# ESTIMATING THE RETURNS TO SCHOOLING: A COMPARISON OF FIXED EFFECTS AND SELECTION EFFECTS MODELS FOR TWINS

A. Agyeman

*(CSIR - Science and Technology Policy Research Institute (STEPRI), Ghana).*
*Email: nesiagyeman@yahoo.com*

**ABSTRACT**

Strong empirical links exist between the number of years spent schooling and earnings. However, the relationship may be masked due to the effect of unobserved factors that influence both wages and schooling. Two of the main econometric models, namely fixed-effects and selection-effects, used to analyse returns to schooling were compared using monozygotic and dizygotic twins' datasets in Ghana. The efficiency of the models was assessed based on the standard errors associated with the return to schooling estimates. Goodness of fit measures were used as a basis for comparison of the performance of the two models. The results revealed that based on their standard errors, the regression estimates from the selection effects model (MZ = 0.1014±0.0197; DZ = 0.0947±0.0095) were more efficient than the regression estimates from the fixed-effects model (MZ = 0.1115±0.0353; DZ = 0.082±0.0127). However, the AICc values of the fixed effects model ($MZ_{AICc}$ = 57.8 and $DZ_{AICc}$ = 105.4) were smaller than the AICc values of the selection effects model ($MZ_{AICc}$ = 151.6 and $DZ_{AICc}$ = 221.6). Findings from the study indicate that, although, both models produced consistent estimates of the economic returns to schooling, the fixed effects model provided a better fit to the twins' data set.

**Keywords:** Returns to schooling, fixed effects model, selection effects model, ability bias, twins.

## Introduction

Education plays a central role in modern labour markets and is fundamental in promoting economic development and growth. Empirical evidence has over the years shown that an increase in workers' educational level improves their human capital, increasing the productivity of these workers and the economy's output (Canals, 2017; WEF, 2016). One of the most important economic decisions that individuals and policy makers are confronted with is how much to invest in education in order to improve the quality of life of people. Evidence shows that, on average, each additional year of education boosts a person's income by 10% and increases a country's GDP by 18% (UNICEF,

2015; Montenegro & Patrinos, 2014).

Typically, the economic value of investment in education has been estimated in terms of rates of return (Orhan, 2010). The concept of the "rate of return to schooling" has been the subject of continued interest for over fifty years in economics and statistics, and quantifying the returns remains an open question (Broome, 2012). This is because the empirical link between education and earnings may be confounded due to the effect of unobserved factors related to schooling and earnings and failing to account for these factors may result in a biased estimate of the rate of return to schooling. The difficulty in untangling the influence of education on earning outcomes lies

in the fact that an individual's decision about their education is potentially related to their unobserved abilities, tastes or social influences and, in turn, these unobserved qualities have an impact on the individual's earning (Heckman, 2010; Broome, 2012).

A number of empirical methods have been adopted by economists and statisticians to estimate the rate of return to schooling and notable among them is Mincer's 1974 human capital model. Studies using Mincer's model usually report standard estimates of the return to a year of schooling in developed countries of 8-10%, with higher estimates for developing countries and low levels of schooling (Krueger & Lindhl, 2001; Psacharopoulos, 1994). However, in spite of its wide-ranging applications, the model's simple estimation using the Ordinary Least Squares (OLS) technique may be biased due to the presence of unobserved ability. The differences in the average earnings of workers with different levels of education may partly reflect the inherent differences in unobservable characteristics (Willis & Rosen, 1979; Griliches (1977). For example, some evidence for South Africa suggests that OLS estimates of the returns to schooling are likely to suffer from a significant bias due to omitted variables, especially from the omission of ability measures (Hertz, 2003; Case & Yogo, 1999). Consequently, when the unobserved ability differences among workers are correlated with their levels of education (popularly known as ability bias), relying on OLS estimates for the average individual will not be enough to make informed decisions on the percentage of earnings associated with acquiring more education. As specified by Kaymak (2009), it is not clear whether the rising earnings ratio between more educated workers and those with less education reflects a higher return to formal education or a larger ability bias. The distinction he indicates is essential for human capital policy.

"Ability bias" may bias upwards the observed returns to schooling because high-ability people find it easier to take on more education or bias downwards the observed returns to schooling if low-ability people compensate by completing more education (Leigh & Ryan, 2008). The problem of ability bias arises basically from the fact that people with high ability are more likely to pursue more schooling because they are more academically capable and will therefore benefit more from studying and are more likely to earn more money (Bonjour *et al.,* 2002). Failing to control for ability bias therefore, will cause the earnings-education regression model to attribute all wage differences to schooling and none to ability and this will ultimately lead to an overestimate of the effect of an additional year to schooling. It is important therefore, to control for ability bias in order to have a better understanding of the "true", unbiased returns to education.

Among the recent approaches that have been proposed for addressing the problem of ability bias are the Fixed Effects (FE) and the selection-effects models on a sample of twins. '*Twins*' studies exploit the idea that it is possible to estimate the "true" effect of schooling on earnings by comparing the earnings received by twin pairs who obtain different amounts of schooling, but are assumed to have similar ability levels. Identical twins are assumed to have the same inherent ability and share the same family background and non-identical (fraternal) twins share on average, 50% of their inherent ability and if reared together will have a common family background (Miller *et al.,* 1995). By comparing the results from identical twin pairs and fraternal twin pairs, it is possible to separately parse out the components of ability bias that are due to genetic characteristics and family background (Leigh & Ryan, 2008). While, the fixed-effect approach relates the differences in the earnings of a set of twins to differences in characteristics of the same

