2007) and in China (Chen and Hamori, 2009).

The paper is structured as follows. Section 2 provides a review of previous literature examining rates of return to education in Kenya. Section 3 explains the Mincerian framework used to estimate the rate of return to an additional year and level of schooling in Kenya, and the empirical strategy adopted to simultaneously correct for endogeneity of schooling and sample selection bias. Section 4 describes the dataset and variables used in the analysis. The main results and discussion are presented in Section 5, and Section 6 gives some concluding remarks.

2. Literature Review on Returns of Education in Kenya

Several studies have been undertaken to analyze the returns to additional years of schooling in Kenya. Among the earliest studies is the World Bank sponsored study conducted by Thias and Carnoy in 1969, with few following in 1970's while the recent studies initiated around 2000. A detailed list of the studies is presented in the Table A1. Rates of return obtained from various studies cannot be directly compared due to variations in methodology and data coverage. However it is useful to review the literature in order to understand previous empirical strategy and trends of the results.

A World Bank cost-benefit analysis by Thias and Carnoy (1969) estimated returns to levels of education using cross-sectional urban earnings data collected in a 1968 Labor Force Survey of private and public sector employees in the cities of Nairobi, Mombasa, and Nakuru. Calculating the rate of return that set the discounted time value of education costs equal to the stream of its benefits, they found overall returns to education for urban males to be quite high: 32.7% for primary, 36.1% for lower secondary, 23.8% to higher secondary, and 27.4% to university education, respectively. For females, returns to primary education were much lower at primary level i.e., 9.5%, but returns to lower secondary education were comparable to that of males (33.6%). Johnson (1972) estimated continuous returns to education based on earnings function of log hourly wages on categorical form of years of education expressed in quadratic, by using data from a survey of 1970 wage activity for low- to middle-income African households in Nairobi. Controlling for potential human capital accumulated through experience, union membership, type of employment (government or self), gender, and major tribal variables, Johnson found increasing marginal returns to education, with a base return of 1%, and each additional year of schooling increasing it further to 2.2. Using 1980 data on wage-workers in Nairobi, Armitage and Sabot (1987) estimated far lower private returns to the completion of secondary education of 14.5% for government-supported institutions, and just 9.5% for community-funded *harambee* (Kiswahili for "let's pull together") schools.

Using data from three labor force surveys conducted in 1978, 1986, and 1995, Appleton, Bigsten and Manda (1999) found fairly high private returns. They estimated returns to education using both Mincerian regression and CBA. For example, returns in 1995 are 25%, 7%, and 35%, for primary, secondary, and university education respectively. The study also shows that there is a trend that the return to the tertiary education is higher and increasing over the years whereas the returns to the primary education is decreasing. Manda, Mwabu & Kimenyi (2002), another recent study which emphasized on incorporating human capital externality, also show that the returns to tertiary education are estimated highest in 1995, but returns to primary and secondary education are much lower.

The latest study so far -Manda, Mwabu & Kimenyi (2002) computed returns to education in 1994 using Welfare Monitoring Survey of Kenya, and further disaggregated according to location and sex of workers. The study showed that the returns to education are higher for males across all education levels; the returns to schooling for males are 11.0%, 17.8%, 35.2% and females stand at 5.7%, 15.8%, 32.2%, for primary, secondary and tertiary education. This study so far provides the most comprehensive returns to education by level of education and sex. Other studies did not provide such detailed estimates.

There are two major insights from the literature review. First, earlier studies such as Thias and Carnoy (1969), Johnson (1972), some results using old data in 1978 and 1986 of Appleton, Bigsten and Manda (1999) showed the diminishing pattern of the rates of return to education. However more recent results such as the results using 1995 data from Appleton, Bigsten and Manda (1999) and Manda, Mwabu & Kimenyi (2002) suggest increasing returns to higher levels of education in Kenya, as primary education becomes a social norm for the majority of Kenyans. Second, the studies undertaken to assess returns to education in Kenya differ according to three dimensions namely- geographical coverage, disaggregation by the sex variable, and control for endogeneity and selection biases. The latest study was conducted using 1994 data, prior to major education reforms to achieve universal primary education in Kenya. Except Manda, Mwabu & Kimenyi (2002) previous studies did not differentiate rates of returns to education and the varying rates of retunes to different levels of education. It is important to evaluate returns to additional year of schooling according to sex so that policymakers are better equipped to chart strategies to educate girls, as women in Kenya have dearth of employment opportunities arising due to lack of education (World Bank, 2004). Finally, none of the published studies adopts methods to correct the estimates of returns to education for endogeneity and selection bias.

