UNIVERSITY OF CAPE COAST

ESTIMATING THE EFFECT OF EDUCATION ON EARNINGS IN GHANA: MODELS FOR TWINS

BY

ADELAIDE AGYEMAN

Thesis submitted to the Department of Mathematics and Statistics of the School of Physical Sciences, University of Cape Coast, in partial fulfilment of the requirements for the award of Doctor of Philosophy Degree in Statistics

DECEMBER 2012

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result of my own original work and that no part of it has been presented for another degree in this University or elsewhere.

Candidate's Signature:.. Date: 06/09/13

Candidate's Name: ADELAIDE AGYEMAN

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature:.................................. Date: 06/09/13

Principal Supervisor's Name: PROFESSOR N.N.N. NSOWAH-NUAMAH

Co-Supervisor's Signature:.. Date: 6/09/13

Co-Supervisor's Name: PROFESSOR (MRS.) N. MENSAH

ABSTRACT

This thesis estimates the economic returns to an additional year of schooling in Ghana using data from a twins' survey. Consequently, the relative importance of the roles of genetics and family background in determining earnings and returns to schooling in Ghana is examined. A number of models and estimation methods were utilized to illustrate the sensitivity of different estimators to model specification.

The results indicate that the economic return to schooling in Ghana using Mincer's Human Capital model is about 10%. Estimates of the economic returns to schooling using fixed effects and selection effects regression models and incorporating an instrumental variables approach to correct for measurement error in self-reported schooling levels was also assessed. The measurement error corrected return to schooling for monozygotic twins was larger than the standard ordinary least squares return to schooling estimate indicating a downward bias in the ordinary least squares return to schooling.

Finally, the Restricted Maximum Likelihood approach was adopted to identify unobservable differences in the returns to schooling for twins'. The analysis revealed significant unobservable differences ($p<0.05$) in the REML returns to schooling for dizygotic twins' whiles, unobservable differences in the REML returns to schooling for monozygotic twins was not significantly different from zero. The estimated "pure" rate of return to education in Ghana could therefore be used as an indicator for considering policies related to education.

ACKNOWLEDGEMENTS

The author would like to express appreciation to her academic supervisors, namely; Professor N.N.N. Nsowah-Nuamah and Professor (Mrs.) N. Mensah for their guidance and direction.

The author would also like to express her gratitude to Dr. E. K. Essel, Head of the Department of Mathematics and Statistics, University of Cape Coast and Professor K. Mensah, Dean, School of Physical Sciences, University of Cape Coast for their immense contribution to the academic development of the author.

Finally, the author wishes to express her appreciation to Dr. Hans Adu-Dapaah, Director, Crops Research Institute of the Council for Scientific and Industrial Research (CRI-CSIR) for his support for the development of this thesis.

DEDICATION

To my Family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

FCUBE	Free Compulsory Universal Basic Education
FGLS	Feasible Generalised Least Squares
GDP	Gross Domestic Product
GLSS	Ghana Living Standard Survey
GSS	Ghana Statistical Service
HCEF	Human Capital Earnings Function
HLM	Hierarchical Linear Model
ICC	Intraclass Correlation Coefficient
ICT	Information and Communication Technology
INHEA	International Network of Higher Education in Africa
IQ	Intelligence Quotient
IV	Instrumental Variable
JSS	Junior Secondary School
NAS-NRC	National Academy of Science - National Research Council
OECD	Organisation for Economic Co-Operation and Development
OLS	Ordinary Least Squares
REML	Restricted Maximum Likelihood
SSA	Sub-Saharan Africa
SURE	Seemingly Unrelated Regression Equations
WAEC	West African Examination Council
WASSCE	West African Senior Secondary Certificate Examinations

CHAPTER ONE

INTRODUCTION

1.1 Background

After fifty years of independence from colonial rule, education in Ghana is still undergoing reforms and is therefore one of the central topics in the public policy debate. Ghana like most West African countries, struggles to provide its citizens with even basic educational facilities. The adult illiteracy rate in Ghana is over 40% (Population Census, 2000) and due to the restrictions of poverty and the familial necessity of child's assistance around the home or farm,the majority of children are restricted from attending lower level schools. Education in Ghana plays a central role in modern labor markets and one of the most important economic decisions that individuals and policy-makers have to face is how much to invest in education.

A number of studies in Ghana and in many different countries and time periods have confirmed that better-educated individuals earn higher wages, experience less unemployment, and work in more prestigious occupations than their less-educated counterparts (Card, 1999). In the absence of experimental evidence, it is very difficult to know whether the higher earnings observed for better-educated workers are caused by their higher education, or whether individuals with greater earning capacity have chosen to acquire more schooling.

Findings over the years in Ghana have established a strong relationship between schooling and income (Glewwe, 1996; Kingdon and Soderbom, 2007 and Sackey, 2008). The relationship between schooling and income is however influenced by the effect of unobserved factors (such as measures of intellectual ability, family background etc.). However, one of the major problems affecting the development of a relationship between wages and schooling is the lack of rich data sets that can be used to control more extensively for measures of intelligence, family background, etc. Consequently, previous studies have omitted ability due to the complexities in its assessment. In order to address this problem, (Card, 1999; Ashenfelter and Krueger, 1994; Griliches, 1977 and 1979) proposed the use of data on schooling and wage variation between identical twins as a powerful tool to assess the variation between wages and schooling. However, such an approach is largely unexplored because of the difficulty in controlling for genetics and family environment when studying the effects of schooling on earnings, and that failing to do so may cause a large bias, up to two-thirds of the non-controlled coefficient (Taubman, 1976).

It is widely realized that an increasingly complex society and rapid technical change requires highly educated workforce, if the country wishes to succeed in the international competition. Interestingly enough, most of the arguments in this debate are cast in economic terms (Conneely and Uusitalo, 1998). To estimate the economic returns to schooling, omitted ability has been found to introduce bias in ordinary least squares estimates. Identical twins and siblings studies have however, been utilised to control for ability bias in estimating the impact of schooling on income in recent studies. This is based on the presumption that omitted ability is entirely made up of a genetic effect and a family effect which therefore disappears with differencing between family members with the same genes.

The Education System in Ghana

Currently, the Ghanaian education system consists of pre-school, primary school, secondary school and higher learning institutions. The main purpose of pre-school is to provide a basic education for young children before they go on to formal education. The objectives of pre-school education are to provide a home substitute for young children and offer them opportunities for overall personal development, provide opportunities for holistic development of the child through organized individual and group play activities, create awareness in the children of their national heritage and culture, pre-dispose the child to conditions of formal education in order to accelerate the learning process during formal schooling and, finally lay a solid foundation for all-round learning (Ministry of Education, 2001). Pre-school education begins at the age of 3 or 4 at a government kindergarten, a non-government agency or a private sector kindergarten. This level of education covers the ages of 2-6 years. It is made of 1 1/2 to 2 years Nursery and 1 1/2 to 2 years Kindergarten, which together constitute pre-school education. It is not compulsory and is mainly enjoyed by urban children.

Basic Education starts at age six and continues for nine years. The nine-year basic Education programme is made up of six years Primary Education and three years Junior Secondary Education. Primary Education constitutes the foundation of the educational system. During the six years, students will acquire numeracy and literacy skills (i.e. the ability to count, use numbers, read, write and communicate effectively, lay the foundation for inquiry and creativity, develop a sound moral attitude and a healthy appreciation of Ghana's cultural heritage and identity, develop the ability to adapt constructively to a changing environment, lay the foundation for the development of manipulative and life skills that will prepare the individual to function effectively to their own advantage as well as that of their community and inculcate in good citizenship education as a basic for effective participation in national development. The Junior Secondary School forms an integral part of compulsory Basic Education. It is both terminal and continuing. The curriculum of the junior Secondary School has been reviewed and expanded to include practical skills orientation. Consequently, in addition to the general subjects, the curriculum has been designed to provide opportunities for students to acquire basic pre-technical, pre-vocational and basic life skills which will enable the pupils:

1. To discover their aptitudes and potentialities so as to induce in them the desire for self improvement;
2. To appreciate the use of the hand as well as the mind and make them creative.

Over the nine years of basic education, students are assessed by continuous school-based assessment until at the end of Year nine they experience the first National Examination known as the Basic Education Certificate Examination (B.E.C.E.) to evaluate their performance. More than 35 percent of the students who complete the Junior Secondary Schools enter second-cycle institutions, of which there are two kinds, namely:

1. Senior Secondary Schools (public and private).
2. Technical Institutes (public and private).

Education at the Senior Secondary School level is designed to cater for students of ages 16 to 18 years and lasts for three years after the nine years of Basic Education. As part of the educational reforms, the Senior Secondary School has been designed to offer the students the opportunity to build on the foundation laid at the Basic Education level and to strengthen the general intellectual knowledge and skills that are required for occupations and for further education. The three-year programme is to cater for various aptitudes in the fields of technical, vocational, agricultural, business and general education.

A curriculum relevant to the socio-economic development and manpower requirements of the country has been developed for the Senior Secondary School. The Senior Secondary School system has replaced the traditional 7-year (i.e. 5-year Secondary + 2-year Sixth Form) course. It has the following objectives:

1. To reinforce and build on knowledge, skills and attitudes acquired at the Junior Secondary School level;
2. To produce well-developed and productive individuals equipped with the qualities of responsible leadership capable of fitting into a scientific and technological world and to contribute to the socio-economic development of their own areas and the country as a whole;
3. To increase the relevance of the content of the curriculum to the culture and socio-economic problems of the country.

Students sit for the West African Senior Secondary Certificate Examinations (WASSCE) after completing secondary education. WASSCE is a type of standardized test in West Africa. It is administered by the West African Examinations Council (WAEC). It is only offered to candidates residing in Anglophone West African countries.

Successful candidates from the Senior Secondary Schools are evaluated for admission to the various departments of the national universities. The system of education in Ghana now consists of up to 12 years of pre-university and 4 years of university education. The attainment of university education is the ultimate goal of most Ghanaian students. However, the nation's five universities are able to admit only a small fraction of qualified applicants because of limited facilities and faculty. Besides university education, the nation provides opportunities for public higher education through other avenues. For example, there are 7 diploma-granting institutions, 21 technical colleges, 6 polytechnics, and 38 teacher training colleges. In addition, a number of private computer-training schools have opened at the major urban centers in the country. This education system of Ghana has worked as a result of a series of reforms.

1.2 Statement of the Problem

Education in Ghana is one of the key factors in promoting economic development. Investments in education may be more profitable than other types of investments and as such education is promoted as a means of improvement in productivity and economic growth. Investments in education in Ghana therefore can be judged in terms of economic rates of return. Consequently, a number of studies (Shultz, 2003; Glewwe, 1996) estimating the returns to education in Ghana have established that the better educated workers tend to have higher incomes and less poverty than the less educated. The average Ghanaian worker receives a 7.1% pay increase for each additional year of education acquired (Hall and Jones, 1999). At the aggregate level, Lau, et al (1991) estimate that a 1-year increase in the average education level of the adult can lead to increases of 3-5% in real GDP in Psacharopoulos, 1985, 1994; Sackey, 2008; Kingdon and Söderbom, 2008 Sub-Saharan Africa. Human capital theory provides a methodology for estimating economic rates of return to education (Mincer, 1974; Becker, 1975). Application of this methodology to Ghana has produced apparent high rates of return, which are often cited as evidence for further investments in education, particularly primary education (Asafu-Adjaye, 2013; Psacharopoulos, 2004; Schultz, 2003). However, a central concern in such a methodology, relates to its estimation using least squares regression. When estimating the returns to education by ordinary least squares (OLS), there are two potential sources of bias due to unobservable characteristics and factors, such as a person's ability that would certainly be correlated with the observable characteristics, such as schooling, resulting in biased return to schooling estimates for Ghana.

First, there is the difficulty of extracting education's effect on income. That is, a worker's natural ability, his family background, and his innate intelligence are all possible confounding factors that must be controlled for in order to estimate the effect of education on income accurately. Thus if individuals with high absolute earnings capacity both acquire more education and earn higher wages, schooling will be positively correlated with the presence of an unobservable factor and yet also correlated with wages. This ability bias induces an upward bias in the estimated average return (Griliches, 1977). The second difficulty in measuring the effect of income on education has to do with the false reporting of education levels. This measurement error in the schooling variable induces a downward bias in the case of classical measurement error (Griliches, 1977; Blackburn and David Neumark, 1995). Thus, the question of whether the estimated returns to education for Ghana reflect the true productivity-enhancing effects of education or whether they reflect some other unobserved factors like ability or family background. There is a need therefore to control for these biases caused by endogeneity of schooling in the earnings equation which will allow for a more accurate description of the true effect of additional schooling on earnings for Ghana, thereby providing reliable information for economists and policy makers.

A number of approaches to deal with this problem have been proposed. Recent work have recommended an approach using twins (or siblings) in order to eliminate endogeneity bias though majority of the literature on the return to schooling employs instrumental variables (IV) to handle the endogeneity of schooling problems. Differences between twins in levels of schooling and earnings are exploited based on the fact that this eliminates differences in innate ability or motivation (Taubman, 1976; Ashenfelter and Krueger, 1994). Under the key identifying assumption that ability is common among siblings (particularly monozygotic (MZ) twin pairs) consistent estimates are obtained as long as problems of measurement errors in the schooling variable can be dealt with adequately. Notwithstanding, endogeniety bias from measurement error is also likely to be greater in any method that identifies the return to education from differences in education. In order to address such errors, Ashenfelter and Krueger, (1994) recommend that the potential for measurement-error-induced bias can be reduced by instrumenting the education of one twin with an estimate obtained from responses from the other twin.

Another approach by Ashenfelter and Krueger (1994) and Ashenfelter and Rouse (1998) provide estimates of the returns to education and the resulting endogeneity bias (to which they refer as a "selection effect"). The model of optimal schooling choices that they used suggests that measures of the education of a twin's sibling, the average education of the twins, or father's education could be employed as an additional regressor to control for any "family" effect that affect the absolute level of earnings. The selection effects model of Ashenfelter and Krueger (1994) allows a direct assessment of the magnitude of these effects (ability and shared family characteristics) through an explicit modeling of the family effects factor.

Furthermore, estimation of returns to education using Ordinary Least Squares (OLS) disregards variation in the returns for workers in the same education group. Knowing the extent individual-specific rates of return vary will be important to policy makers and economists in Ghana, not only because it gives an indication of the benefits of schooling individuals accrue, but because rates of return give an indication of supply and demand shifts in the labor market

(Freeman, 1977; Card and Lemieux, 2001); and because rates of return have implications for technological change (Goldin and Katz, 2008). Significant variation in returns of Ghanaian workers with higher returns for those with higher levels of income (assumed to indicate high ability individuals), and investment in education will generate more inequality. This could challenge the conventional view of investment in education, which is that education promotes equality in the long run, other things being equal.

In order to design effective economic policies for Ghana, it is important to have an idea about the individual economic benefits from education and in the variance in the return across individuals and families. This is because the monetary benefits of an additional year of schooling vary largely across the population due to heterogeneity in the returns to schooling. Variation in the returns to schooling is related to individual (ability) and family background differences and may help to explain why returns to schooling differ across individuals. Roy (1951); Willis and Rosen (1979), and Willis (1986), views human capitals as heterogeneous multidimensional attributes, and people choose their educational attainment based on the comparative advantage of their different attributes of abilities. The varying returns imply a random coefficient model of earnings determination based on the restricted maximum likelihood estimates.

1.3 Focus of this Thesis

The main focus in this thesis lies with the accurate estimate of the impact of an additional education on earnings in Ghana. However, there are some key issues in a study on the relationship between education and earnings which must be effectively addressed.

The key problems identified with the estimation of the economic returns to schooling are endogeneity bias, ability bias and measurement error. These three well-known arguments explain why OLS may render inconsistent return estimates. In order to eliminate or reduce these biases and obtain the precise rate of return to education, this study seeks to use a new survey of twins and siblings.

Twins share common or similar genes and, to a large extent, common family background and by relating within-twin-pair differences in education to within-twin-pair differences in earnings, the study is able to difference out the influence of unobserved genetic traits and common family background that may otherwise bias the schooling coefficient. This study uses the Twins/Sibling methods to deal with unobserved abilities since socioeconomic background, genetic traits and, to some extent personal characteristics are more likely to be similar between siblings than between randomly selected individuals (Griliches, 1979). According to Behrman and Rosenzweig (1999) the magnitude of the ability bias can be estimated by comparing the effect of education on earnings from a sample of randomly selected individuals with a sample of MZ twins. In the wage returns to education literature, a number of studies use twin methods to be able to control for and estimate a potential ability bias (Behrman and Taubman (1976), Taubman (1976), Ashenfelter and Krueger (1994) and Bonjour *et al.* (2003). As these studies are closely related to our choice of subject, we find twin methods to be useful also to examine the returns to education in Ghana.

In order to effectively address the key issues in a study on the relationship between education and earnings using twins, this thesis focuses on three main areas, namely:

1. Modeling returns to education in Ghana using twins.
2. Endogeneity of education.
3. Heterogeneity in the returns to schooling.

1.3.1 Modeling returns to education in Ghana using twins

The first area of focus in this thesis is a straightforward attempt to estimate the rate of return to years of education in Ghana using statistical and econometric models. The study of returns to education has a long tradition in labor economics and this tradition is based on standard human capital theory. Human capital refers to the stock of skills and knowledge relevant to performing labor to produce economic value. It is the skills and knowledge gained through education and experience that was first defined as such by Adam Smith (1776). Thus, schooling is viewed as an investment in human capital (Mincer, 1958; Becker, 1964), implying that the returns to schooling may be measured in terms of the extra income due to additional schooling. In modeling the returns to education in Ghana, this study first adopts Mincer's human capital returns to education approach which is line with traditional mainstream empirical human capital research.

There are circumstances where the only estimates of the Mincerian return to schooling available are obtained using standard statistical techniques. It is therefore important to understand whether estimates of the Mincerian return to schooling obtained with least-squares techniques are systematically different from estimates relying on twins or an IV approach.

The growing literature on this issue suggests that, overall, the estimates obtained using twins or an IV approach are somewhat larger than estimates using least-squares techniques Ashenfelter, Harmon and Oosterbeek (1999). Hence, the question of whether these differences are significant in the returns to education in Ghana is analyzed in this study.

The key idea behind the strategy of studying the relationship between education and earnings for twins/siblings is that some of the unobserved differences that bias a cross-sectional comparison of education and earnings are reduced or eliminated within families. Recent analyses of data on twins has produced interesting insights into the roles of genetics and family background as mediating influences in the relationship between schooling and income (Behrman et al. 1977, 1994; Ashenfelter and Krueger 1994; Miller et al. 1995). Although the findings of the various studies have not been unanimous, all but Ashenfelter and Krueger (1994) ascribe some role to the influence of family background. Important studies of the return to education using US twins include Ashenfelter and Krueger (1994), Ashenfelter and Rouse (1998), Behrman *et al.*, (1994).

The major issues encountered in the estimation of the return to schooling using Mincer's human capital model are potential biases in the estimates caused by measurement errors in education, ability bias and the endogeneity of schooling. However, while a twin design have some distinct advantages, it also brings problems of its own. Griliches (1979) cautions that twins are “not a panacea”, for a number of good reasons and renewed reservations about the use of twins for estimating the returns to education have been expressed in Bound and Solon (1999). Consequently, important issues that make twins potentially problematic for estimating the return to education on earnings are measurement error and endogeneity of differences in schooling.

This study employs twins-based estimates to control for individual ability differences (omitted ability bias) in the returns to education. This approach is in line with recent approaches for correcting potential biases in the return to education estimates, which include estimating earnings functions from differences within twins or siblings. It is also worth noting that within-twin differences of the return to schooling as used in this study hold out the promise of eliminating unobservable ability and family effect, which causes the omitted variable bias in the OLS estimation. Earlier studies relied heavily on test scores in an attempt to remove ability bias from the return to schooling estimates. Although the rate of return to education varies significantly in response to various influencing factors, the average estimate for developed economies generally ranges from 5% to 10% (Wilson, 2001).

Generally, it was found that failing to account for (pre-school) ability differences leads to an overestimation of the return to schooling. This conclusion was largely refuted by a number of studies in the 1990's that relied on various natural experiments and instrumental variable techniques (Angrist and Krueger, 1991; Card, 1995; Harmon and Walker, 1995 and Conneely and Uusitalo, 1997).

The instrumental variable estimates were systematically, though often insignificantly, higher than comparable ordinary least square (OLS) estimates. Until just a few years ago this empirical evidence was limited to the United States of America (US) data but in recent years, several studies have appeared in the United Kingdom (Harmon and Walker 1995; Dearden 1995), Sweden (Meghir and Palme 1997), Australia (Miller, Mulvey and Martin 1995) and Netherlands (Levin 1997). The instrumental variable estimates in these studies were quite similar to the US findings.

Regarding measurement error, Griliches (1979) notes that the use of estimates obtained from differencing in general, and differencing within twins in particular, exacerbates measurement error in schooling and so increases the tendency for estimates to be attenuated (i.e. biased towards zero) because of this larger measurement error. The solution to a pure measurement error problem is to use a second measure of the variable that is measured with error. Based on the assumption that the measurement error is classical (i.e. that the errors are independent of the truth), and that the two measures are correlated, the second measure can be used as an instrument for the first. Thus, one way to solve the problem of measurement error bias is to use the instrumental variable method. Ashenfelter and Krueger (1994) asked each twin about their own schooling and their co-twin's schooling. The difference in twin cross-reported schooling is used as an instrument for difference in self-reported schooling. This innovation has largely been responsible for the subsequent revival of the use of twins to estimate the returns to schooling. In this study, we follow the innovative approach of

Ashenfelter and Krueger (1994) to obtain good instrumental variables. More specifically, in the survey for this study, each twin reported both their own education and their co-twin's level of education. In the presence of measurement error in self-reported education, cross reported education is a potential instrument, as the report of the other twin should be correlated with the true education level of a twin but uncorrelated with any measurement error that might be contained in the self-report.

The empirical estimates show that, accounting for measurement error, endogeneity and ability differences, the estimates for the return to additional years of schooling are between 11 and 13% (Uusitalo et. al, 1999). This study adds one more piece to this accumulating international evidence by examining the potential biases in OLS estimates of the returns to schooling and suggests alternative methods to correct them. Thus, the positive ability bias in the ordinary least squares estimates is more than offset by a negative bias caused by endogeneity or measurement error.

1.3.2 Endogeneity of education

The second key area of focus in this study is addressing the potential endogeneity of schooling associated with the estimation of the return to schooling in Ghana. The problem of endogeneity arises if individual unobserved traits, which are in the error term, are systematically correlated with both included independent variables (e.g., education) and the dependent variable (earnings) in a regression model. This implies that the regression coefficient in an Ordinary Least Squares (OLS) regression is biased. For instance, if individual ability is positively correlated with both education and earnings, then any positive coefficient on education in the earnings function may simply reflect the cross-section correlation between ability, on the one hand, and both education and earnings, on the other. While it is unlikely that unobserved traits are identical across family members, it is likely that they are much more similar within a family than across families and, as such, family fixed effects estimation gives an estimate of the return to education that reduces endogeneity bias without necessarily eliminating it entirely. A recent solution to this endogeneity problem has been found in identifying exogenous sources of variation in schooling to build a new set of instrumental variables for years of education attained (Angrist and Krueger 1991; Card 1998). A twin-differencing strategy (Miller et al, 2006; Ashenfelter and Krueger, 1994; Bingley et al., 2005) which relies on the existence of differences in schooling within identical twin pairs have also been discovered to be helpful in overcoming the endogenity of schooling problem.

However, a major criticism has recently been leveled at twin-based methods by Neumark (1999), as well as by Bound and Solon (1999). Building on earlier work by Griliches (1979), they argue that whilst within-pair differencing removes genetic variation, differences might still reflect ability bias to the extent that ability is affected by more than just genes. In other words, within-twin-pair estimation may not completely eliminate the bias of conventional cross-sectional estimation, because the within twin-pair difference in ability may remain in the differences in the error, which may be correlated with the differences in observed individual variables that affect earnings, which includes education, age, age squared, gender, marital status, and job tenure between the twins. If endogenous variation in education comprises as large a proportion of the remaining within-twin-pair variation as it does of the cross-sectional variation, then within twin-pair estimation is subject to as large an endogeneity bias as cross-sectional estimation. In this event, IV estimates to correct for measurement error in reported schooling may exacerbate upward omitted ability bias in the estimated education effect (Bound & Solon, 1999; Neumark, 1999).

Consequently, in addition to a family fixed effects regression of earnings and instrumental variables regression, the selection effects model proposed by Ashenfelter and Krueger, (1994) are employed to estimate the returns to education in order to address the problem of endogeneity of within-twin pair schooling differences in this study. The Hausman (1978) specification test has also been used to measure the extent of the endogeneity bias.

1.3.2 Heterogeneity in returns to education

The third area of focus in this thesis covers the issue of whether heterogeneity exists in the returns to schooling relationship and, if so, what is the best way to model that heterogeneity. Differing abilities alter returns to education so that there exists a family of returns to education. This is precisely what we mean by heterogeneity in the returns to schooling. Empirical results indicate that heterogeneity is present in returns to education. Previous studies have found strong evidence that the heterogeneity follows a continuous rather than discrete distribution, and that bivariate normality provides a very reasonable description of individual-level heterogeneity in intercepts and returns to schooling.

The major issue is that unobserved ability induces heterogeneity in the distribution of earnings conditional on education through its effect on both the intercept and the education coefficient. In this case the labor market cannot be well characterized by a single rate of return to education. This study follows the idea of heterogeneous returns to education in a random coefficient model and finds out whether, there are unobservable differences in returns by employing multilevel modeling to estimate a mixed model. This essentially estimates a random coefficient (on education) model and decomposes the variance around the mean return into family heterogeneity, individual heterogeneity, and luck or risk.

Estimation of a random coefficient model allows one to control for unobserved heterogeneity at the cluster level (twins as individuals clustering in families). Secondly, the solution to cluster confounding satisfies the controversial statistical assumption associated with the "random effects" (RE) approach that level-1 independent variables be uncorrelated with the random effects term. Thirdly, unlike the fixed effects (FE) approach, the proposed method allows for the inclusion of level-2 variables (panel data), thus not limiting the types of hypotheses one can test. And fourthly, the method allows for statistical tests of cluster confounding, i.e., whether differences between within- and between-cluster effects are statistically significant.

The model has an individual-specific intercept and slope that may depend on observable variables and unobservable heterogeneity. The heterogeneity components capture influences from gender, family background, age, preferences, ability, etc. We are interested in estimating the heterogeneous effects of schooling on log of earnings in the random coefficient model. In this model, the return to education varies across individuals in both, observable heterogeneity in returns and unobserved individual-specific returns to schooling. Hence, there is no single parameter for the return to schooling, i.e. there is a distribution of effects across individuals. Heterogeneity is incorporated by allowing the intercept and slope of this linear relationship to vary across individuals but impose a degree of similarity across individuals by assuming that effects are drawn from the same normal population. We also estimate other more general specifications for the heterogeneity distribution, and introduce explanatory variables into the second stage of our hierarchical model to make out if observable characteristics might help to explain the unobserved heterogeneity across individuals.

Traditionally, unobserved heterogeneity enters exclusively the intercept of the wage equation but not the slope coefficient (Gebel and Pfeiffer, 2007). One appealing feature of the random coefficient model is that variation in unobserved heterogeneity affects the slope as well, i.e. unobserved heterogeneity influences the wage effect of education. Such hierarchical or "random effects" models have been suggested by or employed in past work in the schooling literature by Becker and Chiswick (1966) and Chiswick (1974), among others. More recently, theoretical issues in an elaborated version of this model were described in Heckman and Vytlacil (1998).

Recent studies of the association between schooling and earnings have emphasized the heterogeneity in the economic return to an additional year of education across otherwise comparable individuals (Card and Krueger (1992), Heckman, Layne-Farrar, and Todd (1996); Altonji and Dunn (1996), and Ashenfelter and Rouse (1998). Heckman and Vytlacil (1998) and Card (1999) discuss theoretical models of heterogeneous returns to education). Despite increased attention to the possibility of heterogeneous returns to education across individuals, there is still considerable uncertainty about the mechanism generating this heterogeneity. Part of this uncertainty is attributable to the absence of a formal model that explicitly recognizes the possibility that the return to schooling varies with observable characteristics, like family background variables.

The mixed model emphasizes the potential role of unobservable ability influencing both schooling and earnings and concludes that even though, returns to schooling of genetically identical individuals should be the same, we find that some statistical tests associated with the model provide little evidence not consistent with this hypothesis. The results obtained in this study suggest that heterogeneity in returns to education for identical twins are negligible but significant for fraternal twins and thus provide a further contribution to the literature on returns to schooling.

An assumption made in most empirical studies when estimating the standard Mincerian wage equation is that the return to schooling is homogenous (i.e., returns are constant across individuals). This assumption ignores the heterogeneity that is found in returns to education, though observed and unobserved factors can lead to heterogeneity in returns, (i.e. returns vary across individuals) which may influence the effect of education. Thus, it follows that there is no single effect of education but rather a whole distribution of individual effects (Blundell et al., 2005; Heckman et al., 2006). The varying returns imply a random coefficient model of earnings determination. This study empirically examines these effects by allowing the coefficients (intercept and slope of this linear relationship) of the random coefficient model to vary across individuals.

Heterogeneity refers to differences in the returns to education across individuals due to factors unobservable to the econometrician, but known to the individual at the time of their decisions. More generally each individual faces a distribution of returns that is conditional (in its mean and possibly higher moments) on individual characteristics observable to them, but some of which are not observable by the econometrician. Thus, even if each individual were to receive their expected return to education with absolute certainty, returns would still appear to vary randomly across the population. Ability and family background variables may help to explain baseline differences in earnings across individuals, and may also help to explain why returns to schooling differ across individuals.

A two-level hierarchical linear model is used to model the unobservable differences in the returns to schooling in this study. The hierarchy in the data is described by the observations of siblings (the individual level) within families. While similar to ordinary regression analysis with interaction terms added, the hierarchical model differs in that it allows for a more complex error structure to be analyzed. In the hierarchical model, stochastic terms are included at both levels to account for random variation in the returns to schooling. The empirical results suggest that there exists significant heterogeneity in the data, which is reflected in variation in the estimated intercept and slope coefficients.

1.4 Justification of Study

The main purpose of this study is to provide a consistent estimate of the return to education and in so doing report new estimates of the economic returns to education using data on twins. Twin studies help unravel the relative importance of genetic and family background influences in the relationship between schooling and earnings.This study is of interest for five main reasons. First, given the interest in genetics and economic success (Richard J. Herrnstein and Charles Murray, 1994), data on genetically identical individuals are of particular value. Second, while there are many earnings/education studies, there are comparatively few based on identical twins.

Thus we contribute to this literature by conducting a systematic investigation on the returns to education in Ghana using twins' data. In this regard, the study aims at addressing the constraints of earlier education and schooling models in Ghana by controlling for endogeneity of schooling since our data set provides information on twins or siblings which also include information on variables that can be used in the analysis of schooling attainment function such as family background (e.g., parental education).

Third, our study is the first for Ghana to present within-twin-pair estimates using identical twins and fraternal twins. Some studies have been done on the returns to schooling in Ghana but none so far have used twins data in their estimation. Kingdon and Söderbom (2007) investigate the education-earnings relationship in Ghana, drawing on the Ghana Living Standards Survey for 1998-99. Fourth, we have followed Ashenfelter and Krueger's (1994) innovation of asking one twin to report on the schooling of the other, in order to examine possible measurement error. Though some criticisms of within-twin-pair estimates have been set out by (Bound and Solon, 1999; Neumark, 1999), indicating that within-twin estimation cannot completely eliminate the bias of the OLS estimator it can help tighten the upper bound on the return to education.

This study is also important because recent studies of the association between schooling and earnings have emphasized the heterogeneity in the economic return to an additional year of education across otherwise comparable individuals (Card and Krueger, 1992; Heckman *et al.*, 1996; Altonji and Dunn, 1996; Ashenfelter and Rouse, 1998; Heckman and Vytlacil, 1998; and Card, 1999). In spite of the increased attention to the possibility of heterogeneous returns to education across individuals, there is still considerable uncertainty about the mechanism generating this heterogeneity. Part of this uncertainty is attributable to the absence of a formal model that explicitly recognizes the possibility that the return to schooling varies with observable characteristics, like family background variables.

1.5 Objectives of the Study

This study is basically planned to evaluate the nature of returns to higher education using twins data in Ghana with a view to controlling for genetics and family background effects in the rate of returns to education. Specifically, the study has the following objectives:

1. To determine the socio-demographic characteristics of twins in Ghana.
2. To estimate the rate of return to education in Ghana for an additional year of schooling.
3. To investigate the potential role of unobservable ability in the determination of income.
4. To determine heterogeneity in returns to education by adopting hierarchical linear model regression framework.
5. To determine whether neglected endogeneity imparts serious biases to the returns to education estimates in Ghana by adopting an instrumental variable approach and the selection-effects model.
6. To recommend policy actions to decision makers in Ghana.

1.6 Research Questions

In an attempt to achieve the stated objectives, the study addresses the following research questions:

1. What is the economic return to an additional year of schooling?
2. Is endogeneity of schooling an important issue in the estimation of the returns to schooling?
3. Is there evidence of individual heterogeneity in returns to education?
4. What are the roles of genetics and family background in the relationship between education and earnings?

1.7 Significance of Study

In addition to identifying a "pure" return to education in Ghana (i.e., controlling for the effects of family background in the returns to education using twins data), this study will also be helpful in informing policy makers and educationists on the importance of upgrading the education and skill levels of the work force in Ghana. This study is therefore of interest for four main reasons, namely:

Firstly, in evaluating the effect of education on earnings the study will seek to find out whether the rates of return to education provided by Shultz (2003) still holds for the labour market conditions prevailing in Ghana when ability and family background effects are accounted for.

Secondly, twin-based estimates for the return to education will serve as a reliable guide for designing educational policies in Ghana whilst considering the enormous cost of running education. Consequently, beneficiaries of education, especially higher education, could be made to pay at least a portion of the cost of educating themselves, centering on the principle that the returns (private) to education increase as a result of higher levels of educational attainment (Psacharopolous, 1994). In addition, the government of Ghana could also support private schools (e.g., Universities) and non-governmental organizations to be involved in the quest to expand education in Ghana.

Thirdly, as this study will identify the highest level of education attained by majority of Ghanaians, government subsidies or credit systems should be introduced if this level of education is low to encourage people to attain higher levels of education as a more educated, better-trained person is capable of supplying a larger amount of useful productive effort than one with less education and training (McConnell et. al., 2006).

Finally, the estimated "pure" rate of return to education in Ghana could be used as an indicator for:

1. Guiding policy makers advocating the use of educational services as part of the plan for poverty alleviation;
2. Suggesting strategies for reducing the incidence of poverty in Ghana.

1.8 Limitations of the Study

In this study, data on the examination performance and literacy test scores of twins were not collected and therefore ability tests scores were not included in the earnings equations as an observed variable. Some studies have found out that ability test scores have a strong effect on the choice of education and on subsequent earnings (Conneely and Uusitalo, 1998). Sandewall et al., 2009, also observed that within-pair differences in Intelligence Quotient (IQ) have a statistically and economically significant effect on within-pair differences in schooling and inclusion of IQ reduces within-pair estimates of returns to schooling by about 15% across various specifications.

Secondly, the effect of school quality on educational attainment and ultimately earnings was also not considered. School quality is hard to define and measure. It is influenced by not only school expenditures, but also characteristics that are hard to measure like norms, attitudes and peer effects among teachers and pupils (Hoxby, 2000). The Coleman Report in 1966 (Coleman et al., 1966) found that measured school quality had very little effect on pupil achievement once family background and school composition effects had been taken into account. The subsequent U.S. literature looking at this issue has, on the whole, tended to confirm this somewhat surprising finding, or at best found only weak effects of school quality on pupil achievement, Hanushek, (1986) and Hanushek, Rivkin, and Taylor, (1996).

Finally, this study does not address the nonmonetary benefits of education. Education provides benefits in such areas as personal growth, occupational choice, longer lifespan, better health, and benefits related to consumption and savings (Vila, 2000). Higher wages, higher income and other purely economic benefits are an important part of the individual benefits to education.

CHAPTER TWO

LITERATURE REVIEW

2.1 Schooling Reforms in Ghana

Ghana has since independence made significant strides in its education system. The educational system in Ghana is undergoing a slow but consistent reform process. The government's focus lay in expanding primary education and increasing teacher training, with positive results in regard to enrollment. On average it takes about 20 years for a child to complete their education in Ghana. Children from wealthy families usually benefit from attending private schools while children who are from poor families attend public schools.

The Education Reform Programme introduced in 1987/88 and the free Compulsory Universal Basic Education (FCUBE) 1996 programme, have contributed immensely to the structure of Basic Education that we have today and the achievements so far made. A major motivation for the FCUBE was the recognition of unsatisfactory enrolment rates for children at the primary and junior secondary school levels. Recent evidence seems to suggest an increase in demand for schooling by Ghanaians, and at the tertiary level, there are even concerns about supply constraints such as infrastructure problems and absorption capacities. In 2006/2007 academic year, the total enrolment in both the public and private crèches and nurseries was 184,574 while enrolment in public primary schools increased to 3,365,762, showing an increase of 7.8% when compared to the 2005/2006 enrolment of 3,122,903, (Ministry of Education, Science and

Sports Statistics, 2007). In addition, the total enrolment in both the public and private junior secondary schools was 1,132,318, an increase of 8.8% when related to the 2005/2006 enrolment of 1,041,002. Between 1990 and 2000, the higher education sector in Ghana which is composed of 6 public and 30 private universities and 10 polytechnic institutions recorded a 162 percentage increase in total enrolment (i.e. , from 13,415 to 44,389 students), Gadzekpo, (2008). However, much of this growth was due to dramatic increases in enrolments at the polytechnics (INHEA, 2006). The implementation of the education reform programme in 1987 reduced the pre-tertiary years of schooling from 17 to 12. The reform involved a phasing out of the existing middle and secondary schooling (i.e., ordinary and advanced levels) components. The reform programme also saw a phasing in of a three-year junior secondary schooling and a three-year senior secondary school component. This reform programme, in a way, synchronized the pre-tertiary years of schooling in both the public and private sectors of education. Although, The 1987 Education Reform Programme succeeded in solving some of the problems confronting the sector, including the reduction of the duration of pre-tertiary education from 17 to 12 years and expanding access to education, the sector was however, still beset with a number of problems. These included, poor quality teaching and learning, weak management capacity at all levels to the educational system and inadequate access to education. The FCUBE educational reform was therefore designed and implemented for a ten year period (1996-2005) to address some of the shortcomings of the educational reforms. The main objectives were to:

1. Expand access to good quality basic education.
2. Promote efficient teaching and learning.
3. Improve teacher moral and motivation through incentive programmes.
4. Ensure adequate and timely supply of teaching and learning to schools.
5. Improve teacher community relations.