individuals in a regression framework, the selection effects model (Ashenfelter & Krueger, 1994) alternatively incorporates unobservable family effects in the earnings equation and the unmeasured family effects are modelled as depending on the educational accomplishments of each twin member (Miller *et al.,* 1995). Yet, amongst these two models, it is essential to compare their relative performance, given a common data set to obtain robust or consistent and unbiased rates of return to schooling estimates. The best model should be able to effectively reduce or eliminate the unobserved differences that bias a cross-sectional relationship between education and earnings in order to obtain a "true" or "pure" return to schooling. In a study on returns to schooling, Nakamuro & Inui (2012), employed the fixed-effects and the selection effects models by making use of data from a sample of twins in Japan to control for ability bias in returns to schooling estimates. Their findings indicate that the OLS estimate of the return to schooling (10%) is about twice that of the fixed effect (FE) and generalised least squares (GLS) estimates of between 4.5 and 5.4%, suggesting that their standard OLS estimate is upward biased. However, the within-twin pair differences in their FE and GLS rate of return to schooling estimates were less than 1%. According to a 1999 study by David Card, studies of identical twins suggest that standard OLS estimates of education's returns are ability-biased upward by 10%. Similarly, Taubman (1976), examined within-pair differences in annual earnings and schooling among male twin veterans in the NAS-NRC dataset and found evidence of substantial upward ability bias in traditional cross-sectional estimates of the returns to schooling. His estimates decreased from 8.8 to 4.8% when the OLS regression and the fixed-effects models were compared, suggesting that the OLS estimate is biased upward by the omission of family and ability effects. In

another related study, Behrman *et al.* (1977), analysed data from a sample of male World War II veteran twins and observed that, of the overall rate of return to schooling of 8%, about 2.7% could be attributed to schooling, 3.2% to genetic factors and 2.1% to shared family environment. This is a confirmation that genetic and family factors explain a portion of the effect of education on earnings.

Contrary to these findings, empirical evidence from studies using a sample of young twins from Ohio in the United States and Australia by Ashenfelter & Krueger (1994); Miller *et al.* (2006, 1995) indicate that estimates of the rate of return to schooling obtained using conventional models such as OLS regression model are similar to that of either the fixed-effects or the selection effects models. According to a study on the return to an additional year of schooling for identical twins in the US labour market, Ashenfelter & Krueger (1994) concluded that family and genetic effects make virtually no contribution to the returns to schooling. Bound & Solon (1999) and Neumark (1999) however, suggest that Ashenfelter & Krueger's (1994) findings may stem from ability differences between twins that are not removed in the fixed effects model, and from between-twin's differences in schooling that are chosen endogenously (Miller *et al.,* 2006). Notwithstanding, Miller *et al.* (2006) found out from their selection effects model estimates that, family background and genetic factors account for around one-half a percentage point of the total return to education of between 6.5 and 7% and genetic factors account for a further one-half of a percentage point.

This study seeks to compare two econometric models of the return to an additional year of schooling, highlight their strengths and weaknesses and discuss how these factors might relate to obtaining a consistent and unbiased estimate of the return to an additional year of schooling. We hope the findings will be

able to provide an appropriate basis for further empirical research.

The objectives of this study are, therefore, (i) to estimate the returns to education using the twins fixed-effects approach and selection-effects approach, (ii) to compare the efficiency of the regression estimates and the performance of the fixed-effects model and the selection-effects model (iii) to assess the robustness of the estimates of the returns to schooling to the choice of estimating model and (iv) to provide an overall view of the relevance of standard estimates of rates of return to schooling for education policy and linking this policy to labour market outcomes.

## Experimental

*Survey data*

In the absence of a national twins' database in Ghana, the study used data from a labour market survey, which was carried out by a team of five questioners in December 2007 and January 2008 in three cities in Ghana, namely Accra, Kumasi and Takoradi. Based on a modification of existing twins' questionnaires by Ashenfelter & Krueger (1994) and others, the survey covered a unique and rich source of information on the socio-economic characteristics (age, gender, marital status, earnings, education, family background characteristics such as sibling education, father's and mother's education etc.) of '*twins*' in Ghana. Questionnaires were administered through face-to-face personal interviews to gainfully employed adult twins aged between 18 and 65. Twins were identified through various channels including twins registered at the twin's clubs, various work places, markets, shops, colleagues, friends, relatives, and households. Overall, these channels permitted a roughly equal probability of contacting most of the twins in the three cities. In Kumasi 404 respondents were identified, whereas in Accra and Takoradi a total of 96 respondents were

identified. Altogether, 500 respondents were identified. Fifty percent (50%) of twins identified were randomly selected and interviewed giving a total of 250 respondents made up of 125 twin pairs. Out of the 250 respondents, 144 individuals were dizygotic (DZ) twins and 106 were monozygotic (MZ) twins. Data analysis is performed using three samples of twins (Pooled, Monozygotic and Dizygotic) to control for genetic and other family background unobserved factors in the rate of the economic return to schooling.

*Models for estimating the returns to schooling - statistical approach*

The association between educational attainment and annual earnings were investigated using the ordinary least squares, fixed-effects and the selection-effects estimating equation models. STATA 14.0 (StataCorp LP, 2015) programmes were used to analyse the data.

*Ordinary Least Squares (OLS) regression model*

The Ordinary Least Squares regression model in this study was based on the human capital model by Mincer (1974). In Mincer's model, the natural logarithm of earnings is expressed as a linear function of years of schooling and a quadratic function of potential experience:

$$LnY_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \varepsilon_i \quad (1)$$

where $LnY_i$ is the natural logarithm of the observed earnings for individual $i$, $\beta_0$ is a constant term, $S_i$ is the number of years of schooling for individual $i$ and $\beta_1$ is the average rate of return to one additional year of schooling, $\sigma_i^2$ potential experience (age and its square) of the individual worker $i$ which is entered in linear and quadratic forms, $\varepsilon$ is assumed to be normally distributed with mean zero and variance $\sigma_i^2$.