3. Methodology

The empirical framework adopted to estimate the private returns to education use the Mincerian semi-logarithmic basic earnings function (Mincer, 1974),

$$\ln W_i = \alpha + \beta S_i + \gamma_1 A_i + \gamma_2 {A_i}^2 + \delta \boldsymbol{X}_i + \varepsilon_i \qquad (1)$$

where $\ln W_i$ is the natural logarithm of the hourly wage reported by each individual i; α is a constant; S_i is years of schooling; A_i is a measure of worker experience, entered in linear and quadratic forms; $\boldsymbol{X}_i$ is a vector of other observed exogenous explanatory variables, and an error term, ε_i. The wage-earning specification is examined for males and females separately. The coefficient is interpreted as the private rate of return to education, that is, the relative change in wages for each additional year of schooling, averaged across all sampled individuals and levels of education.

It is a recognized fact that returns obtained from simple ordinary least squares (OLS) estimation of the Mincerian earnings function may be biased due to endogeneity of the educational variable. As an example, if years of schooling are positively correlated with an unobserved or otherwise omitted ability factor which also affects earnings in a positive way, OLS will tend to overestimate, since high-ability individuals both complete more years of schooling and earn higher wages in the labor market. Here, the potential endogeneity of years of schooling is addressed by adopting a conventional instrumental variables (IV) approach, where an observable covariate that affects schooling but not earnings is used to instrument for schooling in the following two-equation model:

$$S_i = \zeta \boldsymbol{Z}_i + \mu_i \qquad (2)$$

$$\ln W_i = \beta S_i + \gamma_1 A_i + \gamma_2 {A_i}^2 + \delta \boldsymbol{X}_i + \nu_i \quad (3)$$

where $\boldsymbol{Z}_i$ is a vector of the instrument and other observed exogenous explanatory variables A_i, ${A_i}^2$, and those already in $\boldsymbol{X}_i$; and μ_i and $\nu_i$are error terms.

This study chooses the commonly-used instrument of maternal education (e.g., Trostel, Walker, and Woolley, 2002). In low-income countries where women are mostly poorly educated, unemployed and have comparatively little standing or bargaining power over household decisions such as educational expenditure and enrollment, while mothers with more education are suggested to enhance their children's educational outcomes largely directly. They do so by improving the time children spend on educational activities outside school, helping children with schoolwork, and encouraging educational assistance from other family members (Andrabi, Das, and Khwaja, 2012). Indeed, Kabubo-mariara and Mwabu (2007) found maternal education to be an important determinant of both educational enrollment and overall grade attainment in Kenya. Paternal education is not selected as a second instrument, because fathers' education and its related outcomes (his socioeconomic status and occupation) are far more likely to influence their children's employment status as well as education, especially as African youths frequently find it necessary to utilize familial social capital and contact networks during the job-hunting process (Filmer and Fox, 2014).

In practice, IV estimates of returns to education in the literature typically exceed those obtained from OLS by a magnitude of 20% of more. Measurement error may bias returns to education either downwards. If the overall effect of the measurement error outweighs, it leads to overestimation of returns to schooling using OLS. Nonetheless, Card (2001) estimates the impact of measurement bias to be relatively of the order of 10%, and suggests that the large gap reflects downward bias in OLS estimates due to heterogeneous returns to education, where individuals with high discount rates choose to complete less schooling (Lang, 1993).

OLS estimates of returns to education may also be biased due to sample selectivity, if the wage-working sample is not fully representative of the working population. Accounting for females in estimation of returns to education addresses concerns of selectivity bias, as it is observed that education does influence employment of females in a positive way and as better-educated individuals earn higher salaries, returns to education for females are expected to be biased upwards. In Kenya, however, where there is substantial labor market heterogeneity, possibly owing to male workers in the informal or small-scale agricultural sectors who generally earn less than their formally-employed counterparts (Nyaga, 2010) and do not report an official wage, such that OLS returns to education for males are also biased upwards. Consequently, this potential sample selection bias is corrected for by applying Heckman's two-step method (Heckman, 1979) on both males and females, wherein individual's participation in wage-earning activity is modelled as being determined by a selection equation

$$D_{W_i} = 1[\theta \boldsymbol{T}_i + \eta_i > 0] \qquad (4)$$

where the dummy variable indicating selection D_{W_i} equals 1 if wage-earning is observed ($W_i > 0$), and 0 otherwise; $\boldsymbol{T}_i$ is a vector of additional observed exogenous explanatory variables for participation. In addition to S_i, A_i, ${A_i}^2$ and the explanatory variables already in $\boldsymbol{X}_i$, this study includes as selectivity variables the natural logarithm of the individual's household expenditure, $\ln HHE_i$; and drawing on the determinants of Kenyan participation in employment reported by Nyaga (2010), household size, disaggregated into the number of children in the household aged below 6 years (primary-school age), HHChildren6-_i, and the number of elderly in the household aged over 65 (above the working-age threshold), HHAdults65+_i ; and dummy variables indicating whether an individual is household head, D_{headship_i} ; and the household owns its present dwelling, $D_{\text{HHownshouse}_i}$. η_i is an error term.