In spite of the fact that Ghana's education system has come far and made the nation what it is today, the increasing challenges of the twenty-first century demand an improvement of the education system to make it more responsive to national goals and aspirations as well as global demands. As a result of this demand, the new 2007 educational reform was initiated to strengthen the existing education system. The salient highlights of this reform were:

1. A universal basic education system which includes two years kindergarten.
2. At the basic level, emphasis shall be on literacy, numeracy, creative arts and problem solving skill.
3. Secondary education extended to four years.
4. Teacher Training Colleges will be upgraded and conditions of service for teachers improved, with special incentives for teachers in rural areas.
5. Greater emphasis will be put on Information and Communication Technology (ICT) and Science and Technology.

A major motivation for the FCUBE was the recognition of unsatisfactory enrolment rates for children at the primary and junior secondary school levels (Sackey, 2008). In particular, the low enrolment rates for girls put them in a disadvantaged position – their chances of getting formal sector jobs are low, the potential to realize a steady stream of income is diminished, and in cases where they assume headship of their households, the ability to cater for the needs of their children is stifled. Though the educational attainment of workers since the FCUBE educational reform program in 1995 has increased, there are gender gaps in education. This tendency has a potential adverse implication for policy efforts geared towards reducing the incidence of poverty in Ghana (Sackey, 2008).

Female school enrolment has become a topic of increasing interest as the importance of girls' education to a wide spectrum of socio-economic outcomes has become apparent. According to Lloyd and Gage-Brandon (1994), high fertility rates have a negative impact on school enrolment of girls in Ghana. Mothers depend on girl children to take care of younger siblings while they are raising a family. Furthermore, mothers expect their sons to take care of them in their old age, and therefore are interested in their education leading to a higher income. The study implies that if social prejudices were lessened, the subsequent decrease in fertility could bring about greater participation of girls in school. But interestingly, the trends are changing now. Enrolment of girls in public primary schools increased from 1,281,780 in 2005/2006 to 1,366,476 in 2006/2007 representing a corresponding increase of 0.7% in percentage share of girls' enrolment from 47.7 % in 2005/2006 to 48.4% in 2006/2007. Enrolment of girls in public junior secondary schools also increased from 406,989 in 2005/2006 to 438,517 in 2006/2007, at a rate of 0.1%. The share of girls'enrolment in private junior secondary schools increased from 76,752 in 2005/2006 to 88,715 at 0.6%

in 2006/2007. This represent an increased from 46.0% in 2005/2006 to 46.6% at 0.6% in 2006/2007. Female workers with lower levels of education outnumber their male counterparts with similar education levels and vice versa. The incidence of female workers between 18 and 47 years of age show higher incidence of primary school attainment than male workers (GSS, 2008). Beyond the primary school level, more male workers have attained middle and secondary school than had female workers in (GSS, 2008). In other words, although improvements occurred for female workers, their higher levels of education have not yet caught up with the educational attainment of their male counterparts. Generally, there is an upward trend in the schooling attainment of both male and female workers and major implications of the rise in the educational status of workers are the potential impacts on productivity and earnings (Sackey, 2008). Thus, not only is the Ghanaian economy likely to benefit from increased output, but also there is potential for a narrowing of the gender earnings gap.

Education Policy Implications

Education policies in Ghana are directed to meet the skill needs of the modern workplace and to improve the performance of the individuals in the labor market. In fact, education is seen almost as a universal cure to some of the most severe economic problems such as unemployment and poverty. As such, Barro and Lee (2001) observed that a greater amount of educational attainment implies more skilled and productive workers, who in turn increase the output of goods and services. In addition, an abundant well-educated human resource also helps to facilitate the absorption of advanced technologies. Policy interest in education in

Ghana is therefore linked to its potential to raise earnings and reduce poverty. Consequently, education in Ghana is a significant determinant of earnings as it is one of the most important components of individual human capital (Becker 1993). Generally, it has been noted that a number of factors influence how education contributes to labour market development and earnings. These factors include (i) Human Capital Earnings Function (ii) Return to schooling. Economic research strongly suggests that rates of return to education have been growing over the last decades Ashenfelter, Harmon, and Oosterbeek, (2000). The high relative individual returns suggested by these studies clarify that investing in education is high payoff and low risk relative to most other investment options. These findings suggest that understanding the benefits of education to individuals is an important area of exploration for both individuals and policymakers. As such, an important first step in evaluating education policy is the measurement of returns to schooling. Understanding the magnitude of benefits is important for analyzing choices about the quantity and type of education pursued by individuals. Also, an ultimate hope is that by understanding the benefits to education, we can better help policymakers understand what education does for individual citizens of Ghana.

2.2 Education and Earnings

The relationship between education and earnings is one of the most studied topics. Generally, there is a strong belief that achievement of higher levels of education is a well established path to better jobs and better earnings. This empirical link between education and earnings can however, be masked due to the effect of unobserved factors related to schooling. More specifically, failure to account for the effect of unobserved "abilities" or family background factors on both earnings and attained schooling levels would lead to incorrect inferences regarding the effect of education on earnings. The concept of "ability" in this study refers to those marketable unobservable factors that make up an individual's initial endowment of human capital and translate into higher earnings. These may vary across families as well as individuals.

The literature concerning the relationship between education and earnings, in recent years, has expanded significantly mainly due to improved access to better-quality data as well as to a more appropriate methodology of investigation. The studies devoted to establishing this relationship deal in an explicit way with methodological issues and try to disentangle the effects of education from those of other (mainly unobservable) factors affecting economic returns (such as individual ability). This group of studies includes Cannari and D'Alessio (1995), Colussi (1997) and Brunello and Miniaci (1999). Most of these studies have established a positive relationship between education and earnings by using data from national surveys. However, recent analyses on this relationship have used data on twins which produces interesting insights into the roles of genetics and family background as mediating influences in the relationship between schooling and income (Behrman et al. 1977, 1994; Ashenfelter and Krueger 1994; Miller et al. 1995). Although the findings of the various studies have not been unanimous, all but Ashenfelter and Krueger (1994) ascribe some role to the influence of family background. The availability of twins data (with multiple measures of schooling) helps to deal with endogeneity of education arising from measurement error while indirectly controlling for any ability bias arising from "family effects". In the twins literature, estimates rely crucially on the assumption that any ability bias is due to unobservable family factors.

In addition, the links between education and earnings are of deciding factor to decisions about the efficient allocation of resources. However, due to omitted variables, interpretation of such estimates is usually qualified by comment on possible upward biases. Behrman & Deolalikar's (1993) criticism is that the studies which typically attribute the association between years of schooling and earnings do not include a host of other factors that plausibly may be correlated with years of schooling that affect wages. Sackey, (2008) emphasized both gender and age cohort effects in the educational credentials of workers in Ghana. He further indicated that there are more female than male workers with lower levels of education and more males than females at upper levels. This observation is consistent with schooling trends in sub-Saharan Africa. For both female and male workers, the age cohort effect is seen in terms of younger workers having higher educational attainment than older ones. Thus, at all levels of education, there is a tendency for school attainment to decline as people move from younger to older age cohorts. This trend is consistent with global patterns in educational attainment. Younger age cohorts tend to have more schooling than older people.

Moll (1998) stressed the point that years of schooling, as an input measure of human capital, may influence the wage if it captures other elements. Topel (1991) has concluded that, other things remaining constant, 10 years of job tenure raise the wage of the typical worker by over 25%. The strong positive relationship between tenure and wage rates was also assessed by Altonji & Williams (1997). The strong long term employer-employee relationship conditioned by promotion provisions was mentioned by Theodossiou (1996) to specify the significant effect of tenure on wages. Firms, in order to discourage labour turnover and inter-firm mobility, establish long-term employment relationships with their most highly valued employees. Thus, employees with longer tenure with their current employer have higher earnings than other employees with the same total work experience but relatively shorter tenure.

Opposing the significant effect of tenure on wages, Altonji & Shakotko (1987) argued that the partial effect of tenure on wages was small because the strong relationship between tenure and wages was due primarily to heterogeneity bias across individuals and across job matches. Similarly, Jacobson, Lalonde, & Sullivan (1993) have found that high tenure workers separating from distressed firms suffer long term losses averaging 25% per year. Re-examining the wage-tenure relationship, Williams (1991) has found that tenure increases wages only in the first several years of employment.

The occupation in which a worker is employed has an important effect on the level of his/her wages and salaries. Disparities in earnings between different occupations have been often noticed in less developed countries than in developed countries (Kothari, 1970). Earnings differentials would not indicate compensating differentials but rather signal enlarged inequalities because some individuals not only are denied the possibility of working at high and satisfied job levels but also have to accept lower wages (Hartog, 1986). For that reason the reward for education differs substantially by the job level at which an individual is occupied. The argument against the above assertion is that occupation and jobs are irrelevant entities in explaining earnings differentials because market forces tend to equate rates of return throughout and thus equilibrium situation will exist in the long-run.

It is widely accepted that family background affects education by influencing the amount of education individuals obtain. Family background and influences are more important in determining education and earnings (Krishnan, 1996). Altonji & Dunn (1996) explored the possibility that the education slopes of wage equations are influenced by family background as measured by father's and mother's education. Beach & Finne (1988) have also researched the positive effects of parents' education on son's educational attainment, and found an increasing importance of indirect and total effects of the family background variables on earnings. Controlling for workers' own schooling and the schooling of other relatives, Lam & Schoeni (1993) have discovered the relationship between having a father with a university education and getting a 20% wage advantage when compared with illiterate father, and a 9% wage advantage when compared to a father with 4 years of schooling. Sahn & Alderman (1988) have pointed out that the wage offer in developing countries is influenced by other genetic and environmental influences captured in the wage of one's father. Thus, the significant impact of family background on earnings could mean that family background determines the quality of education and learning environment at home (as educated parents can improve the educational opportunities of their children through their absorption of attitudes and acquisition of human capital) or would indicate that individuals from a better family background are able to get the better jobs through family connections and influences.

The partial cause of earnings differentials may also be sector of employment. Mann & Kapoor (1988) have explored that, on the average, public sector workers are paid much higher wages than the private and joint sector workers. Rees & Shah (1995) have reasoned that the private wage determination is subject to profit constraint, whereas the public sector wage determination is subject to an ultimate political constraint. Thus, wages in the public sector are higher than in the private sector. Pritchett (1999) highlighted the situation in which governments are taking resources away from non-governmental activity in the form of taxes so as to pay additional workers whose marginal product in the public sector is very low but are paid much higher wages than workers in the private sector.

2.3 The Human Capital Theory

The study of returns to education has a long tradition in labor economics and this tradition is based on standard human capital theory. Human capital refers to the stock of skills and knowledge relevant to performing labor to produce economic value. It is the skills and knowledge gained through education and experience that was first defined as such by Adam Smith (1776). Thus, schooling is viewed as an investment in human capital (Mincer, 1958; Becker, 1964), implying that the returns to schooling may be measured in terms of the extra income due to additional schooling.

The conventional theory of human capital developed by Becker (1962) and Mincer (1974) views education and training as the major sources of human capital accumulation that, in turn, have direct and positive effect on individuals' life time earnings. Recent studies of education and earnings determination are almost always embedded in the framework of Mincer's (1974) human capital earnings function (HCEF). In the Mincerian earning function, the logarithm of the hourly observed earnings of an individual is explained by schooling years, potential labor-market experience and experience squared. The coefficient of school years indicates the returns to education, *i.e.*, how much addition in earnings takes place with an additional school year. In his review of the literature on alternative specifications of the earnings function to determine whether the simple structure in Mincer's model is the most appropriate, Lemieux (2002), states that "the human capital earnings function remains a parsimonious and relatively accurate way of modeling the relationship between earnings, schooling and experience. Its status as the 'workhorse' of empirical labour economic research on earnings determination is well deserved".

A finding that is virtually universal across countries and years is the concave nature of the earnings function. This is as a result of the negative experience squared coefficient found when estimating the earnings equation. This means for those continuously attached to the labor market, earnings rise at a decreasing rate throughout one's life until depreciation exceeds human capital accumulation. Mincer, (1974) tested this proposition using OLS regression with cross-sectional data. Gautier and Teulings (2003) also present strong evidence for the concavity for six OECD countries. Indeed, this earnings function has been used to estimate the partial effect of schooling on the log of earnings (interpreted as rates of return from schooling) for about 100 different countries (Psacharopoulos and Patrinos, 2002).

There are circumstances where the only estimates of the Mincerian return to schooling available are obtained using standard statistical techniques. It is therefore important to understand whether estimates of the Mincerian return to schooling obtained with least-squares techniques are systematically different from estimates relying on twins or an IV approach. The growing literature on this issue suggests that, overall, the estimates obtained using twins or an IV approach are somewhat greater than estimates using least-squares techniques. The question of whether these differences are significant is analyzed in Ashenfelter, Harmon and Oosterbeek (1999).

2.4 Returns to Education

The earnings premium associated with additional education can be thought of as a 'rate of return' on that educational investment. According to the micro labor literature, the rate of return to education measures the extra earnings of a worker for an additional year of schooling and training. Economists and policy makers are interested in obtaining accurate measures of the percentage of earnings associated with acquiring more education. From a "private" point of view, under certain conditions, it provides a measure of the "return" to investment in additional schooling. From a social standpoint, the return to education could give an indication of the relative scarcities of people with different levels of education and hence it may provide a guide for educational policies (Arias et al. 2001). This study's focus is on the micro literature because it is the rate of return to education that determines the amount of human capital investments at the individual level. The rate of return to schooling, in the modern/human capital sense of the term, plays an important role in the determination of educational attainment and ultimately on earnings received by workers in the labour market (Harmon and Walker, 1995). Increasingly, governments and other agencies are providing financial resources towards "studies on returns to education along with other research, to guide macro-policy decisions about the organization and financing of education reforms" (Psacharopolous and Patrinos, 2004).

In common usage, the coefficient on schooling in a regression of log earnings on years of schooling is often called a rate of return. In fact, it is a price of schooling from a hedonic market wage equation. It is also a growth rate of market earnings with years of schooling. The justification for interpreting the coefficient on schooling as a rate of return derives from a model by Becker and Chiswick (1966). It was popularized and estimated by Mincer (1974) and is now called the Mincer model. This model is widely used as a vehicle for:

1. Estimating "returns" to schooling quality.
2. Measuring the impact of work experience on male–female wage gaps.

3. Economic studies of returns to education in developing countries.

The "returns to schooling" is a much-studied parameter in labor economics. Knowing the effect of schooling on earnings and other economic outcomes has important implications for educational policy, for efforts to better understand the evolution of inequality and for studies examining the sources of economic growth (Card, 1995; Katz and Autor, 1999).

Jacob Mincer's model of earnings (1974) which is a cornerstone of empirical economics and ultimately the framework used to estimate returns to schooling in recent studies is widely accepted in this discussion. It is the basis for economic studies of education in developing countries and has been estimated using data from a variety of countries and time periods.

Returns to education remain of central policy concern in both developed and developing countries. In developed countries, observed rises in returns to education have been imputed to skill biased technical change (Katz and Autor, 1999). In poorer countries such as those of Sub-Saharan Africa (SSA) it has been argued that returns may have been falling as a result of rapid expansion of education. As educational supply grows without a commensurate rise in demand the probability of getting a job for any given level of education declines and, among those with jobs, returns may fall.

Average returns to schooling are highest in the Latin America and the Caribbean region and for the Sub-Saharan Africa region. Returns to schooling for Asia are at about the world average. The returns are lower in the high-income countries of the OECD. Interestingly, average returns to schooling are lowest for the non-OECD European, Middle East and North African group of countries.

In his analysis of whether conventional patterns of rates of return to education prevail in sub-Saharan Africa, Bennell (1996) concluded that this is not the case. He noted that country case studies did not support the notion of consistently higher returns to primary education than either secondary or higher education. His findings were contrary to the broad picture provided on sub-Saharan Africa by Psacharopolous (1994). Differences in data and methodology used were highlighted.

Keswell and Poswell (2004), estimating the returns to schooling in South Africa for the periods 1995, 1997 and 2000, found that after accounting for censoring (via the use of Tobit models) the range for the private returns to schooling was 15–26%. Their study showed a decline in the returns to an additional year of schooling from 23.2% in 1993 to 18.2% in 2000. The variables used in their analysis were the years of schooling, age and its quadratic terms. In their alternative uncensored models (based on OLS regression), the authors found the returns to an additional year of schooling to be between 17% and 26%. In this framework, the returns to schooling had fallen from 24.5% in 1993 to 20.2% in 2000. The authors noted, however, that controlling for race in their models altered the results remarkably: (including race dummies, the estimated Mincerian rate of return in all years considered is less than half of that indicated), Keswell and Poswell, (2004).

Mincer (1974) finds that an additional year of schooling yields a net increase of 11.5% in annual earnings by estimating the Mincerian wage equation on cross-sectional data from the 1960 census for the US. Subsequently, the Mincerian wage equation has been estimated for many countries by using OLS. The results generally yield estimates of β_1 between 5% and 15%, with slightly larger estimates for women than men, Psacharopoulos (1994). By equating discounted costs and benefits, Becker (1964) estimates an internal rate of return to college and high school education of 13% to 28%. However, Solow (1965) argues that these large estimates are not corrected for correlations between education and ability. In order to solve this problem, Ashenfelter and Krueger (1994) estimate the return to schooling by contrasting wage rates of twins with different levels of educational attainment. They find that an additional year of schooling generates a wage increase of about 12% to 16%. In a similar manner, by analyzing a cross-section of twins, Rouse (1999) concludes that the rate of return to education is about 10% per year of schooling. Furthermore, Arias and McMahon (2001) estimate dynamic and expected dynamic rates of return to college and high school in the US. They find average returns of 13.3% in real terms or 11.7% after correcting for ability, family factors, and measurement errors.

Empirical evidence for developed western economies suggests that the average estimate of the return to an additional year of education ranges from 5% to 10%, Wilson (2001). For example, for the UK, Mincer (1974) converts his 16.2% gross increase in annual earnings to a net increase of 11.5% by factoring out increased labor force participation associated with an increase in education.

Dearden (1998) finds that the average annual return to an additional year of full-time education is 5.5% for men and 9.3% for women. Comparisons with less-developed countries show that the rate of return to education tends to be higher in latter countries, Acemoglu (2002). However, at least some of these countries show estimated returns to human capital investments of nearly the same magnitude, for example Belarus with 10.1% (Pastore and Verashchagina, 2006).

Table 2.4: The Rate of Return to Education

Author	Sample	Models and Methods	Schooling Coefficients
Becker (1964)	1949 Census Income data	HCEF; OLS	13 - 28%
Mincer (1974)	1960 Census for the US	HCEF; OLS	11.5%
Ashenfelter and Krueger (1994)	Twinsburg Twins Survey, August 1991	Linear Model, Instrumental Variable Regression; OLS, 2SLS	12 - 16%
Psacharopoulos (1994)	U.S. Studies	HCEF; OLS	5 - 15%
Dearden (1998)	The NCDS Census Survey, 1991	Instrumental Variable Regression; 2SLS	5.5 - 9.3%
Ashenfelter, Harmon, and Oosterbeek (1999)	Review 96 estimates from 27 studies, 9 Countries	HCEF, Instrumental Variable Regression; OLS, 2SLS	6.6 - 9.3%
Arias and McMahon (2001)	US CPS data from 1967 through 1995	Full method; OLS	11.7 - 13.3%
Wilson (2001)	Michigan-PSID, 1970 and 1980 U.S. Census, Common Core of data	Structural model – utility maximization model; OLS	5- 10%

Source: A Review of Human Capital Theory: Microeconomics. Prof. Jörg Baumberger, University of St. Gallen, Department of Economics

In a meta-analysis of the literature on returns to education, Ashenfelter, Harmon, and Oosterbeek (1999) review 96 estimates from 27 studies regarding different countries. They find that the average OLS estimate of the return to schooling is 6.6%, whereas the average IV estimate is 9.3%. Even after adjusting for a possible publication bias (because the probability of being published is higher for statistically significant results), the average IV estimate is 8.1% and still exceeds the average OLS estimate.

Twins-Based Returns to Education

Twin and sibling studies have contributed to the education field of research considerably in recent years. Monozygotic twins have used in the returns to schooling analysis to isolate the influence of family background and ability (Miller, Mulvey and Martin 1995). Using monozygotic (identical) twins, this method is effective in observing the correlation between education and the wage rate as it removes the influence of natural ability or other characteristics (Ashenfelter and Krueger 1994). Ashenfelter and Krueger (1994) in their study of US twins data concluded that the omission of ability from the earnings equation did not significantly bias estimates of the return to education upwards. In the past, there was nothing or little in the literature that focused on returns to schooling in Africa using twins data. This apparent dearth in the early literature might be linked with the fact that most estimates of returns to schooling have used data from national surveys which do not have well documented information about twins. Ashenfelter and Krueger's (1994) method of collection data on twins was therefore an appropriate one to follow in Africa. Earnings are influenced by a wide range of factors, though the main ones of interest appear to be ability, environment, schooling and experience. Various approaches have been taken to assess the roles these play and one that has stimulated considerable interest in the recent literature exploits samples of twins.

Behrman and Taubman, (1976) and Taubman, (1976) pioneered the use of data on twins for studying the returns to schooling. Examining within-pair differences in annual earnings and schooling among male twin veterans in the NAS-NRC dataset, Taubman (1976) found evidence of substantial upward ability bias in traditional cross-sectional estimates of the returns to schooling. Taubman's (1976) estimates decreased from 8.8% to 4.8% when moving from regression on the cross-section to within-pair estimation, despite correcting for an assumed 10% measurement error in the schooling data. The results in Behrman and Taubman (1976) imply similarly that standard OLS estimates are considerably upward biased.

The co-twin approach experienced a revival in the 1990s, following the innovation by Ashenfelter and Krueger (1994) to collect data on both own schooling and co-twin's schooling from each individual in the sample. Having two measures of schooling, they then used the first-difference of schooling reported by one member of a pair as an instrument for the first-difference reported by the other member. If measurement errors are uncorrelated, this allows for a correction of the problem of measurement error in the schooling variable. Under the equal ability assumption, their approach thus provides a consistent estimate of the returns from schooling.

Ashenfelter and Krueger's (1994) within-pair IV estimates were, surprisingly, considerably higher than standard least squares estimates on the cross-section. However, later studies strongly suggest that these initial results were due to sampling variation, as analyses of extensions of this sample produced within-pair IV estimates that were not higher than conventional cross-sectional estimates (Ashenfelter and Rouse, 1998; Rouse, 1999). These later findings are consistent with most other co-twin studies (Miller et al. 1995; Isacsson, 1999; Behrman and Rosenzweig, 1999, Bonjour et al. 2003), who likewise find only a small upwards ability bias.

Two recent additions to the co-twin literature are Isacsson (2004) and Zhang et al. (2007). Isacsson (2004) has the benefit of working with a representative dataset comprising education and income data for a very large number of Swedish monozygotic twins born 1926-1958, 2,609 pairs in total, and is therefore able to provide precise estimates of non-linearities in returns to schooling, and to allow for non-classical errors in the measurements of schooling. Zhang et al. (2007) analysed a dataset of 914 pairs of Chinese monozygotic twins and found that the returns to schooling during the Cultural Revolution (defined as 1966-1976 in their study) was roughly the same as that of later cohorts. In both these studies, the implied ability bias in cross-sectional estimates is positive.

David Card (1994) summarizes five recent studies that compare the education and earnings of twins. Two features of these studies contrast with the earlier literature on twins surveyed by Griliches (1979). First, the samples in the recent literature are relatively large, and tend to include a broader range of age and family background groups. Second, following the lead of Ashenfelter and Krueger's (1994) innovative paper, most of the recent studies squarely address the problem of measurement error. For each study reports on a cross-sectional (OLS) return to education, and two within-family differenced estimates: one estimated by OLS and the other corrected for measurement error are given.

The Ashenfelter and Rouse (1998) study utilizes three years of data collected in the Princeton Twins Survey (PTS): their sample includes 340 pairs of identical twins, 60 percent of whom are women. Ashenfelter and Rouse's (1998) within-family estimates of the return to education are about 30 percent lower than their corresponding OLS estimates. This finding contrasts with the results in Ashenfelter and Krueger (1994) based on only one year of data from the PTS, which indicated a bigger within-family than OLS estimate. The PTS questionnaire asked each twin their own education and their sibling's education. This extra set of responses allows Ashenfelter and Rouse (1998) to use one twin's responses about the difference in schooling for the pair as an instrument for the other twin's responses. The IV estimates, presented, are 25 percent larger than the simple differenced estimates, and about 10 percent below the corresponding OLS estimates. Rouse (1999) extends the analysis in Ashenfelter in Rouse (1998) with one further year of data from the PTS. Her findings are generally consistent with those in Ashenfelter and Rouse (1998), although Rouse's (1999) IV estimate is somewhat above the estimate reported by Ashenfelter and Rouse (1998), and actually exceeds the OLS estimate for the same sample.

The study by Miller et al. (1995) uses data for 1,170 Australian twin pairs (about one-half female). The advantage of the large sample size is offset by the absence of useable income data: Miller et al. (1995) have to impute incomes based on two-digit occupation. Thus, twins with the same two digit occupation are coded as having the same income. For identical twins Miller et al. (1995) found that the within-family estimate of the return to education is almost 50% lower than the cross-sectional estimate; for fraternal twins, the within-family estimator is 40% lower. Like the PTS, the Australian twins data set includes multiple reports of each twin's education. Miller et al. (1995) follow Ashenfelter and Krueger's (1994) procedure of using one twin's responses on the difference in schooling for the pair as an instrument for the other's responses. For identical twins, the resulting IV estimate is about 40 percent above the differenced OLS estimate, but still 25 percent below the cross-sectional estimate. For fraternal twins the IV estimate is actually slightly above the OLS estimate.

Behrman et al. (1994) analyze a data set that pools the NAS-NRC sample of white male World War II veterans with data on men from the Minnesota Twins Registry. While the main focus of their paper is on models of inter-familial resource allocation, an appendix table reports cross-sectional and within-family estimates of the return to schooling. For identical twins, Behrman et al. (1994) found that the within-family estimate of the return to schooling is about 50% as large as the cross-sectional OLS estimate, while for fraternal twins the relative ratio is 80%. Although they do not actually estimate IV models to correct for measurement error, Behrman et al. (1994) report that the reliability of the within-family difference in schooling for identical twins in the NAS-NRC sample is 0.62. They however, do not give a comparable estimate of the reliability ratio for fraternal twins. Results in Miller et al. (1995) and Ashenfelter and Krueger (1994), also suggest that the reliability of within family differences in schooling for fraternal twins is about 0.8. Using this estimate, a corrected estimate of the within-family return to schooling for fraternal twins is 0.071. The relative magnitudes of the OLS and within-family estimators for identical and fraternal twins in Behrman et al. (1994) and Miller et al. (1995) are therefore very comparable.

Finally, Isacsson (1997) analyses earnings and schooling differences among a large sample of Swedish twins. For a sub sample of the data he has information on two measures of schooling: one in a register held by Statistics Sweden; another based on self-reported education qualifications. Isacsson (1997) finds that the within-family estimate of the return to schooling for identical twins in the sub sample with two schooling measures is less than 50% as large as the corresponding OLS estimator, while for fraternal twins the ratio is 80%. He constructs IV estimates for the within-family model using the difference in the survey measures of schooling as an instrument for the differences in the registry measures. For identical twins, the within-family IV estimator is only marginally above the within-family OLS estimate, implying almost no measurement error bias. For fraternal twins, on the other hand, the IV procedure raises the within-family estimate by 35 percent. Since one would have expected bigger measurement error attenuation for identical twins than fraternal twins, the patterns of Isacsson's (1997) findings are somewhat puzzling.

Isacsson (1997) also constructs measurement-error-corrected estimates of the return to education for a broader sample of twins, assuming "low" and "high" estimates of the reliability of his main schooling measure (reliabilities of 0.85 and 0.95, respectively). For fraternal twins the corrected within-family estimates lay in a fairly tight range (0.044 to 0.060) that brackets the within-family IV estimate based on the two schooling measures (0.054). For identical twins the range of the corrected estimates is wider (0.027 to 0.060) and lies above the within-family IV estimate based on the two schooling measures (0.024). Taken as a whole, Isacsson's (1997) results suggest that the measurement-error-corrected within family estimate of the return to education for fraternal twins in Sweden is about as big as or even bigger than the corresponding OLS estimate. The precise relative magnitude of the measurement-error corrected within-family estimate for identical twins is more uncertain, and seems to be very sensitive to assumptions about measurement error. A cautious interpretation of Isacsson's (1997) findings is that there may be some upward bias in OLS estimates of the return to schooling relative to the within-family estimate for identical twins.

2.5 Gender-Based Analysis of Returns to Schooling

Using household survey data from 1996/97 to 1998/99 for Nigeria and ordinary least squares method, Aromolan (2006, 2004) found the returns to an additional year of schooling at the post-secondary were 10.4% for male and 12.2% for female wage earners; and 13.7% for self-employed male and 15.4% for self-employed females. Generally, the wage returns to an additional year of post-secondary education were found to be between 10% and 15% for workers in the labour market in Nigeria. At the primary and secondary levels, however, these returns were quite low, ranging between 2% and 4%. On the basis of his empirical results, the author concludes that "increasing public investment to encourage increased attendance in basic education is not justifiable on grounds of private efficiency, unless investments to increase school quality have higher private returns" (Aromolan, 2004).

In a different study on the Nigerian economy, Okuwa (2004) used data from the 1995 Nigerian labour market survey to examine the private returns to higher education. For all levels of education, the returns to schooling were higher for private sector workers than public sector workers. The returns to schooling also increased as higher levels of schooling are attained. The return to an additional year of secondary schooling was - 0.5% for males and 3.5% for females. At the university level, schooling returns were 16.3% for males and 10.7% for females. In the private sector the returns to additional year of university education brought returns of 16.8%, while in the public sector this was 12.6%. On the bases of these findings, the author provides a policy recommendation that "the university, which attracts the highest magnitude of returns, should be properly funded and equipped with modern technology, especially the laboratory, library, information system and infrastructure" (Okuwa, 2004).

In a similar study in Kenya, Kimenyi et al. (2006) examined human capital externalities and private returns to education using 1994 data sets from a national welfare monitoring survey. They found a positive relationship between the level of education and the associated returns. Taking into account human externalities, the returns to an additional year of schooling increased from about 8% for primary school to 23% for secondary school and then to 25% for university level of education. At the university level, the returns to schooling were higher in urban than rural areas (about 61% for urban females versus 21% for rural females, and 35% for urban males versus 17% for rural males). The authors conclude that: "... public policies that expand schooling opportunities for underprivileged social groups benefit the whole society via the externality effects of education. The benefits are in terms of improved productivity and earnings" (Kimenyi et al. 2006).

The case for Zambia is presented in the analysis by Nielsen and Westergard-Nielsen (2001). Providing an empirical analysis based on 1993 survey data, these authors found the return to schooling to be higher in rural than in urban areas. They associated this trend with the apparent low quality of schooling, noting that it was possible for people to complete primary schooling without being able to read and write. The implication of a rural-based economy with a relatively high illiterate rural population is that "having some education probably works as a signal for some underlying favourable unobservable characteristics" (Nielsen and Westergard-Nielsen, 2001).

In a gender-focused study, Dougherty (2005) explains the difference in the male-female schooling coefficient in terms of job characteristics (i.e., composition effect) and occupation choice (i.e., occupation effect). With regard to the "composition effect", the difference is explained by the under-representation of females in jobs where schooling is a relatively unimportant factor in the determination of earnings. He cites the example of female under representation among union workers, where schooling is subordinated to seniority as a determinant of earnings, or in self-employment where entrepreneurial skills are relatively highly valued. In terms of the "occupation effect", the tendency for women to be segregated in occupations with relatively low pay generates an earnings gap between men and women. Another plausible cause of earnings differential between male and female workers is the quality of educational attainment.

In the UK, the average annual return to a first degree in terms of hourly wages (compared to just A- levels) is in the range of 5% to 8% for men and 10% to 13% for women (Blundell, Dearden, Meghir, and Sianesi, 1999). Studies from other countries also find that investments in women's education tend to yield higher rates of return than investments in men's education. For example, Butcher and Case (1994) find higher returns for women in the US. In this context, Mincer and Opek (1982) suggest that the restoration of human capital - after labor market interruptions associated with the depreciation of human capital - is more efficient than the accumulation of new human capital by men who stay inside the labor market the whole time.

This gender difference in the returns to education arises because the earnings of women are considerably lower than those of men (Blundell, Dearden, Meghir, and Sianesi, 1999). The gender wage gap can be decomposed into three different parts: gender differences in human capital accumulation, occupational sex segregation, and discrimination (as residuum) (Kanazawa,2005). According to Winter-Ebmer and Zweimüller (1992), occupational sex segregation can have three different reasons: different preferences for various occupations, crowding (i.e. disadvantages in "male jobs" leading to oversupply in the more "female jobs"), and human capital theory. With respect to human capital theory, Blackburn (2004) empirically finds that men perform better in math-oriented tests and women better on speed-oriented tests. However, he argues that test score differences explain only a small part of the gender wage gap.

Another explanation for the gender wage gap may be the fertility decision of women which leads to labor market interruptions. This gives rise to gender differences in the turnover rate and thus in employment and wages (Erosa, Fuster, and Restuccia, 2002). Polachek (1981) extends this argument to a rationale for occupational sex segregation and suggests that women tend to choose jobs with low penalties for intermittent employment. Although this reasoning is rejected by England (1982), Schumann, Ahlburg, and Mahoney (1994) find that the male-female wage differential can be partially attributed to job characteristics. In an empirical study for apprentices in West Germany, Kunze (2005) verifies a gender wage differential of about 25% that is attributed to occupational segregation. However, Blau (1998) suggests that the convergence in male and female college majors may be responsible for a reduction in the gender wage gap during the 1980s.

Although, returns to education have been estimated in studies using the HCEF in Ghana, some persistent econometric problems still persist that may lead to heterogeneity in returns to schooling and endogeneity of schooling. As Ghana is experiencing a rapid growth in its primary and post-secondary school enrollment rates, due to its educational reforms, economists and policy makers are concerned about the relative costs and benefits of higher education for those who were not previously receiving it and these issues may give inaccurate estimates of the economic return to schooling in Ghana that may misguide policy makers. There is the need therefore to examine these issues in relation to the returns to schooling in Ghana.

2.6 Heterogeneity in Returns to Education

There is much controversy regarding the measurement of returns to education, especially because, of heterogeneity in returns by Card (1999, 2001), Carneiro and Heckman (2002), Carneiro, Heckman and Vytlacil (2001) and Blundell, Dearden and Sianesi (2002). The main problem in measuring returns to education is that the decision to take more education is a complex process. Factors such as individual ability, financial constraints, family background and preferences are usually unobserved by the researcher. This creates a problem that relates to observed and unobserved heterogeneity in the return parameters of education and the interpretation of different return parameters, Lang (1991), Willis and Rosen (1979), Card (1995, 1999) and Heckman and Vytlacil (1999). This heterogeneity arises if individuals select their education on the basis of their comparative advantages; Roy (1951), Garen (1987) and Willis and Rosen (1979).

Heterogeneity refers to differences in the returns to education across individuals due to factors unobservable to the econometrician, but known to the individual at the time of their decisions. More generally each individual faces a distribution of returns that is conditional (in its mean and possibly higher moments) on individual characteristics observable to them, but some of which are not observable by the econometrician. Thus, even if each individual were to receive their expected return to education with absolute certainty, returns would still appear to vary randomly across the population. Ability and family background variables may help to explain baseline differences in earnings across individuals, and may also help to explain why returns to schooling differ across individuals.

An assumption made in most empirical studies when estimating the standard Mincerian wage equation is that the return to schooling is homogenous (i.e., returns are constant across individuals). This assumption ignores the heterogeneity that is found in returns to education, which may influence the effect of education. Furthermore, unobserved ability induces heterogeneity in the distribution of earnings conditional on education through its effect on both the intercept and the education coefficient. Thus, it follows that the labor market cannot be well characterized by a single rate of return to education and that there is no single effect of education but rather a whole distribution of individual effects (Blundell et al. 2005; Heckman et al. 2006).

A number of researchers (Koop, 2002; Cainero, 2002; Godde and Schnabel, 1998; Altonji and Dunn, 1996) have accounted for heterogeneity in the returns to education by using fixed effects approaches to control for unobserved family-specific heterogeneity. One potentially significant limitation of fixed effects models is that it cannot estimate effects of variables which vary across individuals but rather accounts for limited heterogeneity in the data by allowing for group specific intercepts. Although differencing within twin pairs surely eliminates much of the omitted ability, Griliches (1979) notes that if the "ability" that is rewarded in labor markets has more than a purely genetic component, then even among monozygotic (MZ) twins ability differences will remain. Bingley et al. (2005) modeled the unobservable differences in the returns to education by decomposing unobserved heterogeneity in returns to schooling into individual and family effects and quantifying the relative importance of both effects.

Arias et al, 2001; Mwabu and Schultz, 1996; Fitzenberger and Kurz, 2003 and Machado and Mata, 1999 also use quantile methods and obtain varying returns across quantiles. Quantile regression methods estimate returns to schooling for individuals at different quantiles of the conditional distribution of earnings which is viewed as reflecting the distribution of unobservable ability. Quantile techniques also characterize the effect of education on the whole conditional distribution of earnings. Estimating the effect of education at conditional quantiles allows for heterogeneity in the returns to education.