*Fixed-effects model*
The return to education was estimated based on a linear model by Chamberlain (1982) which has also been frequently used in the twins' literature as the "within-twins" model (Ashenfelter & Krueger, 1994; Ashenfelter & Rouse, 1998). In this model, the difference in the earnings between two twins in a pair is related to differences in the educational attainment and other observable characteristics of each twin. This fixed effects model provides an estimate of the impact of education on the outcome variable which is not biased by the omission of family background variables (Miller, Mulvey & Martin, 1995). The earnings equations for a pair of twins are specified as follows:

$$LnY_{2i} = \alpha X_i + \beta Z_{2i} + \mu_j + \varepsilon_{2i} \quad (2)$$

$$LnY_{2i} = \alpha X_i + \beta Z_{2i} + \mu_j + \varepsilon_{2i} \quad (3)$$

where, $LnY_{ji}$ ($j = 1, 2$) represents the natural logarithm of the earnings of both twins in the pair. These earnings are assumed to depend on an unobservable component that varies by family, $\mu_i$, observable components, such as age and gender (difference is 0 in each pair of twins except for dizygotic twins), that vary by family but not across twins, $Z_{ji}$, observable components, such as schooling and marital status, that may vary across the twins in a pair, $Z_{ji}$ ($j = 1, 2$), unobservable individual components, $\varepsilon_{1i}$ and $\varepsilon_{2i}$ which are assumed to be independent of $Z_{ji}$ and $\mu_i$. The terms $\beta$ and $\beta$ are the common intercept and the return to education respectively.

In order to estimate the return to schooling efficiently, the difference between equations (2) and (3) are taken and both the observable and unobservable family effects (i.e., $X_i$ and $\mu_i$) are eliminated resulting in the first differenced, or fixed-effects model. This is specified as:

$$LnY_{1i} - LnY_{2i} = \beta(Z_{1i} - Z_{2i}) + \varepsilon_{1i} - \varepsilon_{2i} \quad (4)$$

The fixed-effects model is estimated by the Ordinary Least Squares (OLS) method. The within-pair difference in earnings is related to differences in observable components that vary across the twins in a pair. The major feature of the fixed effects model is that genetic resemblance and common environment influences are reduced or eliminated.

*Selection-effects model*
Although the fixed effects model provides a simple and intuitive framework for estimation, it does not allow us to directly estimate the correlation between the unobservable family effects and schooling. It is this correlation which may cause usual estimates of earnings equations to suffer omitted variables bias. The "selection effects" model, introduced by Ashenfelter & Krueger (1994), does allow explicit consideration of the correlation with unobservable family effects in the earnings equation. If the correlation between the unobservable family effects $\mu_i$ and the observables is:

$$\mu_j = \gamma Z_{1i} + \gamma Z_{2i} + \delta X_i + \omega_i \quad (5)$$

where, the common $\gamma$ shows that the correlation is the same with each twin's observables, and $\omega_i$ is an uncorrelated, random error, then substituting the $\mu_i$ term into (2) and (3) results in equations (6) and (7):

$$LnY_{2i} = [\alpha + \delta]X_i + \gamma Z_{1i} + [\beta + \gamma]Z_{2i} + \varepsilon'_{2i} \quad (6)$$

and

$$LnY_{2i} = [\alpha + \delta]X_i + \gamma Z_{1i} + [\beta + \gamma]Z_{2i} + \varepsilon'_{2i} \quad (7)$$

where, $\varepsilon'_{1i} = \omega_i + \varepsilon_{1i}$ and $\varepsilon'_{2i} = \omega_i + \varepsilon_2$.

The selection-effects model is estimated by the feasible generalised least squares (FGLS). FGLS estimates are the seemingly unrelated regression method (Zellner, 1962). We use

FGLS to increase efficiency by exploiting cross-equation restrictions and to ensure correct computation of sampling errors.

In this framework, the coefficient on the co-twin's observable characteristics $\gamma$ provides an estimate of the impact of family effects (Miller, Mulvey & Martin, 1995) (i.e., the earnings of twin 1, depends on the observable characteristics of the other twin, $Z_{2i}$ because those co-twin characteristics act as an indicator of the unobserved family background that is common to the two twins). The estimated coefficients on own-education are comprised of two parts $[\beta + \gamma]$ , where $\beta$ captures the returns to schooling and $\gamma$ captures the effect due to family unobservable characteristics. Ordinarily, these two effects cannot be untangled, so the estimated coefficient will be a biased estimate of $\beta$ unless $\gamma = 0$. The coefficient on the co-twin's educational level can be subtracted from the coefficient on the own-education variable to derive an estimate $[\beta + \gamma] - \gamma$ of the pure effect of schooling for monozygotic twins, or of the effect of schooling biased by the omission of a direct measure of ability for dizygotic twins (Miller, Mulvey & Martin, 1995).

*Model Selection and Inference*
Measures of goodness of fit statistics: the corrected Akaike information criteria (AICc) (Hurvich & Tsai, 1989), the Bayesian information criterion (BIC) (Schwarz, (1978), and the likelihood ratio test were used to establish model credibility and confidence. The formulae for goodness of fit statistics are as follows:

$$\textbf{A}\ \mathcal{E} = -2\log(L) + 2k + \frac{2k(k+1)}{n-k-1}, \text{ and}$$
$$(8)$$

$$\textbf{B}\ C = -2\log(L) + k\log(n) \qquad (9)$$

where, $n$ is the number of observations, $k$ is the number of model parameters and $-2\log(L)$ is the model log-likelihood estimate.