Using the parameter $\hat{\theta}$ estimates obtained from the probit $P\left(D_{W_i} = 1 \middle| \boldsymbol{T}_i\right) = \Phi(\theta \boldsymbol{T}_i)$ over the entire working-age subsample, the inverse Mills ratio $\lambda_i(\hat{\theta}\boldsymbol{T}_i) = \frac{\phi(\hat{\theta}\boldsymbol{T}_i)}{\Phi(\hat{\theta}\boldsymbol{T}_i)}$ is computed for each observation and included as an additional exogenous explanatory variable in the selectivity-corrected Mincerian

$$\ln W_i = \beta S_i + \gamma_1 A_i + \gamma_2 {A_i}^2 + \delta \boldsymbol{X}_i + \iota \lambda_i(\hat{\theta} \boldsymbol{T}_i) + {\nu_i}' \quad (5)$$

estimated for the selected subsample, where the coefficient ι measures the covariance of the residuals in the selection and earnings equations $\sigma_{\eta_i, \varepsilon_i}$, and its statistical significance and sign indicates the existence and, if so, direction of the sample selectivity bias, which is expected to be negative.

To adjust for both endogeneity of education and sample selectivity simultaneously requires the combination of the Heckman and IV procedures. Following Wooldridge (2002), $\boldsymbol{T'}_i$ in the joint Heckman-IV first-stage selection probit $P\left(D_{W_i} = 1 \middle| \boldsymbol{T'}_i\right) = \Phi(\theta' \boldsymbol{T'}_i)$ estimated over the entire working-age subsample incorporates *all* exogenous explanatory variables, i.e., the instrument and those already in $\boldsymbol{T}_i$**,** omitting S_i . Similarly, $\boldsymbol{Z'}_i$ in the second-stage IV equation for S_i is a vector of the newly estimated $\lambda'_i(\widehat{\theta'} \boldsymbol{T'}_i)$ and $\boldsymbol{T'}_i$ for all observations in the selected subsample:

$$D_{W_i} = 1[\theta' \boldsymbol{T'}_i + \eta_i > 0] \quad (6)$$

$$S_i = \zeta \boldsymbol{Z'}_i + \mu_i \quad (7)$$

$$\ln W_i = \beta S_i + \gamma_1 A_i + \gamma_2 {A_i}^2 + \delta \boldsymbol{X}_i + \iota \lambda'_i(\widehat{\theta'} \boldsymbol{T'}_i) + \nu_i \quad (8)$$

OLS, IV, Heckman-corrected, and joint IV-Heckman-corrected returns to education are estimated on the entire sample of wage-workers. This overall sample is then disaggregated into subsamples of wage-workers whose highest grade was bounded by primary (i.e., those born in or before 1971 who had completed up to Standard 7, or born in or after 1972 and had completed up to Standard 8), secondary (those who had completed from Standard 8 to Secondary 4), and tertiary education (those who had completed Secondary 4 or higher). Doing so permits the slope of the earnings function (the rate of return to education) to vary across the three levels of education.

4. Data

The study uses the 2005-2006 Kenya Integrated Household Budget Survey (KIHBS), collected information from a nationally representative sample of 13,430 households on a wide range of socioeconomic indicators relating to demographics, education, employment, expenditure, and consumption. The analysis is restricted to wage-earners of working-age (15 to 65 years) at the time of the survey, excluding full-time students. The sample sizes of male and female workers are 5,406 and 3,146, respectively.

The labor module in the KIHBS household questionnaire asked household members their average daily working hours and earnings for the previous month. Assuming 20 working days per month, this information is used to calculate each wage-worker's hourly wage. Age is substituted as a proxy for potential work experience A_i, primarily because years of prior working experience or job tenure were not directly surveyed. However, as noted by Barouni and Broecke (2014), Mincer's traditional expression for A_i, age minus schooling minus primary entry age, is also less relevant in African countries where late primary matriculation, repetition, and dropping out are relatively common. Educational capital was recorded as the highest grade completed, from which the continuous variable for years of schooling S_i was computed, adjusted for the different systems pre- and post-1985 educational reform. S_i is subsequently used to define subsamples of wage-workers by highest participatory education level. The descriptive statistics is presented in Table A2.

5. Results

Average returns to an additional year of schooling for the overall sample of wage-workers are statistically significant at the 1% level for both sexes and at mean value of independent variables. OLS returns to schooling for males (14.0%) and females (13.8%) are nearly equivalent. However, IV returns to schooling sharply diverge between males (11.5%) and females (14.2%). The large Cragg-Donald F-statistic (218.8, males; 93.9, females) and high F-statistics of the first stages (77.0 males; 34.7 females) and Shea Partial statistics (0.218.8 males; 0.197, females) suggest mother's education to be an adequate instrument for years of schooling for both males and females. In contrast, returns to schooling estimated by the Heckman two-step method (12.7%, males; 11.2%, females) show evidence of an expected significant upward selectivity bias in the OLS returns to schooling, although the magnitude of the selectivity bias is comparatively small. The Heckman returns to schooling are 10% and 23% lower than OLS estimates for males and females.