Conneely and Uusitalo (1998) also investigate the question of heterogeneous returns in the context of a random coefficients model of wage determination. They use data on ability test scores and family background variables on a sample of Finnish men and parameterize potential heterogeneity in the mean return to education by relating these factors with education. They find stronger evidence of variations in returns to education most of which, nevertheless, cannot be explained by observable individual heterogeneity.

Bingley et al. (2005) examine unobservable differences in returns to education by exploiting panel data to show that there are large and significant variances to the returns to education. They employed multi-level modeling to estimate a mixed model with both fixed and random effects. This essentially estimated a random coefficient (on education) model and decomposed the variance around the mean return into family heterogeneity, individual heterogeneity, and luck or risk. They discovered that individual variance in returns is smaller for monozygotic than for dizygotic twins.

2.7 Endogeneity of Schooling

The conventional wisdom in the literature is that OLS estimates of returns to schooling are biased and inconsistent due to endogeneity in the schooling variable (Griliches, 1977; Wooldridge, 2002). Schooling or the level of education is treated as an endogenous variable when the error term presumptuously includes a host of unobservable individual characteristics that might affect the return to schooling. The possible presence of these unobservable characteristics that influence both earnings and the likelihood of completing school, means that estimates of the effect of years of schooling on earnings may be biased. Statistical endogeneity of education in the earnings equation may therefore result from

1. Unobserved determinants (such as ability, family and individual characteristics, school and teacher quality, etc.) of education that also influence earnings and/or

2. Measurement error.

Various estimation methods have been implemented to tackle these biases. Among these, the IV technique using a wide range of instruments has become one of the predominantly employed alternatives. The majority of the literature on the return to schooling uses instrumental variables (IV), an econometric method to handle the endogeneity of schooling problems. This involves jointly estimating the factors that influence the level of schooling attained and the wage outcome. To this end, the instrumental variable has to be correlated with schooling and should be uncorrelated with unobserved individual earning capacities (i.e., a source of variation in the level of schooling can be identified that is unrelated to wages, then in theory this can be used to gain unbiased estimates of the effect of schooling on wages). To determine whether any endogeneity in the schooling variable has a significant effect on the schooling return parameter the Hausman test is used.

It is important to address potential biases in the OLS and IV estimates as these biases could offer competing explanations for the gap between the OLS and IV estimates. The return to schooling literature has identified a variety of potential biases. Ability bias can arise in OLS estimates because higher ability individuals can invest more in schooling and also have higher earnings. Ability bias is known to be consistent with human capital models of schooling. Card (1999) gives a range of ability bias across studies of 0% to 50% of the private return, but suggests putting more weight on a range of 10% to 15%. An additional factor that could lead to bias is measurement error bias. Previous studies emphasize that OLS estimates may be biased downward by roughly 10% to 20% due to measurement error. Existing estimates for the ranges of these biases are based on various IV strategies, and different instruments may lead to different estimated combinations of private returns. Estimates based on law changes (e.g. compulsory schooling) that affect large fractions of the population may yield estimates of the social returns to schooling. In contrast, estimates based on family background, twins or access to schooling (e.g. distance to schooling) may yield estimates of the private returns as these instruments are based on private characteristics relating to schooling decisions.

Most previous twin studies on returns to education have also addressed endogeneity of schooling in two ways. One approach (Ashenfelter and Krueger, 1994; Ashenfelter and Rouse, 1998) treats ability as an unobserved family effect and estimates a "fixed effects model" based on a version of Mincer differenced equation for each twin pair. Ashenfelter and Krueger (1994) and Ashenfelter and Rouse (1998) also provide estimates of the returns to education and the resulting endogeneity bias (to which they refer to as a "selection effect"). The model of optimal schooling choices that they used suggests that we use measures of the education of a twin's sibling, the average education of the twins, or father's education as an additional regressor to control for any "family" effect that affect the absolute level of earnings. However, twin-based methods have also been criticized by Neumark (1999), as well as by Bound and Solon (1999). They argue that there may still be endogeneity that causes bias in the wage difference equation because the within differences in schooling may be correlated with the error term - equation – that is, the wage difference equation also suffers from some endogeneity because differencing has not removed all of the ability bias. The presumption in the twins literature is that the omitted ability is entirely made up of a genetic effect and a family effect which therefore disappears with differencing between family members with the same genes. There is, in general no strong reasons for thinking that this is necessarily the case – for example, birth weight differs between twins (actually by more than between non-twin siblings) and there is substantial evidence that birth weight has real effects. Neumark, 1999 and Bound & Solon, 1999, note that if differencing does not remove *all* of the omitted ability then the within-twin estimator may still be biased, and may even be more biased than least squares applied to individuals.

The corresponding results have generally led to a consensus that OLS estimates of the returns to education are biased downwards. Griliches (1979) suggested the possibility of both upward and downward bias. Upward bias remains the conventional wisdom (Ehrenberg-Smith 1991), even if some authors imply that this is true only if one analyses wages of mature workers (Blackburn-Neumark 1993). These contrasting empirical results are due to econometric problems: omitted variables and measurement errors could involve opposite distortions without specifying which one prevails. A general framework to describe the problem could be found if we focus on the potential endogeneity of education in the earnings equation. The interpretation of these estimates in understanding the effect of education on earnings then becomes an issue of significant concern to policymakers.

A wide body of literature attempted to examine how severely OLS estimates of the returns to schooling might be biased. The evidence from this literature is, however, mixed. It suggests that whether OLS estimates are upward or downward biased depends on how ability differences are accounted for. For example, studies where endogeneity is accounted for via the inclusion of an explicit measure of ability report an upward bias in OLS estimates (Blackburn and Neumark, 1993) whereas those based on panel data and where ability is captured by individual fixed effects conclude to a downward bias in OLS estimates (Guillotin and Sevestre, 1994). Another approach consists in eliminating differences in innate ability by exploiting differences between twins or siblings in the levels of schooling and earnings. Using U.S. siblings and twins data, respectively, Ashenfelter and Zimmerman (1993) and Ashenfelter and Krueger (1994) report estimates that are much higher than typical OLS ones. In contrast, using U.K. twins data, Blanchflower and Elias (1993) find evidence of an upward bias in OLS estimates. However, studies using instrumental Variables (IV) by exploiting natural variations in data caused by exogenous influences on the schooling decision systematically conclude to a downward bias in OLS estimates. (Angrist and Krueger, 1991; Card, 1993; Kane and Rouse, 1995; Dearden, 1995; Harmon and Walker, 1995; Uusitalo, 1999). Even though, there is no unanimity in these studies about the importance of the endogeneity bias. For example, while

Angrist and Krueger (1991, 1992) conclude to a limited impact of endogeneity, the results in Butcher and Case (1994) or Kalwij (1996) suggest such an impact is rather large. Furthermore, Lauer and Steiner (2001) estimate homogeneous returns to education for Germany by IV-methods using different family background variables as instruments. The results depend on the instruments used. The estimated returns to education vary between 6.6 and 14.8 percent. Jochmann and Pohlmeier (2004) use different instruments in the case of heterogeneous returns as for example the number of siblings, secondary school density or the unemployment rate at graduation. Again, the results vary to some degree with the chosen instruments.

Bound (1999) noted that in spite of the enormous value of using twins in the relationship between earnings and schooling a difficulty arises since monozygotic twins, notwithstanding their remarkable similarity and identical genetic endowment, are not exactly identical. They differ in temperament and abilities. As a result of this, measured schooling variation between monozygotic twins, like that between families, is contaminated both by endogenous determination of which twin goes to school longer and by measurement error. In addition, there is a notion that originally motivated the siblings-based estimation approach—that the empirical association between earnings and schooling confounds the causal effect of schooling with other factors that influence both wages and schooling.

Ashenfelter and Krueger (1994) have exploited the presumed similarity of twins and the availability of multiple measures of schooling to explicitly model the relationship between family ability and education parametrically, while addressing the measurement error and endogeneity biases using standard panel data methods. They found some evidence of the existence of a negative relationship between ability and returns to education, suggesting that less able individuals benefit more from additional schooling. Twins (or siblings) data have also been employed to attempt to eliminate endogeneity bias by exploiting the differences between twins in levels of schooling and earnings, on the grounds that this eliminates differences in innate ability or motivation.

2.8 Recent Twins Research

What general conclusions can be drawn from the recent twins literature? Suppose on a priori grounds one believes that identical twins have identical abilities. Then the within-family estimator for identical twins, corrected for measurement error biases, is consistent for the average marginal return to schooling in the overall twins population. Assuming that this is the case, the estimates in David Card's (1998) paper suggest that a cross-sectional OLS estimator yields a slightly upward-biased estimate of the average marginal return to education: the magnitude of the bias ranges across studies from 50 percent (Isacsson, 1997) to zero (Rouse, 1997). Given the limitations of the imputed earnings data used by Miller et al. (1995) and Behrman et al. (1994) and the uncertainties in the measurement error corrections for Isacsson's (1997) study, more weight can be put on the Ashenfelter and Rouse (1998) and Rouse (1999) studies, which suggest a smaller range of biases (more like 10-15 percent).

A second conclusion emerges from studies that present results for fraternal twins. In these studies the measurement-error-corrected within-family estimator of the return to education for fraternal twins is about equal to the corresponding OLS estimator. Interestingly, Ashenfelter and Zimmerman's (1997) measurement-error-corrected estimate of the return to schooling for brothers (constructed under the assumption that brothers have identical abilities) is also about equal to the corresponding OLS estimate. Since fraternal twins are essentially brothers (or sisters) with the same age, the similarity of the findings for fraternal twins and brothers is reassuring. Assuming that OLS estimates are upward-biased relative to the true average causal effect of education, the within-family estimates based on fraternal twins or brothers must also be upward-biased. Moreover, since the OLS estimator is downward biased by measurement error, whereas the corrected within-family estimates for fraternal twins or brothers are not, one can conclude that the ability bias in within-family estimators for fraternal twins or brothers are smaller than the ability bias in cross-sectional OLS estimators: on the order of one-half as large. This implies that ability differences between brothers or sisters are relatively less important determinants of within-family schooling outcomes than are overall ability differences in the determination of schooling outcomes for the population as a whole.

Such a finding opens up the interesting question of how and why families affect the schooling decisions of children with differential abilities. Behrman et al. (1982) present a model incorporating parental preferences in the distribution of education resources across siblings that is consistent with either reinforcing or compensatory behavior (i.e., families may spend more educating either their more- or less-able children). Their empirical findings support the notion of compensatory parental behavior (behavior that would lead to a reduction in the relative importance of ability differences in determining education outcomes within families than between families). If one does not believe that identical twins have identical abilities, then even the within-family estimator of the return to education for identical twins may be biased by ability differences.

Ashenfelter and Rouse (1998) present a variety of indirect evidence in support of the hypothesis that identical twins are truly identical, and that differences in their schooling levels are attributable to random factors rather than to ability differences. For example, they report that schooling differences among identical twins are uncorrelated with birth order and with their spouse's education. Despite this evidence, and the strong intuitive appeal of the "equal abilities" assumption for identical twins, however, I suspect that observers with a strong a priori belief in the importance of ability bias will remain unconvinced.

CHAPTER THREE

MODELS, ESTIMATION METHODS OF MODELS AND ECONOMETRIC ISSUES

Human capital literature suggests that higher levels of education have a positive relationship with earnings and, by implication, productivity. However, it may also be the case that high earnings contribute to higher levels of education as they provide the funding to access related goods and services. This chapter describes the multivariate models and estimation methods that were used to estimate the effects of education on earnings. It also sets out some of the econometric issues associated with this type of research.

3.1 The General Modelling Framework

The effects of education on earnings are commonly estimated using a human capital earnings function based on the model specified by Mincer, (1974). To assess the extent of bias in conventional rates of return to schooling associated with the failure to control adequately for genetics and common environment, two models have been used by economists in recent times. Firstly, a fixed effects model wherein the differences in the earnings of members of a set of twins are related, in a regression framework, to differences in characteristics of the same individuals. Estimation of separate equations for monozygotic (MZ) and dizygotic (DZ) twins is equivalent to holding constant, in the first instance, genetic endowments and common environment, and, in the second instance, common environment influences only. This amounts to an implicit control for these factors. Secondly, a structural model (selection effects) that explicitly accounts for family effects (genetic endowments and common environment) through the inclusion in the estimating equation of information on a respondent's co-twin has also been applied. In Ashenfelter and Krueger (1994), information on the respondent's co-twin's educational attainment is used to compute family effects. Both of these economics models have been used in this study. To address the problem of endogeneity of schooling, the Instrumental Variable (IV) regression model was used. Finally, to account for unobservable differences in the returns to education, this study employs multilevel modelling to estimate a mixed model with both fixed and random effects. This essentially estimates a random coefficient (on education) model and decomposes the variance around the mean return into family heterogeneity, individual heterogeneity, and luck or risk or decomposes unobserved heterogeneity in returns to schooling into individual, family effects and risk and quantifies the relative importance of both these effects.

3.1.1 The Human Capital Earnings Function

The effects of human capital characteristics on wages are commonly estimated using a human capital earnings function based on the model specified 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:

$$Y_i = \beta_0 + \beta_1 S_i + \beta_2 Exp_i + \beta_3 Exp_i^2 + \varepsilon_i \qquad (3.1)$$

where Y_i is the natural log of earnings, S_i is years of schooling, Exp_i is the potential experience of individual i, $\beta's$ are the regression coefficients, ε_i is a normally distributed homoskedastic residual or a well-behaved error-term.

Experience is included as a proxy for the accumulation of human capital that occurs after formal education (such as on-the-job training). Potential experience was used because of a lack of reliable data on actual labour market experience. Mincer (1974) proposed the alternative of “potential experience”, i.e. the number of years an individual could have worked after completing schooling. Assuming that he/she starts schooling at 6 years old and begins working immediately after S_i years of schooling, Exp_i is equal to Age-Years of schooling-6 (A-S-6). Exp_i^2 represents ‘experience squared’ to capture a concavity of the observed earnings profile. The quadratic term is included to allow for a possible decline in the returns to this form of human capital over the individual’s life. (For example, technological change can render redundant the skills accumulated early in a person’s working life). The quadratic function of potential experience implies that over time returns to experience diminish and eventually could become negative. Conveniently, coefficients in the log-linear wage equation can be interpreted as approximations of percentage effects. That is, β_1 can be read as an approximation of the effect on earnings of an additional year of schooling in percentage terms.

3.1.2 Fixed Effects Model

The return to education in this study is estimated based on a linear model by Chamberlain, (1982) which specifies wage rates as consisting of an unobservable component that varies by family, observable components that vary by family, observable components that vary across individuals and an unobservable individual component. This linear equation for the returns to education is specified as follows:

$$Y_i = \alpha X_i + \beta Z_i + \mu_i + \varepsilon_i, \qquad (3.2)$$

where subscript i refers to individual i, Y_i is the logarithm of earnings, X_i is the set of family variables, Z_i is a set of individual variables that affect earnings (e.g., education, age, gender, marital status, and job tenure), α is the intercept , β is the return to education, μ_i represents a set of unobservable variables (i.e., the effect of ability or family background) and ε_i is the error term, which is assumed to be independent of Z_i and μ_i. When the data used in the analysis are supplied by twins, the earnings equations for a pair of twins are written in the form:

$$Y_{1i} = \alpha X_i + \beta Z_{1i} + \mu_j + \varepsilon_{1i}, \qquad (3.3)$$

$$Y_{2i} = \alpha X_i + \beta Z_{2i} + \mu_j + \varepsilon_{2i}, \qquad (3.4)$$

where Y_{ji} (j = 1, 2) is the natural logarithm of the earnings of both twins in the pair, X_i is the set of observed variables that vary by family but not between the twins, Z_{ji}(j = 1, 2) is a set of variables that vary between the twins (eg., marital status, educational levels), α is the intercept and β is the return to education.

In order to estimate the return to schooling (and to other twin-specific characteristics), the difference between Equations (3.3) and (3.4) are taken to eliminate the ability effect, obtaining the first differenced, or fixed-effects estimator specified as:

$$Y_{1i} - Y_{2i} = (Z_{1i} - Z_{2i})\beta + \varepsilon_{1i} - \varepsilon_{2i} \tag{3.5}$$

The first difference removes both the observable and unobservable family effects (i.e., X_i and μ_i). The major feature of the fixed effects model is that genetic resemblance and common environment influences are held constant implicitly. This model will also net out of the estimated impact of schooling, the compounding effects of any other fixed effects that affect earnings (e.g., race, possibly some affective characteristics such as motivation).

3.1.3 Seemingly Unrelated Regressions

The return to education for twins is also estimated in the framework of a system of linear equations using the seemingly unrelated regression (SUR) model, proposed by Zellner, (1962). This allows for the direct comparison of the returns to education between twins and takes into account the correlated error terms of twins. The SUR model is a generalization of a linear regression model that consists of several regression equations, each having its own dependent variable and potentially different sets of exogenous explanatory variables. Each equation is a valid linear regression on its own and can be estimated separately, which is why the system is called seemingly unrelated (Greene, 2002). A seemingly unrelated regression (SUR) system comprises of several individual relationships that are linked by the fact that their error terms are assumed to be correlated across the equations. The correlation among the equation error terms could come from several sources such as unobservable heterogeneity, reflecting unobserved genetic and family determinants (Hougaard, et al. 1992).

There are two main motivations for use of SUR. The first one is to gain efficiency in estimation by combining information on different equations. The second motivation is to impose and/or test restrictions that involve parameters in different equations.

The SUR model can be viewed as either the simplification of the general linear model where certain coefficients in matrix B are restricted to be equal to zero, or as the generalization of the general linear model where the regressors on the right-hand-side are allowed to be different in each equation. The SUR model can be further generalized into the simultaneous equations model, where the right-hand side regressors are allowed to be the endogenous variables as well.

The SUR model is given by

$$Y_i = X_i\beta_i + \varepsilon_i, \qquad i=1,\ldots\ldots,m, \tag{3.6}$$

where, Y_i and ε_i are $T\times 1$ vectors, X_i is a $T\times k_i$ matrix, and β_i is a $k_i \times 1$ vector. These m vector equations are stacked on top of each other and the system takes the form proposed by Zellner, (1962) as

$$\begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{vmatrix} = \begin{Vmatrix} X_1 & 0 & & 0 \\ 0 & X_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & X_m \end{Vmatrix} \begin{Vmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{Vmatrix} + \begin{vmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{vmatrix} \Leftrightarrow \; y = X\beta + \varepsilon \qquad (3.7)$$

where, each, y_i is a $T \times 1$ vector of sample values on the dependent variable(s), X_i is a $T \times k_i$ matrix of sample values on the k_i independent variables and β_i is a k_i x 1 vector of coefficients

The assumption of the model is that error term ε_i is independent and normally distributed with:

$$E(\varepsilon_i) = 0\,;\; E(\varepsilon_i \varepsilon_i') = \sigma_{ii} I_T\,,$$

and the variance-covariance matrix $\Omega = E[\varepsilon\varepsilon']$ is defined as:

$$\Omega = \begin{vmatrix} E[\varepsilon_1\varepsilon_1'] & E[\varepsilon_1\varepsilon_2'] & \cdots & E[\varepsilon_1\varepsilon_m'] \\ E[\varepsilon_2\varepsilon_1'] & E[\varepsilon_2\varepsilon_2'] & & E[\varepsilon_2\varepsilon_m'] \\ \vdots & & \ddots & \vdots \\ E[\varepsilon_m\varepsilon_1'] & E[\varepsilon_m\varepsilon_2'] & \cdots & E[\varepsilon_m\varepsilon_m'] \end{vmatrix} = \begin{vmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \cdots & \sigma_{1m}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & & \sigma_{2m}I_T \\ \vdots & & \ddots & \vdots \\ \sigma_{m1}I_T & \sigma_{m2}I_T & \cdots & \sigma_{mm}I_T \end{vmatrix}, \qquad (3.8)$$

where $E[\varepsilon_m \varepsilon_p'] = \sigma_{mp} I_T$ and σ_{mp} is the covariance of disturbances between the m^{th} and p^{th} equations, contemporaneously (which is assumed to be constant across all equations). The regression relations for the different individuals are only related via the correlation of the error terms, but the error covariance across individuals is unrestricted.

3.1.4 Selection Effects Model

In contrast to the fixed effect model, the selection effects model explicitly incorporates family effects (genetic endowments and common environment) in the earnings equation. Information on the respondent's co-twin's educational attainment is used to compute family effects. Consideration of this alternative model provides a means for assessment of the robustness of the findings obtained from the fixed-effects model.

In the selection effects model, the earnings (Y_{ji}) of twin i, who is a member of family j, depends on variables that vary across families but not between twins (in this instance age), on individual-specific variables (education), and on unmeasured family effects (μ_j). The unmeasured family effects are modeled as depending on the educational attainments of each twin member, and on the age of the twins. Hence the selection effects model is given as:

$$\begin{aligned} Y_{ji} &= \alpha X_j + \beta S_{ji} + \mu_j + \varepsilon_{ji} \\ \mu_j &= \gamma S_{j1} + \gamma S_{j2} + \delta X_j + \omega_j \end{aligned}, \tag{3.9}$$

i=1, 2: j=1, n

Substituting for the μ_j term in the earnings equation results in the reduced form:

$$Y_{ji} = (\alpha + \delta) X_j + (\beta + \gamma) S_{ji} + \gamma S_{j-i} + \varepsilon_{ji}, \tag{3.10}$$

i=1, 2: j=1, n

In this equation the coefficient on the co-twin's educational attainment (γ) provides an estimate of the impact of family effects which can be subtracted from the coefficient on the own-education variable ($\beta + \mu$) to derive an estimate of the pure return to schooling.

3.1.5 Instrumental Variable (IV) Regression Model

Instrumental variables methods were developed to overcome the problem of regressor-error dependencies in regression models. The regressor (independent variable) that is correlated with the error term can be called an endogenous regressor. Regressors that are uncorrelated with the error term are exogenous. An IV approach to getting consistent estimates when there is an endogenous regressor (or regressors) requires that there are some variables available that are correlated with the endogenous regressor, but are not correlated with the error term in the model. These variables are called instruments. Instruments are variables that only influence the dependent variable through their effect on the endogenous regressor. The IV approach assumes that a set of variables Z, called instruments or instrumental variables is available. These instruments should be uncorrelated with the error term ε, i.e., $E(\varepsilon \mid Z) = 0$ and explain part of the variability in the endogenous regressors. Hence, the instruments Z cannot have a direct effect on y (the instruments Z are exogenous). In this study our interest is in getting consistent estimates of β (i.e., the effect of an additional year of schooling on earnings) when we regress individual labour earnings y on years of education X in a sample of twins, with some other demographic variables (w). The standard IV regression model is obtained by augmenting the standard linear regression model with a model for the endogenous regressors and the instruments. The two-equation model describing natural logarithm of earnings and years of schooling is applied:

$$\begin{aligned} Y_i &= \beta S_i + \gamma_i X_i + \varepsilon_{1i} \\ S_i &= \gamma_2 Z_i + \varepsilon_{2i} \end{aligned} \quad , \tag{3.11}$$

where Earnings Y_i of individual i are determined by schooling S_i, X_i is a vector of exogenous variables that influence Y_i, Z_i is a vector of instrumental variables that influence the schooling decision, γ_2 represents the effect of the instruments on the endogenous regressors, ε_{2i} is a vector containing the error terms and $\text{cov}(Z_i, \varepsilon_{1i}) = 0$. The correlation between S_i and ε_{1i} (the degree of endogeneity), arises because of nonzero covariances between ε_{1i} and ε_{2i}. The errors are assumed to have mean zero. The most influential variables in the Z vector are the ability and family background variables. The vectors X_i and Z_i overlap with ability variables appearing in both equations.

Ashenfelter and Krueger (1994) determined the extent of measurement errors in the education variable on the earnings function by the use of a fixed effects IV regression model for twins as:

$$\begin{aligned} Y_1 &= \beta S_1^1 + \gamma_1 X_1 + \varepsilon_{11} \\ S_1^1 &= \gamma_2 S_1^2 + \varepsilon_{21} \end{aligned} \quad \text{for twin1} \tag{3.12}$$

$$\begin{aligned} Y_2 &= \beta S_2^2 + \gamma_1 X_2 + \varepsilon_{12} \\ S_2^2 &= \gamma_2 S_2^1 + \varepsilon_{22} \end{aligned} \quad \text{for twin2} \tag{3.13}$$

$$\begin{aligned} Y_1 - Y_2 &= \beta(S_1^1 - S_2^2) + (\varepsilon_{11} - \varepsilon_{12}) \\ S_1^1 - S_2^2 &= (S_1^2 - S_2^1) + (\varepsilon_{21} - \varepsilon_{22}) \end{aligned} \quad , \tag{3.14}$$

where $S_1^2 - S_1^1$ is the difference in the sibling reported estimates of the schooling levels and it is used as an instrumental variable in the fixed effect IV model, S_1^2 - twin 1's report of twin 2's educational level , S_2^1 - twin 2's report of twin 1's educational attainment and $\varepsilon_{21} - \varepsilon_{22}$ represents measurement error.

3.1.6 Linear Mixed Model – Hierarchical Linear Model

The hierarchical linear model (HLM) is also referred to as variance or covariance components model (Dempster et al. 1981), random coefficients model (Rosenberg, 1973), multilevel linear model (Mason et al. 1983), mixed-effects and random-effects model (Laird & Ware, 1982), and mixed linear model (Goldstein, 1986). Hierarchical linear models are listed under the rubric of the linear mixed model (Davidian & Giltinan, 1995) and can be considered as an extension of the standard linear regression model (Paterson & Goldstein, 1991).

The HLM was used to uncover unobservable differences in the returns to schooling for both monozygotic and dizygotic twins by analyzing the variance in the returns to schooling. It includes both fixed and random effects which take into account both individual and family unobserved heterogeneity. The HLM essentially estimates a random coefficient (on education) and decomposes the variance around the mean return into family heterogeneity, individual heterogeneity, and risk. The two-level HLM by Lindley & Smith, (1972) estimates the mean effect of education on earnings and the variance in returns to education around this mean. The linear modeling framework is assumed as follows:

$$\text{Level - 1: } Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij} \tag{3.15}$$

$$\text{Level - 2: } \beta_{0j} = \gamma_{00} + \gamma_{01} Z_{1j} + \mu_{0j} \tag{3.16}$$

Substitution of Equations (15) into (16) results in the equivalent mixed-model representation

$$Y_{ij} = \gamma_{00} + \beta_1 X_{1ij} + (\mu_{0j} + \gamma_{01} Z_{1ij} + \varepsilon_{ij}) \tag{3.17}$$

In this combined model, Equation (3.17), the fixed part of the model is the segment $(\gamma_{00} + \beta_1 X_{1ij})$, where the intercept is γ_{00}, the slope coefficient of X_{1ij} is β_1, X_{1ij} is the matrix of the individual level variables (e.g., age, sex), the random error now has three components $(\mu_{0j} + \gamma_{01} Z_{1ij} + \varepsilon_{ij})$, where μ_{0j} is the random effect of the j^{th} family, γ_{01} is the random effect of the j^{th} individual (twin), Z_{1ij} are the family-level variables, $\varepsilon_{ij} \sim N(0, \sigma^2)$, where σ^2 is the residual variance at level-1 after controlling for X_{1ij}.

The model as specified in Equation (3.17) will result in the estimation of the following parameters - Fixed-effect coefficients: γ_{00}, β_1, the level-1 variance component: σ^2, the level-2 variance-covariance components: $\mu_{0j}, \gamma_{01}, \varepsilon_{ij}$. The linear mixed model given in Equation (3.17) can be expressed by the following matrix notation, in which the fixed part component is the average or expectation $X_j\beta$; the random part component is $Z_j\mu_j$; and the individual-level random part is ε_j:

$$Y_j = X_j\beta + Z_j\mu_j + \varepsilon_j, \tag{3.18}$$

where $\mu_j \sim N(0,G)$ and $\varepsilon_j \sim N(0,R)$, Y_j is the vector for the observed values of the response variable; X_j is the known matrix of the predictors; β is the unknown fixed-effects parameter vector; ε_j is the unobserved vector of the level-1 random errors, which are assumed to be independent and follow a normal distribution, $\varepsilon_j \sim NID(0,\sigma^2 I)$. Z_j is the design matrix for random effects and μ_j is the vector of random-effect parameters, assumed to have an independent normal distribution, $\mu_j = NID(0,T)$, Henderson, (1975). The fixed effect would refer to the overall expected effect of an individual's education on income: the random effect gives information on whether or not this effect differs between families. The covariance matrix for the measurement or residual errors $R = Var(\varepsilon)$ has dimension $n \times n$, in most examples $R = \sigma^2 \mathrm{I}$. The covariance matrix for the random effect coefficients $G = \mathrm{var}(\mu)$ has dimension, $q \times q$, where q is the number of random effect coefficients. G contains two variance components, 1) Family and 2) Individual. The covariance matrix describing the covariance between any two observations in the data set can be calculated directly from the matrix representation of the model in the following way:

$$V = \mathrm{var}(y) = ZGZ' + R \tag{3.19}$$

Another fundamental outcome of HLM analyses is the intraclass correlation coefficient generated by running an unconditional model. The unconditional model expresses the individual-level outcome Y_{ij} by combining two linked models: one at the individual level (level 1) and another at the family level (level 2). The model at level 1 expresses an individual's outcome as the sum of the intercept for the individual's family and a random error term associated with each individual. At level 2, the individual intercept is expressed as a sum of the grand mean and sequences of random deviations from such mean. The HLM unconditional model (no covariate predictor) is written in the form:

Level-1 model:

$$\mathrm{Y}_{ij} = \beta_{0j} + \varepsilon_{ij}, \tag{3.20}$$

$$\mu_{0j} = N(0, \sigma_{\mu}^{2})$$

Level-2 model:

$$\beta_{0j} = \beta_{0} + \mu_{0j}, \tag{3.21}$$

$$\mu_{0j} = N(0, \sigma_{\mu}^{2})$$

Combined model:

$$\mathrm{Y}_{ij} = \beta_{o} + \mu_{j} + \varepsilon_{ij}, \tag{3.22}$$

$$Var(\mathrm{Y}_{ij}) = Var(\mu_{0j} + \varepsilon_{ij}) = \sigma_{\mu}^{2} + \sigma_{\varepsilon}^{2}$$
$$Cov(\mu_{0j}, \varepsilon_{ij}) = 0$$

The proportion of variance or the percentage of observed variation in the dependent variable attributable to family-level characteristics is known as the intraclass correlation. The intraclass correlation coefficient (ICC) is specified as:

$$\rho = \frac{\sigma_{\mu}^{2}}{\sigma_{\mu}^{2} + \sigma_{\varepsilon}^{2}}, \tag{3.23}$$

The percentage of variance attributable to individual level traits is easily found according to $(1-\rho)$. A researcher who has found a significant variance component for σ_{μ}^{2} may wish to incorporate macro level variables in an attempt to account for some of this variation. The average correlation (expressed in the so called intra-class correlation) between variables measured on siblings from the same family will be higher than the average correlation between variables measured on siblings from different families.

In addition, the intraclass correlation measures the extent to which observations are dependent on a grouping variable (e.g., families). The presence of a significant intraclass correlation is an indicator of the need to employ multi-level modeling rather than conventional regression (OLS). To pursue OLS regression modeling anyway in the face of lack of independence and lack of homoscedastic error variance, will mean that significance tests will not be accurate or OLS significance tests (and standard errors and confidence limits) are not at all robust when the assumption of independence is violated.

3.1.7 Return to Schooling by Ability Model

The simplest optimizing model of school choice is consistent with the well known stylized facts about the determinants of schooling choice and the relationship of earnings to schooling. These stylized facts capture 1) the family background of twins has a strong influence on educational attainment, and 2) the relationship of log earnings to schooling is essentially linear (Mincer, 1974). Individual heterogeneity in the optimal schooling choice arises from two sources: differences in the costs of (or tastes for) schooling and differences in the economic benefits of schooling also known as heterogeneity in the marginal return to schooling.

Following Ashenfelter and Rouse (1998), this study employs a simple structural model based on Becker's (1967) optimizing model of school choice focusing on the link between earnings, ability and education with the assumption that the marginal return to schooling varies by family and is correlated with the unobservable ability. The optimal level of education is likely determined endogenously as a function of the level of ability and other factors such as family background. It is assumed in the model that individuals attempt to maximize utility which is a function of income and schooling.

$$U(y,S) = \ln(y) - f(S) = \ln[g(S)] - f(S), \tag{3.24}$$

where $y = g(s)$ represents the observable relationship of earnings (y) to schooling (S), and $\ln[g(s)]$ and $f(s)$ are increasing convex functions that represent the (log) benefits and costs of schooling.

Maximizing utility in Equation (3.24) requires that optimal schooling (S^*) satisfy the first-order condition,

$$\frac{g'(S)}{g(S)} = f'(S) \tag{3.25}$$

In order to implement this model empirically, functional forms for the marginal (proportional) benefits are chosen. To capture the well known stylized fact that (log) earnings is a linear function of schooling that may vary across individuals, for individual i, in family, j, the marginal benefit MB_{ij} of schooling is represented by,

$$MB_{ij} = \frac{g'(S_{ij})}{g(S_{ij})} = b_j + \Theta A_{ij}, \tag{3.26}$$

where A_{ij} is unobserved "ability" of the individual or twin i.

Assuming that the marginal cost MC_{ij} of schooling has the simple form,

$$MC_{ij} = f'(S_{ij}) = r_j + r_0 S_{ij}, \tag{3.27}$$

It follows that the optimal level of schooling is,

$$S_{ij} = \frac{b_j - r_j}{r_0} + \frac{\Theta}{r_0} A_{ij} = S_j^* + \frac{\Theta}{r_0} A_{ij}, \tag{3.28}$$

which varies across families S_j^*, and may also vary by individual ability if Θ differs from zero. It is clear from Equation (3.28) that schooling varies within the family and is correlated with within-family differences in ability. It follows that the key assumption identifying the return to schooling from within-family variability in schooling levels is that $\Theta = 0$; that is, any differences in schooling are determined by differences in tastes or other characteristics that are uncorrelated with the unobserved determinants of earnings, i.e., "optimization errors". The model we fit assume that $\Theta = 0$, has the (theoretically) testable proposition that individuals in the same families will have the desired schooling levels. That is, observed differences in schooling levels, S_{ij} are due to measurement or optimization errors ε_{ij} so that

$$S_{ij} = S_j^* + \varepsilon_{ij}, \tag{3.29}$$

where, the errors ε_{ij} have mean zero and must be independent of the optimal desired schooling level S_j^*.

Integration of MB_{ij} with respect to S_{ij} and assuming $\Theta = 0$ gives the well known log wage equation

$$Y_{1j} = A_j + b_j S_{1j} + dX_j + \varepsilon_{1j}, \tag{3.30}$$

$$Y_{2j} = A_j + b_j S_{2j} + dX_j + \varepsilon_{2j}, \tag{3.31}$$

where Y_{1j} and Y_{2j} are the logarithms of the wage rates of the first and second twins in a pair, S_{1j} and S_{2j} are the schooling levels of the twins (or, more generally, all attributes that vary within families), X_j are other observable determinants of wages that vary across families, but not within twins (such as race and age), and ε_{1j} and ε_{2j} are unobservable individual components. A_j is an unobserved family component that represents an unspecified combination of innate (inherited) "ability", family environment, or general unobserved skills, and may be correlated with attained schooling levels. The return to schooling is b_j. According to this model, there may be two types of ability: A_j, which confers higher earnings at all levels of schooling ("absolute advantage"), and b_j, which confers higher net returns to schooling and may also be correlated with ability ("comparative advantage). Under this scenario, the marginal return to schooling according to Ashenfelter and Rouse (1998) can be written as:

$$\beta_i = \beta_0 + \beta_1 A_i, \tag{3.32}$$

where, the parameter, β_1 indicates the degree of heterogeneity in the return to schooling that results from the distribution of "abilities" or "learning environments" across families. If families with higher levels of innate 'ability' or more favourable learning environments for their children gain more benefit more from schooling, then β_1 should be positive. Ashenfelter and Rouse (1998) propose that the earnings for twin 1 and twin 2 where schooling returns varies with ability may be written as:

$$Y_{1i} = \alpha X_i + \beta_0 S_{1i} + \gamma\left[\frac{1}{2}(S_{1i} + S_{2i})\right] + \beta_1\gamma\left[\frac{1}{2}(S_{1i} + S_{2i})\right]S_{1i} + \varepsilon_{1i} \tag{3.33}$$

$$Y_{2i} = \alpha X_i + \beta_0 S_{1i} + \gamma\left[\frac{1}{2}(S_{1i} + S_{2i})\right] + \beta_1\gamma\left[\frac{1}{2}(S_{1i} + S_{2i})\right]S_{2i} + \varepsilon_{2i} \tag{3.34}$$

where ε_{1i} and ε_{2i} are the error terms. Equations (3.33) and (3.34) include an interaction term between the individual's schooling level and the family's average schooling level. The coefficient of this interaction term, $\beta_1\gamma$ is the product of the two types of ability bias: the correlation of the marginal benefits of schooling with ability β_1 and the correlation between the level of ability and schooling, γ.

3.2 Estimation Methods of the Models

In order to estimate the returns to education in this study, the properties of estimation procedures for each of the model specifications that were considered above are outlined. The method of least squares is used to estimate the regression coefficients in Mincer's human capital model in Equation (3.1) and the fixed effects model in Equation (3.5). These two models take the form of a multiple linear regression model. The second estimation method is the two-stage least squares (2SLS) estimator also known as the Instrumental variable (IV) technique. The IV approach was used to obtain consistent estimates when solving the endogeneity of schooling problem. In addition to an IV method of estimation, estimates were obtained using Feasible Generalised Least Squares (FGLS). This method is the best estimator that allows cross-equation restrictions on the coefficients apparent in the fixed effects model (Equations (3) and (4)) and can also produce unbiased estimates in the selection effects model proposed by Ashenfelter and Krueger (1994). Finally, we use the Restricted Maximum Likelihood (REML) method which is one of the recommended methods for estimating variances and covariances in the hierarchical linear model. Residual maximum likelihood (REML) is often preferred to maximum likelihood (ML) estimation as a method estimating covariance parameters in linear models because it takes account of the loss of degress of freedom in estimating the mean and produces unbiased estimation equations for the variance parameters (Wu et al. 2001; Smyth and Verbyla, 1996).