The likelihood ratio test (LRT) which is presented as the difference in the log likelihoods was also used to compare the goodness of fit of the fixed-effects and selection-effects models. The likelihood ratio test is

$$\tau = \frac{l(\beta_{fe})}{l(\beta_{se})} \qquad (10)$$

Under the null,

$$LRT = -2\operatorname{h}(\tau) = 2\ln\{l(\beta_{se}) - l(\beta_{fe})\} \ (11)$$

The test is chi square distributed. Standard errors denoted as

$$SE = \frac{\sigma}{\sqrt{n}}$$

were also used to evaluate the efficiency of the regression estimates. The STATA 14.0 statistical and economic software was used for the data analysis. The post-estimation procedure of STATA 14.0 reported goodness of fit statistics (e.g., AICc, BIC and -2LL).

**Results**

*Descriptive statistics*
Summary descriptive statistics of monozygotic and dizygotic twins are presented in Table 1. On average, the twins had almost 13 years of education (Table 1). This indicates that a good number of the twins in this study did not have very high educational qualifications. MZ twins have 14 years of education, which is about three extra years of education on average than DZ twins (Table 1). On average, DZ twins were almost 2 years older than MZ twins (Table 1). Monozygotic twins earned more on average than DZ twins, a difference of about 460 US dollars (Table 1). This might probably be due to the fact that MZ twins in this study were more highly educated than DZ twins.

**TABLE 1**
*Descriptive Statistics of Selected Variables: Sample of Twins in Ghana*

| Variable | Means and Standard Errors | | |
| --- | --- | --- | --- |
| | **Pooled sample** | **Monozygotic twins** | **Dizygotic twins** |
| Own education (years) | 12.576 | 14.009 | 11.521 |
| | (0.343) | (0.535) | (0.427) |
| Co-twins education (years) | 12.692 | 13.840 | 11.847 |
| | (0.345) | (0.550) | (0.429) |
| Male (proportion) | 0.488 | 0.509 | 0.472 |
| | (0.032 | (0.049) | (0.042) |
| Age (years) | 32.816 | 31.887 | 33.500 |
| | (0.649) | (0.905) | (0.907) |
| Married (proportion) | 0.432 | 0.321 | 0.514 |
| | (0.031) | (0.046) | (0.042) |
| Log of annual income | GH¢7.184  (0.054) | GH¢7.368 | GH¢7.049 |
| | | (0.084) | (0.068) |
| Sample size | 250 | 106 | 144 |

Note: Standard errors in parentheses below the means. Average exchange rate of GH¢0.94 to the US dollar prevailing in June 2007.

*Ordinary Least Squares (OLS) estimates of the returns to schooling*

Twins studies of the returns to schooling typically begin with OLS estimates as a way of replicating conventional cross-sectional estimates. A benchmark set of results of the OLS regressions using the pooled, MZ and DZ twins' data are set out in Table 2. The results for the pooled sample of twins showed a strong and positive association between earnings and education for the pooled, MZ and DZ twins (Pooled = 10.1%; MZ = 9.4%; DZ = 10.4%, $p < 0.0001$) and also revealed that the return to schooling was quite high according to average world estimates. However, OLS return to schooling estimates for MZ twins, were smaller than that of DZ twins. Average occupational earnings also increased with age

(0.05, $p = 0.028$) and this result was statistically significant at 5% level indicating that experienced workers earn more as productivity-related skills are improved. Male workers were found to be at a substantial 10.3% earnings advantage compared with female workers for the pooled sample of twins (Table 2). Although, the effect of the proportion of married twins on earnings was negative, it was however significant (-19.4%, $p < 0.05$), indicating that being married does not guarantee an individual an increase in earnings. Education, potential experience, gender, and marital status explained about 49% of the variance (R-Squared) in the log annual earnings for all twins (Table 2). This indicates that about half of the variance in the OLS regression model may be attributable to other factors such as family background (e.g., parental education and income) and unobservable genetic traits.

**TABLE 2**
*OLS estimates of Equation (1) for Pooled, MZ and DZ Twins*

| Parameter | Pooled | MZ | DZ |
|---|---|---|---|
| Intercept | 4.8132*** | 4.0642*** | 5.5757*** |
|  | (0.4094) | (0.6553) | (0.5228) |
| Years of Schooling | 0.1014*** | 0.0943*** | 0.1038*** |
|  | (0.0074) | (0.0107) | (0.0104) |
| Age | 0.0515** | 0.0936** | 0.0078 |
|  | (0.0233) | (0.0363) | (0.0302) |
| Age squared | -0.00049 | -0.00083* | -0.000036 |
|  | (0.0003) | (0.00048) | (0.00039) |
| Male | 0.1033 | 0.030069 | 0.2102 |
|  | (0.0777) | (0.1185) | (0.1023) |
| Married | -0.1936** | -0.2634** | 0.1457 |
|  | (0.0789) | (0.1215) | (0.1020) |
| $R^2$ | 0.4913 | 0.5542 | 0.4636 |
| N | 250 | 106 | 144 |

Note: **=$p < 0.05$, ***=$p < 0.001$. Standard errors in parentheses below means

*Fixed-effects analysis*

A regression of the within-twin difference in earnings on the within-twin difference in education levels (i.e. fixed effects estimate) was estimated for the MZ and DZ twins' samples. The results in columns 3 and 4 of Table 3 corresponds to fitting equation 4 under the Materials and Methods Section using the Ordinary Least Squares (OLS) estimator. These are the results where unobserved and family characteristics such as innate ability and family background shared by twins are differenced out. Estimates of the return to schooling in this study revealed that the "true" impact of education was 11% and 8% for MZ and DZ twins', respectively (Table 3). These estimates are also highly significant ($p < 0.0001$) suggesting that an additional year of schooling increases an individual's earnings on average by about 8% or 11%. It is however noted that, the fixed effect estimate of the return to schooling for MZ twins was about three percentage points higher than the estimate for DZ twins (Table 3). This difference may be due to the explicit control for both ability and family background in the estimations for MZ twins, but only for family background in the estimations for the DZ twins who share 50% of genes and were reared together.

