Essays in Labour Economics: Thailand's Labour Market Adjustment during the Structural Transformation Process

La-Bhus Fah Jirasavetakul

A thesis submitted for the degree of Doctor of Philosophy in Economics

Worcester College
University of Oxford

Trinity Term 2014

Essays in Labour Economics: Thailand's Labour Market Adjustment during the Structural Transformation Process

La-Bhus Fah Jirasavetakul, Worcester College

A thesis submitted for the degree of Doctor of Philosophy in Economics,

University of Oxford, Trinity Term 2014

Abstract

I examine the importance of human capital for economic development in Thailand during the period of high economic growth and structural transformation (1985-2000), using labour force survey data. The three main chapters attempt to estimate the effects of education, as a measure of human capital, on three major outcomes in the Thai labour market, namely (i) earnings; (ii) sector of employment; and (iii) earnings inequality. I address the endogeneity problem of education using an education policy shift—the change in the compulsory schooling law—that produces exogenous variation in education. The three main chapters adopt distinct modelling frameworks. The details of each of the main chapters are as follows.

The third chapter investigates how education increases earnings and the probability of being in the non-agricultural sector. As the education policy shift influences educational attainment in a discontinuous way, a regression discontinuity (RD) framework is adopted to identify the average returns to education and the effect of education on the sector of employment. It is important to emphasise that the RD technique constrains the effects of education on the two outcomes to be linear and to be applicable only to sub-populations. My results confirm significant effects of education on both earnings and the sectoral sorting process. In addition, there are heterogeneous effects of education by gender.

The fourth chapter is an extension of the previous chapter. I allow the returns to education to be heterogeneous across education levels and sectors of employment, while attempting to estimate the returns for the entire population. I use a control function (CF) approach and a double selection correction to estimate the sectoral earnings process, while jointly accounting for the choice of education and the selection into sectors and paid employment. I find that the returns to education are non-linear and higher in the non-agricultural sector especially for medium and highly educated workers. This suggests that human capital plays a crucial role in facilitating a structural transformation towards the non-agricultural sector.

In the final chapter, I study how the increased primary education completion rate affects earnings inequality. While there exists a burgeoning literature on the average returns to education, less attention has been devoted to estimating the effects of education on the distribution of earnings. I identify the effects of primary education completion on earnings at different points of the distribution, and thus earnings inequality, using a recently developed approach, called regression discontinuity distributional treatment effects. My results suggest that the increased primary education completion rate reduces earnings inequality as the returns to primary education are larger for the poor than the rich.

Word Count and Format

The computerised word count is 70,553. Taking page 2 as a representative page, the thesis has 400 words per page. Multiplying this by 199 pages (excluding bibliography) results in an overall word count of 79,600.

This thesis consists of six chapters, including the introduction and conclusion chapters. The second chapter provides background on the relevant literature on Thailand and data for the following three chapters of the thesis. A typical chapter contains a number of sections. Some chapters contain appendices. All relevant tables and figures are at the end of each chapter. Tables are presented first, followed by figures.

Declaration

Chapter 4 of the thesis builds on my MPhil thesis, which was submitted in Trinity Term 2010 as part of the requirements for the degree of MPhil in Economics, University of Oxford. The title of my MPhil thesis was "Labour Markets in Models of Economic Development: Empirical Evidence from Thailand".

All of the data were analysed using Stata Software, version 12.1. The analyses in Chapter 5 were based on my own Stata routine "*rddite.ado*" which was developed from the command "*rdqte.ado*" (Frandsen et al., 2012). While the latter is the first to implement non-parametric regression discontinuity quantile treatment effects in Stata, it does not allow for a discrete running variable which is appropriate for my case. In addition, "*rdqte.ado*" estimates only the quantile treatment effects, not the inequality treatment effects.

Acknowledgements

Although the thesis has only one author, it has only been possible with the support from a number of people and institutions. First and foremost, I would like to thank both of my thesis supervisors, Francis Teal and Debopam Bhattacharya, for their invaluable supervision, approachability, and generosity with their time. Francis has been extremely supportive and helpful, as well as providing me with constructive advice on developing economies at both the macro and micro levels. Over the years, I have learnt tremendously from him especially about working on developing country data. His attempts to encourage me to question my empirical results at a deeper level have always been appreciated. Debopam has inspired me to take up challenges on the econometric front and provided full support and guidance as I navigated my way through. I am very grateful for his comments and patience in answering numerous econometric and programming questions. I have gained so much technical knowledge of economics from working with him. I would also like to thank my examiners, Margaret Stevens and Christopher Heady, for their very helpful comments and suggestions.

I cherish the opportunities I have had at Oxford, where I have learnt immensely from the demanding coursework and fruitful discussions during and outside various seminars and workshops. The Centre for the Study of African Economies (CSAE) has also created a stimulating environment to conduct empirical research. I would like to thank participants of the Gorman Workshop, the CSAE Workshop and Conferences, the EUDN PhD Workshop, the Singapore Economic Review Conference, and the Pacific Trade and Development Conference, for insightful comments on my thesis chapters. In particular, I would like to express my gratitude to Erlend Berg, Stephen Bond, Marcel Fafchamps, James Fenske, Simon Quinn, and Anisha Sharma. Financial support from the Oxford Economic Papers and George Webb Medley research funds, as well as the Doctoral Studentship from Department of Economics, is gratefully acknowledged.

I would like to extend my warmest regards to Chayawadee Chai-anant, William McCleary, Sarayut Nathaphan, Chaipat Poonpatpibul, Daungporn Rodpengsangkaha, Phurichai Rungcharoenkitkul, and Nitinant Wisaweisuan. Without their support and advice at various stages during my previous study and work, the opportunity to pursue postgraduate studies in Oxford would not have arisen.

It would be impossible to mention all fellow economists and friends in and outside of the Department of Economics, yet my special thanks go to Jackie Cheng, David Deregowski, Antony Hoyte-West, Naureen Karachiwalla, Zsoka Koczan, Sang Lee, and Yuwi Manachotphong for their company and support as well as helpful discussions related and unrelated to economics. I wish to extend my thanks to Rebecca Scott, who kindly helped me run important errands on my behalf while I was away in Washington DC. Christoph Lakner deserves a very special thank you for enjoyable and never-ending debates on labour force data, Stata programming, and inequality.

Finally, I wish to thank my parents and brother, to whom this thesis is dedicated. Without their love, support, and encouragement, the journey towards the completion of this thesis would have never begun and certainly would have never come to an end.