3.2.1 Least Squares estimation (OLS) of the Regression Coefficients

The method of least squares is the most common procedure used to estimate the regression coefficients in a linear regression model. The OLS estimator is specified as

$$\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\varepsilon, \quad (3.35)$$

where the variance of the sampling distribution of $\hat{\beta}$ is $Var(\hat{\beta}) = \sigma_{\varepsilon}^2(X'X)^{-1}$. The variance of the error term, σ_{ε}^2, is estimated by calculating the regression residuals $\hat{\varepsilon} = y - X\hat{\beta}$ and their sample variance $\sigma_{\varepsilon}^2 = \frac{\hat{\varepsilon}'\varepsilon}{n-k}$. The n-k in the denominator is the model 'degrees of freedom' and is only relevant as a bias correction when k is large, or n is small, otherwise n is acceptable. The diagonal elements of (square) matrix $(X'X)^{-1}\sigma^2$ are the variances of each estimated parameter in $\hat{\beta}$. The square roots of the elements are the standard errors. Hypothesis tests regarding the parameter values are based on these standard errors. Considering the linear models in Equations (3.1 and 3.2), an unbiased estimator of the parameter vector β is an estimator $\hat{\beta}$ that produces estimates that are, on average, equal, β the true parameter. The OLS estimator has the following desirable properties

1. $\hat{\beta}$ is a Linear function of the sample values of Y, where $Y_i : (i = 1,....,N)$,

   $\hat{\beta} = \Sigma_i k_i Y_i$, where $k_i = \frac{x_i}{\Sigma_i x_i^2}$, $(i = 1,....,N)$,

2. $\hat{\beta}$ is Unbiased: $E(\hat{\beta}) = \beta$ which means that $E\left|\hat{\beta} - \beta\right| = 0$,

   $E\left|\hat{\beta} - \beta\right| = (X'X)^{-1}X'\varepsilon$, which requires that $E[X'\varepsilon] = 0$,

3. $\hat{\beta}$ is Consistent, $E(\varepsilon) = 0; Cov(\varepsilon, X) = 0$,

4. $\hat{\beta}$ is Efficient, i.e., It has minimum variance - $Var(\varepsilon, X) = \sigma^2$,

5. $\hat{\beta}$ is Asymptotically normal, $\varepsilon \sim N(0, \sigma^2 I)$.

3.2.2 Two stage least squares (2SLS) or IV estimation

Two-stage least squares (2SLS) regression analysis technique is used when the error terms are correlated with the independent variables. 2SLS is a method of extending regression to cover models which violate ordinary least squares (OLS) regression's assumption of recursivity.

A 2-stage least squares method or 2SLS is only performed when $l > k$, where k = number of independent variables and l is a set of variables of a matrix (Z). Otherwise when $l = k$ then only one stage is enough. Two stages in 2SLS refer to

1. A stage in which new dependent or endogenous variables are created to substitute for the original ones,
2. A stage in which the regression is computed in OLS fashion, but using the newly created variables.

When $l = k$	When $l > k$
$\hat{\beta}_{2SLS} = (X'ZZ'X)^{-1}X'ZZ'y =$ solution vector	<u>Stage1:</u> • X is regressed on Z $\hat{X} = Z(Z'Z)^{-1}Z'X$
	<u>Stage 2:</u> • Y is a regressed on X $\beta_{2SLS} = \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}X'Z(Z'Z)^{-1}Z'y$

The purpose of the first stage is to create new dependent variables which do not violate OLS regression's recursivity assumption. The IV estimates is computed in two stages using least squares regressions. Hence, this method is sometimes called two-stage least squares (2SLS). For an instrumental variable (an "instrument") Z to be valid, it must satisfy two conditions

1. Instrument relevance: $Cov(z, x) \neq 0$

2. Instrument exogeneity: $Cov(z, \varepsilon) = 0$

3.2.3 Generalized least squares (GLS)

Generalized least squares (GLS) estimator is a technique for estimating the unknown parameters in a set of equations in a linear regression model. The GLS is used in an attempt to exploit the information in the correlated errors simultaneously in order to achieve greater efficiency in the estimates. OLS will yield unbiased & consistent estimates for each separate equation. However, because the approach ignores the correlation of the disturbances the estimates will not be efficient. The GLS is applied when the variances of the observations are unequal (heteroscedasticity), or when there is a certain degree of correlation between the observations. In these cases ordinary least squares can be statistically inefficient, or even give misleading inferences.

Although Equation (3.10) can be fitted by ordinary least squares (OLS), generalized least squares (GLS) is the optimal estimator for these equations because of the cross-equation restrictions on the coefficients. Generalized least squares also provides the appropriate estimates of standard errors for the estimated coefficients and can directly estimate both the return to education and the ability or family background effect.

The SUR model is usually estimated using the GLS method. Estimation via Generalized Least Squares takes into account the variability across equations and will yield BLU (best, linear, unbiased) estimates. This is a two-step method where in the first step OLS is used to estimate Equation 3.6. The residuals from this regression are used to estimate the elements of matrix Σ:

$$\hat{\sigma}_{mp} = \frac{1}{T}\hat{\varepsilon}'_m \hat{\varepsilon}_p$$

In the second step if Σ is known,

parameter estimates can be obtained by using the generalized least squares (GLS) estimator a

$$b_{GLS} = \left[X'\left(\Sigma^{-1} \otimes I_N\right)X\right]^{-1} X'\left(\Sigma^{-1} \otimes I_N\right)y \qquad (3.36)$$

This estimator is unbiased in small samples assuming the error terms ε_{mp} have symmetric distribution; in large samples it is consistent and asymptotically normal with limiting distribution

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \left(\frac{1}{T}X'\left(\Sigma^{-1} \otimes I_T\right)\right)X\right)^{-1}$$

In practice, however, Σ is rarely known and for this case feasible generalized least squares (FGLS) estimators have been proposed. The FGLS estimator is given as:

$$\tilde{\beta} = \left(X'\hat{\Omega}^{-1}X\right)^{-1}\left(X'\hat{\Omega}^{-1}y\right), \text{ where, } \Omega = E[\varepsilon\varepsilon'] \qquad (3.37)$$

and the variance - covariance (VCV) estimator as

$$VCV = \left(\tilde{\beta} - \beta\right)\left(\tilde{\beta} - \beta\right)^{-1} = \left(X'\Omega^{-1}X\right)^{-1} \tag{3.38}$$

The inclusion of Ω^{-1} improves the efficiency of the estimates, especially when the disturbances are highly correlated, but the independent variables are not.

Equivalence of SUR to OLS. There are two important cases when the SUR estimates turn out to be equivalent to the equation-by-equation OLS, so that there is no gain in estimating the system jointly. These cases are

1. When the matrix Σ is known to be diagonal, that is, there are no cross-equation correlations between the error terms. In this case the system becomes not seemingly but truly unrelated.
2. When each equation contains exactly the same set of regressors, that is, $X_1 = X_2 = \ldots = X_m$. The estimators turn out to be numerically identical to OLS estimates follows from Kruskal's theorem, or can be shown via the direct calculation.

3.2.4 Restricted Maximum Likelihood (REML) estimation of Variance Components

In multilevel estimation, restricted maximum likelihood (REML) estimation is one of the most commonly used procedures for estimation. In statistics, restricted (or residual) maximum likelihood (REML) is a method for fitting linear mixed models. In contrast to conventional maximum likelihood estimation, REML can produce unbiased estimates of variance and covariance parameters (Smyth and Verbyla, 1996) and relies on classical asymptotic theory and assume normally distributed errors. The method was first described by Patterson and Thompson (1971), although they did not use the term REML. A review of the early literature was given by Harville (1977). REML estimation is available in a number of general-purpose statistical software packages, including Genstat (the REML directive), SAS (the MIXED procedure), SPSS (the MIXED command), Stata (the xtmixed command), and R (the nlme package), as well as in more specialist packages such as MLwiN and ASReml.

Restricted maximum likelihood (REML) estimators were developed to correct for the downward bias found in the usual maximum likelihood estimates. In contrast to the ML estimators, REML estimators do take into account the loss in degrees of freedom which occurs from estimating the fixed effects parameters. It is also known that the REML procedure produces estimates that are at least as good as the ML and sometimes better. The variance-covariance components are also estimated with less bias using the restricted maximum likelihood procedure rather than the maximum likelihood in many situations (Wu et al. 2001).

The problem of finding residual maximum likelihood (REML) estimate of variance components in the mixed linear model reduces to computing the maximum of the marginal likelihood function or log-likelihood function corresponding to a maximal invariant (Rao, 1997).

$$l_{REML}(G,R) = -\frac{1}{2}\log|V| - \frac{1}{2}\log|A^T V^{-1} A| - \frac{(N-p)}{2}\log r^T V^{-1} r - \frac{(N-p)}{2}\left[1+\log\frac{2\pi}{(N-p)}\right], \quad (3.39)$$

where $r = Y - A\left(A^T V^{-1} A\right)^{T} A^T V^{-1} Y$ and $p = rank(A)$

3.3 Econometric Issues

There are a number of econometric issues that may lead to bias in estimates of returns to education. These include significant issues like ability bias or unobserved heterogeneity, endogeneity bias and measurement error. Several approaches to these problems have been developed and this study will address this issue following the methodology of Ashenfelter and Krueger, 1994 and Miller et al. (1995).

The rate of return to schooling is an important factor in determining educational attainment and participation and, ultimately, earnings. There are a variety of sources of bias associated with ordinary least-squares (OLS) estimates of the return to schooling. The three well-known arguments that explain why OLS may render inconsistent return estimates are endogeneity bias, ability bias and measurement error in the schooling variable.

3.3.1 Endogeneity Bias

Endogeneity bias arises where the dependent variable (e.g., annual earnings) has a causal effect on one (e.g., education) or more of the explanatory variables. This could occur if higher levels of education lead to higher earnings and, at the same time, higher earnings contribute to higher levels of education. Failing to account for the feedback effects of earnings on education can lead to biased estimates of the effects of education on earnings. Endogeneity between education and earnings can also arise because of the possible presence of unobservable characteristics that influence both earnings and the likelihood of completing school. Estimates of the effect of years of schooling on earnings may therefore be biased as a result of these unobserved factors that affect both education and earnings. A recent solution to this endogeneity problem has been found in identifying exogenous sources of variation in schooling to build a new set of instrumental variables for years of education attained (Angrist and Krueger 1991; Card 1998).

In some cases IV estimation results suggest a downward bias in OLS estimates (i.e. the rate of return for the IV estimate is much larger than the OLS estimate). Card (1993) suggests that the estimates of schooling returns using IV methods are almost double those found using OLS. Butcher and Case (1994) also show that IV methods produce double the OLS estimates. Findings in Ashenfelter and Krueger (1994) and Ashenfelter and Zimmerman (1993) which use twins and siblings data, also show evidence of much larger rates of return to education than OLS suggests.

Furthermore, in an optimizing investment model we would expect a positive correlation between schooling and its return, which would imply that OLS estimates of the rate of return would be biased upward. Ashenfelter and Krueger (1994) and Ashenfelter and Rouse (1998) provide estimates of the returns to education and the resulting endogeneity bias (to which they refer to as a "selection effect"). The model of optimal schooling choices that they used suggests that we use measures of the education of a twin's sibling, the average education of the twins, or father's education as an additional regressor to control for any "family" effect that affect the absolute level of earnings.

Testing for Endogeneity - Hausman Specification Test.

Hausman's specification test (Hausman, 1978), or *m*-statistic, can be used to test hypotheses in terms of bias or inconsistency of an estimator. This test was also proposed by Wu (1973). Hausman's *m*-statistic is as follows.

Given two estimators, $\hat{\beta}_0$ and $\hat{\beta}_1$, where under the null hypothesis both estimators are consistent but only $\hat{\beta}_0$ is asymptotically efficient and under the alternative hypothesis only $\hat{\beta}_1$ is consistent, and the *m* statistic is

$$m = \hat{q}'\left(\hat{V}_1 - \hat{V}_0\right)^{-1}, \tag{3.40}$$

where $\hat{V}_1$ and $\hat{V}_0$ represent consistent estimates of the asymptotic covariance matrices of $\hat{\beta}_1$ and $\hat{\beta}_0$, and $\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$ The *m*-statistic is then distributed χ^2 with *k* degrees of freedom, where *k* is the rank of the matrix $\hat{V}_1 - \hat{V}_0$. A generalized inverse is used, as recommended by Hausman (1982). The Hausman's *m*-statistic can be used to determine if it is necessary to use an instrumental variables method rather than a more efficient OLS estimation.

3.3.2 Unobserved Heterogeneity - Ability Bias

Despite several decades of research in labor economics and dramatic improvements in data collection, a large fraction of the variation in earnings levels among individuals still remains unexplained. As a consequence, the study of unobserved sources of heterogeneity is important not only for the well-known purpose of correctly estimating earnings effects of key observed factors such as schooling, but also for our understanding and assessment of some of these less tractable determinants of earnings. In econometric terms, 'unobserved heterogeneity' describes a situation where some unobserved characteristics (such as a person's innate ability or their work ethics) are related to both the dependent variable (e.g., earnings) and one or more independent variables (e.g., education). Statistical inferences may be erroneous if, in addition to the observed variables under study, there exist other relevant variables that are unobserved, but correlated with the observed variables. The source of heterogeneity is typically attributed to data limitations and the unobservability of certain productivity related factors. A second source of heterogeneity is related to preferences. For example, differences in the relative valuation of earnings with other on-the-job and off-the-job amenities are likely to affect occupational and job choices, and thus give rise to earnings differentials. Unobserved heterogeneity could arise in the context of the relationship between education and earnings. If an unobserved variable (e.g., innate ability, including drive, motivation, discipline, focus, charisma and communication skills) leads to better education and higher earnings, estimated coefficients for the effects of education on earnings might be biased and not reflect the true underlying effects of education on wages. Unobserved heterogeneity is a potential problem when estimating the relationship between education and earnings. 'Ability bias' is a specific form of unobserved heterogeneity that refers to the possibility that some people have innate abilities (such as cognitive ability) that would make it easier for them to complete education. "Ability" bias is due not only to the impact of endowed ability on schooling but also includes effects of correlated family background factors that directly affect schooling. Even in the absence of formal education, these characteristics would be sought after in the labour market and rewarded with higher wages. Therefore, some of the benefits that are associated with education might have more to do with the person's innate characteristics than their level of education, and estimates of the effects of education on wages might be biased. The impact of any ability biases in the returns to education is therefore a major issue arising in estimating results. The difficulty rests with measuring the effects of ability on income independent of education levels (Wei, 2008).

It is often assumed that natural ability and educational attainment are positively correlated but it is possible that the correlation goes the other way with the more able leaving education in pursuit of earnings opportunities while those with lower ability continuing in the system (Leigh, 2008). However, many US studies summarized by Card (1999 and 2001) show that ability biases to estimates of the return to education are not large.

The role of family background in the returns to education has also been discussed in the literature as it is likely to confound the returns to education. Weale (1993), for example, reports British results which show that once the occupation of the parents is taken into account, the private return to an additional year of education is lower for people from a higher occupational background than for those from less privileged backgrounds. Twin studies of the return to schooling are one attempt to try to control for the effects of family background and to identify a "pure" effect of education on earnings. Methods for obtaining valid statistical inferences in the presence of unobserved heterogeneity include the instrumental variables method; multilevel models, including fixed effects and random effects models; the Heckman correction for selection bias .

OLS estimates of the effect of education on earnings are consistent only if, for example, unobserved variables are not correlated with both education and earnings. However, if an unobserved characteristic, (e.g., ability) has a positive effect on earnings and schooling, then OLS estimates of the returns to schooling will be biased upwards. Ability has an effect on earnings independent of education but is positively/negatively correlated with schooling and usually not controlled for in regression analysis. Using the ordinary least squares (OLS) method to estimate β results in a bias such that

$$E(\beta_{OLS}) = \beta + \gamma \frac{\sigma_{AS}}{\sigma_S^2} \neq \beta ,$$

where β is the return to education, A denotes ability and S is the level of education attained.

As ability is thought to be positively correlated with education and earnings, the estimate of β_{OLS} is assumed to be upward biased. Hence, there is a risk that estimations of the education effect on earnings in previous literature are upward biased, or even that the entire effect of education on earnings is due to ability bias. Controlling for ability bias has a twofold importance. First of all, we need to have a better understanding of the true, unbiased returns to education. Second, we would like to know how much of earnings differential is due to ability selection and how much of the differential still persists even after controlling for ability.

Solutions to ability bias include: Using the twins (or siblings) approach to attempt to eliminate ability bias by exploiting the differences between twins in levels of schooling and earnings, on the grounds that this eliminates differences in innate ability or motivation. Ashenfelter and Krueger, 1994 and Ashenfelter and Rouse, 1998) treats ability as an unobserved family effect and estimates a "fixed effects model" based on a version of Mincer differenced equation for each twin pair. Taubman (1976) provided an early example of such a study, which found that the return to schooling was only 3 percent. Laplagne et al. (2007) used HILDA data to estimate the effects of education on labour force participation. They used a series of econometric tests to test for the presence of unobserved heterogeneity, and found statistically significant evidence of unobserved heterogeneity in the data. They concluded that 'unobserved heterogeneity means that the coefficients from the standard multinomial logit model are likely to be biased upward' (Laplagne et al. 2007, p. 45). To the extent that labour productivity is explained by inherent ability (rather than by education), the ability of governments to increase labour productivity by increasing average education levels is lower than would be implied by estimates of the effects of education on wages (as a proxy for productivity).

Leigh (2007) estimated the returns to education in Australia using HILDA data. As part of his analysis Leigh reviewed Australian and overseas literature on ability bias — that is, the extent to which unobserved characteristics account for both the level of education and the measure of performance. Depending on the method used, Australian estimates of ability bias were between 9 per cent and 39 per cent. Overseas estimates ranged from 10 per cent to 60 per cent. For the purposes of his analysis, Leigh assumed that ability bias meant that estimates of the returns to education were biased upward by 10 per cent.

3.3.3 Measurement Error

A third source of potential bias in the estimation of the returns to education is measurement error. Measurement error is the variation between measurements of the same quantity (in our case this refers to measurement error in the schooling variable) on the same individual. Measurement error arises from mis- reporting or mis-measurement of educational attainment. Griliches (1977) argued that measurement errors in schooling would lead to a downward bias in the OLS estimate of the effect of schooling on earnings that could partially offset any upward ability biases. A conventional assumption is that observed schooling S_i^* differs from true schooling S_i by an additive error such that

$$S_i^* = S_i + \varepsilon_i,$$

where ε_i is a random variable that satisfies $E[\varepsilon_i] = 0$ and $E[\varepsilon_i^2] = \sigma_i^2$ and uncorrelated with earnings.

A main drawback of OLS estimates of returns to education is that they suffer from omitted variable bias. But correcting for omitted variable bias using fixed-effects greatly exacerbates the measurement-error bias. Thus, in order to obtain credible fixed-effects estimates, they must be corrected for measurement errors in reported schooling. If schooling is measured with error, this would imply that the bias from measurement error in schooling is likely to increase by forming differences between twins, and even more so when differencing between identical twins, causing the fixed-effects estimates to be biased downwards (Griliches, 1979). Classical measurement error in the fixed effects model is assumed and the method of instrumental variables is applied using the independent measures of the schooling variables as instruments (each twin's schooling report of the other twin) resulting in the fixed effects IV model written as

$$\begin{aligned} Y_{1i} - Y_{2i} &= \beta\left(S_1^1 - S_2^2\right) + \varepsilon_{1i} - \varepsilon_{2i} \\ &= \beta \Delta S' + \Delta\varepsilon \end{aligned} \tag{3.41}$$

The classic solution to the measurement error problem in twin studies proposed by Ashenfelter and Krueger (1994), was to instrument for one's own education, using the co-twins report on one's own education. This yielded estimates of schooling returns of about 16% per year of schooling; a threefold increase over typical findings. Other studies emphasize that OLS estimates may be biased downward by roughly 10% to 20% due to measurement error (Griliches (1977) and Blackburn and David Neumark, 1995). Existing estimates for the ranges of these biases are based on various IV strategies, and different instruments may lead to different estimated combinations of private returns. Hertz (2003) finds that measurement error corrected estimates in South Africa are considerably lower than the OLS estimates. Yet, Duflo (2001) shows that ability bias and measurement error bias approximately offset each other in Indonesia.

Hence the debate on the size and the direction of the bias in the returns estimate still continues, with mixed evidence from different countries at different times.

CHAPTER FOUR

METHODOLOGY

This chapter describes the design and research methodology that was implemented to describe the relationship between education and earnings and outcome expectancies (economic benefits and issues) in Ghana. It also includes a description of the sample size and characteristics, the research settings, the procedures for sampling and data collection. Finally, this chapter describes the instruments used as well as the data analysis procedures.

4.1 Study Area

The study areas are three Administrative Regions in Ghana, namely; Greater Accra Region, Ashanti Region and Western Region. The three regions were purposively selected due to the fact that over 60% of the population of the country lives in these 3 out of the total 10 Regions of the country. The study limited itself to only the city centers of these regions. The reason for selecting city centers was that people who live in rural areas were mostly not very much educated and primarily involved in subsistent farming and therefore data on individual earnings would be difficult to acquire.

4.2 Sources of Data

This study obtained primary data collected from a sample of Ghanaian adult twins aged between 18 and 65 years who were gainfully employed. These twins were identified by a team during a 2007 and 2008 labour market twins' survey, through various channels, including colleagues, friends, relatives, members of twins clubs in Greater Accra, Ashanti and Western Regions, twins at various work places, markets, shops and a number of households. Overall, these channels permitted a roughly equal probability of contacting all of the twins in these cities, and thus the twins sample that was obtained was approximately representative.

Reliable and up-to-date information (secondary data) available in the published literature on the relationship between earnings and education was also accessed using grey literature, internet sources, journals, e-library etc. This was done in order to obtain some better insight of the economic benefits of additional education on earnings and to support the theoretical as well as the methodological part of this thesis.

4.3 Data Collection

The data that was collected is the first twins' dataset in Ghana with the objective of finding the socioeconomic relationship between earnings and education. Altogether there are 250 individuals, of which 144 individuals are non-identical twins and 106 individuals are identical twins. Data was collected on a wide range of socio-economic characteristics of respondents by a team of 5 interviewers between December 2007 and January 2008. A total of 130 pairs of twins were interviewed during the survey, but 5 pairs were not included in the data analysis because they were below 18 years of age. As with the data analysed by Ashenfelter and Krueger (1994), Miller et al. (1995) and Ashenfelter and Rouse (1998), each twin provided reports on both their own level of education and on that of their co-twin. This permits application of the instrumental variable (IV) estimators proposed by Ashenfelter and Krueger (1994) and Black et al. (2000).

The thesis data collection methods were structured such that where both twins respond that they have identical facial color, looks, gender, and age they were classified as being identical twins. If not the twin was termed as being fraternal or dizygotic or non-identical. Qualitative methods that were used to collect information for analysis, included:

1. Structured Questionnaires
2. Informal or key informant interviews

4.3.1 Structured Questionnaire Administration

A questionnaire was developed and administered to twins (see Appendix 1 for details of the questionnaire. The questionnaire was developed based on experiences gathered and results of previous studies in the area. Excerpts from Ashenfelter and Krueger's Twinsburg Questionnaire in the United States" were also included. The questionnaire addressed issues related to genetics, family background, schooling and earnings in the labour market since they are all important determinants of earnings in the labour market. The influence of genetics on earnings has been studied using monozygotic twins. Such a study involving monozygotic twins who are genetically identical especially raised together and with the same family background is a powerful tool for examining the roles of genetics and family background as mediating influences in the relationship between schooling and earnings (Ashenfelter and Kruegar, 1994), (Behrman et al. 1977; Miller et al. 1995). A study of determinants of earnings in the labour market is important since a number of studies have confirmed that better-educated people are more successful in the labour market, (Asplund and Pereira 1999; Ashenfelter and Rouse 1999). The most reliable determinant of genetic constitution is DNA-tests. However one of the key constraints in this study is the difficulty of undertaking DNA-tests in the country. Such tests are currently expensive and only done in a few places. Therefore in this study twins were rather asked to give information on whether they are monozygotic or dizygotic. Variables that were assessed by the questionnaire included:

1. Highest level of educational attainment
2. Family Background characteristics (e.g. mother's age)
3. Parent's educational attainment, (years of schooling completed)
4. Parent's occupational status
5. Type of assets owned
6. Ethnicity
7. Demographic characteristics (age, gender, marital status, etc.)
8. Personality traits

9. Zygosity diagnosis (Do the twin pairs look as alike as two peas in a pod? Is it hard for strangers to tell them apart? etc.)

The questionnaire was pre-tested in the Kumasi Metropolis before large-scale administration. Most of the questionnaires were completed through face-to-face personal interviews.

4.3.2 Informal Interviews

Informal meetings were also held with twin groups and their families using the Participatory Rapid Appraisal (PRA)[1] method to determine their perception of the contribution of genetics, family background and schooling to income generation.

4.4 Data Analysis

4.4.1 Variables used in the Estimation

The variables used to estimate the earnings equations are defined in Table 1 (Chapter 2, section 2.4). The independent variables are education, including father and mother's educational level, age, gender and marital status, while the dependent variable is log earnings. These independent variables relate to characteristics that can affect an individual's earning capacity. Broadly, these can be considered as labour market and demographic variables. The education variable denoted by 'Years of schooling' was assumed to be a continuous variable and was constructed from the education and training qualifications individuals

[1] Participatory Rapid Appraisal (PRA) is an intensive, systematic and semi-structured learning experience carried out in the community by a multi-disciplinary team including local community members

report that was completed by the twins. The representation of an individual's education is provided by their highest level of education attained (degree or higher; diploma or certificate, etc.). Number of years of schooling is considered a reliable indicator of the level of educational attainment in accordance with a number of other studies (Card, 1995; Mincer, 1974; Becker, 1964). A linear relationship between educational levels and years of education completed is assumed. The construction of the education measure is further presented in the appendix.

The age variable, which denotes "age distributions of earnings" provide a clearer picture of when the higher earnings are received by the individual who has invested in education. Age is included in the model to control for age effects on returns to education. There was also the need to include a gender variable in the regression equation because studies have shown that there are significant differences in returns to education between men and women (Fitzenberger et al. 2004; Schnabel and Schnabel, 2005). A dummy variable for gender was therefore included in the regression equation whereby women were assigned a value of one and men a value of zero.

The dependent variable which is the natural logarithm of annual earnings was derived from gross wage or salary income (from all jobs). The wage rate is an indicator of an individuals' productivity. The distribution of log earnings is close to normal and all things being equal this study models the natural logarithm of annual earnings.

4.4.2 Models, Estimation Methods and Econometric Issues

A number of models and estimation methods were used in this study to determine the returns to schooling. The models and estimation methods are used to explain econometric issues associated with the relationship between schooling and earnings. These econometric issues addressed in this study include ability bias, endogeneity bias and measurement error bias.

Table 4.2: Description of Model Variables

Variable	Description
Dependent variable Log earnings - Y_i	Natural log of the annual earnings
Independent Variables	
Number of Years of Completed Education	No education = 0years, Primary = 1-6years, Middle/JSS = 7-10years, Secondary = 11-17years and Higher = 18-25years
Age	15<age<70
Age squared	-
Gender	Male = 0, Female = 1
Marital Status	Not married = 0, (constitutes living together, separated and divorced) Married = 1
Father's education	No education = 0years, Primary = 1-6years, Middle/JSS = 7-10years, Secondary = 11-17years and Higher = 18-25years
Mother's education	No education = 0years, Primary = 1-6years, Middle/JSS = 7-10years, Secondary = 11-17years and Higher = 18-25years

Models

The models used establish the relationship between education and annual earnings by employing the

1. Mincer's Human Capital Earnings Function,
2. Linear Regression Model,
3. Instrumental Variable (IV) Regression Model,
4. Fixed Effects Model,
5. Selection Effects Model,
6. Linear Mixed Model.

These models were used to estimate the returns to schooling using three data sets, namely; pooled twins sample, MZ twins sample and DZ twins' sample. Furthermore, this study assumes the relationship between education and earnings to be linear, even though other functional forms are possible.

Estimation Methods and Econometric Issues

Mincer's Human Capital Earnings Function (HCEF) is the primary economic model that economists use to measure returns to education and therefore recent studies of education and earnings determination are almost always embedded in the HCEF framework developed by Mincer (1974). Mincer's HCEF was used to obtain the economic return to schooling (β_1) for male and female twins by applying the simple Ordinary Least Squares (OLS) estimator.

The return to education in this study was also estimated based on a linear model (LM) derived by Chamberlain, (1982). It is often argued that age may be a more important determinant of earnings than potential experience, since it may be better able to capture elements of a worker's personality, such as maturity, that are valued by the employers. Moreover, in many award systems seniority plays a very important role – especially in the public sector and some large private sector corporations. For this reason using the linear model, the standard Mincer model is re-estimated using age instead of potential experience. In addition, age rather than potential work experience as proposed by Mincer, 1974 in the HCEF is used as a control variable because of possible measurement error in education. The linear model was fitted by the ordinary least squares (OLS), feasible generalized least squares (FGLS) and two-stage least squares (2SLS) estimators to obtain the return to education. A test for equality in the coefficients between MZ and DZ twins of explanatory variables used in the Mincer's equation was performed to justify the regression analysis on the pooled sample.

Ashenfelter and Krueger (1994) noted that the conventional economic models (HCEF and LM) used above do not expressly incorporate the influence of natural ability or family environment in the estimation of the returns to schooling. Ashenfelter and Krueger (1994) therefore proposed the application of the fixed-effects model to assess the extent of bias in conventional rates of return to schooling. Separate analysis for monozygotic (MZ) and dizygotic (DZ) twins were performed. Genetic endowments and common environment were accounted for using the MZ twins. Whereas common environment influences only were controlled for using DZ twins (Behrman, Taubman, 1976). To obtain the true return to education devoid of any bias, the fixed effects model was fitted with the

OLS to control for omitted ability (made up of a genetic effect and a family effect

which therefore disappears with differencing between family members with the same genes) bias. On the other hand, the 2SLS method of estimation was used to account for measurement errors. Consequently, the difference between cross-sectional estimates (OLS) of the linear model and the fixed effects twin-differencing estimates by OLS (FEOLS) was used as a measure of ability bias in the education coefficient. If the estimates do not differ there is no indication of ability bias, but if the twin-differencing coefficient is smaller then there is evidence of ability bias in the cross-sectional estimates.

The Instrumental Variable (IV) Regression Model was adopted to overcome the potential bias of the OLS estimates caused by measurement errors in schooling reports and endogeneity. To incorporate the effect of measurement errors on the returns to education on the twins' data used in this study, the innovative approach of Ashenfelter and Krueger (1994) was adopted to obtain a good instrumental variable (i.e., the co-twins report on one's own education). This variable is clearly independent of measurement error and it is expected to correlate with the true schooling level of a twin, but uncorrelated with any measurement error that might be contained in the self-report. The IV regression model was then fitted by the two-stage least squares (2SLS) method of estimation to obtain the economic returns to education corrected for endogeneity.

In addition, in this study the FEOLS estimates were compared to the fixed effects two-stage least squares estimates (FE2SLS) to determine the extent of measurement error in the education coefficient. Subsequently, the 2SLS/IV and the first difference IV (FEIV) estimates were also comparatively evaluated as they provide an estimate of the magnitude of ability bias and both control for measurement error. This study also accounted for endogeneity of schooling by comparing the OLS estimation method of the Linear Model and 2SLS estimation method of the Instrumental Variable Model (i.e., the difference between OLS and 2SLS). Endogeneity of schooling may also impart a downward bias on the conventional OLS rate of return estimates. The Hausman specification test by Wu, (1973) was made use of to determine if the education variable was endogenous.

The selection effects model, which is an alternative structural model to the fixed effects model, and developed by Ashenfelter and Krueger (1994) was also employed in this study to explicitly account for unmeasured family effects (i.e., genetic endowments and common environment). The unmeasured family effects were modeled using the educational attainment of each twin. The selection effects model was estimated using the feasible generalized least squares (FGLS) method. The coefficient on the co-twin's educational attainment provided an estimate of the impact of family effects. In addition the selection effects model, an alternative structural model developed by Ashenfelter and Krueger (1994) was used to explicitly account for family effects (i.e., genetic endowments and common environment). The unmeasured family effects were modeled as depending on the educational attainment of each twin. The selection effects model was estimated using the generalized least squares (GLS) method. The coefficient on the co-twin's educational attainment provided an estimate of the impact of family effects. This family effects estimate also shows the extent of endogeneity bias (which is referred to as a "selection effect") in the returns to education.

This study further utilizes multi-level modeling to estimate the two-level hierarchical linear model (HLM) by Lindley & Smith, (1972) to account for the hierarchical structure of the twins' data and the lack of independence of observations. The reason for performing a hierarchical linear modeling analysis is to test hypotheses about intercepts and slopes. The HLM model consists of both fixed and random effects which essentially estimates a random coefficient (on education) model. Individual and family unobserved heterogeneity in the returns to education were controlled for through the estimation of the hierarchical linear model (HLM) of earnings determination. Estimation of the intercept and the random coefficient was done by the restricted likelihood estimation method (REML). The REML directive in SAS estimated the variance components using a ridge-stabilized Newton-Raphson algorithm (Searle, Casella, and McCulloch, 1992) to maximize the residual likelihood function. These were then used to construct an estimate of variance-covariance matrix. This study identified and characterized the extent of unobserved heterogeneity in returns to schooling by comparing OLS estimates in the linear model to the fixed effects REML estimates in the hierarchical linear model to control for unobserved heterogeneity. The difference in these estimates shows how unobserved heterogeneity influences the earnings effect of education. Yet again, the variance in the returns to education in this study was estimated with the random effects REML. This variance was decomposed into family heterogeneity, individual heterogeneity, and the residual error.

The optimal schooling model developed by Becker, (1967) was used to replace the assumption of a homogeneous return to schooling with the assumption that the marginal return to schooling, varies by family and is correlated with the unobservable ability. Based on this optimal schooling model, the earnings equations proposed by Ashenfelter and Rouse (1998) were used to obtain the return to schooling by ability. The product of the individual's (twin *i*) schooling level and the average schooling level of the family (i.e., average of twin 1 and twin 2's educational attainment) was included in the return to schooling by ability model (Ashenfelter and Rouse, 1998) to remove any absolute ability bias that may arise in the estimation of the economic return to schooling. Estimations of the proposed equations were performed by the GLS method. Furthermore, the correlation between schooling and ability was estimated by including the average schooling level term in the model. Subsequently, the degree of heterogeneity in the return to education for twins was also estimated where schooling varies with ability. The coefficient of the interaction term of the product of the individual's schooling level and the family's average schooling level was estimated to obtain the product of the correlation between ability and schooling and the heterogeneity in the return to schooling. The effect of ability on the marginal benefit of schooling was subsequently determined by dividing the coefficient of the interaction term by the coefficient of the family's average schooling level.

4.4.3 Statistical Package used for Data Analysis

The SAS (Statistical Analysis System) version 9.1 and STATA version SE 11 are the statistical software packages used in analyzing the data in this study.

CHAPTER FIVE

RESULTS

5.1 Labour Market and Demographic Characteristics

Labour market and demographic characteristics have been observed in several studies to have a statistically significant effect on earnings. Therefore, in estimating the returns to education, it is appropriate to augment the earnings equation with a number of control variables that relate to characteristics that can affect an individual's earning capacity.

A summary of the general characteristics of twins used in this study is presented in Table 5.1.1. The Table highlights information on earnings, education, and other variables including demographic characteristics for MZ and DZ twins. The results are presented at both individual and type of twins (MZ, DZ) levels. This helps to create a larger picture about each respondent and their genetic make-up. It also describes some of the outcomes that are used as a basis for analyses in other chapters and provides a starting point for research questions, including comparative studies that rely on a comparison between twins in Ghana and other survey samples in other countries. Furthermore, the summary statistics of the pooled twins sample, MZ twins sample and DZ twins sample are reported in Table 5.1.2.

5.1.1 Demographic Characteristics

In estimating the relationship between education and earnings, this study considered demographic factors such as age, gender, marital status and ethnicity. Based on the Ghanaian twins sample, 48.8% are males and 51.2% are females. Females slightly outnumber males in a population due to high survival rates. This reflects the population structure of Ghana which is what all the population censuses depicts (Nsowah-Nuamah, 2007). The youngest age group was between 15-19 years and the oldest was between 60-65 years. The largest proportion of the survey population is in the 25-29 age group, (29.6%) followed by 20-24 age group, (18.40%), Table 5.1.1. The age of twins in the data set ranged from 18 to 65 years. The mean age is 33 ± 10.3 years (Table 5.1.2). The twins sample distribution in this study is more skewed to the left considering their age structure. This distribution reflects the pyramidal age structure of the Ghanaian population and indicates a higher percentage of young people within the Ghanaian twins sample. The mean age of MZ twins is 31.9 years, while the mean age of DZ twins is, on average, almost 2 years older than the MZ twins (Table 5.1.2). In this study, the age group of the mothers with the highest proportion of twin births was between 30 to 34 years (Table 5.1.1). Women who fall within age group 30 to 40 usually face age-specific fertility issues and ovulation-inducing hormones are administered when they seek for help at hospitals which usually results in multiple births.