Using the fixed effects approach to address the problem of ability bias in returns to schooling, the OLS estimates for the returns to schooling for MZ and DZ twins in Table 2 were compared with the fixed effects OLS (FEOLS) estimates in Table 3. The results revealed that the OLS return to schooling estimate was less than the fixed-effects OLS estimate for MZ twins by about 0.02% points, indicating a downward ability bias to the OLS estimate. The return to schooling for MZ twins increased by almost 15% in the fixed effects regression model and remained highly significant ($p < 0.01$). The results further revealed that the constant term of the fixed effects OLS regression was significant ($p < 0.05$), indicating the presence of unobservable fixed effects.

This is a confirmation that OLS squares returns to schooling for MZ twins may be biased. On the other hand, OLS return to schooling (10.4%) for DZ twins (Table 2) was higher than the fixed effects OLS estimate (8.2%, Table 3), indicating an upward ability bias to OLS estimate for DZ twins. This positive ability bias suggests that higher ability individuals invest more in schooling and also have higher earnings.

*Selection-effects analysis*
Estimates from the selection effects model using the model proposed by Ashenfelter & Krueger (1994) are also presented in Table 3. This alternative model provides a means for assessment of the robustness of the findings obtained from the fixed-effects model. The results in Table 3 correspond to fitting equations (6 and 7), under the Materials and Methods Section, using the feasible generalised least squares (FGLS) estimator. These are the results that include the sibling's education level as an additional regressor to control for any "family" effects that affect the absolute level of earnings in each twin's wage equation. The coefficient of this variable is a measure of the selection effect, ($\gamma$ ). These estimates take account of the cross-equation restrictions apparent in Equations (6 and 7), which can directly estimate both the return to education and the ability or family background effect.

Based on the selection effects model and using the FGLS estimator, the education coefficient which comprises of the returns to education and the family effect or ability [$\beta + \gamma$], the results revealed that, for each additional year of schooling, the rate of return to schooling for MZ twins was large (10.1%, $p < 0.05$, Table 3). The results in Table 3 also show that the coefficient ($\gamma = -0.007; p>0.05)$ on the co-twin's educational attainment was negative and not significant. This coefficient is the estimated ability or family effect and

also provides an estimate of the correlation between the education levels of the twins and the unobserved family background effect (residual). The negative coefficient (selection effect) is an indication that ability or family effects are either negligible or uncorrelated with educational decisions for MZ twins. The negative effect also suggests that using the MZ twins sample, the better educated families may receive a slightly lower benefit to education. This result also implies that a regression estimator of the returns to schooling that does not adjust for the selection effect might be downward biased. Average occupational earnings also increased with age and was statistically significant ($p < 0.05$) indicating that experienced MZ twin workers earn more as productivity-related skills are perfected. The results further revealed that, the effect of socio demographic characteristics such as gender (i.e., male or female) and marital status (i.e., married) on earnings of MZ twins was not significant ($p>0.05$, Table 3).

Results from the FGLS estimation showed that the return to an additional year of schooling for the DZ twins was 9.5% and also highly significantly ($p < 0.05$, Table 3). This estimate is however about 0.6 percent points lower than that of the MZ twins. The results also showed that the selection effect for DZ twins was small (0.02%) but positive (Table 3). Contrary to the results of the selection effect for MZ twins, the highly educated families of DZ twins were those who would be the most highly compensated in the labour market (i.e., highly educated DZ twins would receive higher returns to schooling). This result also implies that a regression estimate of the returns to education that is not adjusted for the selection effect may be upward biased for individuals that are not genetically identical. The effect of an additional year of schooling using the selection effects model was highly ($p < 0.05$) significant for male DZ twins, how

ever, age and subsequently experience did not have any significant impact on earnings.

The structural or net effect of schooling which is given by the coefficient on own years of education minus the coefficient on sibling's years of education in the selection effects model was estimated to be about 11% for MZ twins (Table 3). This was achieved by subtracting the coefficient (-0.007) of the co-twin's educational attainment from the struc-

tural estimate (0.1014) of the return to schooling that controls for omitted variables bias (i.e. the coefficient on the own education variable). The net effect of schooling (0.1115) was similar to the fixed effects schooling estimate for MZ sample of twins (Table 3). This estimate of the net effect of schooling gives an indication of the robustness of the selection effects model. A similar trend was also observed for DZ

twins in the selection effects model.

**TABLE 3**

*Estimates of Twins Models of Log Annual Earnings*

| | Monozygotic Twins | | Dizygotic Twins | |
|---|---|---|---|---|
| Variable | Selection effects | Fixed effects | Selection effects | Fixed effects |
| Intercept | 5.0683*** | 0.1146** | 5.500*** | 0.0602 |
| | (0.3695) | (0.0538) | (0.2772) | (0.0599) |
| Own Education | 0.1014*** | 0.1115*** | 0.0947*** | 0.0820*** |
| | (0.0197) | (0.0353) | (0.0095) | (0.0127) |
| Co-twin's Education | -0.0071 | | 0.0160 | |
| | (0.0197) | | (0.0096) | |
| Age | 0.0321*** | | 0.0038 | |
| | (0.0099) | | (0.0071) | |
| Male | -0.0301 | | 0.2181*** | |
| | (0.1649) | | (0.0689) | |
| Married | -0.0969 | | 0.0215 | |
| | (0.1689) | | (0.1243) | |
| R² | **0.3908** | 0.1664 | **0.4505** | 0.3768 |
| N | 106 | 53 | 144 | 72 |