The estimates of returns to education yielded from the joint IV-Heckman procedure are 14.9% and 13.5% for males and females, respectively. The estimate of male workers using IV-Heckman procedure is slightly higher than the OLS estimates, but that of female worker is almost same as the OLS estimate. In fact, the selectivity term is statistically insignificant in the joint IV-Heckman estimate of returns to schooling for females (cf. エラー! 参照元が見つかりません。A5, column 4), which is indicative that the sample selection bias would not be a serious issue in estimation returns of schooling in Kenya. Table 1 below summarizes the findings from different methods.

Table 1. Private Returns to an Additional Year of Schooling by Means of Estimation, Overall Sample

	Male	**Female**
OLS	14.0%***	13.8%***
IV	11.5%***	14.2%***

Heckman	12.7%***	11.2%***
Joint IV-Heckman	14.9%***	13.5%***

*Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*
[a] selectivity term not statistically significant.
Source: Table A3

The previous analysis using a continuous education variable as the dependent variable shows that there is little gender difference in returns to additional year of schooling across all the methods. However, the analysis to estimate varying returns to education by the level of education unpacks the nature of returns and shows a significantly different picture i.e., the returns to education vary significantly between different levels of education, between male and female workers, and the analysis using a continuous education dependent variable may lead to wrong policy implications. The results are summarized in Table 2. Major findings are the following. First, the rates of return at primary level are statistically insignificant for both male and female samples when both biases are controlled. Second, the returns to an additional year of schooling increases progressively i.e., the return of the secondary education is higher than that of primary education, and the return of the tertiary education is higher than that of secondary education. The coefficients of rate of return at secondary and tertiary levels are statistically significant at the 1% level across sexes and means of estimation. Using the IV-Heckman approach, the rates of return of secondary and tertiary education for males are estimated as 20.1% and 71.2%, respectively. For female workers, they are 36.8% and 87.6%, respectively. Last, the rates of return of schooling among female workers are higher than male workers consistently, except the primary education where the return for female workers is insignificant. These results unpack the nature of returns to education and seem to provide a better understanding such as increasing returns. The results in the previous section show that the return to education among male workers is higher than female, but this analysis shows that the female's returns of secondary and tertiary education are higher than males ones by 59.0% and 53.5%. These results would rather encourage the government to take more active roles in promoting girl's secondary and post-secondary education. Consistently, OLS returns to schooling are biased in both upwards and downwards with respect to IV estimates while the Heckman returns demonstrate the presence of statistically significant upward selectivity bias in the OLS estimates.

Table 2. Private Returns to an Additional Year of Schooling Disaggregated by Level of Education and Estimation Methodology

	Male			**Female**		
	Primary	*Secondary*	*Tertiary*	*Primary*	*Secondary*	*Tertiary*
OLS	6.3%***	18.2%***	23.3%***	0.90%	23.2%***	33.3%***
IV	0.90%	14.7%	90.1%***	-0.94%	36.4%***	91.9%***
Heckman	4.5%***	16.5%***	22.3%***	-1.1%	18.1%***	31.1%***
Joint IV-Heckman	0.60%	20.1%***	71.2%***	-0.73%	36.8%***	87.6%***

Observation	*544*	*412*	*162*	*229*	*252*	*110*

*Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*
Source: Table A4

6. Conclusion

The study calculates returns to additional year of schooling based on gender, almost five decades after the first study of this kind; Thias and Carnoy (1968) undertook analysis, and after 13 years since the latest estimates were made by Manda, Mwabu and Kimenyi (2002). The paper extends its analysis for calculation of returns to education, by using various other approaches such as IV, Heckman and joint IV-Heckman with a purpose to simultaneously correct for two potential sources of bias in OLS estimates of the Mincerian earnings function: the endogeneity of schooling and earnings; and potential non-representativeness of the wage-working sample, which only contains individuals who reported wages and may exclude workers employed in the informal or small-scale agricultural sectors with different earnings profiles. The returns to schooling on the overall sample of wage-workers estimated by the IV and Heckman methods separately demonstrate statistically significant biases as expected and in line with previous literature, with OLS returns biased both upwards and downwards compared to IV results depending on the level of education, and biased slightly upwards from Heckman results, due to sample selectivity. The paper makes a case for employing IV and Heckman methods simultaneously to estimate returns to education more precisely. In addition the returns of education are estimated at the overall and by the level of education to unpack the nature of returns of education. The study finds that the joint IV-Heckman estimation finds the returns to an additional year of schooling as 14.9% for males and 13.5% for females, using a continuous education dependent variable.

The estimation of the returns to additional years of primary, secondary, and tertiary education reveals the situation to be more nuanced. In particular, the joint IV-Heckman estimates of returns to primary education are not statistically significant for both males and females, demonstrating a minimal wage differential between workers with and no primary education. The relatively small selectivity correction to OLS returns to schooling implies that wage structures for Kenyan workers with primary or less than primary education have become fairly similar. The spurt in primary enrollment in Kenya[1] seems to have led to a phenomenal increase in number of individuals with primary schooling thereby causing decline in returns to primary education. It would also be worthwhile to undertake future efforts to evaluate the selection bias over time in order to explore trends in Kenya which has a sizable informal employment sector.