Table 5.1.1: Demographic Characteristics of Twins in Ghana

Twins Characteristics	All Twins (%)	Monozygotic Twins (%)	Dizygotic Twins (%)
Sex			
Male	48.80	21.60	27.20
Female	51.20	20.80	30.40
Age			
15-19	2.40	0.80	1.60
20-24	18.40	8.00	10.40
25-29	29.60	12.80	16.80
30-34	12.00	6.40	5.60
35-39	14.40	7.20	7.20
40-44	8.00	3.20	4.80
45-49	7.20	1.60	5.60
50-54	4.00	0.80	3.20
55-59	1.60	0.80	0.80
60-65	2.40	0.80	1.60
Highest level of education			
No education (0)	2.40	0.00	2.40
Primary (1-6)	4.40	1.20	3.20
Middle/JSS (7-10)	53.20	19.60	33.60
Secondary (11-17)	8.80	2.40	6.40
Higher (18-25)	31.20	19.20	12.00
Occupation			
Professional	26.00	13.20	12.80
Clerical	6.00	3.20	2.80
Business	35.60	13.20	22.40
Agriculture	6.80	1.20	5.60
Production & Labourer	25.60	11.60	14.00
Marital Status			
Married	44.00	13.60	30.40
Living together	0.80	0.80	0.00
Separated	1.60	1.20	0.40
Divorced	0.80	0.40	0.40
Never married	52.80	26.40	26.40
Mothers Age			
15-19	4.00	1.60	2.40
20-24	24.00	8.80	15.20
25-29	16.00	6.40	9.60
30-34	34.40	14.40	20.00
35-39	17.60	8.80	8.80
40-44	4.00	2.40	1.60

This marital pattern is expected in this sample, where over 76% of the sample is below 40 years of age (Table 5.1.1). On the other hand, 30.4% of DZ twins are married, whilst only 13.6% of MZ twins are married (Table 5.1.1). This suggests that there might be a stronger bond between MZ twins than DZ twins and therefore MZ twins tend to remain single for much longer than the DZ twins.

All the twins were employed on a full-time basis. Looking at the occupational classification, the business category emerged with the highest percentage (35.6%) followed closely by the professional and the production and labourer category (26% and 25.6% respectively). The percentage of MZ twins who are in the professional category was higher than that of the DZ twins (Table 5.1.1). The annual average earnings for all the twins was GH¢7,000 (Table 5.1.2), where earnings include wages, bonuses, and subsidies. The mean incomes of the monozygotic (GH¢7.368) and dizygotic (GH¢7.049) twins are similar, though MZ twins earn more on average than DZ twins. This might probably be due to the fact the MZ twins in this study were more highly educated than the DZ twins.

5.1.2 Labour Market Characteristics

In this study, about 53% of the twins have completed elementary or basic education (MSLC/JSS), and slightly above 30% have tertiary school qualifications (Table 5.1.1). Twins with virtually no educational qualifications are the smallest group (2.4%). This same trend is identified in both the monozygotic twins group and the dizygotic twins group.

Table 5.1.2: Means and Standard Errors of Selected Variables: Ghanaian Twins Survey

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

Note: Figures in parentheses are standard errors

On average the twins have almost 13 years of education (Table 5.1.2). This indicates that a good number of the twins in this study do 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 5.1.2). The report of the respondent's level of education (12.6 years) is similar to that reported by his or her co-twin ((12.7 years) Table 5.1.2).

It was observed that around 77% of twins report the same own level of education, 1% with one year's difference in education, 5% with two years' difference, and the remaining 17% with a difference of more than two years (Figure 5.1). It is clear that many twins report identical education levels, so that many within-twin education differences are zeros. There is also some amount of

variability in the reported wage differences of twins with the same educational levels. Higher earnings were also found to be associated with high education levels (Figure 5.1)

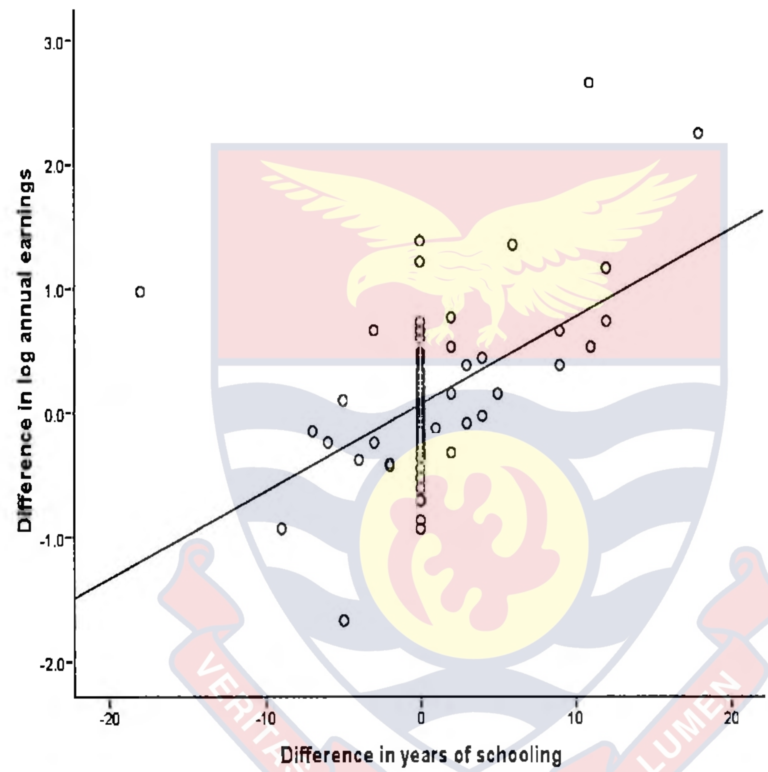

Figure 5.1: Within-Twin Pair Differences in Years of Schooling and Annual Earnings

5.1.3 Correlations between Earnings and Education Levels

The correlations among the (logarithmic) income, (self reported and co-twin reported) education levels and mother's and father's education levels are reported in Table 5.1.3 for Monozygotic twins and in Table 5.1.4 for dizygotic twins. In this analysis the twin who was the first to come out of the womb was chosen to be twin 1 in each pair. The correlation between the educational attainments of MZ twin pairs is (0.963) and only (0.339) for DZ twin pairs (Tables 5.1.3 and 5.1.4). This indicates that MZ twins are more alike in terms of their educational attainments than DZ twins. The correlation between the self-reported measure of educational attainment and the report on this educational attainment by the co-twin is the same for both MZ twins and DZ twins (0.999). The simple correlation coefficient between the self-reported and co-twin-reported measures of educational attainment shows the extent of variation in reported measures of schooling.

The correlation between the self-reported measure of educational attainment and co-twin-reported education of the same twin, that is, $corr(S_1^1, S_2^1)$ and $corr(S_2^2, S_1^2)$ are (0.999) and (0.999) for MZ twins, (Table 5.1.3). This is positive and very high indicating that MZ twins are more likely to report the same own-level of educational attainment than DZ twins, and are characterized by having a greater similarity between the own-report on educational attainment and the co-twin's report. On the other hand $corr(S_1^1, S_2^1)$ and $corr(S_2^2, S_1^2)$ for DZ twins are (0.999) and (0.697). They indicate that between 1% and 30% of the measured variation in educational attainment for DZ twins' is error and allows for direct estimates of the extent of measurement error in (the cross-sectional) reported schooling in the twins data. The co-twin's report is used as an instrument to accommodate the problem of measurement errors in the own-report on educational attainment. It does not matter whether the instrumental variable is measured without error so long as the respective measurement errors are uncorrelated.

Table 5.1.3: Correlation Coefficients between Selected Variables for MZ twins

A. Pearson Correlation Coefficients for Monozygotic twins, N = 53
Prob > |r| under H0: Rho=0

Parameter	S_1^1	S_2^2	S_1^2	S_2^1	M_1	M_2	F_1	F_2	Y_1	Y_2
S_1^1	1.0000									
S_2^2	0.9628 <.0001	1.0000								
S_1^2	0.9632 <.0001	0.9997 <.0001	1.0000							
S_2^1	0.9999 <.0001	0.9628 <.0001	0.9632 <.0001	1.0000						
M_1	0.2960 0.0314	0.2571 0.0631	0.2565 0.0638	0.2960 0.0314	1.0000					
M_2	0.4655 0.0004	0.4820 0.0003	0.4812 0.0003	0.4655 0.0004	0.6028 <.0001	1.0000				
F_1	0.4808 0.0003	0.4791 0.0003	0.4777 0.0003	0.4808 0.0003	0.4585 0.0006	0.4165 0.0019	1.0000			
F_2	0.5436 <.0001	0.5565 <.0001	0.5563 <.0001	0.5436 <.0001	0.4324 0.0012	0.6524 <.0001	0.8326 <.0001	1.0000		
Y_1	0.6396 <.0001	0.6101 <.0001	0.6058 <.0001	0.6396 <.0001	0.1027 0.4645	0.1913 0.1701	0.3300 0.0158	0.3820 0.0048	1.0000	
Y_2	0.6385 <.0001	0.6625 <.0001	0.6547 <.0001	0.6385 <.0001	0.0608 0.6657	0.2175 0.1177	0.3504 0.0101	0.4009 0.0029	0.8829 <.0001	1.0000

Notes: S_1^1 = total years of schooling of twin 1; S_1^2 = sibling 1 reported years of schooling of sibling 2;
S_2^2 = total years of schooling of twin 2; S_2^1 = sibling 2 reported years of schooling of sibling 1;
M_1 =twin 1's report of mother's educational level; M_2 = twin 2's report of mother's educational evel;
F_1 = twin 1's report of father's educational level; F_2 = twin 2's report of father's educational level;
Y_1 = twin 1's earnings; Y_2 = twin 2's earnings.

Table 5.1.4: Correlation Coefficients between Selected Variables for DZ Twins

B. Pearson Correlation Coefficients for Dizygotic twins, N = 72
Prob > |r| under H0: Rho=0

Parameter	S_1^1	S_2^2	S_1^2	S_2^1	M_1	M_2	F_1	F_2	Y_1	Y_2
S_1^1	1.0000									
S_2^2	0.3391	1.0000								
	0.0036									
S_1^2	0.5860	0.6970	1.0000							
	<.0001	<.0001								
S_2^1	0.9999	0.3391	0.5860	1.0000						
	<.0001	0.0036	<.0001							
M_1	0.3366	0.4460	0.5250	0.3366	1.0000					
	0.0038	<.0001	<.0001	0.0038						
M_2	0.3283	0.3233	0.4520	0.3283	0.6362	1.0000				
	0.0049	0.0056	<.0001	0.0049	<.0001					
F_1	0.3496	0.3807	0.4843	0.3496	0.7377	0.6008	1.0000			
	0.0026	0.0010	<.0001	0.0026	<.0001	<.0001				
F_2	0.3470	0.3272	0.4770	0.3470	0.6267	0.7250	0.7849	1.0000		
	0.0028	0.0050	<.0001	0.0028	<.0001	<.0001	<.0001			
Y_1	0.5123	0.3163	0.4033	0.5123	0.4355	0.3873	0.2868	0.3050	1.0000	
	<.0001	0.0068	0.0004	<.0001	<.0001	0.0008	0.0146	0.0092		
Y_2	0.3612	0.4898	0.6190	0.3612	0.4209	0.3726	0.2726	0.2968	0.7103	1.0000
	0.0018	<.0001	<.0001	0.0018	0.0002	0.0013	0.0205	0.0114	<.0001	

Notes: S_1^1 = total years of schooling of twin 1; S_1^2 = sibling 1 reported years of schooling of sibling 2;
S_2^2 = total years of schooling of twin 2; S_2^1 = sibling 2 reported years of schooling of sibling 1;
M_1 = twin 1's report of mother's educational level; M_2 = twin 2's report of mother's educational level;
F_1 = twin 1's report of father's educational level; F_2 = twin 2's report of father's educational level;
Y_1 = twin 1's earnings; Y_2 = twin 2's earnings.

These correlation coefficients provide a measure of the reliability ratio of the measure of educational attainment. On the whole, the high correlations suggest that the co-twin-reported level of education is a good instrumental variable for self-reported level of education in our sample. The correlation coefficients between parental education and twin education for both MZ and DZ twins were also very significant (p<0.05) though lower than the estimates for the twins themselves. The level of parental educational attainment could be a good predictor for the schooling success of children and shows the relevance of family background level effects of education on earnings and the returns to education. Furthermore, the within-twin pair earnings among MZ twins are more highly correlated than that of dizygotic twins.

5.2 Returns to Education in Ghana

In this section, the estimated return to education are reported using different samples (MZ twins sample, DZ twins sample and the pooled sample of MZ and DZ twins) and different methods. Firstly, the relationship between education and earnings is examined with Mincers' earnings function. Secondly, OLS regressions are performed using the pooled sample, monozygotic twins and dizygotic twins samples and comparisons are made between the estimated coefficients for the three samples. This comparison may serve as a way to check the representativeness of the monozygotic twins and the dizygotic twins samples. Thirdly, the return to education is estimated for the pooled sample of twins and for MZ and DZ twins with the use of the 2SLS-IV estimator. Fourthly, the within-twin-pair fixed effects and GLS estimations are conducted using the MZ and DZ twins sample, followed by examinations of possible bias in fixed effects estimates and the impact of measurement error. Finally, the REML estimator was utilized to detect possible unobserved heterogeneity in the return to education for twins caused by the influence of individual and family background characteristics on earnings.

5.2.1 The Mincerian Return to Schooling by Twins and Gender

The empirical results were derived from the OLS estimator using equation 3.1 as presented by Tables 5.2.1 and 5.2.3. The Mincerian earning functions is estimated by assuming that the schooling variable is exogenous, to indicate the bias that might be introduced by neglecting the endogeneity issue. It estimated the Mincerian earnings equations where the natural log of annual earnings received by an individual is a function of years of schooling, potential experience and its square. Demographic characteristics such as gender and marital status have been observed in other studies to have a statistically significant effect on earnings and were therefore included in the earnings equation as dummy variables. Mincer's earnings function also assumes that the return to schooling is a single parameter which does not vary across individuals.

Coefficients from Mincer's log earnings regression were estimated by OLS for the MZ twins data set only, the DZ twins data set only and the pooled twins data set. Subsequently, the same procedure was carried out for the pooled twin data set divided into males and females. This is to find out whether there are any significant differences between the returns to schooling for males and females in Ghana. In Table 5.2.1, the OLS regression estimates for the pooled sample, MZ and DZ twins indicated that education has a positive and significant effect on earnings (0.098; 0.095; 0.098: $p<0.0001$). Estimated coefficients on the return to schooling tend to be lower for MZ (0.0949) than for dizygotic (0.0977) twins. Although, experience, gender and marital status were also positively related to earnings, their effect was not significant except (Table 5.2.1). However, education, potential experience, sex and marital status explained about 49% of the variance (R-squared) in the log annual earnings for all twins (column 1, Table 5.2.1). This indicates that about half of the variance in Mincer's earnings OLS regression model is attributable to other factors like family background (e.g., parental education and income) and unobservable genetic traits.

Table 5.2.1: Estimated Coefficient from Mincer's Log Earnings Regression by Twins (Monozygotic and Dizygotic)

Variables	Pooled Sample	Monozygotic Twins	Dizygotic Twins
Intercept	5.02582*** (0.34009)	5.28421*** (0.56474)	5.07099*** (0.43944)
Education	0.09807*** (0.00759)	0.09491*** (0.01191)	0.09774*** (0.01027)
Experience	0.06188 (0.03531)	0.03880 (0.06058)	0.05657 (0.04503
Experience-squared	-0.000856 (0.00083)	-0.000056 (0.00148)	-0.000905 (0.00102)
Gender	0.12634 (0.07857)	0.02902 (0.12755)	0.21246** (0.10214)
Marital status	0.04293 (0.08262)	0.09077 (0.13626)	0.04070 (0.10706)
MSE	0.37848	0.40001	0.36365
DF	244	100	138
R-squared	0.4872	0.4918	0.4726
N	250	106	144

***, ** - significant at 0.01 and 0.05 probability level respectively

Likewise, education, potential experience, sex and marital status explained about 49% and 47% of the variance in the log annual earnings for MZ and DZ twins respectively. Also, one additional year of experience which signifies an individual's human capital accumulation through job training increased earnings by 6% for DZ twins and 4% for MZ twins. Table 5.2.2 reports the difference in OLS regression coefficients between MZ and DZ twins and the associated t-value for each comparison. In all five comparisons (last five rows of Table 5.2.2), the *t*-tests of the OLS coefficient differences were not significantly different from zero ($p>0.05$).

Consequently, OLS regression analysis was also performed using the pooled sample. In Table 5.2.1, control for gender and marital status gives results which are in line with the basic theory. This theory specifies that men have a higher slope in the education earnings relationship than women but a higher intercept in the earnings function for women indicates that their earnings do catch up with men at higher levels of education. Thus the gender earnings gap is lower at higher levels of education. Further, as demonstrated in Figure 5.2, the earnings-potential experience profiles of males and females differ considerably, with those of females being much flatter than those of males. In Table 5.2.3, regression on Mincers' earnings function does not control for gender because the analysis is done by gender. The effect of marital status on annual earnings appears to impact negatively ($\beta=-0.0943$, $p>0.05$) on women twins (Table 5.2.3). Moreover, the estimate was also not significantly different from zero. Table 5.2.3 also reports a rate of return to schooling of around 11% for men and 9% for women indicating that earnings are relatively higher for male twins than for female twins.

Table 5.2.2: Test for Differences in the Rates of Return to the level of Schooling between MZ and DZ twins using Mincer's Model

Parameter	Coefficient	Standard error	t value	Pr > |t|
Intercept	5.2842	0.5496	9.61	<.0001
Years of Schooling	0.0949	0.0116	8.19	<.0001
Experience	0.0388	0.0590	0.66	0.5111
Experience –squared	-0.00006	0.0014	-0.04	0.9690
Gender	0.0290	0.1241	0.23	0.8154
Marital status	0.0908	0.1326	0.68	0.4944
Twin type (intercept)	-0.2132	0.7095	-0.30	0.7640
Years of Schooling *twin type	0.00283	0.0156	0.18	0.8561
Experience *twin type	0.0178	0.0748	0.24	0.8124
Experience-squared*twin type	-0.0008	0.0018	-0.48	0.6335
Gender *twin type	0.1834	0.1621	1.13	0.2590
Marital status *twin type	-0.0501	0.1719	-0.29	0.7711

The coefficient on experience (0.1264) was higher for men than for females (0.0489) with a difference in coefficient of about (0.078). This indicates that there are more experienced males than females in the labour market and that females tend to be at a disadvantage when experience is a determining factor for employment in the Ghanaian labour market. The earnings-age profile was concave as reflected in the increasingly more negative estimated coefficients for experience squared. The earnings-age profiles for MZ and DZ twins were -0.00006 and -0.0009 respectively and for male and female twins were -0.0027 and

-0.0005 respectively. The impact of work experience on male twins was however significant (Table 5.2.3), indicating that probably, the males are better educated and more skilled than the females.

Table 5.2.3: Estimated Coefficient from Mincer Log Earning Regression by Gender

Variables	Males	Females
Intercept	4.4761*** (0.57675)	5.2702*** (0.0455)
Years of Schooling	0.1070*** (0.01088)	0.0907*** (0.0105)
Experience	0.1246** (0.06301)	0.0489 (0.0459)
Experience-squared	-0.0027 (0.0016)	-0.0005 (0.0010)
Married	0.2222 (0.11811)	-0.0943 (0.11641)
MSE	0.3598	0.3909
DF	121	127
R^2	0.5406	0.4549
N	122	128

***, ** - significant at 0.01 and 0.05 probability level respectively

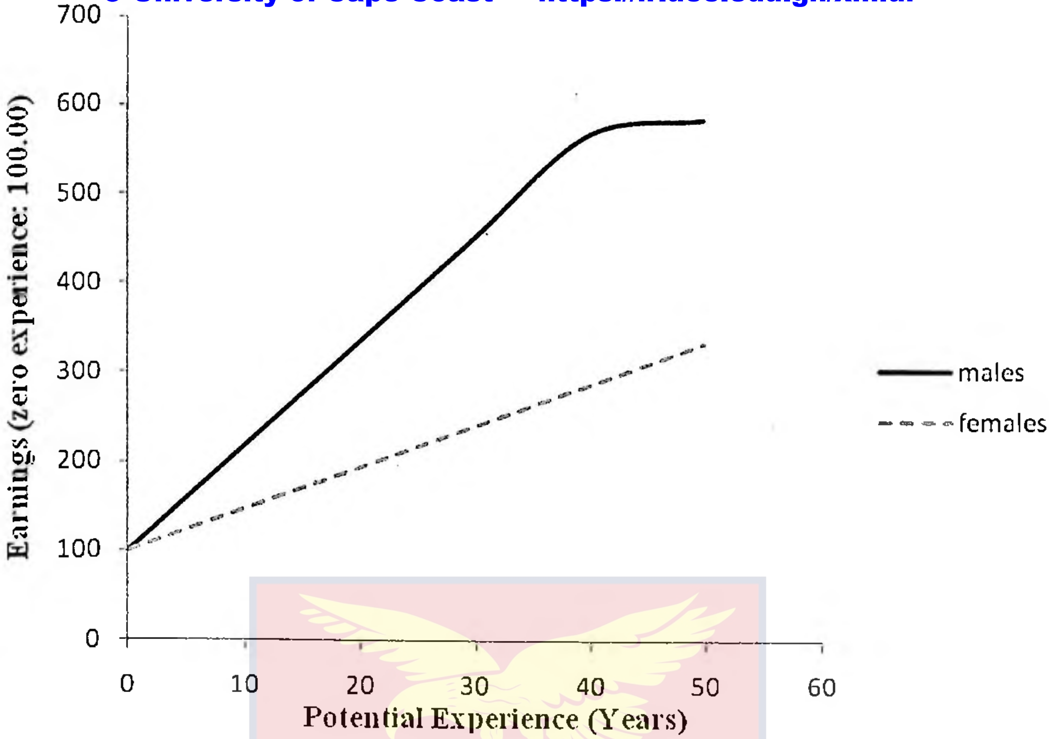

Figure 5.2: Experience-Earnings Profiles of Males and Females

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

Presented in Table 5.2.4 are the results of the regression function (equation 3.2) with education, age, age squared, gender, and marital status as independent variables that may be considered strictly exogenous and the logarithm of annual earnings as the dependent variable. Twins studies of the return to schooling typically begin with OLS estimates as a way of replicating the 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 5.2.4. The results in Table 5.2.4 for the pooled sample show a strong and positive association between earnings and education (β_1=0.10; p<0.0001) and also indicates that the return to education is quite large. Average occupational earnings also increased with age (β_2=0·05, p=0.028) and this result was statistically significant ($p<0.05$) indicating that experienced workers earn more as productivity-related skills are perfected. Male workers were found to be at a substantial 10.3% earnings advantage compared with female workers (Table 5.2.4).

Table 5.2.4: OLS estimates of Equation (3.2) for Pooled, MZ and DZ Twins

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

***, ** - significant at 0.01 and 0.05 probability level respectively

Although, the effect of the proportion of those who are married on earnings was negative, it was however significant (β_5=-0.19; $p<0.05$), indicating that being married does not guarantee an individual an increase in earnings.

OLS regressions are also estimated for the monozygotic and dizygotic twins samples and presented in Tables 5.2.4. The effect of age on earnings for every additional life year was significant ($p<0.05$) for MZ twins but insignificant ($p>0.05$) for DZ twins, however, the difference between the two regression coefficients was not significant. This finding may be associated with age being a better proxy for actual work experience for MZ twins than it is for DZ twins. Moreover, the MZ twins sample are on average younger than the DZ twins and therefore a decline in earnings could come about at older ages.

Although, positive effects on earnings were found for both MZ and DZ male twins, these effects were not significantly different ($p>0.05$) (Table 5.2.4). In the cross-sectional regression on MZ twins, marital status is significant and has the predicted negative sign for MZ twins, which means that unmarried individuals earn less than their married counterparts. On the other hand, the effect of marital status on earnings is also negative but insignificant ($p>0.05$) for DZ twins. Testing for equality of the estimated coefficients in the linear regression model, the OLS results of the monozygotic (MZ) twins were compared with that of the dizygotic (DZ) twins. The estimates show that the differential intercept and differential return to education coefficients are not significant, implying that the intercept and return to education coefficients for the DZ twins are statistically equal to that of the MZ twins, (Table 5.2.5).

Table 5.2.5: Test for Differences in the Rates of Return to the Level of Schooling between MZ and DZ twins using the General Linear Model

Parameter	Coefficient	Standard error	t value	Pr > |t|
Twin type (intercept)	1.5115	0.8427	1.79	0.0742
Years of Schooling *twin type	0.0095	0.0149	0.64	0.5234
Age*twin type	-0.0857	0.0475	-1.81	0.0723
Age-squared*twin type	0.0008	0.0006	1.28	0.2005
Gender*twin type	0.1801	0.1572	1.15	0.2532
Marital status*twin type	0.1177	0.1593	0.74	0.4608
R^2	0.5217			
N	250			

***,**,*-significant at the 0.01,0.05 and 0.1 probability level respectively

5.2.3 Feasible Generalized Least Squares (FGLS) estimates of the returns to schooling

As an alternative to the OLS estimator, estimates of the FGLS estimates are also reported to check the robustness of the OLS estimates. FGLS estimate of the returns to schooling for the pooled sample of twins is 0.097 (Table 5.2.6). This estimate ignores the family effect or the potential correlation between schooling level and family background by setting the selection effect (γ=0). Though numerically, smaller, the FGLS estimate differs slightly from the corresponding OLS estimate (0.1014) for the pooled twins sample (Table 5.2.4).

Unlike the OLS estimates (Table 5.2.4) of the returns to education where the proportion of males was not a significant determinant of the relationship between earnings and education, the proportion of male earnings rather benefited from an additional year of education in the FGLS estimates (Table 5.2.6) of the returns to education ($p<0.10$). For the MZ twins the return to education is estimated to be 9.6% (Table 5.2.6) when the FGLS estimator is employed. Though similar to the OLS result 9.4% (Table 5.2.4), the FGLS result is slightly higher by 0.2%. Age is also found to significantly ($p<0.05$) affect FGLS estimates of the returns to education (Table 5.2.6) though the OLS estimates for MZ twins in Table 5.2.4 were smaller in magnitude. This confirms the fact that GLS estimates are known to improve the efficiency of regression estimates. Subsequently, Table 5.2.6 reports the GLS estimates of the return to education to be 9.8% for DZ twins. This FGLS result is also smaller than the OLS estimate (10.4%) by about 0.6%. It is worth noting that both GLS and OLS estimates of the returns to education for all the samples were similar. In effect, either the OLS estimator or the FGLS estimator could be used for cross sectional estimations of the returns to education.

5.2.4 Instrumental variable (IV) estimates of the returns to schooling

The results of the application of the two-stage least squares (2SLS) IV estimator using equation (3.11) is presented in Table 5.2.7 for the pooled twins sample, MZ and DZ twins samples. They contain the IV results which relate earnings to education, gender, marital status and age. In this model the report on twin 1's schooling by twin 2 is used as an instrument for twin 1's self-reported level of educational attainment. The use of the co-twin's report of education as an instrument yields a 2SLS estimate of the return to schooling of about 0.104 (standard error = .008).

Table 5.2.6: FGLS Estimates of Equation (3.2) for Pooled, MZ and DZ Twins

Parameter	Pooled	MZ	DZ
Intercept	2.1914*** (0.4820)	1.2074* (0.6782)	3.274*** (0.6502)
Years of Schooling	0.0973*** (0.0095)	0.0963*** (0.0146)	0.0977*** (0.0128)
Age	0.1962*** (0.0275)	0.2490*** (0.0388)	0.1329*** (0.1329)
Age squared	-0.0023*** (0.0036)	-0.00285*** (0.00053)	-0.00158*** (0.00048)
Gender	0.1628* (0.0947)	0.2176 (0.1575)	0.2268*** (0.1339)
Married	-0.1009 (0.0948)	-0.1790 (0.1307)	-0.0328* (0.1302)
R^2	0.4351	0.5060	0.4312
N	250	106	144

***,**,*-significant at the 0.01,0.05 and 0.1 probability level respectively

Earnings is positively related to age and this relationship is significant ($p<0.05$) but negatively related to the proportion of married twins. Although, males earned about 10.4% more than females, there is no evidence that being male contributes significantly ($p>0.05$) to the education and earnings relationship. Estimation of the standard specification of the earnings equation by 2SLS yields a return on schooling of around 9.4% for MZ and 11.0% for DZ twins. Whiles the effect of age on earnings was positive and significant (β_2=0.094; p=0.01) for MZ twins, its impact was small and insignificant (p>0.05) for DZ twins. The earnings of male DZ twins are estimated to be significant (p<0.05) and 21% higher than the earnings of female DZ twins.

Table 5.2.7: Instrumental Variable 2SLS Estimates of Equation (3.11) for Pooled, MZ and DZ Twins

Parameter	Pooled	MZ	DZ
Intercept	4.8123***	4.0662	5.5929
	(0.4095)	(0.6554)	(0.5235)
Years of Schooling	0.1043***	0.0940	0.1100
	(0.0077)	(0.0108)	(0.0110)
Age	0.0494**	0.0937	0.0028
	(0.0234)	(0.0363)	(0.0304)
Age squared	-0.00046	-0.00084	0.000028
	(0.0003)	(0.00048)	(0.00039)
Gender	0.1036	-0.0298	0.2089
	(0.0778)	(0.1185)	(0.1024)
Married	-0.1930**	-0.2634	-0.1455
	(0.0790)	(0.1215)	(0.1021)
R^2	0.4910	0.5542	0.4622
1st Stage F-statistic	748.06***	972.81***	259.14***
Hausman Test of exogeneity	3.68	0.02	4.90
N	250	106	144

***,**,*-significant at the 0.01,0.05 and 0.1 probability level respectively

In contrast, there were no significant (p>0.05) earnings/gender differences for male MZ twins. The analysis also indicates a significant, negative relationship between earnings and marital status (β=-0.2634; p<.05). This means that the proportion of MZ twins who are married have significantly lower returns to education than those who are not married. However, the effect of marital status on earnings for DZ twins was negative but not significant. The variability of earnings for MZ, DZ and the Pooled twins (measured by the R-squared) explained by the IV model ranges from 32 to 55 percent, while the variability explained by the OLS model varies from 49 to 55 percent. This is not surprising and, of course, consequence of adding other explanatory variables (e.g., gender, marital status) to the standard Mincer specification. Nevertheless, the outcome of an increased explained variability of wages remains a desirable outcome in many circumstances, including the analysis of the impact of education on within-twins earnings.

Testing for the relevance the instrumental variable (twins report of education on the other) in the first-stage regression, the F-statistics for the three different samples (MZ, DZ, Pooled) were all greater than 10 (Table 5.2.7). This suggests that "twins reported education" is highly correlated with the endogenous regressor (self reported education) even after controlling for the exogenous regressors and is a relevant instrumental variable. Furthermore, the test of exogeneity (Table 5.2.7) of the "twins reported education" in the second-stage regression, shows that the coefficients for the MZ, DZ and Pooled twins are zero indicating that the instrumental variable is not correlated with the error term. This suggests that "twins reported education" is a valid instrument satisfying both the relevance and exogeneity requirements in instrumental variable regression.

5.2.5 Fixed Effects Estimation of the Returns to Schooling

5.2.5.1 Fixed-Effects Ordinary Least Squares (FE-OLS) Returns to Schooling

A regression of the within twin difference in earnings on the within twin difference in education levels (i.e. fixed effects estimate) is estimated for the pooled sample of twins, MZ and DZ twins. Using equation 5 sibling differences are taken to estimate fixed effects, using ordinary least squares (OLS). In the fixed effects model genetic resemblance and common environment influences are held constant implicitly.

This method of estimation also nets out of the estimated impact of schooling the compounding effects of any other fixed effects that affect earnings (e.g., race, possibly some affective characteristics such as motivation). Estimating the return to education using the pooled twins sample indicates that the true impact of education is 8% (Table 5.2.8). Thus, the within-twin-pair estimate of the return to education is smaller than the OLS estimate in Table 5.2.4. This shows that a part of the return to education that is found by the OLS estimate is the result of the effects of unobserved ability or family background. Similar patterns are also observed with the results of the MZ and DZ twins samples (Table 5.2.8). It is however noted that, the fixed effect estimate of the return to schooling for MZ twins is about two percentage points higher than the estimate for DZ twins (Table 5.2.8). This difference is due to the explicit control for ability and family background in the estimations for MZ twins, but only for family background in the estimations for the DZ twins who were reared together.

Table 5.2.8: Twin-differencing OLS of Equation (5.5) for Pooled, MZ and DZ Twins

Parameter	Pooled	MZ	DZ
Intercept	0.0818** (0.0407)	0. 1146** (0.0533)	0.0600 (0.0595)
Years of Schooling	0.0835*** (0.0110)	0.1115*** (0.0350)	0.0823*** (0.01268)
R^2	0.3138	0.1664	0.3754
F-value	57.44***	10.18***	42.07***
N	125	53	72

***,**,*-significant at the 0.01,0.05 and 0.1 probability level respectively

The returns to education were (0.08; 0.11; 0.08); for the pooled twins sample, MZ and DZ twins samples respectively using the fixed effects model that utilize the twins nature of the data. These results are all positive and highly significant (p<0.0001), though the MZ twins estimate is bigger in magnitude compared with the linear model result 0.094 (Table 5.2.4). This indicates that the fixed effects model could be considered a better model for estimating the returns to education since higher returns are obtained when genetic and environmental factors are considered or taken into account.

5.2.5.2 Fixed-Effects Instrumental Variable (FEIV) Returns to Schooling

Estimates of the fixed effects model using 2SLS are also reported in Table 5.2.9 for the pooled sample of twins, MZ and DZ twins. Here the difference in actual (i.e. self-reported) schooling levels is instrumented by the difference in co-twin-reported schooling levels. This was carried out in order to account for any measurement error bias in the estimates. The returns to education are estimated to be 11%, 14% and 10% for the pooled twins sample, MZ and DZ twins samples respectively. These estimates are significant ($p<0.0001$) and similar to the 2SLS results 10%, 9% and 11% (Table 5.2.9) for the pooled twins sample, MZ and DZ twins samples respectively that do not take ability and family background differences into consideration. The FE-2SLS estimate for MZ twins increased by about 36%. Furthermore, the difference in these two estimates (2SLS and FE-2SLS) for the MZ twins is observed to be higher than the estimates for the pooled and DZ twins. This result may be a reflection of correlations in the reported measures (own-level of education, co-twin's level of education).

Table 5.2.9 Twin-differencing 2SLS of equation (3.14) for Pooled, MZ and DZ Twins

Parameter	Pooled	MZ	DZ
Intercept	0.0712* (0.0415)	0.1162** (0.0537)	0.0404 (0.0611)
Years of Schooling	0.1053*** (0.0131)	0.1412*** (0.0465)	0.1043*** (0.0149)
R^2	0.2966	0.1547	0.3484
1st Stage F-statistic	F(1,123)=335.90***	F(1,51)=68.20***	F(1,70)=214.64***
N	125	53	72

***,**,*-significant at the 0.01,0.05 and 0.1 probability level respectively

5.2.6 Selection effects Estimation of the Returns to Schooling

Estimates from the selection effects model proposed by Ashenfelter and Krueger (1994) are presented in Table 5.2.10. Consideration of this alternative model provides a means for assessment of the robustness of the findings obtained from the fixed-effects model. The selection effects model uses measures of the education of a twin's sibling as an additional regressor to control for any "family" effects that affect the absolute level of earnings. The results in Table 5.2.10 correspond to fitting equation (10) using the feasible generalized least squares (FGLS)[1] 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 and estimates of the covariates that are used in the OLS estimates. The coefficient of this variable is a measure of the selection effect, (γ). These estimates take account of the cross-equation restrictions apparent in Equation (10), which can directly estimate both the return to education and the ability or family background effect. Based on the selection effects model, the education coefficient which comprises of the returns to education (γ_1) and the family effect or ability (γ_5), indicates 9.3% earnings per year of education for the pooled twins sample (Table 5.2.10). The coefficient (0.011) on the co-twin's educational attainment (Table 5.2.10) is the estimated ability or family effect and also provides an estimate of the correlation between

---

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.

the education levels of the twins and the unobserved family background effect (residual). The positive coefficient found here indicates that the income rewards for members of families with above-average educational attainments are relatively high, though not significant (p>0.05).

Table 5.2.10: Selection Effects FGLS Estimates of Equation (3.10) for Pooled, MZ & DZ Twins

Parameter	Pooled	MZ	DZ
Intercept	4.7876*** (0.5478)	4.1501*** (0.9169)	5.5982*** (0.6954)
Own education	0.0933*** (0.0076)	0.1014*** (0.0189)	0.0950*** (0.009949)
Co-twin's education	0.0114 (0.0076)	-0.0076 (0.0189)	0.0164 (0.0100)
Age (years)	0.0446 (0.0314)	0.0857* (0.0507)	-0.0021 (0.0405)
Age-squared (years)	-0.00041 (0.000408)	-0.0008 (0.0007)	0.00008 (0.0005)
Male (proportion)	0.1773*** (0.0599)	0.01257 (0.1657)	0.2172*** (0.0699)
N	250	106	144
R^2	0.4269	0.3875	0.4499

***,**,*-significant at the 0.01,0.05 and 0.1 probability level respectively

Subsequently, the structural or net effect of schooling in this model is estimated to be 8.2%. This is achieved by subtracting the coefficient (0.011) of the co-twin's educational attainment from the structural estimate (0.093) of the return to schooling that controls for omitted variables bias (i.e., the coefficient on the own education variable). In addition, the net effect of schooling is similar to the fixed effects OLS estimate for the pooled sample of twins (Table 5.2.8). This gives an indication of the robustness of the selection effects model. The effect of additional education on male earnings is significant (p<0.01) when the selection effects model is used, however, age and subsequently experience did not have any significant impact on earnings.

FGLS estimations are repeated for MZ and DZ twins (Table 5.2.10). The FGLS estimation again shows that the return to education is large (0.101) and the family effect for MZ twins is small (-0.0076) and negative (Table 5.2.10). This negative selection effect is an indication that ability or family effects are either negligible or uncorrelated with educational decisions for MZ twins. It was observed 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. It is also noted that all other covariates included in the model did not have any significant (p>0.05) effect on earnings of MZ twins.

Table 5.2.10 also presents estimates of the selection effects model outlined in (10) for DZ twins. The selection effect for DZ twins is found to be similar to that of the pooled sample of twins. The result (0.02) indicate that this effect is also small but positive. It was however observed that contrary to the selection effect for MZ twins, the highly educated families are 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 estimator of the returns to education that does not adjust for the selection effect might be upward biased for individuals that are not genetically identical.