Note: **=$p < 0.05$, ***=$p < 0.001$. Standard errors in parentheses below means

*Model Efficiency and Comparison - Goodness of fit Statistics*

The results in Table 3 also show the standard errors associated with the return to schooling estimates. The standard errors of the return to schooling estimates obtained using the FGLS equations (MZ = 0.0197; DZ = 0.0095) were lower than the standard errors obtained using the fixed-effects OLS estimations ((MZ = 0.0353; DZ = 0.0599), Table 3). In other words, FGLS coefficients were seemingly

more accurate and offered better estimates. This confirms the fact that FGLS estimates improve the efficiency of regression estimates by exploiting cross-equation restrictions and ensuring correct computation of sampling errors. To select the best fitting model which is able to effectively reduce or eliminate the unobserved differences that bias a cross-sectional relationship between education and earnings, the fixed effects and selection effects models were compared using three measures of goodness of fit

statistics namely, the Akaike information criterion with a correction for small sample sizes (AICc), the Bayesian information criterion (BIC) and the likelihood ratio test (LRT). The AICc, BIC and the log likelihood function are calculated for both the fixed-effects and the selection effects model and the results are presented in Table 3. The results revealed that the AICc values of 57.8 and 105.4, for MZ and DZ twins respectively in the fixed effects model were smaller than that of the selection effects model (MZ = 151.63; DZ = 221.59). Based on the concept that 'smaller AICc values are better when comparing two or more models',

the smaller AICc values in the fixed effects model suggest that for both sets of twins the fixed effects model effectively controlled for unobserved factors such as genetic ability and family background than the selection effects model. A similar pattern was also observed using the BIC and log likelihood values for both MZ and DZ twins. The smaller AICc, BIC and loglikelihood values of the fixed-effects model is a clear indication that it is the best fitting model in this study. The likelihood ratio test however, indicated that both models were not statistically different ($p>0.05$) from each other, (Table 4).

**TABLE 4**
*Goodness of Fit Statistics of Fixed-effects and Selection-effects Models*

| Diagnostics | Monozygotic Twins | | Dizygotic Twins | |
|---|---|---|---|---|
| | Selection Effects | Fixed Effects | Selection Effects | Fixed Effects |
| AICc | 151.6307 | 57.79685 | 221.5855 | 105.3938 |
| BIC | 161.4821 | 67.64831 | 232.9688 | 116.7771 |
| -2LogLikelihood | -70.81533 | -23.89843 | -105.7927 | -47.6969 |
| LRT $\chi^2(4)$ | -93.40, Prob > chi2 = 1.0000 | | -110.34 , Prob > chi2 = 1.0000 | |
| N | 106 | 106 | 144 | 144 |

**Discussion**

We provide returns to schooling estimates for both MZ and DZ twins in Ghana using regression models and data that account for unobservable differences in returns to schooling. Furthermore, we compared the performance of the fixed-effects and selection-effects models based on the informational efficiency of the two models and assessed the robustness of the returns to schooling regression estimates. The returns to schooling are thus discussed under the following headings (i) Returns to education using the twins fixed-effects approach and selection-effects approach, (ii) Robustness of the estimates of the returns to schooling to the choice of estimating model and (iii) Relevance of standard estimates of rates of return

to schooling for education policy. Altogether, the analyses presented below provide a better option of efficiently estimating the returns to schooling.

*Returns to education using the twins fixed-effects approach and selection-effects approach*
Fixed-effects and selection-effects regression analysis of the returns to schooling produced very similar results when applied to the MZ and DZ twins' data sets, though the returns to schooling estimates for MZ twins were slightly higher than that of the DZ twins. Either model could therefore be used to provide direct estimates of genetic influences in the context of the estimation of the economic returns to schooling. However, the estimation of fam

ily background effect in the selection-effects model produced contrasting results for the two types of twins' sets, although both selection effects were not statistically significant. While the family effect was negative $\gamma$ = -0.007 for MZ twins, it turned out to be positive $\gamma$ = 0.016 for DZ twins. For the MZ twins therefore, the better educated families are not those who will be the most highly compensated in the labour market. These returns to schooling estimates for MZ twins are similar to results by Ashenfelter & Krueger, (1994) who conclude from their estimates of the selection effects model that family background and natural ability play virtually no role in determining earnings. Contrary, to these results, (Taubman, 1976), examined within-pair differences in annual earnings and schooling among male twin veterans in the NAS-NRC dataset and found evidence of the importance of family effects in his return to schooling estimate of 4.8%. Miller *et al.* (1995) also found evidence that estimated family effects ($\gamma = 0.025$) was positively related to the completed schooling levels for MZ twins and reported a modest role for ability and family background in the relationship between schooling and income.

*Robustness of the estimates of the returns to schooling to the choice of estimating model*
Given the similarity of the findings, for the purpose of obtaining robust estimates of the returns to schooling in the two econometric models, special attention is placed on standard errors of the coefficients of the estimated returns to schooling. In both the MZ and DZ twins' samples, the standard errors obtained using the FGLS results ($SE_{MZ}$ = 0.0197; $SE_{DZ}$ = 0.0095) are lower than the standard errors in the FEOLS estimations ($SE_{MZ}$ = 0.0353; $SE_{DZ}$ = 0.0127). In other words, FGLS coefficients seem to be more accurate and yield better estimates. Accordingly, the FGLS returns to schooling estimate gained efficiency over that

of the FEOLS returns to schooling estimate as expected. These results are similar to those of Ashenfelter & Krueger (1994) and Ashenfelter & Rouse (1998) in that they also obtained lower standard errors for their FGLS coefficients. One possible explanation associated with these lower standard errors when using the FGLS estimator instead of the FEOLS estimator is that, the FGLS estimator is more efficient than the FEOLS estimator and this is a consequence of the Gauss-Markov theorem which infers that when the residuals are correlated across observations, OLS standard errors can be biased and either overestimate or underestimate the true variability of the coefficient estimates.