Contents

List of Tables

List of Figures

List of Abbreviations

2SLS	Two-Stage Least Squares
ATE	Average Treatment Effects
CF	Control Function
CPI	Consumer Price Index
FES	Family Expenditure Survey
FIML	Full Information Maximum Likelihood
FRD	Fuzzy Regression Discontinuity
GDP	Gross Domestic Product
GHS	General Household Survey
GOF	Goodness-of-Fit
HSES	Household Socio-Economic Survey
IMR	Inverse Mill's Ratio
ISI	Import-Substituting Industrialisation
ISIC	International Standard Industrial Classification
ITE	Inequality Treatment Effects
IV	Instrumental Variables
LATE	Local Average Treatment Effects
LFS	Labour Force Survey
LITE	Local Inequality Treatment Effects
LQTE	Local Quantile Treatment Effects
MLE	Maximum Likelihood Estimation
NESPD	New Earnings Survey Panel Dataset
NSO	National Statistical Office of Thailand
OLS	Ordinary Least Squares
QTE	Quantile Treatment Effects
RCM	Rubin Causal Model
RD	Regression Discontinuity
SBTC	Skill-Biased Technical Change
SDL	Standard Deviation of Logarithms
SLA	School Leaving Age
SSE	Error Sum of Squares
SUTVA	Stable Unit Treatment Value Assumption
THB	Thai Baht
Theil-L	Theil's Entropy Index L
Theil-T	Theil's Entropy Index T
WDI	World Development Indicators

Chapter 1

Introduction

During the 1980s and 1990s, Thailand experienced rapid economic progress compared with other developing economies within the region and beyond (Figure 1.1a). Gross Domestic Product (GDP) per capita increased by a remarkable 6.5 per cent per annum over the one and a half decades ending in 1996. At the same time, structural transformation occurred at both the production and employment levels. Non-agricultural production increased, while the contribution of agricultural production to the economy declined substantially (Figure 1.1b). The share of agricultural employment fell from 67 per cent in 1986 to 48 per cent in 2000 and non-agricultural employment increased accordingly (Figure 1.1c).

The labour market also experienced a profound change. First, a significant gap in labour productivity between the agricultural and non-agricultural sectors persisted throughout this period (Figure 1.1d). Non-agricultural output per worker was 7 to 10 times greater than agricultural output per worker. A similar pattern holds for labour income. On average, labour income was 4 to 6 times larger in the non-agricultural sector. As analysed in more detail in a subsequent chapter, human capital endowments improved substantially but differed greatly across sectors. Among the labour force, the primary completion rate increased from 19.3 per cent in 1985 to 51.4 per cent in 2000, while the lower secondary and high school completion rates rose by 17.3 and 11.2 percentage points respectively (National Statistical Office of Thailand, 2013). Average years of education increased from 5 years in 1985 to 7 years in 2000. In addition, average educational attainment observed in the non-agricultural sector was nearly twice that observed in the agricultural sector over this period (National Statistical Office of Thailand, 2013).

Understanding how these labour market conditions facilitated economic growth – which was driven by structural transformation – and affected other aspects of economic development would be useful for other developing countries experiencing similar economic structures. This thesis examines the importance of human capital for economic development in Thailand during this period of high economic growth and structural transformation (1985-2000), using labour force survey data. Human capital is measured by educational attainment, and the three labour market outcomes of interest are earnings, sector of work, and earnings inequality. I address the endogeneity of education using an education policy shift – the change in the compulsory schooling law – that produces exogenous variation in education. Owing to different outcomes of interest related to various types of heterogeneity, the three main chapters of the thesis adopt distinct modelling frameworks. The thesis consists of six chapters, including this introduction and the conclusion.

The second chapter provides background for the three main chapters. I review the existing literature on Thailand which is relevant to the thesis. I also explain the education system in Thailand and the education reform, which is used in the identification strategy. Subsequently, I describe the underlying data sources and the construction of the key variables.

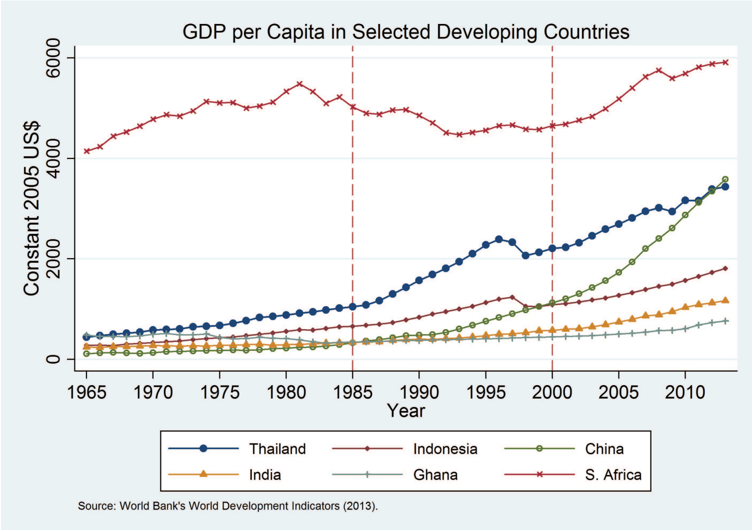
In the third chapter, I examine the overall effects of education on earnings and on the probability of being in the non-agricultural sector. While the change in the compulsory schooling law behaves similarly to an instrument, it influences educational attainment in a discontinuous way by assigning older and younger cohorts to different education regimes. As a result, a regression discontinuity (RD) framework is adopted to estimate the average education effect. It is important to emphasise that the RD technique constrains the estimated effects of education on the two outcomes to be linear. In addition, the estimates only apply to sub-populations whose decisions about educational attainment change as a result of the education policy shift and may thus not be entirely representative at the population level.

In response to these constraints in interpreting the results from the RD estimates, in the fourth chapter, I estimate the effects of education on the sectoral sorting and earnings processes in a more flexible fashion with an objective to compare the returns to education across sectors and levels of education. First, due to a substantial number of unpaid family workers in the agricultural sector, the sectoral sorting process is allowed to be correlated to how workers select themselves into paid or unpaid employment. Second, the effects of education on earnings are allowed to be explicitly heterogeneous across sectors, levels of education, and individuals. These heterogeneous effects allow me to analyse how education facilitates structural transformation. More specifically, if human capital is one of the key factors of growth driven by structural transformation, returns to education – particularly at the medium and high levels of education – should be larger in the non-agricultural sector. Third, I attempt to estimate these effects of education for the entire population. To do so, I estimate the earnings functions for the two sectors separately, using a double selection technique to correct for the sample selection bias and a control function (CF) approach to control for the endogeneity of education and heterogeneous returns across individuals.

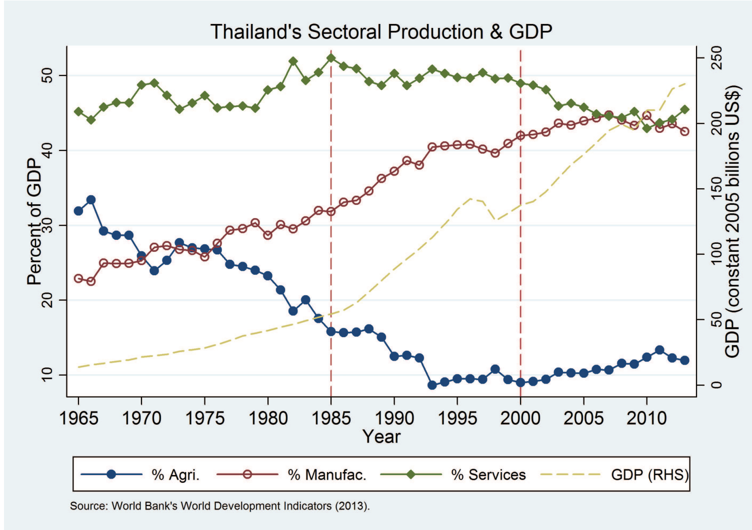
In the fifth chapter, I study how the increased primary education completion rate, induced by the change in compulsory schooling, affects earnings inequality. I seek to identify the effects of primary education completion on earnings at different points of the distribution, and thus earnings inequality, using a recently developed approach—called regression discontinuity treatment effects. The distributional treatment effects can be measured under the RD framework when the effects of primary education completion are allowed to be heterogeneous across the earnings distribution. I investigate whether this type of heterogeneity exists, and if so, whether it reduces or increases overall inequality. Subsequently, I also analyse these distributional effects of primary education completion on earnings for the rural and urban areas separately.