[1] According to the UNESCO Institute of Statistics (various), the national gross enrollment rate in Kenya was 107.7% in 2005.

Conversely, the returns to additional years of schooling increase from secondary (20.1%, males; 36.8%, females) to tertiary education (71.2%, males; 87.6%, females), and the returns to education of females workers are consistently higher than those of male workers. The study renders support to increasing returns to education with increasing levels of education, as presented in previous studies, thereby reinforcing the importance of secondary and tertiary education in Kenya and defying the classical pattern. The study also indicates that measures to enhance access to post-primary levels of education for females might open up windows to avail financially rewarding employment opportunities and consequently reducing their financial vulnerability as females are poorer than men in Kenya as suggested by the World Bank (2004).

The study concludes that implementation of educational policies and investment in education sector in Kenya should facilitate access, and enhance the quality of secondary and tertiary education with special focus to integrate the female population into the education system. Policy measures should not be restricted to raising primary enrollment, but instead, should aim to improve the quality of primary schools in order to curtail factors which give rise to any form of exclusion of children from education. Perhaps such policy measures will maximize students' chances of progressing onwards to post-primary levels of education, only after which they begin to accrue significant returns to additional years of schooling. The results of the paper strongly suggest that education overall remains a favorable sector for public and private investment in Kenya, and that particular attention should be paid to post-primary and girls' education.

Annex

Table A1: Summary of studies undertaken to estimate returns to education in Kenya

Study	Data		Analytical Method	Returns to Education			
	Year	Coverage		Sex	Primary	Secondary	Tertiary
Thias and Carnoy (1969)	1968 Labour Force Survey	Nairobi, Mombasa, and Nakuru	OLS Mincerian Equation and Cost-Benefit Analysis	Male	32.7%	Lower Sec: 36.1%	27.4%
						Higher Sec: 23.8%	
				Female	9.5%	Lower Sec: 33.6%	n.a.
Johnson (1972)	1971	Nairobi	OLS Mincerian wage equation	Overall	Various percentage increments are calculated (e.g., 8.5% from 0 years to 2 years of education).		
					Marginal effect of additional year of education is a convex function of year of education (i.e., 1.0% + 2.2%*year of education)		
Knight and Sabot (1987)	1980	Nairobi	OLS Mincerian Equation and Cost-Benefit Analysis	Overall	n.a.	16%	n.a.
Armitage and Sabot (1987)	The Kenya Survey of Wage Employment and Education 1980	Nairobi		Overall	n.a.	Government: 14.5%	n.a.
						Harambee: 9.5%	
Appleton, Bigsten and Manda (1999)	The 1978 Labour Force Survey, The 1986 Urban Labour Force Survey, The 1995 Regional Programme on Enterprise Development survey	National (1978 and 1986), and Nairobi, Mombasa, Nakuru, and Eldoret (1995)	OLS Mincerian wage equation and Cost Benefit Analysis	Overall	**Mincerian**		
					1978: 8%	42%	15%
					1986: 9%	26%	30%
					1995: 2%	12%	69%
					Cost-Benefit		
					1978:24%	Lower Sec:23%, Higher Sec: 28%	13%
					1986: 22%	Lower Sec:17%, Higher Sec:20%	31%
					1995: 25%	Lower Sec:7%, Higher Sec:n.a.	35%
Manda, Mwabu & Kimenyi (2002)	The Welfare Monitoring Survey 1994	National	OLS Mincerian wage equation	Overall	7.9%	17.2%	32.5%
				Male	11.0%	17.8%	35.2%
				Female	5.7%	15.8%	32.2%

Table A2: Descriptive Statistics by Region and Sex

VARIABLES		Male					Female				
		N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
LnW		5406	3.29	1.24	-4.09	8.36	3146	2.90	1.30	-4.94	7.71
Eduyear		5406	8.27	4.04	0.00	18.00	3146	7.80	4.22	0.00	18.00
Married		5406	0.68	0.47	0.00	1.00	3146	0.52	0.50	0.00	1.00
Age		5406	34.76	11.07	15.00	65.00	3146	33.44	10.85	15.00	65.00
Age^2		5406	1331	842	225	4225	3146	1236	804	225	4225
Mother's education		789	2.90	3.69	0.00	17.00	389	3.93	4.17	0.00	18.00
LnHHExp		17166	9.75	1.15	0.00	15.71	17844	9.71	1.06	0.00	15.10
HHChidren6-		17166	1.01	1.14	0.00	9.00	17844	1.17	1.18	0.00	9.00
HHAdults65+		17166	0.16	0.42	0.00	3.00	17844	0.17	0.43	0.00	3.00
Headship		17166	0.47	0.50	0.00	1.00	17844	0.18	0.39	0.00	1.00
HHownhouse		17166	0.62	0.49	0.00	1.00	17844	0.61	0.49	0.00	1.00
Urban		5406	0.50	0.50	0.00	1.00	3146	0.53	0.50	0.00	1.00
Regional Dummy	Central	5406	0.12	0.33	0.00	1.00	3146	0.15	0.36	0.00	1.00
	Coast	5406	0.12	0.33	0.00	1.00	3146	0.11	0.32	0.00	1.00
	Eastern	5406	0.16	0.37	0.00	1.00	3146	0.16	0.37	0.00	1.00
	Northeastern	5406	0.02	0.14	0.00	1.00	3146	0.01	0.10	0.00	1.00
	Nyanza	5406	0.16	0.37	0.00	1.00	3146	0.18	0.39	0.00	1.00
	Rift valley	5406	0.24	0.43	0.00	1.00	3146	0.21	0.41	0.00	1.00
	Western	5406	0.09	0.29	0.00	1.00	3146	0.08	0.26	0.00	1.00