A comparison of the performance of the fixed-effects and the selection-effects models in their ability to explain unobserved effects was measured by goodness of fit statistics (AICc, BIC, log-likelihood ratio statistic). According to model fit statistics AICc and BIC, the fixed-effects model is preferable to the selection-effects model in its ability to reduce or eliminate some of the unobserved differences that bias a cross-sectional comparison of education and earnings. However, the likelihood ratio test indicates that both models were not statistically different ($p > 0.05$) from each other. Similar to findings by Ashenfelter & Krueger (1994), the inclusion of a sibling's education in the earnings model (i.e. selection-effects model) make virtually no contribution to the returns to schooling. Given the similarity of the findings in this study and in other related studies, it appears that estimates of the returns to schooling for both models are robust. Contrary to these findings, the results of a study on '*effects of family background on the returns to schooling*' by Deschenes (2007) indicate that family background variables play an important role in generating variation in the return to schooling across individuals. According to Deschenes (2007), who used a sample size of 17,300, the estimated average causal

effect of education reduces by 20% when the returns to schooling is allowed to vary with family background variables.

*Relevance of standard estimates of rates of return to schooling for education policy*
Empirical evidence about the economic returns to schooling have important implications for policy formulation and decision making within the education and labour market sector. This information is important for policymakers who must decide on education spending, prioritisation of schooling levels, and education financing programmmes (Patrinos, 2016). The economic rate of returns to schooling can inform individuals as to which type of education or training investment will yield them the highest return. The standard returns to schooling estimate in this study is 10%, which is comparable to standard world average estimates (Patrinos, 2016). The results in this study also show that on average the twins have almost 13 years of education. This indicates that a good number of the twins do not have very high educational qualifications and might probably have completed secondary level of education. In most countries, the economic returns to schooling are positive for every educational level and highest at the tertiary level, mostly because of the greater taxes and contributions governments receive from the higher salaries of more educated people (OECD, 2012). Access and quality at the tertiary level should therefore be expanded to ensure future high returns to schooling. Notwithstanding, policymakers should continue to invest in primary and secondary education so that no one will be left behind. For example, in 2017 Ghana introduced a free senior high school (SHS) policy to ensure general access to schooling for all children. This policy was primarily designed to remove financial obstacles to accessing secondary education especially among the poor. The economic returns to schooling can also

be used by governments and other agencies to guide macro-policy decisions about the organisation and financing of education reforms (Psachoropoulos & Patrinos, 2004). For instance, the United Kingdom, through its recent educational reforms, discontinued financing undergraduate university education in public universities and charged tuition fees to all students which hitherto was free and introduced a loan system that allowed students to (annually) borrow up to the fee amount and offered support to low-income students (Azmat & Simion, 2017). The return to an additional year of schooling is therefore a useful indicator of the productivity of education and an incentive for individuals to invest in their own human capital (Patrinos & Psacharopoulos, 2010). For instance, the Organisation for Economic Co-operation and Development (OECD) has adopted the returns to schooling as a key education indicator in their annual Education at a Glance series and other policy documents (OECD, 2001a).

**Conclusion**
The purpose of this paper was to compare the relative effectiveness of the fixed-effects model and the selection-effects model based on their ability to reduce or eliminate some of the unobserved differences that bias a cross-sectional comparison of education and earnings.
The return to schooling among a sample of twins in Ghana is between 8 and 11% when account is taken of genetic and family effects using either fixed-effects models or the selection effects model of Ashenfelter & Krueger respectively. Consequently, it is can be deduced that modeling the returns to education for twins' in Ghana using the fixed-effects and selection-effects models provides a consistent set of findings. The fixed-effects model was however, able to effectively reduce unobserved differences that bias the cross-sectional analysis of the returns to schooling by provid-

better fit to the twins' data than the selection effects model.  Thus, based on AICc estimates, the FE model was able to efficiently control for genetic and family effects in the returns to schooling estimates.

Using a linear regression approach, and controlling for other variables that have the potential to affect annual earnings in the labour market, indicates that family background variables such as the education of a sibling typically exert a small positive effect on earnings for DZ twins and virtually make no contribution to the returns to schooling for MZ twins, although this might probably be as a result of the small sample size. Subtracting the coefficient of a sibling's education from the own schooling coefficient is equivalent to a fixed-effects or within-family estimator. A review of the results suggested that although the fixed-effects model was able to effectively account for some of the unobserved differences that bias a cross-sectional comparison of education and earnings based on the point estimates, the selection-effects model through its lower standard errors produced efficient estimates.  However, taking into account the various measures of goodness of fit statistics, the fixed-effects model gave a better fit to the earnings-education equations.  Additionally, although the selection-effects procedure provides a less powerful statistical test than the fixed-effects model, it provides a check of the robustness of the more restrictive fixed-effects model. Thus, given the limited data, the fixed-effects model seemed to be superior to the selection-effects model. According to findings from this study, the fixed effects model can be considered a better model for estimating the returns to education since higher returns are obtained when genetic and family background factors are considered or taken into account.

## References

ASHENFELTER, O. & KRUEGER, A. (1994) "Estimates of the Return to Schooling for a New Sample of Twins." *American Economic Review* **84,** 1157 – 1173.

ASHENFELTER, O. & ROUSE, C. (1998) Income, Schooling and Ability: Evidence from a New Sample of Identical Twins. *Quarterly Journal of Economics* **113** (1), 253 – 284.