5.2.7 Linear Mixed Model – Hierarchical Linear Model (HLM)

A two-level hierarchical linear model is used in this study to model the unobservable differences in the returns to schooling. The hierarchy in the data used is described by individuals nested within families. To examine the hypothesis that data used are dominated by a hierarchical structure (i.e., individuals nested within families) and to quantify the relative importance of both (individual and family background factors) effects on the returns to education, earnings functions are run using the hierarchical linear model (HLM). Estimations are done using the restricted maximum likelihood estimator (REML) and for individual and family random effects (variance components). In order to reveal how the returns to schooling for individuals are affected by differences among families, the variance was decomposed into two levels (individual and family effects).

Intraclass correlation coefficient

The first step in an HLM analysis is to determine the proportion of the total variance that is between families, called the intraclass correlation (ICC). Using Equation (3.23) the partition of the variance in the dependent variable (annual earnings) across levels according to the ratio of the family-level variance component to the total variance for the pooled sample of twins is $\frac{\sigma^2_{Family}}{\sigma^2_{Family} + \sigma^2} = 0.787$. This indicates that 21% (Table 5.2.11) of the variance in the model can be attributable to family background effect among the individual twins and therefore the linear mixed model/HLM is needed to model these differences. Consequently, the ICC for MZ twins sample is 0.88 and the ICC for the DZ twins sample is 0.70 (Table 5.2.11). This indicates that about 12% and 30% of the variances in the two models are ascribed to family traits of MZ twins and DZ twins respectively, suggesting that MZ twins are more closely genetically related than DZ twins.

Table 5.2.11: Intraclass Correlation Coefficients (ICC) for Pooled, MZ and DZ Twins

Covariance Parameter Estimates			
Covariance Parameters	Pooled	MZ twins	DZ twins
Family variance (σ^2_{Family})	0.5714*** (0.0830)	0.6639*** (0.1396)	0.4687*** (0.0969)
Individual variance (σ^2)	0.1544 *** (0.0195)	0.09250*** (0.0180)	0.2000*** (0.0333)
ICC	0.7873	0.8777	0.7009
N	250	106	144

*** : significant at 0.01 probability level respectively

REML Estimation of the Returns to Schooling

Restricted maximum likelihood (REML) estimation of the return to schooling using the pooled sample is presented in Table 5.2.12. The effect of an additional year spent schooling on individual earnings was 10% and it differed significantly from zero (i.e., p<0.0001). Furthermore, the mean and the variance of 'logarithm of annual earnings' of an individual (5.88 and 0.0757) were also found to be highly significant (p<0.0001). This suggests that the hierarchical structure, in terms of multilevel methodology, affects the earnings of individuals differently and, therefore, the effects of education on earnings vary from one family to another and among individuals.

The results also indicate that 8% and 92% of the variation in the returns to schooling can be attributed to individual effect and family effect, respectively (Table 5.2.12). It demonstrates that, besides the individual characteristics, family characteristics can also explain a large portion of differences in earnings among the individuals. The random part (r_{0ij}, the variation within individuals and across families) at level 2 was very significant (p<0.0001), which provides evidence that covariance of error terms of two individuals (twins) in the same family is not zero. That is, the earnings of individuals in a family are correlated with each other, partially because there are some factors such as parental education, family size and family income etc. that may affect earnings of individuals and more so siblings in the same family, regardless of their educational qualifications. The findings, therefore, indicate that data used are dominated by a hierarchical structure, which may affect both intercept and the slopes (returns to education) of earnings functions. The estimated values of the random effects in the REML analysis also indicate that, the random slopes for number of years spent in school for the pooled twins sample differ significantly from zero (u_{1j} =0.0035, p = 0.0025), Table 5.2.12 , implying that the returns to schooling vary across families.

Table 5.2.12: Parameter Estimates and their Standard Errors from a Hierarchical Linear Model of the Effect of Additional Schooling on Log Earnings in Ghana

Fixed Effect	Coefficient	se	t ratio	p value
Intercept, β_0	5.8844	0.1411	41.70	<.0001
Years of schooling, β_1	0.1003	0.01036	9.68	<.0001
Random effect	**Variance Component**	**se**	**Z**	**p value**
Family variance, σ^2_{Family}	0.9163	0.2247	4.08	<.0001
Slope variance, σ^2_{slope}	0.003532	0.001256	2.81	0.0025
Residual variance, σ^2_{ε}-Level-1 effect	0.07572	0.01038	7.30	<.0001

Using the HLM model and the pooled sample to account for the variation in return to schooling, Table 5.2.13 turns out estimation results from three specifications of the HLM in equation (3.22). Model 1 estimates a multi-level model with a random intercept and fixed intercept coefficients.

Model 2, explains the random intercept earnings coefficient using family-level variables and Model 3 additionally considers a random slope coefficient representing the heterogeneous relationship between educational qualification and earnings. Model 3 further explains the random intercept model by using both individual and family level variables. The coefficient for returns to schooling is positive and statistically significant (p<0.05) in Model 3, implying that higher levels of education lead to higher returns or benefits to education.

The results of the estimation of Model 1 using REML analysis is presented in column 1of Table 5.2.13. The mean return to schooling (i.e. the return to schooling when no covariate is included in the model) across families is 7.1846. The variance component related to the random intercept is 0.5714, which has a corresponding standard error of 0.08295. Because this estimate is more than twice the size of its standard error, there is evidence of significant variation in mean return to schooling across families.

In order to explain some of the family-level variation in return to schooling, two family-level predictors were incorporated into the HLM model statement (i.e., father's and mother's education) resulting in Model 2. The results are displayed in the second column of Table 5.2.13. The mean return (β_0=6.8340) to schooling which corresponds to the expected average earnings for an individual in a family is less than its analogous estimate in Model 1. The results further indicates that a unit increase in the father's education is associated with a 0.02707 unit increase in expected return to schooling and is significant at 5% (p=0.048) and mother's education though not significant is associated with a 0.02186 unit increase in expected return to schooling. This suggests that father's education is associated with higher benefits per year of education than mother's education. Subsequently, Model 3 (Table 5.2.13) adds individual level covariates (e.g., sex, age marital status) in addition to the family level covariates.

Table 5.2.13: Multi-Level Models of the Return to Schooling, Pooled Data

Fixed Effect	Model 1	Model 2	Model 3
Intercept β_0	7.1846*** (0.07203)	6.8340*** (0.1010)	5.6777*** (0.2372)
Mean Schyrs, β_1			0.08775*** (0.01129)
Age (years)			0.01414* (0.006093)
Gender			-0.1488** (0.05701)
Married			-0.00379 (0.08962)
Mother's education		0.02186 (0.01544)	0.01365 (0.01299)
Father's education		0.02707** (0.01358)	0.006098 (0.01120)
Random effect	**Model 1**	**Model 2**	**Model 3**
Intercept	0.5714*** (0.08295)	0.4789*** (0.07187)	0.7652*** (0.2105)
Slope variance, σ^2_{slope}			0.003060*** (0.001187)
Residual variance, σ^2_{ε} - Level-1 effect	0.1544*** (0.01954)	0.1545*** (0.01955)	0.07446*** (0.01019)
Model Fit Statistics	**Model 1**	**Model 2**	**Model 3**
-2 Reml	510.9	421.9	398.1
AIC	514.9	425.9	406.1
BIC	520.6	431.5	417.4

***,** - significant at 0.01 and 0.05 probability level

This was done based on the notion that the return to education is not a single parameter in the population, but rather a random variable that may vary with other characteristics of individuals including family background factors, ability and/or level of schooling. The expected average earnings for an individual in a family was further reduced to (β_0=5.6777) though still significant (ρ<0.001). This suggests that the individual level covariates have explained about 17% of the expected average earnings of an individual. The return to schooling (β_1=0.08775, ρ<0.001) for an individual was positive and significant, though less in magnitude than the return to schooling (β_1=0.1003, Table 5.2.13) in the model that did not include any covariates in its statement.

This indicates that individual and family characteristics explain a part of the differences/heterogeneity in the returns to schooling and that an earnings-education model that does not take into account these characteristics may overestimate the returns to schooling. Age explained a significant (ρ<0.05) portion of earnings in Model 3, implying that the pooled sample may be consisted of mostly experienced or skilled workers in their prime age. The effect of male workers though significant (ρ<0.05) was inversely related to earnings, indicating that most of the males did not earn high incomes. Surprisingly, parental education did not have a significant effect on earnings (ρ>0.05) and marital status was inversely related to earnings, implying that marriage was not associated with an earning premium in the labour market.

Furthermore, the sources of variability that are associated with the returns to schooling in this study are also presented at the random effects section of Table 5.2.13. The variance component corresponding to the random intercept has now decreased from 0.5714 in Model 1 to 0.4789 in Model 2 (Table 5.2.13), demonstrating that the inclusion of the father and mother's education explains a good deal of the family-level variance. Albeit, the estimate remains more than twice its standard error of 0.07187, suggesting that some of the family-level variance remains unexplained. Contrary to the trend expectation, the variance component corresponding to the random intercept ($\sigma_0^2 = 0.7652$) in Model 3 is larger than the corresponding estimates in Models 1 and 2 (Table 5.2.13). Additionally the random intercept is positive and significant ($p<0.01$), suggesting that there remains some variation in the expected earnings of individuals unaccounted for by the variables in the model. Moreover, the variance component ($\sigma_1^2 = 0.00306$) for the random slope is slightly higher than its standard error (0.00116), suggesting that the model picks up most of the variance in the returns to schooling that exist across families, though this variation is still significant. This indicates that observed rates of returns may vary across individuals and is due to factors unobservable to the econometrician or statistician, but known to the individual at the time of their decisions. According to the smaller-is-better rule for the information criteria, Model 3 has a smaller Akaike's information criterion (AIC), smaller Bayesian Information criterion (BIC), and lower Restricted log likelihood (-2RLL) compared to the other two models and is therefore considered the best model. The probability chi-square of the difference in the log likelihood test of the three models, showed that there were significant (p<0.01) differences between the three models (Table 5.2.14). The results of the HLM in this study therefore suggest that there exists significant heterogeneity in the returns to schooling, which is reflected in significant variation in the estimated intercept and slope coefficients.

Table 5.2.14: Testing the significance of 3 HLM Models

Parameter	Difference Log likelihood (-2LL)	Difference in df	Probability chi square
Model1&2	6.9	1	0.0086
Model1&3	105.9	4	0.000
Model2&3	391.2	2	0.000

Results of the estimation of the variance in the returns to schooling according to type of twin (MZ and DZ twins) and using the HLM model of the effect of education on earnings is presented in Table 5.2.15. The model included both family and individual level variables (gender, marital status, age, father's education and mother's education). Gender and marital status are dummy variables coded as 1 for males and 0 for females and 1 for married and 0 otherwise. The fixed effects expected average earnings ($\beta_0 = 5.626$) and the returns to schooling ($\beta_1 = 0.0706$) was significant ($\rho < 0.001$), though this effect was less than that of the pooled sample in Tables 5.2.11 and 5.2.12. With the exception of age which turned out as a significant ($\rho < 0.001$) determinant of earnings, all the other covariates which were included in the twins model did not have any significant ($\rho > 0.05$) effect on earnings. Mother's education and marital status were inversely related to earnings.

In Table 5.2.15, the variance component for the random intercept (0.2914) for dizygotic twins is 2 times greater than that of monozygotic twins (0.1403). This confirms the fact that monozygotic twins are genetically identical, and any difference in returns to education within twin pairs comes from environmental factors. The fixed effect refers to the overall expected effect of an individual's (twins) number of years spent in school on his/her earnings. The random effect gives information on whether or not this effect differs between families. In contrast, dizygotic twins share only half of their genes, and differences in returns to education are as a result of both environment and genetic factors. For the MZ twins the variance due to environmental factors (error variance) was 8% compared to about 10% for dizygotic twins (Table 5.2.15).

5.3 Effects of Endogeneity of Schooling

The problem of endogeneity occurs when the regression coefficient in an OLS regression is biased and inconsistent. A set of estimation results are presented under this section to identify a variety of potential biases associated with OLS estimation in the returns to schooling. Ordinary least squares (OLS) estimates are affected by two biases, namely ability bias, endogeneity bias and measurement error bias.

Table 5.2.15: Fixed Effects and the Variances of Random Effects from a Mixed Model of the Effect of Additional Schooling on Log Earnings of Monozygotic and Dizygotic Twins

Variable	Random effects (variance components)	Standard errors	Fixed effects (coefficients)	Standard errors
Intercept			5.626***	0.2314
Age (years)			0.233***	0.005886
Gender			-0.1804***	0.06085
Married			-0.0422	0.09386
Mother's education			0.0148	0.01275
Father's education			0.01308	0.01095
Schyrs	0.000725***	0.000300	0.0706***	0.009356
$\sigma_{\mu}^{2}(MZ)$	0.1403**	0.08104		
$\sigma_{\mu}^{2}(DZ)$	0.2914***	0.07873		
$\sigma_{\varepsilon}^{2}(MZ)$	0.08057***	0.01570		
$\sigma_{\varepsilon}^{2}(DZ)$	0.09789***	0.01939		
-Res log L	416.2			
N	250			

***,**,*-significant at the 0.01,0.05 and 0.1 probability level respectively

5.3.1 Ability bias

Using the fixed-effects approach to address the problem of ability bias in returns to schooling, the OLS estimate for the returns to schooling was compared with the fixed effects OLS estimate. Consequently, the difference between cross-sectional estimates (OLS) in the linear model and fixed effect estimates (FEOLS) identified the magnitude of the bias in the returns to schooling for the pooled, MZ and DZ twins samples. Where the estimates do not differ there is no indication of ability bias, but if the fixed effect coefficient is smaller then there is evidence of ability bias in the cross-sectional estimates. If the fixed effect estimates of the education coefficient is zero and/or insignificant this provides support that the entire effect of education on earnings is due to ability bias.

Comparison of the MZ twins return to schooling estimates from the OLS estimate (β=0.094), Table 5.3.1 and the FEOLS estimate (β=0.112), Table 5.3.1 indicates a downward ability bias (OLS estimate is less than the FEOLS estimate) to the OLS estimate for MZ twins. This means that a MZ twin with one more year of schooling compared to his/her co-twin has, on average, a return to education that is 0.02 units higher. The return to education for the MZ twins increases by almost 15% in the fixed effects regression, and remains significant ($p<0.0001$). On the other hand, OLS return to education (0.10) for DZ twins is higher than the fixed effects estimate (0.08) in Table 5.3.1. Thus there is an upward ability bias to the OLS estimate for DZ twins. For DZ twins, the intercept becomes insignificant and very small in the twin-differencing, whiles for the MZ twins the intercept remains significant although small. The F-value is significant in both fixed effects regressions, which indicates that the number of years spent schooling has a significant impact on annual earnings in the labour market.

Table 5.3.1: OLS and FEOLS Estimates of the Returns to Schooling

Variable	OLS			FEOLS		
	Pooled	MZ	DZ	Pooled	MZ	DZ
Intercept	4.8132 (0.4094)	4.0642 (0.6553)	5.5757 (0.5228)	0.0818 (0.0407)	0.1146 (0.0533)	0.0600 (0.0595)
Own education	0.1014 (0.0074)	0.0943 (0.0107)	0.1038 (0.0104)	0.0835 (0.0110)	0.1115 (0.0350)	0.0823 (0.0127)
F-Statistic	$F(5,244)=47.14$ ($p<0.00001$)	$F(5,100)=24.86$ ($p<0.00001$)	$F(5,138)=23.86$ ($p<0.00001$)	$F(1,123)=57.44$ ($p<0.00001$)	$F(1,51)=10.18$ ($p<0.00001$)	$F(1,70)=42.07$ ($p<0.00001$)
R^2	0.4913	0.5542	0.4636	0.3183	0.1664	0.3754
N	250	106	144	125	53	72

R-squared (R^2) levels are high for both MZ and DZ twins, more than 50 percent for the MZ twins cross-sectional regressions and about 40 percent for DZ twins differences. The education coefficient is positive both in the cross-sectional and fixed effects twin-differencing regressions on MZ and DZ. Similar to the DZ twins sample, the estimated ability-adjusted rate of return to schooling in Ghana for the pooled sample of twins is 8.3% (Table 5.3.1) which indicates that ability bias is positive (i.e., OLS returns are higher than FEOLS returns). This positive ability bias arising in the OLS estimates suggests that higher ability individuals invest more in schooling and also have higher earnings. Thus, the downward ability bias to monozygotic twins ordinary least squares (MZOLS) estimate implies that abilities do not differ significantly ($p<0.05$) between MZ twins.

5.3.2 Endogeneity bias

The 2SLS estimate is compared to the OLS estimate for MZ twins, DZ twins and the pooled sample to investigate the magnitude of endogeneity bias in the returns to schooling. The 2SLS return to schooling (0.0940, Table 5.3.2) for MZ twins is almost identical to the OLS estimate (0.0943, Table 5.3.2), indicating a slightly negative bias to the OLS return to schooling.

Table 5.3.2: OLS and 2SLS Estimates of the Returns to Schooling

Variable	OLS			2SLS		
	Pooled	MZ	DZ	Pooled	MZ	DZ
Intercept	4.8132 (0.4094)	4.0642 (0.6553)	5.5757 (0.5228)	4.8127 (0.4045)	4.0662 (0.6365)	5.5929 (0.5125)
Own education	0.1014 (0.0074)	0.0943 (0.0107)	0.1038 (0.0104)	0.1043 (0.0076)	0.0940 (0.0105)	0.1100 (0.0107)
First-stage F-Statistic	$F(5,244)=748.1$ ($p<0.00001$)	$F(5,100)=972.8$ ($p<0.00001$)	$F(5,138)=259.1$ ($p<0.00001$)	-	-	-
Hausman test $\chi^2(5)$				3.68 ($p=0.597$)	0.02 ($p=1.000$)	4.90 ($p=0.428$)
R^2	0.4913	0.5542	0.4636	0.4910	0.5542	0.4622
N	250	106	144	250	106	144

This means that, while the point estimates are different it is clear that the IV estimate is not significantly ($p>0.05$) different from the OLS estimate. Two important features of the 2SLS results are worth noting. For this specification, exogeneity tests (Hausman 1978) accept at a 5 percent significance level the null hypothesis that the OLS and two-stage least squares estimates of the coefficient on education are the same. In other words, there are no significant differences between OLS and 2SLS estimates of the MZ, DZ and pooled data and therefore OLS estimates are consistent, (Table 5.3.2) although, the 2SLS estimates were slightly higher than the OLS estimates. This is emphasized by the significant (p-value<0.01) found at the first-stage F-statistic. The 2SLS estimates were also statistically significantly different from zero (p<0.05) and consistent for the pooled, MZ and DZ twins, suggesting that the data support the assumption that education is endogenous to earnings. Additionally, while the instrumental variable appears to have some role in determining earnings its impact on the education coefficient is small especially for MZ twins, suggesting that endogeneity bias is virtually non-existence in MZ twins. It is consistent with the growing view that while there is evidence that education is endogenous to wages the impact of this endogeneity is small. Similarly, the OLS estimate ($\beta_1 = 0.101$, Table 5.3.2) for the returns to schooling in the pooled sample is smaller than the 2SLS estimate ($\beta_1 = 0.104$, Table 5.3.2). The hausman test of endogeneity (Diff=0.003; $\chi^2(1) = 3.68$; p>0.05), however accepts the null hypothesis of no endogeneity bias in the returns to schooling and finds no difference between the OLS estimate and the 2SLS estimate for the returns to schooling. This indicates that number of years of schooling is not affected by endogeneity bias and that the OLS estimates for the pooled, MZ and DZ samples are unbiased and consistent although the 2SLS estimates are still efficient as indicated by their significant p-values (p<0.05).

Table 5.2.8 also provides FGLS estimates of the returns to education and the resulting endogeneity bias (which is referred to as a "selection effect"). These are the results that include measures of the education of a twin's sibling as an additional regressor to control for any "family" effects that affect the absolute level of earnings. In Table 5.2.8, the selection effect for MZ twins is negative (-0.00756) implying that a regression estimator of the returns to schooling that does not adjust for the selection effect might be downward biased. The selection effect for the DZ twins was rather positive (0.016), suggesting an upward bias in the regression estimator of the returns to schooling that does not adjust for the selection effect for individuals who are not genetically identical (Table 5.2.8).

It is worth noting that the endogeneity bias in the returns to schooling for all the three samples of data using the selection effects model was not significant ($p>0.05$). This may infer that unmeasured family effects do not significantly affect earnings of individuals.

5.3.3 Measurement Error Bias

The instrumental variable method was also adopted to solve the problem of measurement error bias in returns to schooling and consequently, endogeneity of schooling. The MZ twins results indicate that the standard OLS estimate (β_1=0.094) of the returns to schooling is not biased downward by measurement error. The measurement error corrected return to a year of schooling (FEIV=0.14) in Table 5.3.4, which is assumed that differencing (FE) together with instrumenting (IV) has removed all ability bias and any measurement error bias is now about 27% higher than (FEOLS=0.11) estimate in Table 5.3.4. The hausman test of endogeneity (Diff=0.03; $\chi^2(1) = 1.02$; p>0.05) however, accepted the null hypothesis that there is no measurement error bias in the returns to schooling for MZ twins and renders FEOLS for MZ twins as efficient. On the contrary, the hausman test of endogeneity rejects the null of no endogeneity for both the pooled (Diff=0.02; $\chi^2(1) = 10.14$; p<0.01) and DZ (Diff=0.02; $\chi^2(1) = 8.77$; p<0.01) twins samples, suggesting that a fraction of the variability in the reported differences in the education levels of twins is due to measurement error (Table 5.3.4). In other words, the conventional fixed effects (FEOLS) method underestimates the economic returns to schooling, thus establishing the FE2SLS estimate as consistent and unbiased for DZ twins and the pooled sample. This shows that the DZ twins FEOLS estimates for return to schooling are likely to be biased downward by measurement error because of endogenous schooling returns.

5.4 Heterogeneity in Returns to Schooling

5.4.1 Returns to Schooling by Ability

The FGLS estimates of Equations (3.33) and (3.34), where the interaction term of the product of the individual's schooling level and the family's average schooling level is included in the equations, are presented in Table 5.4.1. The coefficient of this interaction term is the product of the correlation between ability and schooling (i.e., γ) and the heterogeneity in the return to schooling (i.e., β_1). The FGLS estimate of the return to schooling is -2.5% for MZ twins (Table

5.4.1). The coefficient of the family's average education term (i.e., γ) is estimated to be -0.041, and that of the interaction term (i.e., $\beta_1\gamma$) is estimated at 0.006.

Table 5.3.4: FEOLS and FEIV/FE-2SLS Estimates of the Returns to Schooling

Variable	FE-OLS Pooled	MZ	DZ	FE-2SLS/FEIV Pooled	MZ	DZ
Intercept	0.0818 (0.0407)	0.1146 (0.0533)	0.0600 (0.0595)	0.0712 (0.0412)	0.1162 (0.0526)	0.0404 (0.0603)
Own education	0.0835 (0.0110)	0.1115 (0.0350)	0.0823 (0.0127)	0.1053 (0.0130)	0.1412 (0.0457)	0.1043 (0.0147)
First-stage F-Statistic	F(1,123)=335.90 ($p<0.00001$)	F(1,51)=68.20 ($p<0.00001$)	F(1,70)=214.64 ($p<0.00001$)	-	-	-
Hausman test $\chi^2(1)$	-	-	-	10.14 (p=0.0015)	1.02 (p=0.3129)	8.77 (p=0.0031)
R^2	0.3183	0.1664	0.3754	0.2966	0.1547	0.3484
N	125	53	72	125	53	72

With the availability of these two estimates, the effect of ability on the marginal benefit of schooling (i.e., β_1) may be determined. It is found that the estimate of β_1 is -0.146 = (0.006/-0.041). This means that there is a negative effect of ability on the marginal benefit of schooling (or that the return to schooling is heterogeneous). Genetic endowments and family (i.e., non-genetic) factors combined have a negative effect on marginal benefit of schooling. The negative value for GLS estimate for MZ twins indicate that individuals from higher ability families receive a lower marginal benefit from their investment in human capital, thus suggesting that schooling is compensatory. On the other hand, for the DZ twins, the effect of ability on the marginal benefit of schooling was positive (0.082) indicating that families with higher levels of innate 'ability' or more

Table 5.2.8 also provides FGLS estimates of the returns to education and the resulting endogeneity bias (which is referred to as a "selection effect"). These are the results that include measures of the education of a twin's sibling as an additional regressor to control for any "family" effects that affect the absolute level of earnings. In Table 5.2.8, the selection effect for MZ twins is negative (-0.00756) implying that a regression estimator of the returns to schooling that does not adjust for the selection effect might be downward biased. The selection effect for the DZ twins was rather positive (0.016), suggesting an upward bias in the regression estimator of the returns to schooling that does not adjust for the selection effect for individuals who are not genetically identical (Table 5.2.8).

It is worth noting that the endogeneity bias in the returns to schooling for all the three samples of data using the selection effects model was not significant ($p>0.05$). This may infer that unmeasured family effects do not significantly affect earnings of individuals.

5.3.3 Measurement Error Bias

The instrumental variable method was also adopted to solve the problem of measurement error bias in returns to schooling and consequently, endogeneity of schooling. The MZ twins results indicate that the standard OLS estimate (β_1=0.094) of the returns to schooling is not biased downward by measurement error. The measurement error corrected return to a year of schooling (FEIV=0.14) in Table 5.3.3, which is assumed that differencing (FE) together with instrumenting (IV) has removed all ability bias and any measurement error bias is now about 27% higher than (FEOLS=0.11) estimate in Table 5.3.3. The hausman test of endogeneity (Diff=0.03; $\chi^2(1) = 1.02$; p>0.05) however, accepted the null hypothesis that there is no measurement error bias in the returns to schooling for MZ twins and renders FEOLS for MZ twins as efficient. On the contrary, the hausman test of endogeneity rejects the null of no endogeneity for both the pooled (Diff=0.02; $\chi^2(1) = 10.14$; p<0.01) and DZ (Diff=0.02; $\chi^2(1) = 8.77$; p<0.01) twins samples, suggesting that a fraction of the variability in the reported differences in the education levels of twins is due to measurement error (Table 5.3.3). In other words, the conventional fixed effects (FEOLS) method underestimates the economic returns to schooling, thus establishing the FE2SLS estimate as consistent and unbiased for DZ twins and the pooled sample. This shows that the DZ twins FEOLS estimates for return to schooling are likely to be biased downward by measurement error because of endogenous schooling returns.

5.4 Heterogeneity in Returns to Schooling

5.4.1 Returns to Schooling by Ability

The FGLS estimates of Equations (3.33) and (3.34), where the interaction term of the product of the individual's schooling level and the family's average schooling level is included in the equations, are presented in Table 5.4.1. The coefficient of this interaction term is the product of the correlation between ability and schooling (i.e., γ) and the heterogeneity in the return to schooling (i.e., β_1). The FGLS estimate of the return to schooling is -2.5% for MZ twins (Table

5.4.1). The coefficient of the family's average education term (i.e., γ) is estimated to be -0.041, and that of the interaction term (i.e., $\beta_1\gamma$) is estimated at 0.006.

Table 5.3.3: FEOLS and FEIV/FE-2SLS Estimates of the Returns to Schooling

Variable	FE-OLS Pooled	MZ	DZ	FE-2SLS/FEIV Pooled	MZ	DZ
Intercept	0.0818 (0.0407)	0.1146 (0.0533)	0.0600 (0.0595)	0.0712 (0.0412)	0.1162 (0.0526)	0.0404 (0.0603)
Own education	0.0835 (0.0110)	0.1115 (0.0350)	0.0823 (0.0127)	0.1053 (0.0130)	0.1412 (0.0457)	0.1043 (0.0147)
First-stage F-Statistic	F(1,123)=335.90 (p<0.00001)	F(1,51)=68.20 (p<0.00001)	F(1,70)=214.64 (p<0.00001)	-	-	-
Hausman test $\chi^2(1)$	-	-	-	10.14 (p=0.0015)	1.02 (p=0.3129)	8.77 (p=0.0031)
R^2	0.3183	0.1664	0.3754	0.2966	0.1547	0.3484
N	125	53	72	125	53	72

With the availability of these two estimates, the effect of ability on the marginal benefit of schooling (i.e., β_1) may be determined. It is found that the estimate of β_1 is -0.146 = (0.006/-0.041). This means that there is a negative effect of ability on the marginal benefit of schooling (or that the return to schooling is heterogeneous). Genetic endowments and family (i.e., non-genetic) factors combined have a negative effect on marginal benefit of schooling. The negative value for GLS estimate for MZ twins indicate that individuals from higher ability families receive a lower marginal benefit from their investment in human capital, thus suggesting that schooling is compensatory. On the other hand, for the DZ twins, the effect of ability on the marginal benefit of schooling was positive (0.082) indicating that families with higher levels of innate 'ability' or more favourable learning environments for their children gain more benefit from schooling.

Table 5.4.1: Feasible Generalized Least Squares (FGLS) Estimates of the Returns to Schooling of MZ and DZ Twins

Parameters	FGLS coefficients (MZ)	FGLS coefficients (DZ)
Own Education	-0.025 (0.090)	-0.051 (0.044)
Family Average Schooling Level $(S_1 + S_2)/2$	-0.041 (0.094)	0.049 (0.060)
Own*Average Education	0.006 (0.002)	0.004*** (0.001)
Age	0.096* (0.048)	0.031 (0.039)
Age^2(/100)	0.194 (0.129)	-0.069 (0.100)
Male	0.055 (0.157)	0.186*** (0.067)
R^2	0.431	0.503
N	106	144

***,**,*-significant at the 0.01,0.05 and 0.1 probability level respectively

5.4.2 Unobservable Differences in Returns to Schooling

REML analysis estimated the return to schooling for the pooled and DZ twins sample to be 10% and the MZ twins return to schooling estimate as 9% (Table 5.4.2) using the same variables in the OLS return to schooling. These estimates are similar to the OLS return to schooling estimates in Table 5.2.3a and are also highly significant, (p<0.01). However, the fixed effects return to schooling estimates for both DZ and the pooled sample of twins are lower than the REML estimates but higher for MZ twins. These results indicate the existence of some bias in the REML analysis and confirm the fact that failure to take account of unobserved heterogeneity leads to biased estimates on the returns to schooling. Marital status appeared to affect REML returns to schooling negatively, though the effect was not statistically significant (p<0.05). An extra year of age is estimated to increase REML earnings by about 4-6% for DZ and MZ twins respectively. The coefficient for gender dummy is negative for the pooled, DZ and MZ twins though significant for Pooled and DZ twin's samples, and it suggests that *ceteris paribus* men earn about 14-16% less than women (Table 5.4.2).

Mixed-effects model of earnings determination also presents estimation of variance components. Three estimates are shown. The first is the variance of the intercepts across families estimated at ($\sigma_{\mu 0} = 0.2902; p > 0.05$) for MZ twins and ($\sigma_{\mu 0} = 1.0583; p < 0.01$) for DZ twins, and tested for significance using the Z-statistic (Table 5.4.2). The DZ twins intercept variance is significant, indicating that the intercepts vary across DZ twin families. On the contrary, MZ twins intercept variance did not significantly differ across families. The second is the variance of the slopes, estimated at $(\sigma_{\mu 1} = 0.0025; p > 0.05)$ for MZ twins and $(\sigma_{\mu 1} = 0.0028; p < 0.01)$ for DZ twins. When tested for significance using a Z-test, the results indicates that MZ twins slope variance do not differ significantly and suggests that returns to schooling for MZ twins are not heterogenous (ie., there are no significant unobservable differences in returns to schooling for MZ twins). On the other hand, DZ twins slope variance exhibited some significant differences in returns to schooling, indicating that the slopes in the various families differ more than one could reasonably attribute to chance. That is, for some families the slopes are larger than for others. The final statistic estimated is the residual variance shown to be $(\sigma_{\varepsilon} = 0.0802; p < 0.01)$ for MZ twins and $(\sigma_{\varepsilon} = 0.0691; p < 0.01)$ for DZ twins and statistically significant (Table 5.4.2).

These values are measures of the variance not accounted for by the REML analysis and the likelihood ratio test for MZ twins is $\left(\chi^2(3) = 54.13; p = 0.01\right)$ and $\left(\chi^2(3) = 77.54; p = 0.01\right)$, Table 5.4.2, for DZ twins. The chi-square test rejects the null hypothesis of no individual unobserved heterogeneity quite strongly in the mixed effects model for both MZ and DZ twins and suggests that there are still some unobserved variables to be accounted for in the REML method of analysis.

Furthermore, OLS return to schooling is compared to the REML return to schooling estimate (Table 5.4.2) to identify unobservable differences in the returns to schooling heterogeneity using the MZ twins sample. Comparing the OLS and REML for MZ twins, returns to schooling differ by about 0.5% with a small upward bias in the return to schooling. Thus, accounting for unobserved individual heterogeneity using the REML method lowers the returns to schooling slightly for MZ twins.

Table 5.5.2 Mixed Effects of Pooled, MZ and DZ Twins

Fixed Parameters	Pooled	MZ	DZ
Intercept	4.7509*** (0.6045)	4.6567*** (0.8299)	5.3615*** (0.8414)
Years of Schooling	0.0966*** (0.0101)	0.08928*** (0.0154)	0.0974*** (0.0124)
Age	0.0711** (0.0343)	0.05796 (0.0497)	0.0424 (0.0474)
Age squared	-0.0008 (0.0004)	-0.00032 (0.0007)	-0.0005 (0.0006)
Gender	-0.1424** (0.0567)	-0.0243 (0.1479)	-0.1608*** (0.0584)
Married	-0.0523 (0.0904)	-0.0967 (0.1566)	-0.0209 (0.1052)
Random Parameters	**Variance Components**	**Variance Components**	**Variance Components**
Intercept $(\sigma_{\mu 0})$	0.8350*** (0.2107)	0.2902 (0.3815)	1.0583*** (0.3009)
Slope $(\sigma_{\mu 1})$	0.00295*** (0.0011)	0.0025 (0.0024)	0.0028*** (0.0011)
Residual (σ_{ε})	0.0741*** (0.0101)	0.0802*** (0.0157)	0.0691*** (0.0132)
Likelihood Ratio Test	$\chi^2(3)=131.45$ ***	$\chi^2(3)=54.13$ ***	$\chi^2(3)=77.54$ ***
Intra Class Correlation	0.08	0.22	0.06
N	250	106	144

***,**,*-significant at the 0.01,0.05 and 0.1 probability level respectively

Likewise the DZ twins REML return to schooling estimate (0.097), Table 5.5.2 turns out to be also lower than the OLS return to schooling estimate (0.104),

Table 5.2.3a. This indicates that for individuals who are not genetically identical, the OLS returns to schooling estimate is consistent and the REML return to schooling estimate is biased downward. The OLS returns to schooling increases returns to schooling for DZ twins by about 7% compared with the REML returns to schooling. REML returns to schooling for the pooled twins sample also exhibited similar results compared with the DZ twins estimate (Table 5.4.2). The significant ($p<0.01$) likelihood ratio test in all the three samples, indicates the need to account for unobserved heterogeneity in the returns to schooling in order to obtain consistent and unbiased estimates.

CHAPTER SIX

DISCUSSIONS AND CONCLUSIONS

6.1 Returns to Education

This study investigated the economic return to education in Ghana, using data on twins. A study of the return to schooling using twins is important because of the need to control for ability bias in estimating the effect of education on earnings or to examine the relative contribution of genetic ability and the effects of family background to the observed cross-sectional return to schooling in order to identify a “pure” effect of education on earnings.

The goal of this study was to empirically measure the effect of education on earnings by using twins’ data in Ghana. It is argued in the literature (Ashenfelter and Krueger, 1994; Miller et al. 1995; Behrman and Rosenzweig, 1999; Bound and Solon, 1999; Isacsson, 1999), that monozygotic (from the same egg) twins are genetically identical and have similar family background, and therefore, the effects of unobserved ability or family background should be similar for both twins. Thus, taking the within-twin pair difference will, to a great extent, reduce the unobservable ability or family background effects that cause bias in the OLS estimation of the return to education. Intuitively, by contrasting the earnings of identical twins with different years of education, we can be more confident that the correlation that we observe between education and earnings is not due to a correlation between education and an individual’s ability or family background.

6.1.1 Human Capital Returns to Education in Ghana

The estimated rate of return to schooling using Mincer's human capital model and the pooled sample of twins was 9.8% (Table 5.2.1). This rate of return was 1.5 times the 1995 estimate observed by Jones (2001) in Ghana. For the purposes of separately identifying the components of ability bias that are due to genetic characteristics and family background, this study also estimated returns to additional schooling by type of twins (MZ and DZ) using Mincer's model. 'Twins' studies exploit the idea that it is possible to estimate the effect of schooling on income by comparing the earnings received by twin pairs who obtain different amounts of schooling, but are assumed to have similar ability levels. The estimated rate of return to schooling in Ghana for monozygotic twins is 9.5 % compared to 9.8 % for the dizygotic twins (Table 5.2.1).

The results of this study are however consistent with the worldwide average returns to schooling (about 10 %), as compiled from hundreds of studies, (Psacharopoulos and Patrinos 2004) and at the lower side of conventional average return for Africa, which is between (7-20%), (Uwaifo, 2006). Although, the rate of return for Sub-Saharan Africa is 11.7 percent (Psacharopoulos & Patrinos, 2004) it is low compared to South Africa (Keswell and Poswell, 2004) which ranges from 15-26 % and Ethiopia with an average return of 15.0 % (Girma et.al. 1994). The return to education in Nigeria also ranged from 12 to 13.5 % between 1974 and 1986 (Ryoo et al. 1993). However, in Kenya, which is similar in terms of economic development, the return to education almost equals that for Ghana in this study. Hawley, (2004), estimated an average return in Kenya at between 10.3

and 10.7 % from 1985 to 1998. Both Ghana and Kenya enjoys considerably higher returns compared to some countries in West Africa.