Note: LnW: Log Hourly Wage; Eduyear: Years of Schooling; LnHHExp: Log Household Expenditure

Table A3. OLS and IV, Heckman and Joint IV-Heckman Estimations of Returns to Schooling

	OLS		IV		Heckman		Joint IV-Heckman					
	Male	**Female**	**Male**	**Female**	**Male**	**Female**		**Male**			**Female**	
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
	(eq. 1)		(eq. 3)	(eq. 3)	(eq. 5)	(eq. 5)	(eq. 8)	(eq. 7)	(probit eq. 6)	(eq. 8)	(eq. 7)	(probit eq. 6)
VARIABLES	ln W_i		ln W_i	ln W_i	ln W_i	ln W_i	ln W_i	S_i	D_{wi}	ln W_i	S_i	D_{wi}
S_i	0.140***	0.138***	0.115***	0.142***	0.127***	0.112***	0.149***			0.135***		
	[0.004]	[0.005]	[0.023]	[0.030]	[0.004]	[0.008]	[0.027]			[0.026]		
A_i	0.091***	0.083***	0.064**	-0.037	0.040***	0.018	-0.128	0.219	0.212***	-0.031	-10.088***	0.297***
	[0.009]	[0.012]	[0.031]	[0.045]	[0.012]	[0.019]	[0.094]	[1.184]	[0.014]	[0.186]	[2.867]	[0.022]
A_i^2	-0.001***	-0.001***	0.000	0.001	0.000	0.000	0.002*	-0.003	-0.003***	0.001	0.136***	-0.004***
	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]	[0.001]	[0.016]	[0.000]	[0.003]	[0.039]	[0.000]
Married_i	0.131***	0.064	0.133	0.015	0.083**	0.203***	-0.215	0.366	0.387***	0.01	7.264***	-0.227
	[0.038]	[0.042]	[0.110]	[0.313]	[0.039]	[0.054]	[0.196]	[2.047]	[0.077]	[0.332]	[2.292]	[0.186]
LnHHExp_i								-0.074	0.077***		1.048***	-0.02
								[0.442]	[0.023]		[0.211]	[0.027]
HHChidren6-_i								-0.190*	0.005		-0.310**	-0.004
								[0.113]	[0.020]		[0.147]	[0.025]
HHAdults65+_i								-0.194	0.004		0.981**	-0.027
								[0.224]	[0.042]		[0.39]	[0.059]
Headship_i								-0.514	0.297**		-2.931***	0.022
								[1.415]	[0.122]		[0.94]	[0.201]
HHownhouse_i								-0.161	-0.017		4.236***	-0.123**
								[0.270]	[0.045]		[1.181]	[0.058]
MS_i *(instrument)*								0.521***	-0.021***		0.093	0.006
								[0.121]	[0.006]		[0.076]	[0.007]
λ_i *(selectivity term)*					-0.410***	-0.531***	-1.034**	-2.499		0.005	-44.867***	
					[0.068]	[0.122]	[0.483]	[7.085]		[0.76]	[11.774]	
Constant	0.027	-0.032	0.566	1.771***	1.476***	2.001***	5.079**	6.897	-5.130***	1.719	230.087***	-5.608***
	[0.148]	[0.193]	[0.412]	[0.558]	[0.285]	[0.507]	[2.142]	[33.891]	[0.304]	[4.206]	[62.643]	[0.399]
Observations	5,406	3,146	789	389	17116	17844	789	789	6492	389	389	4835
Censored					11710	14698						
R^2	0.319	0.247	0.151	0.089			0.126	0.287		0.098	0.337	
Pseudo R^2									0.135			0.151
Cragg-Donald F-statistic								28.33			22.31	
Shea Partial R^2								0.179			0.262	
Wald chi2					1067	359.2						
F-test	633.5	257.9	30.53	11.03								