BEHRMAN, J., TAUBMAN, P. & WALES, T. J. (1977) Controlling for and Measuring the Effects of Genetics and Family Environment in Equations for Schooling and Labour Market Success. In: Taubman, Paul (Ed.), *Kinometrics*: The Determinants of Socioeconomic Success Within and Between Families. North-Holland, Amsterdam, pp. 35 – 96.

BONJOUR, D., CHERKAS, L., HASKEL, J., HAWKES, D. & SPECTOR, T. (2002) Estimating Returns to Education Using a New Sample of UK Twins. *Queen Mary and Westfield College, London*.

BOUND, J. & SOLON, G. (1999) Double trouble: on the value of twins-based estimation of the return to schooling. *Economics of Education Review* **18** (2), 169 – 182.

Broome, H. (2012) The Role of Unobserved Heterogeneity in the Returns to Education.

CANALS, C. (2017) Education and economic growth. In: Monthly Report Economic and Financial Market Outlook Number 412. May 2017. Dossier: EDUCATION: MORE CRUCIAL THAN EVER. *CaixaBank Research*.

CARD, D. (1999) "The Causal Effects of Education on Earnings". In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labour Economics, vol. 5. North-Holland, New York, pp. 1801 – 1863.

CASE, A. & YOGO, M. (1999) *Does school quality matter? Returns to education and the characteristics of schools in South Africa* (No. w7399). National bureau of economic research.

CHAMBERLAIN, G. (1982) Multivariate Regression Models for Panel Data. *Journal of Econometrics* **18**, 5 – 46.

DESCHENES, O. (2007) "Estimating the Effects of Family Background on the Returns to Schooling." *Journal of Business & Economic Statistics* **25** (3), 265 – 277.

GRILICHES, Z. (1977) Estimating the returns to schooling: Some econometric problems. Econometrica: *Journal of the Econometric Society*, 1 – 22.

HECKMAN, (2010) Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature* **48** (2), 356 – 398.

HERTZ, T. (2003) Upward bias in the estimated returns to education: Evidence from South Africa. *American Economic Review* **93** (4), 1354 – 1368.

HURVICH, C. M. & TSAI, C. L. (1989) "Regression and time series model selection in small samples", Biometrika, **76** (2), 297 – 307.

INDICATORS, OECD (2012) Education at a Glance 2016. Editions OECD.

KAYMAK, B. (2009) Ability bias and the rising education premium in the United States: A cohort-based analysis. *Journal of Human Capital* **3** (3), 224 – 267.

KRUEGER, A. B. & LINDAHL, M. (2001) Education for growth: Why and for whom? *Journal of economic literature* **39** (4), 1101 – 1136.

LEIGH, A. & RYAN, C. (2008) Estimating returns to education using different natural experiment techniques. *Economics of Education Review* **27** (2), 149 – 160.

MILLER, P., MULVEY, C. & MARTIN, N. (1995) ''What Do Twins Studies Reveal about the Economic Returns to Education? A Comparison of Australian and U.S. Findings'', *American Economic Review* **85** (3), 586 – 599.

MILLER, P., MULVEY, C. & MARTIN, N. (2006) The Return to Schooling: Estimates from a sample of young Australian twins. *Labour Economics* **13** (5), 571 – 587.

MINCER, J. (1974) Schooling, Experience and Earnings. Columbia University Press for National Bureau of Economic Research, New York.

MONTENEGRO, C. E. & PATRINOS, H. A. (2014) *Comparable estimates of returns to schooling around the world*. The World Bank.

NAKAMURO, M. & INUI, T. (2012) Estimating the returns to education using the sample of twins: the case of Japan. In *RIETI Discussion Paper Series 12-E-076*.

NEUMARK, D. (1999) Biases in twin estimates of the return to schooling. *Economics of Education Review* **18** (2), 143 – 148.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD) (2001a) *Education at a Glance*, Paris.

ORHAN, K. (2010) "Comparing two approaches to the *rate of return* to investment in education," *Education Economics, Taylor & Francis Journals* **18** (2), 153 – 165.

PATRINOS, H. A. (2016) Estimating the return to schooling using the Mincer equation. *IZA World of Labour*.

PATRINOS, H. & PSACHAROPOULOS, G. (2010) Returns to Education in Developing Countries. *International Encyclopedia of Education*. 305 – 312.

PSACHAROPOULOS, G. (1994) Returns to Investment in Education: A Global Update. *World Development* **22** (9), 1325 – 1343.

PSACHAROPOULOS, G. & PATRINOS, H. A. (2018) Returns to Investment in Education: A Decennial Review of the Global Literature. Policy Research Working Paper; No. 8402. World Bank, Washington, DC. © World Bank.

*https://openknowledge.worldbank.org/handle/10986/29672* License: CC BY 3.0 IGO.

Schwarz, G. (1978) "Estimating the Dimension of a Model," *Annals of Statistics* **6**, 461 – 464.

Statacorp. (2015) Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

Taubman, P. (1976) 'The Determinants of Earnings: Genetics, Family and other Environments: A Study of White Male Twins.' *American Economic Review* **66**, 858 – 870.

UNITED NATIONS INTERNATIONAL CHILDREN'S FUND (UNICEF) (2015) *The Investment Case for Education and Equity.* UNICEF.

Willis, R. J. & Rosen, S. (1979). "Education and Self-Selection," *Journal of Political Economy* **87**, S7-S36.

WORLD ECONOMIC FORUM (2016) Global Competitiveness Report 2015-2016. *http://reports.weforum.org/global-competitiveness-report-2015-2016/education/#view/fn-53*.

Zellner, A. A. (1962) An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association* **57**, 348 – 368.