Figure 1.1: Overview of Thailand's Structural Transformation

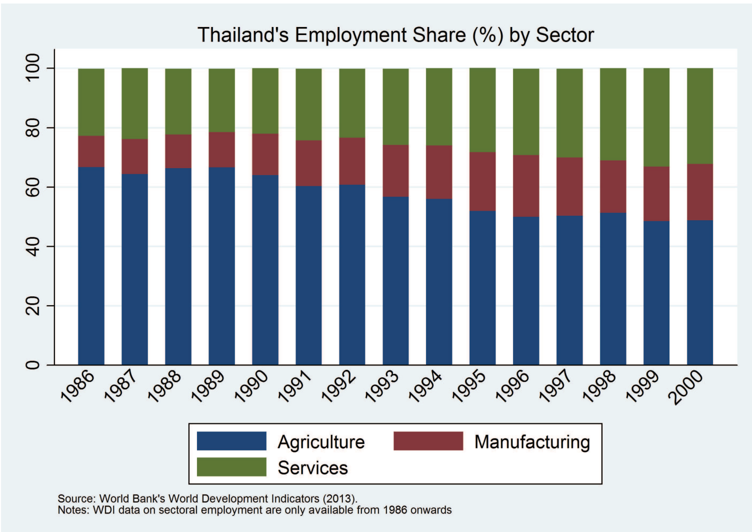
(a) GDP per Capita in Selected Developing Countries



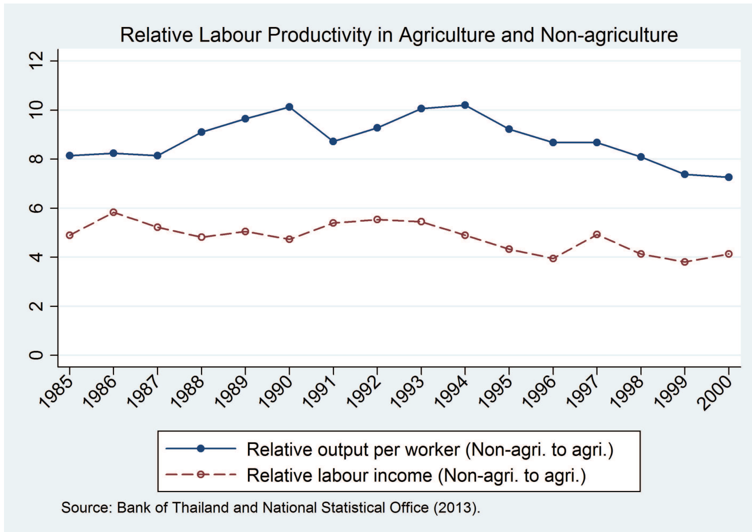
(b) Sectoral Production Share



(c) Sectoral Employment Share



(d) Relative Labour Productivity



Chapter 2

Background and Data

This chapter aims to provide background for the following three chapters of the thesis. First, I review the existing literature on Thailand which is relevant to the thesis. It includes the literature on the effects of education on earnings, sector of work, and earnings inequality in Thailand. Second, I explain the education system in Thailand and the education reform, which is used in the identification strategy. Last, I describe the underlying data sources, discuss the construction of the key variables, and present the overall descriptive statistics of the data to provide an overview of the Thai labour market.

2.1 Relevant Literature on the Impacts of Education on Labour Market Outcomes in Thailand

Among the three labour market outcomes of interest in this thesis – namely earnings, sector of work, and earnings inequality – the literature concentrates on estimating earnings functions for Thailand (see Table 2.1 for a summary of the literature on returns to education in Thailand).[1] While nearly every study recognises the heterogeneity in returns to education across gender, most of them focus only on wage-employed workers and restrict the education effects to be linear. In addition, most papers ignore potential labour supply effects and regional price differentials.[2] Most importantly, only very few papers correct for the endogenous

[1]The discussion here does not include the literature that studies returns to vocational schooling (as opposed to an academic course) as it is not directly relevant to the thesis.

[2]That is, most papers use the log of monthly wage as a dependent variable without controlling for hours of work. The exceptions are Warunsiri and McNown (2010) and Leckcivilize (2013) who use the log of hourly earnings as the dependent variable. Some of the studies that attempt to correct for price differentials only use the national consumer price index (CPI) and disregard potential regional price differentials.

choice of education or non-random sample selection.

Most of the studies investigating the returns to education in Thailand do not attempt to correct for any econometric problems which could result in biased estimates.[3] Chiswick (1977) is one of the first studies estimating an earnings function for Thailand. By using the Household Socio-Economic Survey (HSES) for the year 1971, the paper estimates the returns to education for both wage- and self-employed workers who were non-farm earners and resided in Bangkok.[4] The returns to education are found to be around 9 per cent for men and 13 per cent for women, and they are estimated to be slightly higher when restricting the sample to non-farm wage workers. By using the Labour Force Survey (LFS) and covering a larger sample of wage-employed workers, other more recent studies present slightly higher returns to education, at least for men. Chalamwong and Amornthum (2001) estimate the returns to education to be around 11 per cent for men and marginally lower for women, who were full-time wage earners during the years 1985 to 2000. Hawley (2004), who uses the same data and sample of analysis, also gets quite similar results of around 11 per cent returns to education for both men and women.[5]

Meanwhile, other studies allow for the effects of education to be heterogeneous across levels of education within an ordinary least squares (OLS) framework. They do so by using a spline function approach. In other words, a dummy variable for each of the education levels is included. The returns are found to be larger at higher levels of education for wage-employed workers from both HSES and LFS (Chalamwong and Amornthum, 2001; Patrinos et al., 2006; Tangtipongkul, 2013).

Although most studies ignore any econometric estimation problems, Schultz (1994) attempts to correct for the non-random selection into wage employment using the maximum likelihood estimation (MLE) method. The paper finds that non-random sample selection causes the OLS estimates to be downwardly biased owing to a negative relationship between educational choice and the unobserved wage determining factors among the selected sample of wage-employed workers. In addition, the returns to education increase, from primary to secondary education, but then decline substantially for tertiary education especially for female workers.

[3] That is, they only estimate earnings functions using ordinary least squares regressions (OLS).

[4] By restricting the sample of analysis only to non-farm earners who resided in Bangkok, the study excludes workers in the rest of the country, most of whom were self-employed.

[5] To be precise, Chalamwong and Amornthum (2001) use the Labour Force Survey (LFS) of 1985, 1990, 1995, and 2000, while Hawley (2004) uses the LFS for 1985, 1995, and 1998.

Warunsiri and McNown (2010) is one of the two studies that try to correct for the endogeneity of education. Using a pseudo-panel approach and the LFS during the years 1985 to 2005, they find that the returns to education range from around 13 per cent for male wage-employed workers to 18 per cent for female wage-employed workers. These returns are higher than the OLS estimates and the authors attribute this downward bias to measurement error and a substitution between education and ability. Leckcivilize (2013) addresses the endogeneity of education using an IV approach.[6] The paper uses the LFS during the years 1994 to 2009 to estimate the returns to education for wage earners who were born between 1961 and 1970 and thus aged between 24 to 48 years during the survey years. The instrumental variable used is the intensity of the compulsory education reform during the 1960s and 1970s, which the paper claims varies across localities and cohorts. Leckcivilize (2013) finds the IV returns to be zero for men and around 9 per cent for women, much lower than the OLS estimates. The author explains the vast difference across gender by potential selection into wage employment among women, and a weak instrument problem.