Notes: standard errors in parenthesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.; S_i: Year of schooling; A_i: Age; A_i^2: Age squared; Married: Marital status (1=Married); LnHHExp: Log household total expenditure; HHChildren6-: Living in household with children under age 6; HHAdults65+: Living in household with adults over 65 years old; Headship: Relation to household head (1=head); HHownhouse: household head's ownership of house (1=owned house); MSi: Mother's years of schooling; and λ_i: inverse mill's ratio (heckman's lambda) ; Results of first-stage estimation and probit estimations for column [1] – [6] are available upon request

Table A4. OLS and IV, Heckman and Joint IV-Heckman Estimations of Returns to Schooling, Level of Education Sub-Sample

Male			***Primary***				***Secondary***				***Tertiary***	
	OLS	**IV**	**Heckman**	**IV-Heckman**	**OLS**	**IV**	**Heckman**	**IV-Heckman**	**OLS**	**IV**	**Heckman**	**IV-Heckman**
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
VARIABLES		(eq. 1)	ln W_i			(eq. 1)	ln W			(eq. 1)	ln W	
S_i	0.063***	0.009	0.045***	0.006	0.182***	0.147	0.165***	0.201**	0.233***	0.901***	0.223***	0.712***
	[0.007]	[0.054]	[0.008]	[0.051]	[0.009]	[0.102]	[0.009]	[0.092]	[0.017]	[0.250]	[0.018]	[0.210]
A_i	0.077***	0.122***	0.011	0.178**	0.105***	0.05	0.034**	-0.059	0.152***	0.055	0.080***	-0.074
	[0.011]	[0.037]	[0.017]	[0.086]	[0.012]	[0.043]	[0.016]	[0.095]	[0.026]	[0.126]	[0.030]	[0.154]
A_i^2	-0.001***	-0.001**	0.000	-0.002*	-0.001***	0.000	0.000	0.001	-0.001***	-0.001	0.000	0.000
	[0.000]	[0.001]	[0.000]	[0.001]	[0.000]	[0.001]	[0.000]	[0.001]	[0.000]	[0.002]	[0.000]	[0.002]
Married$_i$	0.200***	0.022	0.213***	0.095	0.091**	0.028	-0.008	-0.198	0.031	0.431	-0.124	-0.264
	[0.048]	[0.124]	[0.050]	[0.160]	[0.046]	[0.152]	[0.050]	[0.237]	[0.075]	[0.365]	[0.084]	[0.488]
λ_i			-0.523***	0.314			-0.524***	-0.58			-0.516***	-1.487*
			[0.101]	[0.431]			[0.072]	[0.460]			[0.106]	[0.817]
Constant	0.810***	0.468	2.643***	-0.886	-0.723***	0.203	1.224***	2.401	-2.437***	-8.651***	-0.532	-1.787
	[0.179]	[0.439]	[0.399]	[1.907]	[0.215]	[1.109]	[0.347]	[2.139]	[0.439]	[2.406]	[0.593]	[3.608]
Observations	2,855	544	10,796	544	3,488	412	9,303	412	1,179	162	2,705	162
Censored			7,941				5,815				1,526	
R^2	0.1	0.066		0.067	0.203	0.101		0.085	0.408	-0.126		0.065
Wald Chi^2			96.68				467.5				486.3	
F-test	78.73	9.189		7.447	221.2	9.908		8.403	201.9	8.085		7.398

Female			***Primary***				***Secondary***				***Tertiary***	
	OLS	**IV**	**Heckman**	**IV-Heckman**	**OLS**	**IV**	**Heckman**	**IV-Heckman**	**OLS**	**IV**	**Heckman**	**IV-Heckman**
	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]
VARIABLES		(eq. 1)	ln W_i			(eq. 1)	ln W_i			(eq. 1)	ln W_i	
S_i	0.009	-0.094	-0.011	-0.073	0.232***	0.364***	0.181***	0.368***	0.333***	0.919***	0.311***	0.876***
	[0.009]	[0.064]	[0.011]	[0.048]	[0.013]	[0.124]	[0.017]	[0.088]	[0.025]	[0.312]	[0.028]	[0.280]
A_i	0.084***	0.049	0.019	0.098	0.057***	-0.048	-0.037	-0.237	0.106***	0.227	0.000	0.043
	[0.012]	[0.050]	[0.022]	[0.113]	[0.016]	[0.071]	[0.025]	[0.206]	[0.032]	[0.189]	[0.048]	[0.194]
A_i^2	-0.001***	-0.001	0.000	-0.001	0.000	0.001	0.001***	0.004	-0.001*	-0.003	0.001	-0.001
	[0.000]	[0.001]	[0.000]	[0.002]	[0.000]	[0.001]	[0.000]	[0.003]	[0.000]	[0.004]	[0.001]	[0.003]
Married$_i$	-0.088*	-0.309	0.124	-0.355	0.071	0.19	0.250***	0.204	-0.025		0.055	
	[0.052]	[0.370]	[0.079]	[0.379]	[0.051]	[0.433]	[0.064]	[0.433]	[0.082]		[0.089]	
λ_i			-0.590***	0.24			-0.680***	-0.748			-0.610***	-1.051**
			[0.164]	[0.454]			[0.135]	[0.815]			[0.199]	[0.480]
Constant	0.931***	2.068***	2.959***	0.776	-0.830***	-0.585	2.134***	3.61	-3.013***	-11.754**	-0.219	-6.558
	[0.213]	[0.609]	[0.603]	[2.515]	[0.284]	[0.977]	[0.659]	[4.788]	[0.575]	[5.219]	[1.096]	[5.103]
Observations	1,747	229	12693	229	1,933	252	8383	252	693	110	2207	110
Censored			10946				6450				1514	
R^2	0.027	-0.024		0.002	0.248	0.065		0.062	0.395	-0.163		-0.058
Wald Chi^2			12.14				283				214.6	
F-test	12.1	1.641		1.539	159.3	6.055		7.448	112.2	9.126		8.523