In Ivory Coast, average returns from education for an additional year of schooling were 4.8 % for the overall sample, and 3.4 and 6.8% for males and females, respectively (Moock et al, 2003), whereas in Burkina Faso young people benefited slightly more than those in Ivory Coast from an additional year at 7.0 % in 1995 (Duflo, 2001). The Mincerian returns to an additional year of schooling by gender reported in this study are around 11% for males and 9% for females. The return to education is slightly higher for men than women and therefore gender differences in our study are not so large indicating that returns to schooling do not differ significantly (*se*=0.01) by gender (Table 5.2.2). This insignificant difference is surprising, since in Ghana, there is a commonly held view that families frequently consider it less important to ensure that girls obtain an education than boys. Boys are conventionally expected to be breadwinners for a family and girls are expected to get married, play the role of housewife and mother and perhaps take a job that is complementary to this role. Consequently, the range of jobs considered appropriate for girls to aspire to has largely been restricted to those traditionally female occupations such as (nursing, office work, teaching, etc.). Although such attitudes may not be as widespread today as they once were, a substantial proportion of workers undoubtedly entered the labour market at a time when they were prevalent. To the extent that this view actually guides families in their attitudes to their children's careers, it will tend to result in girls spending less time in education than boys and undertaking programs of education that lead to employment in the traditionally female occupations. Furthermore, the female wage disadvantage is consistent with findings in the literature that use measures of annual income (Jones, 1983) and on the basis of the studies reviewed in Rummery (1992), around one-half of it is likely to be associated with gender differences in hours worked. Though the returns to schooling for males in this study was similar to findings by Aslam, (2005), the returns to schooling for females was considerably lower in this study and contrary to findings by Behrman and Deolalikar, (1995) and Asadullah, (2006) who report higher returns to schooling for females.

The returns to schooling by gender results in this study are also similar to that of Schultz (1995) whose estimated wage returns to schooling for women are approximately equivalent to those for men, with occasionally the private wage returns being higher for women than men in those economies where women have not attained equal levels of schooling to men.

The negative experienced-squared coefficient found when estimating Equation (3.1) in this study showed that the experience earnings profile in the HCEF model was concave. This finding (earnings function concavity) is virtually universal across countries and years. This means that for those continuously attached to the labor market, earnings rise at a decreasing rate throughout one's life until depreciation exceeds human capital accumulation. Mincer, (1974) shows that the concave experience earnings profile that we observe in the data is implied by declining investment ratios (i.e., investment relative to potential earnings). Polachek, 2007, also depicts cross-sectional Mincer earnings functions for 25 countries and various years and found out that, the experienced-squared quadratic coefficient is always negative. Furthermore, Gautier and Teulings (2003) also present strong evidence for the concavity of experience earnings for six OECD countries.

6.1.2 Modeling Returns to Schooling using Twins

Estimates of the returns to education based on analysis of twins' earnings come to an average rate of return that is very similar to the global average (10%), Psacharopoulos and Patrinos (2002). The estimated rate of return to schooling in Ghana for MZ and DZ twins was 9.4% and 10.4%, respectively. The MZ twins' results in this study are qualitatively similar to that of Ashenfelter and Krueger (1994); Behrman et al (1977) and Rouse (1998) which were 8.4%, 8.0% and10.5%. The results of our study therefore support the findings of Ashenfelter and Krueger (1994) that estimates of the economic returns to schooling may have been underestimated in the past. However, these rates of return to schooling are high compared to similar estimates for other countries. Miller et al. (2006) estimates the mean return to schooling (OLS) for Australia to be 6% and 5.5% for MZ and DZ twins respectively.

The estimated rate of return to schooling for DZ twins is slightly higher than that of MZ twins. This trend is contrary to findings by Miller et al, (2006). The difference in the rate of return may be influenced by the occupations held by the individuals earning the relevant basic wage between the MZ and DZ twins'. The most obvious reason lies in the number of wage-earners working in the state-owned or publicly owned sector. State and local government workers actually are underpaid given that they are equally educated and experienced as their counterparts in the private sectors, according to University of Massachusetts researcher Jeffrey Thompson and John Schmitt, of the Center for Economic Policy Research ***(http://www.hartfordbusiness.com/news14780.html)***. Thompson further stated that state and local government workers in New England are more highly educated and more experienced than their counterparts in the private sector. But once you properly control for education and experience, it becomes evident that public sector workers get lower wages. Thus the lower rate of return to education in the MZ twin group may have been caused in part by the nature of employment in the public sector, which may be less inclined to base wages on the "human capital" variables such as education and experience (Cheung, 1990). The pooled estimate of the rate of return to schooling estimated to be 10.1% is comparable to other studies that are not twin based. These OLS results are slightly higher than earlier Ghanaian OLS results in Jones (2001) which suggest estimated returns of the order of 7% but are comparable to most of the estimates that have appeared in the literature (Psacharopoulos and Patrinos, 2002).

6.1.3 Endogeneity of Schooling

Endogeneity causes OLS estimates of returns to schooling to be biased and inconsistent, Griliches (1977). However, their calculation has not been abandoned in order to evaluate the sign and amount of the bias, but results are ambiguous. Griliches suggested the possibility of both upward and downward bias. Upward bias remains the conventional wisdom (Ehrenberg-Smith 1991), even if some authors imply that this is true only if one analyses wages of mature workers (Blackburn-Neumark 1993).

Calculation by instrumental variables and by the fixed effects model gives, instead, downward bias. Griliches, Hall and Hausman, (1978) found an IV estimate double the OLS estimate using family background variables as instruments for schooling. Angrist-Newey 1991, in comparison to an OLS estimate equal to 0.036, found a fixed-effect estimate equal to 0.080.

The same happens when the natural experiment approach is applied. For example, the works surveyed by Card 1995a and 1998 report an increase in the estimates in a range of 10-100% in comparison to OLS estimations.

These contrasting empirical results are due to econometric problems: omitted variables and measurement errors could involve opposite distortions without specifying which one prevails. It is widely recognized that using OLS to estimate the returns to education from cross-section data is potentially problematic. The standard concern in the literature is that education is an endogenous variable, positively correlated with the earnings residual due to unobserved ability. It is also possible that there is heterogeneity in the returns to education at given levels of education, and that unobserved ability is correlated with the returns (Belzil, 2004). In either case, OLS estimates of the parameters would be biased.

The endogeneity of educational attainment and ability bias which occurs as a result of the bias and inconsistency in the OLS estimator, may be solved by taking advantage of exogenous determinants of schooling decision (i.e, IV method) or compare earnings between genetically identical twins or highly genetic siblings conditional on their education attainment (within-family fixed effect) or utilize panel data. A recent solution to this endogeneity problem has been found in identifying exogenous sources of variation in schooling to build a new set of instrumental variables for years of education attained (Angrist and Krueger 1991; Card 1998). Using the co-twin's report on educational attainment as the instrument for the twin's self-reported educational attainment to account for endogeneity of schooling, an additional year of schooling is associated with an increase in earnings of approximately 9.1% for MZ twins in this study. This estimate of how schooling affects earnings is virtually the same as the OLS estimate (9.2%). However, the t-statistic for the IV estimate indicates that the coefficient for this variable is statistically significantly different from zero indicating that the data support the conjecture that education is endogenous to earnings. Thus while the IV appears to have some role in determining earnings its impact on the education coefficient is small suggesting to some extent that unobserved 'good' characteristics have a positive effect on both earnings and schooling. Hence, the correlation between the residuals in the earnings equation and the schooling equation are positive. It is therefore consistent with the growing view that while there is evidence that education is endogenous to earnings the impact of this endogeneity is small (Rummery, 1999). That is, while the point estimates are different it is clear that the IV estimate is not significantly different from the OLS estimate. It is also interesting that the IV estimate is smaller than the OLS estimate. This is in contrast to many alternative studies which frequently find that the OLS estimate is lower than the IV estimate, Card (1994); Ashenfelter and Krueger (1994) and Harmon and Walker (1995). Grilliche's (1977) use of an IV estimator in place of ordinary least squares resulted in a 50-percent increase in the returns to schooling. The instruments used were family background factors (e.g., mother's education, father's occupation). Contrary to the MZ twins estimate, the DZ twins IV estimate in this study is higher than its OLS estimate. This is similar to most studies (Miller et al (2006); Bonjour et al, (2003) and Angrist and Walker (1991)) that have found that controlling for the endogeneity of education generally leads to increased estimates of the returns to schooling.

However, the direction that the IV estimate will vary from the OLS estimate depends purely on which group of individuals is influenced by the instrument chosen, Rummery (1999). Generally, many studies find that IV estimates are larger than OLS estimates. Card (1995), and in Ashenfelter and Zimmerman (1997), the use of parental education as an instrument leads to estimates that are at least 15% above the corresponding OLS estimates. In Butcher and Case (1994), IV estimation based on a sibling instrument yields an estimated return of 18%, double the OLS result. Card (1993) used geographic proximity to a four-year college education as an instrument for education and again found that the estimated returns to schooling almost doubled, from 7% to 13%. Researchers find that such results are caused by measurement error in schooling levels. More specifically, if an individual's schooling level is measured erroneously and the true value of the returns to schooling is positive, the OLS estimate will be biased toward zero. Thus, the OLS estimate will be too small because of attenuation bias.

Furthermore, the results from IV studies are varied, but majority point towards the presence of a downward bias in OLS estimates. Card (1998) has proposed an explanation for this phenomenon that is based on the hypothesis that the return to schooling is heterogeneous and declines at higher levels of schooling. IV estimates will differ from OLS estimates to the extent that the instrument influences schooling decisions at different levels. If the instrument influences decisions primarily at lower levels of schooling, the IV estimator may be higher than the OLS estimator because it reflects the payoff to schooling at lower rather than higher schooling levels.

Endogeneity of schooling may thus result from unobserved measurement error in education. Measurement errors in the observed education variable may push the estimated return to schooling towards zero, since they lead to variation in the education variable that has no effect on earnings. Measurement error biases estimated returns to education, thereby calling into question the interpretation of these returns in understanding the effect of education on earnings, an issue of significant concern to policymakers. Taking account of measurement errors using an instrumental-variables (IV) estimator resulted in marked changes in the estimates in this study and suggests, however, that ordinary least squares estimates are subject to a downward bias. Similar to the findings of Ashenfelter and Krueger (1994), this study finds that the level of earnings increases from (11% to 14% for MZ twins and 8% to 10% for DZ twins) for each year of education when the IV method is applied to the fixed effects model in the Ghanaian twins sample. Ashenfelter and Krueger (1994) find that measurement error in their sample of US twins increases the estimated rate of return to schooling from (9.2 % to 16.7%) in the fixed-effects model and from 8.8 per cent to 11.6 per cent in the selection effects model). These results suggest that there is evidence of measurement error in the twin differencing estimate. This estimate is also consistent with the study by Rouse, 1999 who find that the within-twin measurement error corrected estimates are also larger than the within-twin differencing estimates. In this study, the instrumental variable estimates that are intended to correct for measurement error in the data are much larger than the least squares estimates and they are consistent with findings in other studies (Bingley et al, 2005) that a considerable fraction of the variability in reported differences in twins educational level is due to measurement error. More specifically, if a twin's schooling level is measured erroneously and the true value of the returns to schooling is positive, the OLS estimate will be biased toward zero. Thus, the OLS estimates of twins in this study are downward biased as a result of some measurement errors. Based on the findings of Card (1999), measurement error bias itself can explain the 10% gap in the estimated returns between OLS and IV estimation. Therefore, two biases generally exist simultaneously in applying the OLS estimation: the upward bias, caused by omitted ability variables, and the downward bias, caused by measurement error in schooling. If the instruments are not correlated with the measurement error in the schooling level, then IV estimates will be free from both biases. The result of IV estimation depends on the relative magnitudes of the omitted ability and attenuation biases. Interestingly, Isacsson, (1997) constructs IV estimates for the within-family model using the difference in the survey measures of schooling as an instrument for the differences in the registry measures.

For MZ twins, the within-family IV estimator is only marginally above the within-family OLS estimate, implying almost no measurement error bias and on the other hand, the IV procedure raises the within-family estimate by 35 percent for DZ twins indicating a bigger measurement error bias. Similar to Isacsson's 1997 findings, Ashenfelter and Zimmerman's, (1993) measurement-error-corrected estimate of the return to schooling for brothers, constructed under the assumption that brothers have identical abilities, is also about equal to the corresponding OLS estimate.

Controlling for the potential endogeneity of schooling by using twins, "natural experiments" or other instrumental variables does not generally reduce the size of the private return (in fact the size of the coefficient usually increases, Card, 1999 and Harmon, Oosterbeek and Walker, 2002). The fixed effects model of the earnings of twins allows the impacts of ability and shared family characteristics to be assessed indirectly (by comparison of the estimates made from samples of MZ twins, DZ twins and individuals). The within-twin estimate of the return to schooling hold out the promise of eliminating omitted ability bias and 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 and Martin, 1995). In our study, the MZ within-twin (or first-differenced) estimate (0.112) of the return to schooling was higher than the comparable cross-sectional estimate (0.92) but followed a similar pattern as in Ashenfelter & Krueger (1994). This pattern is consistent with a downward bias in the OLS estimate due to an omitted family or ability effect. In other words, our estimate indicates that the correlation between omitted ability and schooling may be slightly negative in cross-sectional analyses. Ashenfelter and Krueger (1994) report that within-twin estimates of the return to schooling without correcting for measurement error are high relative to OLS cross-section estimates, compared with what is predicted based on the extent of measurement error in schooling levels and differences. This they state may be attributable to downward, rather than upward omitted ability bias in the cross-section estimates of the return to schooling. In addition, the high within-twin estimates may also be attributable to remaining ability differences between twins that, although smaller, lead to greater upward bias in standard within-twin estimates than in the cross-section estimates (Griliches, 1979). This result is also contrary to earlier studies, (Altonji & Dunn (1996); Ashenfelter & Zimmerman (1997); Behrman et al, (1980) and Rosenzweig & Taubman (1994)) using identical twins that find within-twin estimates considerably lower than cross-section estimates (i.e., there is upward bias in the OLS estimate). However, such findings may be attributable to exacerbation of bias from measurement error in schooling that is caused by differencing across twins (Griliches, 1979). Contrary to the findings in this study, Miller et al. (2006), find that the within-family estimate of the return to education for MZ twins is almost 50 percent lower than the cross-sectional estimate and for

DZ twins, the within-family estimator is 40 percent lower. However, following Ashenfelter and Krueger's (1994) procedure of using one twin's responses on the difference in schooling for the pair as an instrument for the other's responses, Miller et al's, 2006 resulting MZ twins IV estimate is about 40 percent above the differenced OLS estimate, but still 25 percent below the cross-sectional estimate, whiles the DZ twins IV estimate is actually slightly above the OLS estimate.

Behrman et al (1994) analyze a data set that pools the NAS-NRC sample of white male World War II veterans with data on men from the Minnesota Twins Registry and reports that MZ twins within-family estimate of the return to schooling is about 50 percent as large as the cross-sectional OLS estimate, while for DZ twins the relative ratio is 80 percent. These estimates confirms the findings in this study that OLS estimates for MZ twins are lower than the twin differencing estimates, but does not support the smaller within-twin DZ twins estimates in this study. Furthermore, although they do not actually estimate IV models to correct for measurement error, Behrman et al report that the reliability of the within-family difference in schooling for identical twins in the NAS-NRC sample is 0.62. Using this ratio is slightly higher than the ratio reported in earlier work by Behrman et al (1980) for identical twins in the NAS-NRC sample. Griliches (1979) characterized their results as showing a 65 percent reduction in the return to schooling between the OLS and within-family estimators.

Finally, in contrast to the results in this study, Isacsson's (1997) analysis of earnings and schooling differences among a large sample of Swedish twins (about one-half women) shows that the within-family estimate of the return to schooling for identical twins in the subsample with two schooling measures is less than 50 percent as large as the corresponding OLS estimator. However, Isacsson reports similar findings to this study that DZ twins OLS estimates are larger than twin-differencing estimates. Since DZ twins are essentially brothers (or sisters) with the same age, the similarity of the findings for DZ twins is reassuring. This implies that ability differences between brothers or sisters are relatively less important determinants of within-family schooling outcomes than are overall ability differences in the determination of schooling outcomes for the population as a whole.

Based on the strong intuitive appeal of the "equal abilities" assumption for MZ twins this study exploited the within-family fixed effect model to eliminate ability bias and found out that OLS estimates are biased downward for MZ twins by about 15%. This result is similar to estimates by Ashenfelter and Krueger's (1994) reliance on the assumption that abilities do not differ between MZ twins. Interestingly, when comparing the FE estimates with the OLS estimates for the DZ twins and the pooled twins' sample in this study, the OLS estimated effect of education on earnings actually increases contrary to what is theoretically predicted that the FE estimates control for genetic traits and family background. However, other studies indicate that there are ability differences between twins correlated with both education and earnings that will bias the estimates if they are not controlled for (Sandewall, 2009). A further issue, particularly relevant in twin-differencing methods, discussed by Sandewall, (2009), is the importance of variation in the education and earnings. Consequently, a high degree of similarity in the ability, education relationship, twin-differencing does not contribute as much in terms of removing the ability bias (Sandewall, 2009).

Bound and Solon (1999) examine the implications of the endogenous determination of which twin goes to school for longer, and conclude that twins-based estimation is vulnerable to the same sort of bias that affects conventional cross-sectional estimation. The major concern of the within-twin-pair estimate is thus whether it is less biased than the OLS estimate, and is therefore a better estimate (Bound and Solon, 1999; Neumark, 1999). They argue that although taking a within-twin-pair difference removes genetic variation this difference may still reflect an ability bias to the extent that ability consists of more than just genes. In other words, within-twin-pair estimation may not completely eliminate the bias of conventional cross-sectional estimation, because the within twin-pair difference in ability may remain in the between twin-pair difference error which may be correlated with the between twin-pair schooling.

Although within-twin-pair estimation cannot completely eliminate the bias of the OLS estimator, it can tighten the upper bound on the return to education. Ashenfelter and Rouse (1998), Bound and Solon, (1999) and Neumark (1999) have debated the bias with OLS and within-twin-pair estimation at length in recent papers and notes that the bias in the OLS estimator depends on the fraction of variance in education that is accounted for by the variance in unobserved ability that may also affect earnings. Similarly, the ability bias of the fixed effects estimator depends on the fraction of within-twin-pair variance in education that is accounted for by within-twin-pair variance in unobserved ability that also affects earnings.

Some researchers find evidence of significant ability bias, many also find evidence of little ability bias while yet others find that the return to education is biased downwards. Evidence of downward bias has been particularly prevalent in recent clever instrumental variables estimates, leading many labour economists to conclude that ability bias is not a problem (Lang, 1993).

A final point made by Neumark (1999) is that because the OLS within-twin estimates of the return to schooling suffer from both downward and upward biases (due to measurement error and ability biases), whereas the IV within-twin estimates only suffer from upward biases (due to ability bias), the OLS within-twin estimate of the return to schooling may most accurately reflect the 'true' return to schooling. This situation would exist if the within-twin ability bias were large enough to off-set the (relatively large) measurement error bias. This is a theoretical possibility that requires an estimate of the magnitude of the potential ability bias in twins' analyses. While the magnitude of this bias is currently unknown, the upward ability bias in cross-sectional estimates likely provides an upper-bound. Given that recent IV estimates of the return to schooling employing other identifying strategies suggest little ability bias in cross-sectional estimates [Angrist & Krueger (1991); Card (1993); Kane & Rouse (1993)], the overall evidence likely favors the measurement error corrected within-twin estimates.

In contrast to the fixed effect model, the selection effects model explicitly incorporates family effects in the earnings equation. The selection effects model of Ashenfelter and Krueger (1994) allows a direct assessment of the magnitude of these effects (ability and shared family characteristics) through an explicit modeling of the family effects factor. Estimating the selection effects model using the Ghanaian twins' data indicates that the return to schooling net of the effects of natural ability and shared family background for identical twins is about 11%. The results for non-identical twins from the selection effects model are similar to those obtained from the fixed effects model, with the returns to education, including effects due to natural ability, being estimated at 8%. These findings are similar to those of Ashenfelter and 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 our findings, Miller (2001) indicates that family background accounts for the remaining 2 percentage points of the conventionally estimated overall return of 6.5 per cent.

6.1.4 Heterogeneity in the Returns to Schooling

Focusing on the important question of whether heterogeneity exists in the returns to schooling relationship in this study, and, if so, what is the best way to model that heterogeneity, this study yielded additional evidence that genetic endowments and family (i.e., non-genetic) factors (ability) combined have a negative effect on the marginal benefit of schooling for MZ twins' suggesting that the return to schooling is heterogeneous. These results are consistent with previous research on the returns to schooling by ability (Ashenfelter and Rouse, 1998; Lee, 2000). Ashenfelter and Rouse (1998) analyze an expanded version (two additional years) of the sample of genetically identical twins used in

Ashenfelter and Krueger (1994). They exploit the presumed similarity of twins and the availability of multiple measures of schooling to explicitly model the link between family ability and education parametrically, while addressing the measurement error and endogeneity biases using standard panel data methods. They find some evidence of the existence of a negative relationship between ability and returns to education, suggesting that less able individuals benefit more from additional schooling.

In analyzing the extent to which the return to schooling varies with ability level, the main finding by Ashenfelter and Rouse (1998) is that individuals from higher ability families receive a lower marginal benefit from their investment in human capital thus suggesting that schooling is compensatory. These results illustrate the important role unobserved ability plays in determining both schooling and earnings of an individual. Ability contributes to the explanation of differences in schooling, which give rise to differences in earnings between individuals and subsequently influences the return to schooling.

Contrary to these findings Nordin, (2007) shows that there is a strong and positive relationship between returns to schooling and ability for Sweden. He further indicates that for individuals belonging to the lower part of the ability distribution, the return to schooling is about 60 percent lower than for individuals in the middle of the ability distribution. For high ability levels, the return to schooling is about 30 percent higher than for those belonging to the middle of the ability distribution. The same relationship is not evident in the US when using a similar empirical model. Altonji and Dunn (1996) do not find a significant interaction effect between ability and years of schooling when using a standard Mincerian wage equation. Yet, even if the relationship is weaker in the US than in Sweden, there are studies indicating that there is heterogeneity in returns to education in the US as well (Carneiro, 2002; Carneiro et al. 2001, 2003).

The existence of a positive correlation between ability and schooling levels implies that, individuals with higher levels of ability attain higher levels of schooling. This suggests that omission of ability in the earnings estimating equation leads to estimates of the return to schooling that are upward biased. The optimal schooling model outline that returns to schooling may vary with ability when considering family background interactions. The finding of a negative relationship between ability and returns to schooling for MZ twins in this study is in line with the finding by Ashenfelter and Rouse (1998) but in contrast to results reported by Miller *et al*, (2001) based on a sample of genetically identical twins in the U.S, and by Bauer *et al*, (2002) that returns are higher for the more able in Japan. For South Africa, Mwabu and Schultz (1996) report that ability and returns are positively related among white South African who received higher education, whereas returns are homogenous amongst blacks with high education. But at the primary education level, they find that returns to education and ability are negatively related. Following Mwabu and Schultz (1996), we interpret a negative ability-returns relationship as evidence that education is a substitute for ability. This means that maximizing (private) returns to schooling requires the expansion of educational opportunities for the less able or the more disadvantaged.

Krishnan (1996) using a Youth Employment Survey in Ethiopia argues that, the relatively low (but still economically significant) returns at the higher end of the earning spectrum is consistent with the notion that there are important factors (This can take the form of inherent ability, or family connections) leading to high paying employment, which act independently of education-generated human capital. Similarly, Krueger (1999) also argues that the educational production function is concave, so students who are at the lower end of the ability distribution because of their endowments benefit more from additional human capital than students at the higher end.

Our findings are consistent with the findings of Ashenfelter and Rouse (1998) of lower marginal (average) returns for higher ability individuals after controlling for the endogeneity and measurement error in schooling. However, they are at odds with the findings of higher returns for the more able of Conneely and Uusitalo (1998) based on estimation of conditional mean wage functions and the use of test scores to proxy ability.

In estimating the standard model of human capital accumulation, it is usual for the econometrician to assume that the return to schooling is constant across individuals. However, there are good reasons to think that the true return to schooling may vary across individuals. Roy (1951), Willis and Rosen (1979), and Willis (1986) view human capitals as heterogeneous multidimensional attributes, and people choose their educational attainment based on the comparative advantage of their different attributes of abilities. Multilevel models are statistical models used to analyze data that have a hierarchical or nested structure. Thus, to address this issue, using a multilevel approach will be a more accurate model than an ordinary least squares regression (OLS). The mixed model, through its ability to allow for heterogeneity in intercepts and returns to education, has been interpreted as a model that allows marginal costs of and marginal returns to education to vary across individuals in the population (Card, 2001). To address these concerns, the mixed model used in this study might be regarded as a model that captures heterogeneity. Relatedly, we note that some of our analysis also adds controls for potentially important omitted variables such as ability and family characteristics. Hence, variables often argued to be responsible for the endogeneity and omitted variables problem can be controlled for, potentially eliminating the need to analyze a more structural specification. The extent of variability in returns to schooling across individuals is obviously a focus of our study. One possible solution to the problem of unobserved heterogeneity is to introduce a random effect in the hope that it will capture the effects of omitted variables that are independent of the covariates in the model. To control for unobserved heterogeneity this study compared the MZOLS results (0.093) to the MZ-mixed effects results (0.089) and the DZOLS results (0.0896) to the DZ-mixed effects results (0.097). Whiles MZ twins exhibited upward bias in OLS estimate, the DZ twins OLS estimate showed a downward bias. The DZ result is consistent with findings by Walker and Zhu, (2001) whose random coefficient estimates for both men and women in the Labour Force Surveys 1992-2000 were higher than their OLS results. Thus taking into account unobserved individual heterogeneity for MZ twins' lowers returns to schooling when using the mixed model, but increases returns to schooling for DZ twins. This seems to indicate that failure to take into account workplace unobserved heterogeneity will lead to biased estimates on the returns to schooling. Studies have indicated that a downward bias in returns to schooling is as a result of some measurement error, Griliches (1977) and Blackburn and David Neumark (1995) and any upward bias is due to ability and family background differences. Licht and Steiner (1991) estimate human capital earnings functions and try to control for unobserved heterogeneity using panel data, relying on individual changes in education over time to identify the returns to education.

Clustering induces unobserved heterogeneity, which means that the cluster means of the dependent variable will vary across clusters because of unmeasured cluster level factors. Random coefficient models can account for this type of heterogeneity which means that level-1 of the HLM effects vary over clusters due to unmeasured factors. By treating the return to schooling as a random coefficient the variance component of the slope in the MZ twin sample was not significant which is consistent with findings by Ashenfelter and Krueger (1994); Ashenfelter and Rouse 1998, and Behrman *et al* 1980) that MZ twins are supposed to have similar ability and similar family background. Though one problem for their approach is whether the twins' abilities are similar enough and their wages and educations are different enough. Contrary to these results, the variance component (slope) for the DZ twin sample was significant affirming the fact that returns to schooling is heterogenous. Furthermore, the variance components for the intercept and the residual for DZ twins were also significant whiles only the residual was significant for the MZ twins. The empirical results suggest that there exists significant heterogeneity in the data, which is reflected in variation in the estimated intercept and slope coefficients. Similarly, Bingley et al. (2005) exploit panel data using mixed model to show that there are significant variances to the returns to schooling estimates and find that individual variance in returns is smaller for MZ twins than for DZ twins. Since the assumption that any within-family ability differences are either negligible or uncorrelated with schooling decisions, it is important to recognize that MZ twins are formed only when a fertilized egg is accidently split. There can therefore be no genetic differences between identical twins as they have precisely the same DNA. The only way differences might appear between identical twins could be from, where twins are incorrectly classified as identical when they are not or when their level of education are misreported resulting in measurement errors. Fraternal twins are formed when more than one egg is fertilized, and they are no more alike genetically than ordinary brothers and sisters. These results suggest that individuals with higher levels of ability receive slightly higher levels of schooling. As a result cross-sectional estimates of the return to schooling are marginally upward biased by an omitted ability variable. At the same time, the higher ability individuals may receive a slightly lower marginal benefit to schooling (Ashenfelter and Rouse, 1998; Patrinos et al. 2006). Similarly, Chen (2002) used US panel data (NLSY) to separate the variation in the returns to college into heterogeneity and risk components, and found that almost all the variation in returns is accounted for by the heterogeneity component.

6.2 Conclusions

In this study we present estimates of the returns to schooling based on a sample of Ghanaian twins. The data is drawn from a twins' survey of 106 MZ and 144 DZ twins, about 51% of whom are women. This study indicates that the OLS rate of return per year to schooling for Ghana is about 10%. This result is however consistent with the worldwide average returns to schooling (Psacharopoulos and Patrinos 2004). This study also indicates that the economic return to schooling is slightly higher for men than women. The returns to schooling do not differ significantly by gender which is contrary to many studies that find significant differences between earnings of males and females. This study also noted the inverted U-shaped pattern between age and earnings, based on the average earnings by age for all workers at a given time. This finding (earnings function concavity) is virtually universal across countries and years.

The results of the MZ twins in this study suggest that ignoring the endogeneity of education leads to a substantial underestimation of the returns to education. The results further indicate that estimated returns to education are significantly downward biased when the endogeneity of education is ignored. That is, this study presents baseline estimates that suggest that OLS on cross section data is biased downwards, for MZ twins because of measurement error. The extent of bias induced by measurement error is higher for MZ than for DZ twins. The study also finds that the simple fixed effects estimators are also biased downwards for DZ twins but biased upwards for MZ twins. Furthermore, FEIV returns to schooling estimates are about 50% higher than the OLS for MZ twins, indicating that returns increase about 30% when measures are taken to account for endogeneity of schooling. On the other hand the FEIV estimates for DZ twins are 16% higher than OLS returns to schooling. Thus, our modeling resembles the previous earlier US research by Ashenfelter and Krueger that found FEIV MZ estimates that were larger than the corresponding cross section OLS. On the other hand, the measurement-error-corrected within-family estimator of the return to education for DZ twins is about equal to the corresponding OLS estimator. Additionally, OLS returns to schooling were found to be biased upwards for MZ twins and biased downward for DZ twins when the FGLS estimator employed in the selection effects model was used to measure the extent of the bias in the OLS using family background characteristics. Hence it seems that there measurement error in schooling in Ghana, which points to the issue of reliability of education data from developing countries, and the importance of correcting for such biases in order to obtain better estimates of returns to schooling.

In analyzing the extent to which the return to schooling varies with ability level, the main finding of this study is that individuals from higher ability families receive a lower marginal benefit from their investment in human capital. As a result, some evidence of the existence of a negative relationship between ability and returns to education suggests that less able individuals benefit more from schooling. It is also found that there exists a positive correlation between ability and schooling levels. That is, individuals with higher levels of ability attain higher levels of schooling. This suggests that omission of ability in the earnings estimating equation leads to estimates of the return to schooling that are upward biased. Hence, this study illustrates the important role unobserved ability plays in determining both schooling and earnings of an individual. Ability contributes to the explanation of differences in schooling, which give rise to differences in earnings between individuals. It also influences the return to schooling. Thus, results obtained in this study, provide a further contribution to the literature on returns to education.

This study also exploits the twins' data to show that there are significant variances to the returns to schooling estimates. A linear mixed model was adopted to examine the effects that individual and family effects have on return to schooling and observed that individual variance in returns is smaller for MZs than for DZs. The return to schooling for an individual was positive and significant. Consequently, individual and family characteristics were found to explain a part of the differences in the returns to schooling and that an earnings/education model that does not take into account these characteristics may overestimate the returns to schooling.

Finally, viewing the OLS and IV estimates as lower and upper bounds on the value of the return to schooling in the manner of Black et al. (2000), it is seen that in this study the bounds on the return to schooling are similar to that of Ashenfelter and Krueger's (1994) results. There appears little controversy in the general principle underpinning the theory of schooling and earnings - schooling adds considerably to the earnings of individuals. What is at the centre of the debate is that in any context schooling is a choice variable and may not be independent of other factors that affect earnings. This raises the possibility that the observed correlation between schooling and earnings is not a causal relationship, but merely masks a correlation between other factors, such as ability, and earnings.

The findings of this study therefore suggests that

1. Genetics, family background and schooling are all important in determining income in the labour markets.
2. The pure returns to schooling are greater for MZ twins than for DZ twins.
3. The family background effect is considerably greater for MZ than for DZ twins.
4. Genetic effects are not the same for MZ twins and DZ twins.

Accordingly, it is concluded that application of a range of models and estimators to a sample of Ghanaian twins provides a consistent set of findings, and that these findings are in accordance with the evidence reported for twins in Ashenfelter and Krueger (1994) and Rouse (1998). Given the similarity of the findings in this and in related studies, it would appear that the models applied by Ashenfelter and Krueger (1994) are robust.

Further Research

Estimating the economic returns to additional education was the main focus of this study, yet further research is still needed to investigate precisely how a higher level or a faster accumulation of human capital translates into faster growth or higher productivity. Moreover, as the empirical research of the determinants of growth (Barro and Salai Martin, 1995; Krueger and Lindahl, 1999) results indicate that economic growth is positively correlated with education across countries, other sources of identifying information which exogeneously change education need to be examined. Consequently, growth may clearly be the cause of endogeneity of education rather than the consequence of education.

In addition, further research is needed to examine the effect of human capital externality on earnings and returns to education. Investments in human capital may also have external social impacts, which can in turn have indirect economic effects. More education has for instance been found to be associated with better public health, better parenting, lower crime, a better environment, wider political and community participation, and greater social cohesion, all of which is in turn likely to feed back into economic growth (OECD, 1998).

6.3 Implications for Educational Reforms in Ghana

Studies which focus primarily on the economic returns to schooling have important implications for policy formulation and decision making within the education sector. World average rate of return to another year of schooling is 10 percent (Psacharopoulos and Patrinos, 2004) which is comparable to estimates that draw on twins data from this study and other countries. Thus, our findings support the argument of Heckman (2003 and 2005) and Fleisher and Wang (2004) that investing in human capital is worthwhile in Ghana. Furthermore, to sustain the gains realized in educational attainment as a result of the schooling reforms, lingering issues of gender equity need to be addressed by policy makers so that females are not left behind in the intergenerational race for improvements in quality of life.

High returns to education suggest there is an increasing demand for higher educated workers, given that they benefit from an earnings premium when compared to those less educated. Despite the substantial increase in educational attainment of Ghanaians during the FCUBE reform program era, returns rate to education is still low in international terms, and the total employment population ratio is also relatively low. The fifth round of the Ghana Living Standards Survey (GLSS5) revealed that only about a tenth of workforce in various industrial sectors had a secondary or higher qualification. This indicates that it is important to establish a unified labour market with fair competition and freedom of movement. This also illustrates the potentially important interactions between the supply of the educated and labour market outcomes. More specifically, the focus should be on which specific, and also levels of, education are in high demand and should therefore be prioritized in the allocation of public resources.

Policy-makers are particularly interested in the effect of education since it can be influenced by policy measures (European Commission, 2003; OECD, 2009). Establishing its effect, however, is difficult due to endogeneity (Van der Sluis et al. 2008). That is, education appears as a causal variable in an econometric model while it is in fact correlated with the errors in the model. In general, this correlation can be caused by measurement errors or omitted variables. Nobel Laureate Gary Becker theorizes that education provides skills, or human capital, that makes a worker more productive (Becker, 1964). If so, then because a worker's income reflects his or her productivity, education is a key determinant of upward social mobility. It follows that much of the gap between the rich and the poor arises from a lack of skills among the poor—with the policy implication being that education and training should form the cornerstone of programs aimed at reducing income inequality.

The rate of return to schooling measures the extra earnings a worker would get if she or he invests a further year in education. It is therefore an important factor in understanding educational attainment and participation. Moreover, in the framework of human capital theory, it is interpreted as a measure of the impact of education on the productivity individuals bring to the labour market. However, individuals' investment in education is measured and it is indeed hard to assume it is exogenous and hence to estimate the associated return using Ordinary Least Squares (OLS). Education attainment may be endogenous and accounting for the endogeneity of education is an important issue because education is not randomly assigned to individuals, and their choices are heavily reliant on many factors such as their ability, motivation and family background (Card 1995, 1999, and 2001). The pure/true return to education is an important ingredient needed for policy issues and relying on ordinary least squares (OLS) may result in incorrect estimates of the actual return to education.

REFERENCES

Acemoglu, D. (2002). "Technical Change, Inequality, and the Labour Market." *Journal of Economic Literature*, *40*, 7-72.

Altonji, J. G., & Dunn, T. A. (1996). "The Effects of Family Characteristics on the Return to Education." *Review of Economics and Statistics*. *78*, 692-704.

Altonji, J.G., & Shakotko, R.A. (1987). Do Wages Rise with Job Seniority? *The Review of Economic Studies*, *54*(3), 437-459.

Altonji J. G., & Williams, N. (1997). "Do Wages Rise With Job Seniority? A Reassessment." *NBER Working Paper Series 6010*.

Angrist, J. D., & Krueger, A.B. (1991). "Does Compulsory Schooling Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, *106*, 976-1014.

Angrist, J. D., & Krueger, A.B. (1992). "The Effects of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples." *Journal of the American Statistical Association*, *87*, 328-336.

Angrist, J.D., & Newey, W.K. (1991). "Over-identification tests in earnings functions with fixed effects." Journal of Business and Economic Statistics, *9* (3), 317-323.

Arias, O., & McMahon, W. W. (2001). Dynamic rates of return to education in the U.S. *Economics of Education Review*, *20*(2), 121-138.

Arias, O., Hallock, K. F., & Sosa-Escudero, W. (2001). Individual heterogeneity in the returns to schooling: Instrumental Variables Quantile Regression using twins data. *Empirical Economics*, *26*(1), 7-40.

Aromolaran, A. B. (2002), "Private Wage Returns to Schooling in Nigeria: 1996-1999." Discussion Paper No. 849, Economic Growth Center, Yale University, New Haven CT.

Aromolaran, A. B. (2004). "Wage returns to schooling in Nigeria."*African Development Review*, *16*(3), 433–55.

Aromolaran, A.B. (2006). "Estimates of Mincerian Returns to Schooling in Nigeria." *Oxford Development Studies*, *34*(2), 265–89.