While there is no study directly examining the impacts of education on sectoral reallocation of the labour force in Thailand, a few descriptive studies on economic development during the high growth period underline the potential contribution of education to the expansion of manufacturing and service activities. Sussangkarn and Chalamwong (1996) argue that good primary education helped reallocate a substantial proportion of low income agricultural workers towards the manufacturing and service sectors in the late 1980s. Numnak (2006) presents similar facts using data on government expenditure. Since 1985, the share of human development expenditure, which includes government spending on education, has been substantial and increased continuously. The paper shows that education expenditure on its own accounted for around 20 to 25 per cent of total government expenditure over the 1985-2001 period of structural change. By analysing the level of human capital in each of the employment sectors separately, a study by ADB (2007) finds that the increase in education intensity in the 1990s was significantly higher in the non-agricultural sector.[7] Together with the gradual and continuous increase in education of the labour force, most non-agricultural

[6] Relatively similar to this thesis, Leckcivilize (2013) also uses the education reform to identify the effects of education. I was not aware of his contribution when starting my research on this topic in 2009, with the MPhil thesis submission in 2010. In any case, Leckcivilize (2013) is substantially different from the returns to education studied in this thesis in terms of (i) the period of study and (ii) the sample of workers included. In addition, I have serious concerns about the validity of the instruments used in his paper (as discussed in detail in Chapter 3, p. 34).

[7] Education intensity is measured by the share of workers with at least lower secondary education.

jobs created during this period have provided an earnings premium for educated workers. This suggests a sizeable demand for education in the non-agricultural sector.

Regarding the distributional impacts of education on earnings, the existing literature focuses on inequality decompositions and the correlation between education and inequality over time rather than causal effects of education on earnings inequality. Earlier studies using the HSES find that the increase in average education and lower inequality in education are correlated with a decline in income inequality observed in the early and mid 1990s (Fofack and Zeufack, 1999; Motonishi, 2006). However, these correlations are not statistically significant. In addition, they employ only one measure of inequality, namely the Theil-L index, which is more sensitive to the lower part of the distribution (Anand, 1983) and could disguise significant changes at the middle or the top of the distribution. More specifically, Fofack and Zeufack (1999) estimate the key determinants of income inequality using a pseudo-panel technique. As their preferred specification is a fixed effect model, variation in education for a given birth cohort (which is a panel observation) over time is expected to be very low. This may be the reason for the insignificant negative correlation between education and income inequality.[8] On the other hand, Motonishi (2006) constructs panel data at the region level and conducts a regression-based analysis. The paper is interested in quantifying the impacts of inequality of education, measured by a standard deviation of years of education, and financial development on income inequality. However, the inequality measure used is the Theil-L index of household income. By ignoring household size and composition, income inequality could be underestimated. More recently, Lathapipat (2008) studies how changes in the education distribution affect the hourly wage distribution during the years 1987 to 2006, using the LFS. The paper employs the two-stage decomposition technique developed by Firpo et al. (2009, 2011) to estimate the impacts of education on wages at different points of the distribution.[9] Education is found to enhance wage inequality, especially for the rich. While this conclusion differs from the earlier literature and the general perception that education could reduce inequality, it is important to note that the paper only focuses on male wage-employed workers. As a result, these conclusions do no necessary carry out to the overall earnings distribution. In addition, the study does not consider price differentials across time and regions, the endogeneity of education, and the non-random selection into

[8] In addition, it is not straightforward to understand the meaning of this dependent variable which is inequality within a birth cohort.

[9] This technique is a generalisation of the Oxaca-Blinder decomposition that goes beyond the mean.

wage employment.

The review of the relevant literature on the impacts of education on the three labour market outcomes in Thailand suggests that education may have played a crucial role in enhancing earnings and structural transformation, as well as in shaping earnings inequality. However, there is a gap in the literature concerning appropriate econometric methods to obtain causal relationships. Additionally, earlier studies on these issues often exclude self-employed workers from their analysis. In this thesis, I attempt to estimate the effects of education on earnings, sector of work, and earnings inequality for both wage- and self-employed workers in all sectors of work, taking into account various econometrics issues including the endogeneity of education and the sample selection problems where possible.

2.2 Thailand's Education System

Formal education in Thailand consists of primary, secondary (which is divided into lower and upper secondary), and tertiary education (which is divided into three levels: diploma, undergraduate, and graduate levels).[10] Thai children were required by law to start school in the calender year in which they became eight years old.[11] 4 years of lower primary schooling were made free and compulsory at first in 1921 (Ministry of Education of Thailand, 2010). The upper secondary, diploma, and undergraduate levels can be divided further into academic and vocational streams. During that time, Thailand's schooling system was 4:3:3:2. That is, 4 years of lower primary school; 3 years of upper primary school; 3 years of lower secondary school; and 2 years of upper secondary school.

Before the 1990s, the major education policy reform was to raise the compulsory years of schooling from 4 years of lower primary to the full primary education.[12] The policy was firstly and partially implemented in 1960 based on the acceptance of each locality and the readiness of schools in that locality. Instead of a strict rule, municipalities were encouraged to provide education up to the upper primary level free of charge but it was not compulsory to attend to that level.[13] In 1978, the Education Reform Act 1978 was implemented nationwide. First,

[10]In this thesis, I use the words upper secondary and high school, and tertiary education and higher education interchangeably.

[11]In Thailand, the academic year begins in May and ends in March of the following year.

[12]As specified in the first and the second National Economic and Social Development plans (1961-1966 and 1967-1971), Thai policy makers viewed education at primary and lower secondary levels as a requirement for the industrialisation process.

[13]As a result, this 1960 policy might have influenced the decisions about educational attainment of children

the education reform changed the schooling system to 6:3:3, with 6 years of primary school, and 3 years each of lower and upper secondary school.[14] Second, it also raised compulsory education, which was 4 years of lower primary at that time, to 6 years of primary education, free of charge. As the school entry age remained unchanged, the first cohort that was subject to the reform is those who were born in the year 1967, and thus were in the lower primary level (the fourth year of schooling) at the time the reform was implemented. The long preparation time before implementation suggests that a decline in school quality as a result of the extra two compulsory years seems unlikely.

As a result of the Education Reform Act 1978, holding other factors constant, a person who was born in 1966 and only studied for 4 years of lower primary education would have been forced to stay in school for 2 more years, had he been born a year later. It is also expected to affect other children who would continue school beyond the lower primary level regardless of their years of birth because they effectively received additional two years of free education. But their effects may not be exactly equal to 2 years as observed for those who would have quit school right after the compulsory level. There also existed very few children who were born at and after 1967 but were not affected by the 1978 Education Reform due to some exceptions and delays in the implementation of the law in a few remote areas. In particular, exceptions were made for severely disabled children for whom the local schools did not have suitable facilities, and children who lived further than 3 kilometres away from the nearest public school. It is important to recognise that these exceptions remained the same before and after the reform.

2.3 Data

This section provides details of the data used in the thesis. First, I describe the underlying data sources and how the samples are constructed for the analyses. Second, I discuss the construction of the key variables. Subsequently, descriptive statistics are presented to provide an overview of the Thai labour market.

with high ability to a greater extent than those with low ability.

[14]For the upper secondary level, the changes were applied to both vocational and academic education.