Notes: standard errors in parenthesis; *** p < 0.01, ** p < 0.05, * p < 0.1.; S_i: Year of schooling; A_i: Age; A_i^2: Age squared; Married: Marital status (1=Married); Results of first-stage estimation and probit estimations are available upon request

References

Andrabi, T., Das, J., and Khwaja, A. (2012). What Did You Do All Day? Maternal Education and Child Outcomes. *Journal of Human Resources*, 47(4), 873-912.

Armitage, J. and Sabot, R. (1987). Efficiency and Equity Implications of Subsidies to Secondary Education in Kenya. In Newbury, D., and Stern, N. (eds.), *The Theory of Taxation for Developing Countries*. Oxford, UK: Oxford University Press.

Barouni, M. and Broecke, S. (2014). The Returns to Education in Africa: Some New Estimates. *The Journal of Development Studies*, 50(12), 1593-1613. doi:10.1080/00220388.2014.936394.

Buchmann, C. (1999). The State and Schooling in Kenya: Historical Developments and Current Challenges. *Africa Today*, 46(1), 95-117. http://www.jstor.org/stable/4187254.

Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, 69(5), 1127-1160. http://www.jstor.org/stable/2692217.

Chen, G. and Hamori, S. (2009). Economic Returns to Schooling in Urban China: OLS and the Instrumental Variables Approach. *China Economic Review*, 20: 143-152. doi:10.1016/j.chieco.2009.01.003.

Colclough, C., Kingdon, G., and Patrinos, H. (2010) The Changing Pattern of Wage Returns to Educations and Its Implications. *Development Policy Review*, 28(6): 733-747.

Filmer, D., and Fox., L. (2014). *Youth Employment in Sub-Saharan Africa*. Africa Development Series. Washington, D.C.: World Bank. doi:10.1596/978-1-4648-0107-5.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153-161. http://www.jstor.org/stable/1912352.

Kabubo-mariara, J. and Mwabu, D. K. (2007). Determinants of School Enrolment and Education Attainment: Empirical Evidence from Kenya. *South African Journal of Economics*, 75(3), 572-593. doi:10.1111/j.1813-6982.2007.00138.x.

Lang, K. (1993). Ability Bias, Discount Rate Bias, and the Returns to Education. *Mimeo*, Boston University.

Johnson, G. E. (1972). The Determination of Individual Hourly Earnings in Urban Kenya. *Discussion Paper No. 22*, Center for Research on Economic Development, University of Michigan. http://deepblue.lib.umich.edu/handle/2027.42/100801.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. New York, NY: National Bureau of Economic Research.

Nyaga, R. K. (2010). Earnings and Employment Sector Choice in Kenya. *AERC Research Paper 199*, African Economic Research Consortium, Nairobi.

Psacharopoulos, G. (1973). *Returns to Education: An International Comparison*. San Francisco, CA: Jossy-Bass Inc.

Psacharopoulos, G. (1981). Returns to Education: An Updated International Comparison. *Comparative Education*, 17(3), 321-341.

Psacharopoulos, G. (1994). Returns to Investment in Education: A Global Update. *World Development*, 22(9), 1325-1343. doi:10.1016/0305-750X(94)90007-8.

Psacharopoulos, G and Patrinos, P. (2002). Returns to Investment in Education: A Further Update. *Policy Research Working Paper 2881*, World Bank. https://openknowledge.worldbank.org/handle/10986/19231

Thias, H. H. and Carnoy, M. (1972). *Cost-Benefit Analysis in Education: A Case Study of Kenya*. Baltimore, MD: The Johns Hopkins Press.

Trostel, P., Walker, I., and Woolley, P. (2002). Estimates of the Return to Schooling for 28 Countries. *Labor Economics*, 9, 1-16. doi:10.1016/S0927-5371(01)00052-5.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

World Bank (2004). Republic of Kenya Country Assistance Strategy, World Bank, Washington DC

World Bank (2015). World Data Bank: World Development Indicators. http://databank.worldbank.org/.