Asadullah, M. N. (2006). Returns to Education in Bangladesh. *Education Economics*, *14*(4), 453-468.

Asadullah, M. N. (2005). Returns to Education in Bangladesh. QEH Working Paper Series, Working Paper Number 130.

Ashenfelter, O., & Krueger, A. B. (1994). Estimates of the Economic Returns 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." *The Quarterly Journal of Economics*, *113*(1), 253-284.

Ashenfelter, O., & Zimmerman, D. J. (1993). "Estimates of the Returns to Schooling from Sibling Data: Fathers, Sons and Brothers." Working

Paper, 318, Princeton: Industrial Relations Section, University of Princeton.

Ashenfelter, O., & Zimmerman, D. J. (1997). "Estimating of Return to Schooling from Sibling Data: Fathers, Sons and Brothers." *Review of Economics and Statistics*, *79*, 1-9.

Ashenfelter, O., Harmon, C., & Oosterbeek, H. (1999). "A review of estimates of the schooling/ earnings relationship, with tests for publication bias." *Labour Economics*, *6*, 453–70.

Ashenfelter, O., Harmon, C., & Oosterbeek, H. (2000). "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias." NBER Working Papers 7457, National Bureau of Economic Research, Inc.

Aslam, M. (2005). *Rates of Return to Schooling by Gender in Pakistan*. GPRG-WPS-064 University of Oxford.

Asplund, R., & Pereira, P. T. (1999). 'Introduction', In (Eds.), *Returns to human capital in Europe: a literature review*, Helsinki, Finland: ETLA/Taloustieto Oy., (pp. 259-278).

Barro, R., & Lee, J. (2001). "International data on educational attainment updates and implications." *Oxford Economic Papers*, *53*(3), 541–63.

Barro, R. X. S-I-M (1995). *Economic Growth*. New York: McGraw-Hill.

Bauer, T. K. (2000). Educational Mismatch and Wages: A Panel Analysis. *Economics of Education Review*, *2*, 221–229

Beach, C., & Finne, R. (1988). Family background in an extended earnings-generation model: Further evidence. *Eastern Economic Journal, 14* (1), 39-49.

Becker, G.S. (1962). "Investment in human capital: a theoretical analysis." *Journal of Political Economy* (Supplement), *70*, 9-49.

Becker, G.S. (1964). *Human Capital, New York*: Columbia University Press.

Becker, G. S. & Chiswick, B. R., 1966. "Education and the Distribution of Earnings." *American Economic Review*, *56*, 358-369.

Becker, H. (1967). Whose Side Are We On? *Social Problems*, 14, (Winter), (pp. 239-247).

Becker, G.S. (1975). *Human capital: a theoretical and empirical analysis, with special reference to education.* New York: Columbia University Press for the National Bureau of Economic Research, 2nd Edition. (1st edition, 1964.)

Becker, G.S. (1993). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education* (3rd ed.). Chicago: University of Chicago Press.

Behrman, J., & Deolalikar, A. (1993). Unobserved Household and Community Heterogeneity and the Labour Market Impact on Schooling: A case study for Indonesia. *Economic Development and Cultural Change, 41*(3), 41-488.

Behrman, J.R., & Rosenzweig, M.R. (1999). "Ability Bias in Schooling Returns and Twins: A Test and New Estimates." *Economics of Education Review*, *18*(2), 159-167.

Behrman, J., & Deolalikar, A. (1995). "Are there Differential Returns to Schooling by Gender? The Case of Indonesian Labor Markets." *Oxford Bulletin of Economics and Statistics*, *57*, 97-117.

Behrman, J.R., Rosenzweig, M.R., & Taubman, P. (1994). Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment. *Journal of Political Economy*, *102*, 1131–1174.

Behrman, J., Taubman, P. W., & Terence J. (1977). Controlling for and Measuring the Effects of Genetics and Family Environment in Equations for Schooling and Labour Market Success. In: Taubman, Paul (Eds.), *Kinometrics: The Determinants of Socioeconomic Success within and between families.* North-Holland, Amsterdam, (pp. 35– 96).

Behrman, J.R. (1982). "Review Article on Hollis B. Chenery, Structural Change and Development Policy." *Journal of Development Economics*, Elsevier, *10*(3), 313-323.

Behrman, J.R., & Taubman, P. (1976). Intergenerational Transmission of Income and Wealth. *American Economic Review*, *66*(3), 436–440.

Behrman, J.R., Hrubec, Z., Taubman, P., & Wales, T.J. (1980). Socioeconomic Success: A study of the effects of genetic endowments, family environment and schooling. Amsterdam, North-Holland Publishing Company.

Belzil, C. (2004). "On the Specification of Mincerian Wage Regressions with Heterogeneity, Non-Linearity, Non-Seperability and Heteroskedasticity." IZA Discussion papers 1083, Institute for the Study of Labor (IZA).

Bennell, P. (1996). "Rates of Return to Education: Does the Conventional Pattern Prevail in sub-Saharan Africa?" *World Development*, *24*(1), 183-199.

Bingley, P., Christensen, K., & Walker, I. (2005). Twin-based Estimates of the Returns to Education: Evidence from the Population of Danish Twins. Unpublished manuscript.

Black, D. A., Berger, M. C., & Scott, F. C. (2000). Bounding Parameter Estimates with Nonclassical Measurement Error. *Journal of the American Statistical Association*, *95* (451), 739–748.

Blackburn, R. M. & Jarman, J. (2004). "Segregation and Inequality' in European Commission, 25th CEIES Seminar - Gender statistics - Occupational segregation: extent, causes and consequences." Luxembourg: Office for Official Publications of the European Communities.

Blackburn, M. L., & Neumark, D. (1995). Are OLS Estimates of the Return to Schooling Biased Downward? Another Look. *The Review of Economics Statistics*, *77*, 217-29.

Blackburn, M. L., & Neumark, D. 1993. "Omitted-Ability Bias and the Increase in the Return to Schooling." *Journal of Labor Economics*, *11*(3), 521-544. The MIT Press.

Blanchflower, D. & Elias, P. (1993). "Ability, Schooling and Earnings: Are Twins Different?" Mimeo, Dartmouth College, New Hampshire.

Blau, F.D. (1998). "Trends in the Well- Being of American Women, 1970-1995."*Journal of Economic Literature*, *36*(1), 112- 65.

Blundell, R., Dearden, L., Meghir, C., & Sianesi, B. (1999). Human Capital Investment: the returns from education and training to the individual, the firm and the economy. *Fiscal Studies*, *20* (1), 1-3.

Blundell, R., Dearden, L., & Sianesi, B. (2005). "Evaluating the Effect of Education on Earnings: Models, Methods and Results from the National Child Development Survey." *Journal of the Royal Statistical Society* A, Part 3, 473-512.

Blundell, R., Dearden, L., & Sianesi, B. (2002). "Evaluating the impact of education on earnings in the UK: models, methods and results from the NCDS." mimeo, IFS.

Bonjour, D., Cherkas, L.F., Hashel, J.E., Hawkes, D.D., & Spector, T.D. (2003). Returns to Education: Evidence from U.K. Twins. *American Economic Review*, *93* (5), 1799-1812.

Bound, J., & Solon, G. (1999). Double Trouble: On the Value of Twins-based Estimation of the Return to Schooling. *Economics of Education Review*, *8*(2), 169-182.

Brunello, G. & Miniaci, R. (1999). The Economic Returns to Schooling for Italian Men. An evaluation based on Instrumental Variables. *Labour Economics*, *6*(4), 509-519.

Butcher, K., & Case, A. (1994). "The Effect of Sibling Sex Composition on Women's Education and Earnings." *Quarterly Journal of Economics*, *109*, 531-563.

Cannari, L., & DtAlessio, G. (1995). Il Rendimento delltistruzione: alcuni problemi di stima, in Temi di Discussione della Banca d'Italia. Banca dtItalia.

Card, D. (2001). "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica*, *69*(5), 1127-1160.

Card, D., & Lemieux, T. (2001). "Can Falling Supply Explain The Rising Return To College For Younger Men? A Cohort-Based Analysis." *Quarterly Journal of Economics*, *116*(2), 705-746.

Card, D. (1999). 'The Causal Effect of Education on Earnings.' In Ashenfelter, O. and Card, D. (Eds.), *Handbook of Labour Economics*, Elsevier Science B.V.

Card, D. (1998). "The Causal Effect of Education on Earnings." Centre for Labour Economics, Berkeley, Working Paper 2, available at http://violet.berkeley.edu/~iir/cle/card-handbook.pdf.

Card, D. (1995). 'Earnings, Schooling and Ability Revisited.' In S. Polacheck (Ed.), *Research in Labor Economics*, *14*, 23-48, JAI Press: Greenwich, Connecticut.

Card, D. (1994). Earnings, Schooling and Ability Revisited, Working Paper No. 331 Industrial Relations Section Princeton University.

Card, D. (1993). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. National Bureau of Economic Research Working Paper No. 4482, October.

Card, D., & Krueger, A.B. (1992). "Does School Quality Matter: Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy*, *100*, 1-40.

Carneiro, P. (2002). "Heterogeneity in the Returns to Schooling: Implications for Policy Evaluation." Ph.D. Dissertation, University of Chicago.

Carneiro. P., Hansen, K. & Heckman, J. (2003). "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on Schooling Choice." International Economic Review, *44* (2), 361-442.

Caneiro, P., Heckman, J., & Vytlacil, E. (2001). "Estimating the Rate of Return to education When it Varies among Individuals." Working Paper, University of Chicago.

Caneiro, P., & Heckman, J.J. (2002). "The Evidence on Credit Constraints in Post-Secondary Schooling." *Economic Journal*, *112*, 989-1018.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, *18*, 5–46.

Chen, S. (2002). Is investing in college education risky? University of Rochester and University of New York at Albany. Mimeo.

Cheung, K. (1990). Interdisciplinary relationships between social work and other disciplines: A citation study. Social work research and abstracts, *26*(3), 23-30.

Chiswack, B. R. (1974). Income Inequality: Regional Analyses within a Human Capital Framework. New York: Columbia Univ. Press for National Bureau of Economic Research (NBER).

Coleman, J. S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., &. York, R. L. (1966). *Equality of Educational Opportunity* (Washington DC: U.S. Government Printing Office).

Colussi, A. (1997), Il tasso di rendimento delltistruzione in Italia: untanalisi cross-section, Poltica economica.

Conneely, K. & Uusitalo, R. (1998). "Estimating Heterogeneous Treatment Effects in the Becker Schooling Model." Mimeo, Princeton University.

Dearden, L. (1998). "Ability, families, education and earnings in Britain." IFS Working Papers W98/14, Institute for Fiscal Studies.

Dearden, L. (1995). "The Returns to Education and Training for the United Kingdom." Unpublished Ph.D. Dissertation, University College London.

Dougherty, C. (2005). Why Are the Returns to Schooling Higher for Women than for Men? *Journal of Human Resources, 40*(4), 969-988.

Duflo, E. (2001). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy

Experiment." *American Economic Review, American Economic Association, 91*(4), 795-813.

Ehrenberg, R. & Smith, R. (1991), *Modern Labor Economics* (IV ed.), New York: Harper Collins.

Erosa, A., Fuster, L. & Restuccia, D. (2002). "Fertility Decisions and Gender Differences in Labour Turnover, Employment, and Wages." *Review of Economic Dynamics 5*(4), 1354-1378.

Fitzenberger, B., Schnabel, R., & Wunderlich, G. (2004). "The Gender Gap in Labor Market Participation and Employment: A Cohort Analysis for West Germany." *Journal of Population Economics*, *17*, 83-116.

Fitzenberger, B., & Kurz, C. (2003). New Insights on Earnings Trends across Skill Groups and Industries in West Germany. *Empirical Economics 28*, 479-514.

Fleisher, B. M., & Xiaojun, W. (2005). Returns to Schooling in China under Planning and Reform. *Journal of Comparative Economics, 33*, 265–277.

Freeman, R. B. (1977). The Decline in the Economic Rewards to College Education. *Review* of *Economics and Statistics*, *59*, 18-27.

Gadzekpo, V. P. (2008). *Central University.* Retrieved December 10, 2008, from Central University Web site: www.centraluniversity.org/vpyg-10th-anniv.htm.

Gautier P. A., & Teulings, C. A. (2003). "How Large are Search Frictions?" Tinbergen Institute Discussion Papers 03-026/3, Tinbergen Institute.

Garen, J. (1984). The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable Source. *Econometrica*, *52*(5), 1199-1218.

Gebel, M. & Pfeiffer, F. (2007). "Educational Expansion and Its Heterogeneous Returns for Wage Workers." SOEP papers 13, DIW Berlin, The German Socio-Economic Panel (SOEP).

Girma, Sourafel & Kedir, (1994) . Heterogeneity in Returns to Schooling: Econometric evidence from Ethiopia Department of Economics University of Leicester University Road Leicester LE1 7RH.

Glewwe, P. (1996). The relevance of standard estimates of rates of return to schooling for education policy: A critical assessment. *Journal of Development Economics, 51*, 267-90.

Gödde, I., & Schnabel, R. (1998). "Does Family Background Matter? - Returns to Education and Family Characteristics in Germany." Sonderforschungsbereich 504 Publications 98-60, Sonderforschungsbereich 504, Universität Mannheim & Sonderforschungsbereich 504, University of Mannheim.

Goldin, C., & Lawrence, F. K. (2008). *The Race between Education and Technology.* The Belknap Press of Harvard University Press: Cambridge, MA.

Greene, W. H. (2002). *Econometric Analysis*. Pearson Education Press.

Griliches, Z. (1979). Sibling models and data in economics: Beginning of a survey. *Journal of Political Economy*, *87*, S37–S64.

Griliches, Z. (1977). Estimating the Returns to Schooling: Some Econometric Problems. *Econometrica*, *45*, 1–22.

Ghana Statistical Service, 2008. *Ghana Living Standards Survey Report, GLSS 5.* Accra, Ghana Statistical Service.

Griliches, Hall & Hausman (1978). 'Missing data and self-selection in large panels', Annales de l'Insee, *30-31*, 137-176.

Guillotin, Y., & Sevestre, P. (1994), «Estimations de Fonctions de Gains sur Données de Panel: Endogénéité du Capital et Effet de Sélection », *Economie et Prévision*, *116*, 119-135.

Hall, R., & Jones, C. (1999). Why do some countries produce so much more output per worker than others? *The Quarterly Journal of Economics*, *114*(1), 83-116.

Hanushek, E. A. (1986). "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature*, *24*, 1141–1177.

Hanushek, E. A., Steven, G. R., & Lori, L. T. (1996). .Aggregation and the Estimated Effects of School Resources." this REVIEW, *78*(4), 611–627.

Harmon C., Oosterbeek, H., & Walker, I. (2002). The returns to education: A review of evidence, Center for Economics of Education, Working Paper.

Harmon, C. & Walker, I. (1995). Estimates of the Economic Return to Schooling for the United Kingdom. *American Economic Review, 85*(5), 1278-86.

Hartog, J. (1986). Allocation and the earnings function. *Empirical Economics, 11*, 97-110.

Harville, D. A. (1977). "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association*, *72* (358), 320-338.

Hausman, J.A. & Taylor, W.E. (1982). "A Generalized Specification Test." *Economics Letters, 8*, 239-245.

Hausman, J. (1978). "Specification Tests in Econometrics." *Econometrica*, *46*(6), 1251- 1272.

Hawley, J.D., (2004). "Changing returns to education in times of prosperity and crisis, Thailand 1985–1998." *Economics of Education Review*, *23*, 273-86.

Heckman, J. J. (2003). China's investment in human capital. *Economic Development and Cultural Change*, *51* (4), 795–804.

Heckman, J. J. (2005). China's human capital investment. *China Economic Review*, *16*(1), 50-70.

Heckman, J. J. & Vytlacil, E. J. (1998). "Instrumental Variables Methods For the Correlated Random Coefficient Model: Estimating The Average Rate of Return to Schooling When the Return Is Correlated With Schooling." *Journal of Human Resources*, *33*(4), 974-1002.

Heckman, J. J., & Vytlacil, E. J. (1999). Local Instrumental Variables in models with essential heterogeneity. *Review of Economics and Statistics*, *88*(3), 389-432.

Heckman, J.J., Lochner, L. J., & Todd, P.E. (2006). Earnings Equations and Rates of Return: The Mincer Equation and Beyond. In E. A. Hanushek and F. Welch (Eds.). *Handbook of the Economics of Education*, (pp. 307-458). Amsterdam: North-Holland.

Heckman, J., Layne-Farrar, A., & Todd, P. (1996). "Human Capital Pricing Equations with an Application to Estimating the Effect of Schooling Quality on Earnings." *The Review of Economics and Statistics*, MIT Press, *78*(4), 562-610.

Henderson, C. R. (1975). "Best Linear Unbiased Estimation and Prediction under a Selection Model." *Biometrics*, *31*, 423-447.

Herrnstein, R.J. & Murray, C. (1994). The Bell Curve, Intelligence and Class Structure in American Life. Book pp.845, ISBN 0-02-914673-9.

Hertz, T. (2003).Upward Bias in the Estimated Returns to Education: Evidence from South Africa. *The American Economic Review*, *93* (4), 1354-68.

Hougaard, P., Harvald , B., & Holm, N.V. (1992). Measuring the similarities between the lifetimes of adult Danish twins born between 1881–1930, *Journal of the American Statistics, Association*, *87*, 17–24.

Hoxby, C. (2000). "Peer effects in the classroom." NBER working paper No. 7867.

International Network of Higher Education in Africa, 2006. www.bc.edu.

Isacsson, G. (1997). "Estimates of the Return to Schooling in Sweden from a Large Sample of Twins." Unpublished Discussion Paper, Center for Research on Transportation and Society (Borlange, Sweden).

Isacsson, G. (1999). Estimates of the Return to Schooling in Sweden from a Large Sample of Twins. *Labour Economics*, *6*, 471-89.

Isacsson, G. (2004). Estimating the Economic Return to Educational Levels Using Data on Twins. *Journal of Applied Econometrics*, *19*, 99-119.

Jacobson, L. S., LaLonde, R. J. & Sullivan, D.G. (1993). "Earnings Losses of Displaced Workers." *American Economic Review*, American Economic Association, *83*(4), 685-709.

Jochmann, M., & Pohlmeier, W. (2004). "The Causal Effect of Schooling : Empirical Evidence from Germany." Working Papers of the Research Group Heterogenous Labor 04-05, Research Group Heterogeneous Labor, University of Konstanz/ZEW Mannheim.

Jones, P. (2001). Are educated workers really more productive? *Journal of Development Economics*, *64*, 57-79.

Jones, F. L. (1983). Sources of gender inequality in income: what the Australian Census says. *Social Forces*, *62*, 134-53.

Kalwij, A. (1996). Estimating the Economic return to Schooling on the basis of Panel Data. Mimeo, Tilburg University.

Kane, T. J., & Rouse, C.E. (1995). "Labor Market Returns to Two and Four-Year College." *American Economic Review*, *85*(3), 600-614.

Kanazawa, S. (2005). "Is Discrimination Necessary to Explain the Sex Gap in Earnings." *Journal of Economic Psychology, 26*, 269-287.

Kane, T.J., & Rouse, C.E. (1993). *Labor Market Returns to Two and Four-year College: Is a Credit a Credit and do Degrees Matter?* Princeton University Industrial Relations Section Working Paper No. 311.

Katz, L. F., & Autor, D.H. (1999). Changes in the Wage Structure and Earnings Inequality, in O. Ashenfelter and D. Card (Eds.) *Handbook of Labor Economics*, Volume 3A. Amsterdam; New York: North Holland.

Kazianga, H. (2004). Schooling Returns for Wage Earners in Burkina Faso: Evidence from the 1994 and 1998 National Surveys." *Working Papers* 892, Economic Growth Center, Yale University.

Keswell, M., & Poswell, L. (2004). "Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence." *South African Journal of Economics*, *72*(4), 834–60.

Kimenyi, M.S., Mwabu, G. & Manda, D.K. (2006). "Human capital externalities and private returns to education in Kenya." *Eastern Economic Journal*, *32*(3), 493–513.

Kingdon, G., & Söderbom, M. (2008). "Education, Skills, and Labor Market Outcomes: Evidence from Ghana." Education Working Paper Series, no. 12, Washington D.C: The World Bank.

Kingdon, G., & Soderbom, M. (2007). Education, Skills, and Labor Market Outcomes: Evidence from Ghana. World Bank Human Development Network.

Koop, G., & Tobias, J. (2002). Learning about Unobserved Heterogeneity in Returns to Schooling. Mimeo, Dept. of Economics, University of Glasgow.

Kothari, V. N. (1970). Disparities in Relative Earnings among Different Countries. *The Economic Journal*, *80*(319), 605-616.

Krishnan, P. (1996). Family background, education and employment in urban Ethiopia, *Oxford Bulletin of Economics and Statistics*, *58* (1), 167-183.

Krueger, A. B. & Lindahl, M. (1999). "Education for Growth in Sweden and the World." NBER Working Papers 7190, National Bureau of Economic Research, Inc.

Kunze, A. (2005). "The Evolution of the Gender Wage Gap." *Labour Economics*, *12*, 73-97.

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

Lang, K. (1993). "Ability Bias, Discount Rate Bias, and the Return to Education." Unpublished Paper.

Lang, K. (1991). "Persistent Wage Dispersion and Involuntary Unemployment." *The Quarterly Journal of Economics*, MIT Press, *106*(1), 181-202.

Laplagne, P., Glover, M. & Shomos, A. (2007). *Effects of Health and Education on Labour Force Participation*, Productivity Commission Staff Working Paper, Canberra, May.

Lau, L. J., Jamison, D. T., & Louat, F. F. (1991). "Education and Productivity in Developing Countries: An Aggregate Production Function Approach." Policy Research Working Paper Series 612, The World Bank.

Lauer, C., & Steiner, V. (2001). Education and Earnings in Europe. A Cross-Country Analysis of the Returns to Education. In C. Harmon, I. Walker & N. Westergaard-Nielsen (Eds.), (pp. 102-128).

Lee, T.H.C. (2000). *Education in Traditional China. A History.*

Leigh, A., & Ryan, C. (2008). Estimating Returns to Education using Different Natural Experiment Techniques. *Economics of Education Review*, *27*,149–160.

Leigh, A. (2007). 'Returns to education in Australia', *Economic Papers*, *27*(3), 233–249.

Lemieux, T. (2002). "Schooling, Experience and Earnings: Twenty Five Years of "Mincer Equations"." Department of Economics, University of British Columbia.

Levin, J. (1997). "Instrumental Variables Technique and the Rate of Return to Education in the Netherlands." Tinbergen Institute, Amsterdam.

Licht, G., & Steiner, V. (1991). "Stichprobenselektion, unbeobachtete Heterogenität und Humankapitaleffekte bei der Schätzung von Einkommmensfunktionen mit Paneldaten." in *Lebenslagen im Wandel: Zur Einkommensdynamik in Deutschlandseit 1984*, Ulrich Rendtel & Gert Wagner (Eds.), Frankfurt am Main: Campus.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *J. R. Statist. Soc. B, 34*, 1-41.

Lloyd, C. B., & Gage-Brandon, A. J. (1994). High Fertility and Children's Schooling in Ghana: Sex differences in Parental Contributions and Educational Outcomes. *Population Studies*, *48*(2), 293-306.

Machado, J., & Mata, J. (1999). "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression." Preprint.

Manda, D.K. 1997. 'Labour supply, returns to education, and the effect of firm size on wages.' Ph.D thesis, Gothenburg University.

Mann, P., & Kapoor, B. (1988). Earnings differentials between public, private and joint sector in Punjab (India). *The Journal of Development Studies, 25* (1), 97-111.

McConnell, D. A., et al. (2006). Using Concept Tests to assess and improve student conceptual understanding in introductory geoscience courses. *Journal of Geoscience Education*, *54*(1), 61-68.

Meghir, C., & Palme, M. (1997). "Assessing the Rate of Returns to Education Using the Swedish 1950 Education Reform", Unpublished manuscript.

Miller, P., Mulvey, C., & Martin, N. (1995). What do twins studies tell us about the economic returns to education? A comparison of US and Australian findings. *American Economic Review*, *85*(3), 586–599.

Miller, P.W., Mulvey, C., & Martin, N. (2001). Genetic and Environmental Contributions to Educational Attainment in Australia. *Economics of Education Review*, *20*, 211 – 224.

Miller, P. W., 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, New York: National Bureau of Economic Research.

Mincer, J., & Polachek, S. (1974). "Family Investments in Human Capital: Earnings of Women." *Journal of Political Economy*, *82* (2) Part 2, S76-S108.

Mincer, J. (1958). "Investment in Human Capital and Personal Income Distribution." *Journal of Political Economy*, *66*(4), 281-302.

Mincer, J., & Opek, H. (1982). "Interrupted Work Careers: Depreciation and Restoration of Human Capital." *The Journal of Human Resources*, *17*(1), 3-24.

Ministry of Education, Science and Sports. (2007). *Report on Basic Statistics and Planning Parameters for Basic Education in Ghana 2006/2007.* Statistics Research, Information Management and Public Relations (SRIMPR), EMIS Project.

Ministry of Education, 2002. "Meeting the Challenges of Education in the 21st Century." (The Report of the President's Committee on Review of Education Reforms in Ghana.

Moll, P. (1998). Primary schooling, cognitive skills and wages in South Africa. *Economica, 65* (258), 263-284.

Moock, P., Patrinos, H., & Venkataraman, M. (2003). Education and earnings in a transition economy: the case of Vietnam. *Economics of Education Review, 22*, 503-510.

Mwabu, G., & Schultz, T. P. (1996). Education Returns Across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa. *American Economic Review, 86*(2), 335-339.

Neumark, D. (1999). Biases in Twin Estimates of the Return to Schooling. *Economics of Education Review*, *18*(2), 143–148.

Nielsen, H. S., & Westergard-Nielsen, N., (2001). "Returns to Schooling in Less Developed Countries: New Evidence from Zambia." Economic Development and Cultural Change, University of Chicago Press, *49*(2), 365-94.

Nordin, M., (2007). "Ability and Rates of Return to Schooling - making use of the Swedish Enlistment Battery Test." *Journal of Population Economics*, *21*(3), 703-717.

Nsowah-Nuamah, N.N.N. (2007). *Demographic Statistics*: A Handbook of Methodical Measures in Demography, xxii + 410pp Acadec Press.

OECD, (2009). Patent Statistics Manual. OECD, Paris.

OECD, (1998). Human Capital Investment – An International Comparison Centre for Educational Research and Innovation (Paris: Organisation for Economic Co-Operation and Development).

Okuwa, O.B. (2004). *Private Returns to Higher Education in Nigeria.* Research Paper No. 139. African Economic Research Consortium, Nairobi, Kenya.

Pastore, F., & Verashchagina, A. (2006). Private returns to human capital: a case study of Belarus. *Economics of Education Review*, *25*, 91−107.

Patterson, H. D., & Thompson, R. (1971). Recovery of interblock information when block sizes are unequal, *Biometrika*, 58, 545 -554.

Polachek, S.W. (2008). Earnings over the lifecycle: The Mincer Earnings Function and its Applications. *Foundations and Trends in Microeconomics*, *4*, 165–272.

Polachek S.W. (2007). Earnings over the Lifecycle: The Mincer Earnings Function and Its Applications, IZA Discussion Papers, No. 3181, Institute for the Study of Labor.

Polachek, S. (1981). "Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure." *The Review of Economics and Statistics*, *63*(1), 60-69.

Pritchett, L. (1999). Has education had a growth payoff in the MENA region? Middle East and North Africa, Working Paper Series No. 18. The World Bank.

Psacharopoulos, G. & Patrinos, H.A. (2004). "Returns to investment in education: A further update." *Education Economics*, *12*(2), 111–34.

Psacharopoulos, G., & Patrinos, H. A. (2002). "Returns to investment in education: a further update." Policy Research Working Paper Series 2881, The World Bank.

Psacharopoulos, G. (1994). 'Returns to investment in education: a global update.' *World Development*, *22*, 1325–1344.

Psacharopoulos, G. (1985). Returns to education: A further international update and implication. *The Journal of Human Resources, 20* (4), 583-597.

Rao, P.S.R.S. (1997). Variance Components Estimation Mixed Models, Methodologies and Applications Chapman & Hall/CRC Press, London and New York.

Rees, H., & Shah A. (1995). Public-private sector wage differential in the UK. *The Manchester School of Economics and Social Studies, 64* (1), 52-68.

Rouse, C. (1999). Further Estimates of the Economic Return to Schooling from a New Sample of Twins. *Economics of Education Review*, *18*(2), 149-57.

Rouse, C. (1997). "Further Estimates of the Economic Return to Schooling from a New Sample of Twins." Princeton University Industrial Relations Section Working Paper #388, July.

Roy, A. (1951). "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers*, *3*, 135-146.

Rummery, S., V., F., & Verbeek, M. (1999). "Estimating the Returns to Education for Australian Youth via Rank-Order Instrumental Variables." *Labour Economics*, *6*, 491–507.

Rummery, S. (1992). The Contribution of Intermittent Labour Force Participation to the Gender Wage Differential. *Economic Record*, *68*, 351–64.

Ryoo, J.K., Nam, Y.S. & Carnoy, M. (1993). Changing Rates of Return to Education over Time: A Korean Case Study. *Economics of Education Review*, *12* (1), 71–80.

Sackey, Harry A. (2008). Private Returns to Education in Ghana: Implications for Investments in Schooling and Migration. AERC Research Paper 174 African Economic Research Consortium, Nairobi.

Sahn, E., & Alderman, H. (1988). The effect of human capital on wages and the determinants of labour supply in developing country. *Journal of Development Economics, 29* (2), 157-183.

Sandewall, Ö., Cesarini, D., & Johannesson, M. (2009). The Co-Twin Methodology and Returns to Schooling - Testing a Critical Assumption. IFN Working Paper No. 806.

SAS Institute Inc., (2003). SAS (release 9.1) Cary, NC: SAS Institute Inc.

Schnabel, I & Schnabel, R. (2005). "*Family and Gender Still Matter: The Heterogeneity of Returns to Education in Germany*." with Reinhold Schnabel (University of Essen), ZEW Discussion Paper # 02-67.

Schultz, T. (2003). Evidence of returns to schooling in Africa from household surveys: Monitoring and restructuring the market for education. *Discussion Paper* No. 875, Economic Growth Center, Yale University, New Haven CT.

Schumann, P. L., Ahlburg, D.A., & Mahoney, C. B. (1994). "The Effects of Human Capital and Job Characteristics on Pay." *The Journal of Human Resources*, *29*(2), 481-503.

Searle, S. R., Casella, G., & McCulloch, C. (1992). *Variance Components*. New York: John Wiley and Sons.

Siphambe, H. (2000). Rates of Return to Education in Botswana. *Economic of Education Review, 19*(3), 291-300.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*, reprinted in W.B. Todd (ed.) (1976), *Glasgow Edition of the Works and Correspondence of Adam Smith*, vol.1, Oxford: Oxford University Press.

Smyth, G. K., & Verbyla, A. P. (1996). A Conditional Approach to Residual Maximum Likelihood Estimation in Generalized Linear Models. *Journal of the Royal Statistical Society* B, *58*, 565-572.

Solow, R. M. (1965). "Human Capital." *The Journal of Political Economy*, *73*(5), 552-553.

StataCorp. (2009). *Stata Statistical Software: Release 11*. 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.

Theodossiou, I. (1996). Promotions, job seniority and product demand effects on earnings. *Oxford Economic Papers, 48*(3), 456-472.

Topel, R. (1991). Specific capital, mobility and wages, Wages rise with job seniority. *Journal of Political Economy, 99*(1), 145-176.

Uusitalo, R. (1999). "Essays in Economics of Education." University of Helsinki, Department of Economics 79, Department of Economics.

Uusitalo, R. (1997). Return to education in Finland. Paper presented at the EU seminar "Knowledge and work" in Amsterdam on the 24th of April.

Uwaifo, R. (2006). Three essays on the Returns to Education Enigma in Nigeria. University of California Berkeley. Ph.D Dissertation.

Van Der Sluis, J., Van Praag, M., & Vijverberg,W. (2008). Education and Entrepreneurship Selection and Performance: A Review of the Empirical Literature. *Journal of Economic Surveys*, *22*(5), 795–841.

Vila, L. E. (2000). "The Non-Monetary Benefits of Education." *European Journal of Education 35*, 21-33.

Walker, I. & Yu, Z. (2001), "The returns to education: Evidence from the Labour Force Surveys." Research Report RR313, Department for Education and Skills.

Wambugu, A. (2001). Real wages and returns to education in Kenya manufacturing.

http://www.csae.ox.ac.uk/conferences/2001NIRaFBiA/pdfs/wambugu.pdf>

Weale, M. (1993). "A Critical Evaluation of Rate of Return Analysis." *The Economic Journal*, *103*, 729-737.

Wei, LI Qi (2008). *Cognitive Revolution and Second-Generation Cognitive Science.* Acta Psychologica Sinica.

Williams, N. (1991). Re-examining the wage tenure and experience relationship. *TheReview of Economics and Statistics, 73* (3), 512-517.

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

Willis, R. (1986). "'Wage Determinants: A Survey and Reintepretation of Human Capital Earnings Functions." Chapt. 10 in O. Ashenfelter & R. Layard (Eds.): *Handbook of Labor Economics*, Volume I, Elsevier, 525 - 602.

Wilson, K. (2001). "The Determinants of Educational Attainment: Modeling and Estimating the Human Capital Model and Education Production Functions." *Southern Economic Journal, 67*(3), 518-551.

Winter-Ebmer, R. & Zweimüller, J. (1992). "Occupational Segregation and Career Advancement." *Economic Letters*, *39*(2), 229-234.

Wooldridge, J. M. (2002). Unobserved Heterogeneity and Estimation of Average Partial Effects. Michigan State University Working Paper.

Wu, C. T., Gumpertz, M. L., & Boos, D. D. (2001). Comparison of GEE, MINQUE, ML, and REML Estimating Equations for Normally Distributed Data. *The American Statistician*, *55*(2), 125–130.

Wu, D. M. (1973). Alternative Tests of Independence between Stochastic Regressors and Disturbances. *Econometrica*, *41*, 733-750.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *J. Am. Statist. Ass.*, *57*, 348-368.

Zhang, J., Pak-Wai L., & Yung, L. (2007).The Cultural Revolution and Returns to Schooling in China: Estimates Based on Twins, *Journal of Development Economics*, *84*,631-639.

APPENDICES

APPENDIX A

Measuring Earnings

Heckman and Polachek (1974) investigated alternative transformations of earnings and concluded that the log transformation is the best in the Box-Cox class. Finally, and perhaps as important as any other consideration, the log transformation is convenient for interpretation. The studies by Miller et al. (1995) and Isacsson (1999) were also based on annual earnings.

Annual Earnings	**Percent**	**Frequency**
0-2000	67.6	169
2000-4000	20.4	51
4000-6000	6.8	17
6000-8000	2.8	7
8000-10000	0.8	2
10000-12000	1.6	4

APPENDIX B

QUESTIONNAIRE/CHECKLIST FOR EFFECT OF ADDITIONAL SCHOOLING ON INCOME IN GHANA

Introductory Note

This study is an activity that seeks to acquire information on twins and siblings in Ghana and estimate the economic returns to schooling for twins and siblings.

SECTION A: DEMOGRAPHIC DATA

1. Twin/Sibling Code:
2. Name: --
3. Age: ------------------------------------
4. Date of Birth (d/m/y): ------------------------------------
5. Sex: a. Male b. Female
6. Residence: a. Region b. Town
7. Ethnicity (Tribe): --
8. Religion: a. No religion b. Protestant c. Catholic d. Other Christian

   e. Muslim f. Traditionalist
9. Marital Status: a. Married b. Living together c. Separated d. Divorced

   e. Widowed f. Never married
10. Marital Status of your twin: a. Married b. Living together c. Separated

    e. Widowed f. Never married
11. Are both of you registered in any twins club? a. Alone b. He/She alone

    c. Both of us d. None

SECTION B: INDIVIDUAL/FAMILY CHARACTERISTICS

12. Is your twin a brother or sister? a. Brother b. Sister

13. What is his/her name/address/Tel. No.: ------------------------------------

----

14. Are you an identical twin? a. Yes b. No c. Don't know

15. How do you know you are or are not identical twins? a. Blood test
    b. Never Tested

16. Do you have first cousins who are twins? a. Yes b. No c. Don't know

17. If yes, indicate type of twins: a. Identical b. Non-identical c. Don't know

18. As children, did you and your twin look very much alike? a. Yes b. No

19. Who do you resemble (look like)? a. Father b. Mother c. Neither

20. Who does your twin brother/sister look like? a. Father b. Mother c. Neither

21. Does your mother or father mistake one for the other? a. Yes b. No

22. Do other family members mistake one for the other? a. Yes b. No

23. Do strangers have difficulty in telling you apart? a. Yes b. No

24. Do people have difficulty in correctly identifying each twin on photographs? a. Yes b. No

25. Facial colour: a. Fair b. Dark c. Neither

26. What are your personality traits?

Personality Trait	Yes/No
I am the life of the party	
I like being the center of attention	
I am skilled in handling social situations	
I start conversations	
I make friends easily	
I am quiet around strangers	
I don't like to draw attention to myself	
I am a private person	
I enjoy listening to music alone	
I think a lot before I talk	

27. Highest educational level/Main Occupation/Income/salary per month

28. How long have you been in this current job? -------------------

29. Parental Characteristics

Parent	Age at birth of twins	Highest educational level at birth of twins	Highest educational level completed	Occupation at birth of twins	Occupation In which he/she spent most of his/her career
Mother					
Father					

30. If you are 15 years or more, which of the following items do you own?

Item	Tick
Television	
Radio	
Sewing machine	
Car	
Bicycle	
Land/Plot	
Refrigerator/Freezer	
Record player	
Air conditioner	
Furniture	
Stove	
Fan	
Radio cassette	
Radio player	
3-in-one radio cassette recorder	
Video equipment	
Washing machine	
Camera	
Iron (electric)	
House	
Shares	
Boat	
Canoes	
Outboard motor	
Mobile phone	
Motorbike	