2.3.1 Data Sources

The National Labour Force Survey

The main data used in the thesis come from the third-round Labour Force Survey (LFS) of Thailand during the years 1985 to 2000. The LFS has been undertaken by the National Statistical Office of Thailand (NSO), and the raw data is available from 1985 until the present. The primary objective of the survey is to estimate the size and characteristics of the Thai labour force.

The LFS is a continuous cross-sectional household survey. From 1985 to 1997, the survey was conducted in three rounds each year: (1) January to March, which is the dry season; (2) April to June, when a large group of new workers enter the labour force right after completing their education; and (3) July to September, which coincides the agricultural season. Since 1998, a fourth round of the survey has been additionally conducted during October to December. The LFS uses a stratified two-stage sampling design.[15] There are five strata of geographical regions, and 73 to 77 sub-strata of provinces.[16] Each province is divided into two main types of local administration according to population size, namely municipal (urban) and non-municipal (rural) areas. In the first stage, the primary sampling units (PSUs) are blocks and villages for municipal and non-municipal areas respectively. The number of sampled blocks or villages is proportional to the total number of blocks or villages in each province. Second, the ultimate sampling units are households. 9 to 15 households are chosen from each municipal area and the Bangkok Metropolis, while 6 to 12 households are chosen from each non-municipal area. Households covered in the survey are private households and special households, and exclude institutional households such as military installations and diplomatic personnel.[17] Before 1994, the LFS sample accounted for around 0.15 per cent of the total population (around 75,000 to 86,000 people). Since 1994, it has accounted for around 0.3 per cent of the total population (around 178,000 to 217,000 people).

The LFS data is collected through in-depth interviews with household heads. The survey

[15] Therefore, my econometric analysis accounts for both the survey design and population weights.

[16] The geographical regions include North, North East, Central, South, and Bangkok Metropolis. From the years 1985 to 1993, Thailand is comprised of 73 provinces. Three districts (Amnatcharoen, Nongbualamphu, and Sakaeo), which were formally parts of other existing provinces, became new provinces in 1994. One district (Buengkan) became a new province in 2011.

[17] Special households refer to people living in a particular institution such as dormitories and boarding houses inside factory compounds.

questions include detailed information, ranging from economic activities (such as labour force status, occupation, industry, and earnings) to individual productivity-enhancing and socio-demographic characteristics. From 1985 until the present, the questions vary by round. For instance, prior to 1998, questions on earnings and migration are only asked in the first and the third rounds. More importantly, the survey questions as well as the definitions of the labour force have been changed substantially since 2001. First, the survey dropped questions on business and farm incomes for employers and other self-employed workers. As a result, there exists no earnings information for self-employed workers from 2001 onwards. Second, the definition of the labour force has changed. From 1985 to 2000, persons aged 13 years and above are classified as being in the labour force (being employed, unemployed, or seasonally inactive labour force) or outside the labour force (due to, for example, currently studying or being disabled). Due to a change in the legal minimum age for child labour, this cut-off age has been changed to 15 years since 2001.

There are two major reasons for broadly confining the study period to 1985 to 2000. First, as mentioned earlier, this is the period of high economic growth and structural change in sectoral production. It would be very interesting to see how labour market patterns developed to support this outstanding economic development. Second, as the research question posed involves returns to education across sectors and the earnings distribution, it is essential to have labour force information which appropriately represents every sector and the country's economy. Figure 2.1a demonstrates the importance of self-employment in the Thai labour market over the past three decades. Despite a continuous decline in self-employment, more than half of the total employed workers have been self-employed throughout this period. The inclusion of survey data post-2000 would significantly and non-randomly exclude earnings information of self-employed workers whose characteristics may systematically differ from wage-employed workers.

In addition, I use only the third round of each year's LFS data. The third-round LFS is undertaken during June to September, which coincides with the agricultural season. It is expected to capture the actual number of agricultural workers, who may migrate to work temporarily in urban areas during the dry seasons (Sussangkarn and Chalamwong, 1996). The empirical analysis, except for those of earnings inequality, is conducted for men and women separately. The age ranges of the sample differ across chapters due to the identification strategies employed. This will be discussed in detail in each of the chapters.

The Regional Consumer Price Indices

The LFS data report nominal earnings which need to be adjusted for differentials in the price level over time and across regions. Nominal earnings from the LFS are deflated using the regional headline consumer price indices (CPI) from 1985 to 2000 produced by Ministry of Commerce of Thailand.[18] The regional CPI is available for rural and urban areas in five geographical regions, consistent with those of the LFS.

2.3.2 The Construction of the Primary Variables

The four primary LFS variables used in the thesis are hourly earnings, years of education, year of birth, and sector of work. They are defined as follows.

Hourly Earnings

First, the labour earnings variable is constructed from the LFS questions on wages received including over-time pay for wage-employed workers, and net profits on farms and/or businesses for self-employed workers.[19] In cases where there is more than one household member working as a paid worker in the same household farm or business, net profits are then divided equally among them.[20] Second, the measure of earnings used in this thesis is hourly earnings. The survey answers on earnings are in different units (including per hour, per day, per week, and per month) according to how individuals received their labour earnings. Hourly earnings are used to standardise these because other measures such as daily or monthly earnings do not factor out the labour supply effect (Card, 1999; Moenjak and Worswick, 2003). Individuals with higher education are likely to supply more labour (for instance, more hours of work per month), and thus have higher monthly earnings. In this case, the returns to education will be higher for annually, monthly, or weekly earnings than for hourly earnings (Card, 1999). As a result, hourly earnings are more appropriate. Finally,

[18] Prior to 1990, only the headline CPI was reported. Since 1990, the core CPI is also reported. The latter excludes raw food and energy items from the consumer price index basket.

[19] The earnings and over-time pay included here are only those obtained from the primary job.

[20] Household members are identified as working in the same household farm or business if their answers to the following LFS questions are similar: (1) sector of work; (2) type of business; (3) number of employees; and (4) net profits on farm or business. Net profits on household farms and/or businesses are shared only among workers who reported as being an employer or an own-account worker and not among those reported as being an unpaid family worker.

to adjust for inflation and regional price differentials as discussed earlier, hourly earnings are deflated using the regional headline CPI. The earnings variable in all chapters is expressed in Thai Baht, at Bangkok 2007 prices, per hour. One Pound Sterling equals approximately 68.6 Thai Baht at this base year (Bank of Thailand, 2013).

Years of Education

The LFS asks for the highest level of education achieved by an individual. However, it does not ask for the number of years a person has spent in schooling. Levels of educational attainment are converted into years of education according to the LFS codebook with the simplifying assumptions that no-one repeats the same level of education, skips the prerequisite levels, or attains more than one degree.

As discussed in Section 2.2, the schooling system actually changed from 4:3:3:2 prior to 1978 to 6:3:3 at and after 1978. In other words, years in primary school were reduced from 7 to 6 years while years in secondary school were raised from 5 to 6 years. The NSO has already taken this change into account when converting the survey responses to years of education as provided in the LFS codebook. This could be done because the LFS also asks whether an individual was under the old or new education system when studying, and because the NSO has information from the Ministry of Education on how grades (education stages) in the old schooling system compare to those in the new system.

Year of Birth

As year of birth is not reported in the survey, I impute it from information on age and survey year.[21] That is, year of birth is equal to survey year minus age. Figure 2.1b suggests that there exist some rounding-off errors but they do not look particularly severe and there appears no clustering at multiples of five or ten as one might expect.

Sector of Employment

The sector of employment variable is constructed from the LFS questions on mutually exclusive industries of employed persons according to the International Standard Industrial Classification (ISIC-2). I partition sectors of employment into two economic sectors, namely agriculture and non-agriculture. This is because the thesis focuses on the labour market

[21] The cohort imputation will be used later assign individuals to education policy regimes.

implications of Thailand's structural change during the high economic growth period.[22] The agricultural sector consists of agriculture, hunting, forestry, and fishing. The non-agricultural sector consists of industry (including mining and quarrying, manufacturing, utility production, and construction) and services (comprising wholesale and retail trade, restaurants and hotels, transport and communication, finance, real estate and business services, and community and social services).[23]

2.3.3 LFS Overall Descriptive Statistics

The overall descriptive statistics from the LFS during the period of study are presented in Table 2.2 for the entire sample and Table 2.3 for men and women separately. First, the LFS information on labour force participation and employment suggests substantial labour utilisation in Thailand. Despite having declined slightly over time, the labour force participation rate remained above 80 per cent during the years 1985 to 2000, which was relatively higher than the rates observed in other developing countries.[24] Over the same period, unemployment declined, and remained low and stable. The average unemployment rate until 1997 was 1.7 per cent. Owing to the 1997 financial crisis, the unemployment rate rose to 3.1 per cent in 1998. This unemployment rate is still moderate, compared to the large negative economic growth of 10.5 per cent (Bank of Thailand, 2013). It is important to note that the NSO classifies anybody who works at least one hour during the reference week as employed. This may lead to a low unemployment rate, especially during the financial crisis when the labour force was suspected to have migrated to service or agricultural sectors (Sussangkarn and Chalamwong, 1994). However, information on underemployment supports the conclusion of low unemployment. Underemployment, which refers to those who worked less than 35 hours per week and were available for more, remained around 3 to 4 per cent during and a few years after the crisis. Furthermore, nearly 70 per cent of the underemployed workers worked more than 20 hours per week during the same period. While

[22]The important feature of structural change is economic growth accompanied by large changes in the sectoral composition of output and employment. This often refers to a decline in agricultural activities and a corresponding expansion of the non-agricultural sector (Lewis, 1954).

[23]It is important to also note that workers in these sectors may be unpaid if they work for their family farm or business. The LFS also asks for an individual's employment type. The answers are classified into 5 groups, namely private wage employment, public wage employment, employer, own-account worker (sole trader), and unpaid family worker.

[24]During 1985 to 2000, the average labour force participation rates for East Asia and Pacific (developing countries only), South Asia, Sub-Saharan Africa, low-income and lower middle-income countries were 76, 61, 69, 76, and 61 per cent, respectively (World Bank's WDI, 2013)

labour utilisation was high, a substantial number of employed workers were unpaid family workers, especially among female workers and in the agricultural sector. The proportion of unpaid family workers to total employed workers was around 35 per cent in the second half of the 1980s and it decreased to around 25 per cent by 2000. This evidence suggests that, while the selection into the labour force or employment may be relatively unimportant, the self-selection into paid employment must be taken into account in any studies focusing specifically on paid employed workers.

Second, earnings of paid employed workers increased considerably over time. Average real hourly earnings grew by 3.4 per cent per annum from 22 Thai Baht in 1985 to 37 Thai Baht in 2000. While being lower than the mean, median real hourly earnings grew by 3.7 per cent annually, slightly higher than the annual growth of average earnings. In addition, on average, men earned slightly more than women. Over the same period, earnings inequality decreased slightly. Earnings inequality measured by the Gini coefficient of the hourly earnings distribution declined from 0.54 in 1985 to 0.52 in 2000. Meanwhile, the ratio of the 90th to the 10th percentile of the earnings distribution (P90/P10) reduced from 18 times in 1985 to 14 times in 2000. The ratio of the 75th to the 25th percentile of the earnings distribution (P75/P25) also declined by around 1 time over the same period. As the Gini coefficient is relatively more sensitive to changes at the centre of the distribution (Anand, 1983; Cowell, 2011), the movements in these inequality measures suggest that the improvement in inequality in Thailand during the period of study may be attributable more to changes in the tails.

Third, structural transformation is shown by shifts in the sectoral composition of employment towards the non-agriculture sector. The share of agricultural employment decreased from 65 per cent in 1985 to 48 per cent in 2000. Manufacturing and service employment increased in response to this.[25] The shift from agricultural employment to non-agricultural employment was stronger for female workers. Note that, during the two years during and after the financial crisis, employment in the manufacturing sector fell, while the share of workers in the agricultural sector remained rather stable and service employment increased. This evidence does not support the earlier literature which claims that the labour force underwent reverse migration, and hence, the agricultural sector cushioned workers against

[25] The increases in the two sub-sectors were quite equal in terms of percentage points.

unemployment during the crisis (Mephokee, 2003; Krongkaew et al., 2006).[26]

Fourth, educational attainment of the labour force also improved substantially over the same period. Average years of education among employed workers increased from 5.1 years (which was equivalent to less than primary education) in 1985 to 6.9 years (equivalent to higher than primary but lower than lower secondary education) in 2000. The distributions of education for male and female workers are also presented in Figure 2.1c. Over this period, the proportion of employed workers with less than primary education declined greatly from 81 per cent to 49 per cent. In response to this, the proportions of the employed labour force with at least primary and at least upper secondary education increased from 19 and 13 per cent in 1985 to 51 and 31 per cent in 2000, respectively. The education distribution also show that around 90 to 95 per cent of workers have attained less than tertiary education (that is, less than or equal to 12 years of education). Male workers were relatively more educated than female workers but the gap in educational attainment has narrowed over time.

[26] The conclusion from the literature implies that there was underemployment or labour surplus in the agricultural sector. This is not observed in the LFS statistics.

Table 2.1: Summary of Literature on Returns to Education in Thailand

Authors	**Data**	**Year**	**Sample of Analysis**	**Spec.**	**Method**	**Results**
Chiswick (1977)	HSES	1971	Non-farm earners only in Bangkok, aged ≥15	Linear	OLS	• All non-farm earners: Constant returns of 9% for men, and 13% for women. • Non-farm wage earners: Constant returns of 10% for men, and 14% for women.
Schultz (1994)	HSES	1985, 1989	Wage-employed workers, aged 15-60	Spline	OLS and sample selection	• Male workers: 16% for primary, 12% for secondary and 11% for tertiary. • Female workers: 14% for primary, 19% for secondary, and 8% for tertiary.
Patrinos (1995)	HSES	1989	Wage-employed workers, men only	Linear	OLS	• Constant return of 12% for men.
Chalamwong and Amornthum (2001)	LFS	1985, 1990, 1995, 2000	Wage-employed workers, aged 24-35 ⊠	Linear and spline	OLS	• Linear returns of 11% for men and 10% for women. • Male workers: 9% for high school, 10% for vocational school, 14% for vocational college, 18% for university. • Female workers: 7% for high school, 12% for vocational school, 11% for vocational college, 14% for university.
Hawley (2004)	LFS	1985, 1995, 1998	Wage-employed workers, aged 24-35 ⊠	Linear	OLS	• Constant returns of 10% for male workers, and 11% for female workers. • No change in returns over time.
Patrinos et al. (2006)	HSES	2002	Wage-employed workers, aged 25-65, men only	Spline	OLS	• Male workers: 8% for lower primary, 12% for primary, 14% for secondary, 12% for high school, 13% for vocational school, and 22% for university.
Warunsiri and McNown (2010)	LFS	1985-2005	Wage-employed workers, born in 1946-1967 ⊠	Linear	OLS and pseudo-panel	• Constant returns of 13% for men and 18% for women. • OLS returns are downwardly biased.
Tangtipongkul (2013)	LFS	2007-2010	Wage-employed workers, aged 15-60	Spline	OLS	• Increasing returns with vocational education premium of 15% for secondary, and -29% for tertiary. • No change in returns over time.
Leckcivilize (2013)	LFS	1994-2009	Wage-employed workers, born in 1961-1970	Linear	OLS and IV	• Constant returns of 0% for men and 9% for women. • OLS returns are upwardly biased.

Notes:
* HSES is the Household Socio-Economic Survey and LFS is the Labour Force Survey.
**⊠ indicates that a study further confines its sample of analysis to workers who worked at least 30 hours per week.
*** In a spline specification, educational attainment enters the earnings model as dummy variables at various levels.

Table 2.2: LFS Descriptive Statistics

Unit: Million persons, otherwise specified	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Total population	49.5	48.3	50.3	54.2	55.0	56.0	56.9	57.4	58.3	59.2	59.2	59.7	60.4	61.0	61.4	62.1
Male population	50.1%	50.0%	49.9%	50.0%	50.0%	50.0%	50.0%	50.0%	50.1%	49.6%	49.9%	49.9%	49.9%	49.9%	49.7%	49.7%
LFS representativeness	0.15%	0.14%	0.13%	0.13%	0.17%	0.16%	0.16%	0.16%	0.15%	0.30%	0.30%	0.28%	0.29%	0.27%	0.27%	0.26%
Labour Force																
Working population (Age 20-65)	23.6	23.3	25.0	27.4	28.2	29.2	30.1	31.2	32.1	32.9	34.0	34.6	35.6	36.0	36.4	37.0
Of which due to study	2.1%	1.9%	2.0%	1.8%	1.5%	1.8%	1.4%	1.6%	1.7%	1.9%	2.3%	2.3%	2.4%	2.9%	3.2%	3.1%
Labour force (Age 20-65)	20.3	20.1	21.9	24.2	24.8	25.6	26.2	27.4	27.8	28.2	29.0	29.1	30.1	30.1	30.1	30.6
Male labour force (Age 20-65)	54.3%	54.1%	53.3%	53.3%	53.5%	53.6%	54.1%	54.0%	54.4%	54.1%	54.5%	54.9%	54.5%	54.7%	54.7%	54.6%
Labour force participation rate	86.0%	86.2%	87.5%	88.2%	88.1%	87.9%	87.1%	87.9%	86.7%	85.6%	85.3%	83.9%	84.6%	83.7%	82.6%	82.7%
Unemployment rate	2.8%	2.8%	3.5%	2.7%	1.2%	1.5%	2.1%	1.2%	1.1%	1.1%	0.9%	0.9%	0.7%	3.1%	2.6%	2.1%
Underemployment rate	2.2%	2.6%	3.1%	2.4%	1.5%	2.3%	2.2%	2.1%	2.8%	1.9%	1.7%	1.9%	1.9%	2.8%	3.7%	2.9%
Seasonally inactive	0.5%	0.3%	1.9%	0.1%	0.1%	0.4%	0.4%	0.1%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%
Employed workers (Age 20-65)																
Male workers	54.6%	54.3%	54.4%	53.5%	53.6%	53.7%	54.5%	54.1%	54.6%	54.2%	54.6%	55.0%	54.5%	54.7%	54.7%	54.7%
Agricultural workers	65.4%	64.0%	61.8%	63.9%	64.8%	62.1%	58.3%	59.5%	55.6%	54.9%	51.3%	49.9%	49.9%	50.4%	48.0%	48.1%
Unpaid workers	35.5%	35.1%	32.7%	36.4%	37.0%	35.4%	32.4%	33.8%	30.6%	30.7%	28.8%	27.3%	28.7%	27.7%	25.6%	25.4%
Of which are in agriculture	87.9%	87.5%	86.1%	88.0%	88.6%	88.3%	85.9%	88.1%	86.0%	86.0%	82.4%	82.9%	82.9%	81.6%	78.9%	78.8%
Self-employed workers	73.5%	72.7%	71.4%	71.9%	72.5%	71.2%	68.5%	68.4%	65.6%	64.9%	64.5%	62.8%	62.4%	63.2%	61.6%	60.1%
Avg. age	32.1	32.3	32.7	32.8	33.0	33.1	33.3	33.4	33.7	33.9	35.3	35.5	35.9	35.9	36.1	36.4
Avg. years of education	5.1	5.3	5.4	5.5	5.5	5.6	5.7	5.8	6.0	6.1	6.1	6.2	6.3	6.6	6.8	6.9
Avg. hours of work per week	52.6	53.1	51.8	52.2	53.7	50.9	52.3	52.3	51.2	52.0	51.8	51.2	49.9	50.5	49.8	50.3
Earnings																
Avg. hourly earnings (THB)	22.4	21.8	22.2	25.9	24.3	27.6	27.4	31.5	33.5	34.8	38.4	38.6	45.1	37.4	37.6	37.0
Median hourly earnings (THB)	13.3	12.6	13.3	15.0	14.8	16.3	16.9	18.8	20.5	21.0	23.9	24.3	25.3	23.4	23.3	23.0
Earnings inequality																
Gini	0.54	0.54	0.54	0.55	0.53	0.54	0.54	0.54	0.54	0.54	0.53	0.52	0.55	0.51	0.52	0.52
P90/P10	18.2	16.8	17.3	16.1	14.2	15.4	16.0	16.6	17.4	17.1	17.0	15.0	18.1	13.0	14.4	14.0
P75/P25	4.4	4.3	4.2	4.1	3.7	3.7	3.6	3.7	3.8	3.6	3.6	3.1	3.3	3.1	3.0	2.9

Source: Author's calculation from the Labour Force Surveys 1985-2000 (NSO), using the surveys' population weight variable.
Notes: * Labour force includes those who are between 20 and 65 years, do not study, and are currently working or looking for jobs. People in the labour force are categorised into being employed, unemployed, and seasonally inactive.
** Labour force participation is the proportion of the labour force to the overall population aged between 20 and 65 years and not studying.
*** Underemployment refers to those employed workers who work less than 35 hours a week and available for more.
**** Earnings are adjusted for inflation and regional price differences using the regional headline CPI.

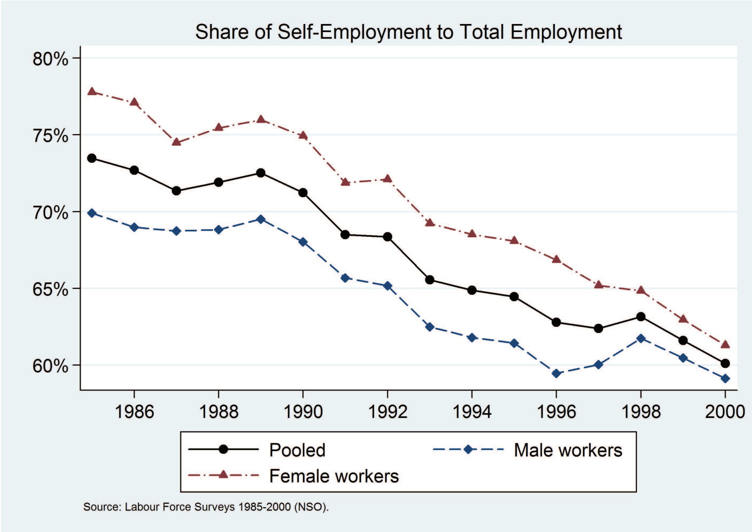
Table 2.3: LFS Descriptive Statistics by Gender

Male labour force	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Labour force participation rate	93.9%	94.1%	94.2%	94.6%	94.9%	94.7%	94.6%	95.0%	94.4%	93.7%	93.3%	92.6%	92.6%	92.0%	91.0%	91.0%
Unemployment rate	2.5%	2.6%	2.7%	2.4%	1.1%	1.4%	1.5%	1.1%	0.9%	0.9%	0.8%	0.9%	0.7%	3.2%	2.5%	2.1%
Underemployment rate	2.0%	2.3%	2.8%	2.2%	1.3%	2.1%	2.0%	1.8%	2.6%	1.7%	1.5%	1.9%	1.8%	2.7%	3.7%	2.9%
Seasonally inactive	0.2%	0.1%	0.9%	0.0%	0.0%	0.4%	0.2%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%
Employed workers (Age 20-65)																
Agricultural workers	63.7%	62.4%	60.9%	62.4%	63.5%	60.7%	57.5%	57.8%	54.4%	53.5%	49.6%	48.3%	48.6%	51.0%	48.8%	48.8%
Unpaid workers	18.2%	17.7%	17.7%	20.4%	20.5%	19.4%	17.6%	18.8%	16.0%	17.1%	14.9%	13.5%	15.5%	15.2%	13.9%	14.1%
Of which are in agriculture	88.3%	87.6%	85.6%	89.8%	89.7%	89.7%	87.7%	89.3%	87.0%	87.4%	83.6%	83.5%	84.3%	83.7%	78.5%	79.1%
Self-employed workers	69.9%	69.0%	68.7%	68.8%	69.5%	68.0%	65.7%	65.2%	62.5%	61.8%	61.4%	59.5%	60.0%	61.7%	60.5%	59.1%
Avg. age	35.5	35.5	35.8	35.6	35.5	35.8	35.7	35.8	35.9	36.0	37.0	37.2	37.3	37.6	37.8	38.0
Avg. years of education	5.1	5.4	5.5	5.6	5.6	5.7	5.8	5.9	6.1	6.2	6.2	6.2	6.4	6.7	6.9	6.9
Avg. hours of work per week	53.9	54.0	52.9	53.5	54.8	53.0	53.8	53.8	52.6	53.4	53.2	52.5	51.1	51.9	51.0	51.5
Avg. hourly earnings (THB)	22.8	22.3	22.9	26.9	25.0	28.0	28.1	32.2	34.2	36.1	40.2	40.0	46.0	37.8	38.8	37.0
Median hourly earnings (THB)	13.6	12.7	13.5	15.6	15.4	16.8	17.4	19.3	20.5	21.7	24.5	25.6	26.0	23.4	23.4	23.0
Female labour force	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Labour force participation rate	78.2%	78.6%	80.9%	81.8%	81.5%	81.2%	79.7%	80.7%	79.0%	77.8%	77.3%	75.3%	76.6%	75.5%	74.3%	74.5%
Unemployment rate	3.2%	3.0%	4.4%	3.2%	1.3%	1.7%	2.7%	1.4%	1.4%	1.2%	1.1%	0.9%	0.8%	3.0%	2.6%	2.1%
Underemployment rate	2.4%	2.9%	3.3%	2.6%	1.6%	2.6%	2.5%	2.5%	3.0%	2.0%	1.9%	1.9%	2.0%	3.0%	3.6%	2.9%
Seasonally inactive	0.8%	0.5%	3.1%	0.2%	0.2%	0.5%	0.7%	0.1%	0.3%	0.2%	0.1%	0.1%	0.1%	0.2%	0.1%	0.1%
Employed workers (Age 20-65)																
Agricultural workers	67.5%	65.9%	62.8%	65.5%	66.5%	63.9%	59.3%	61.5%	57.1%	56.5%	53.3%	52.0%	51.4%	49.7%	47.0%	47.2%
Unpaid workers	56.3%	55.8%	50.7%	54.8%	56.1%	53.9%	50.3%	51.5%	48.0%	46.9%	45.4%	44.1%	44.4%	42.7%	39.7%	39.1%
Of which are in agriculture	87.8%	87.5%	86.3%	87.2%	88.2%	87.8%	85.2%	87.6%	85.6%	85.5%	82.0%	82.6%	82.3%	80.7%	79.1%	78.7%
Self-employed workers	77.8%	77.1%	74.5%	75.5%	76.0%	74.9%	71.9%	72.1%	69.2%	68.5%	68.1%	66.9%	65.2%	64.8%	63.0%	61.3%
Avg. age	35.34	35.42	35.47	35.45	35.4	35.62	35.51	35.67	35.71	35.69	36.75	36.92	37.04	37.24	37.34	37.6
Avg. years of educ.	4.4	4.7	4.9	5.0	5.0	5.2	5.3	5.4	5.6	5.7	5.7	5.8	6.0	6.3	6.5	6.7
Avg. hours of work per week	51.0	51.7	50.5	50.8	52.2	49.1	51.1	51.1	50.1	51.1	51.1	50.2	49.2	50.0	49.3	49.9
Avg. hourly earnings (THB)	21.4	20.8	20.7	23.8	22.8	26.9	26.0	30.0	32.2	32.4	35.2	35.9	43.6	36.7	35.6	37.0
Median hourly earnings (THB)	12.9	12.3	13.3	14.0	13.8	16.0	15.4	17.8	20.0	20.3	22.0	22.9	24.7	23.4	22.8	22.8

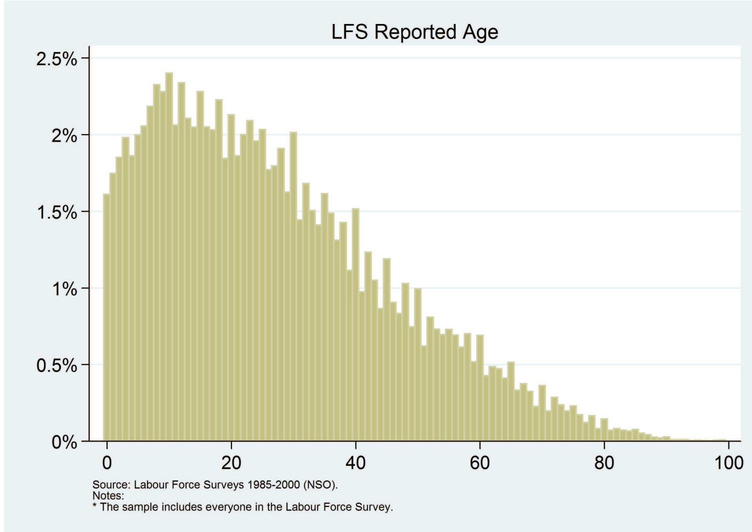
Source: Author's calculation from the Labour Force Surveys 1985-2000 (NSO), using the surveys' population weight variable.
Notes: * Labour force includes those who are between 20 and 65 years, do not study, and are currently working or looking for jobs. People in the labour force are categorised into being employed, unemployed, and seasonally inactive.
** Labour force participation is the proportion of the labour force to the overall population aged between 20 and 65 years and not studying.
*** Underemployment refers to those employed workers who work less than 35 hours a week and available for more.
**** Earnings are adjusted for inflation and regional price differences using the regional headline CPI.

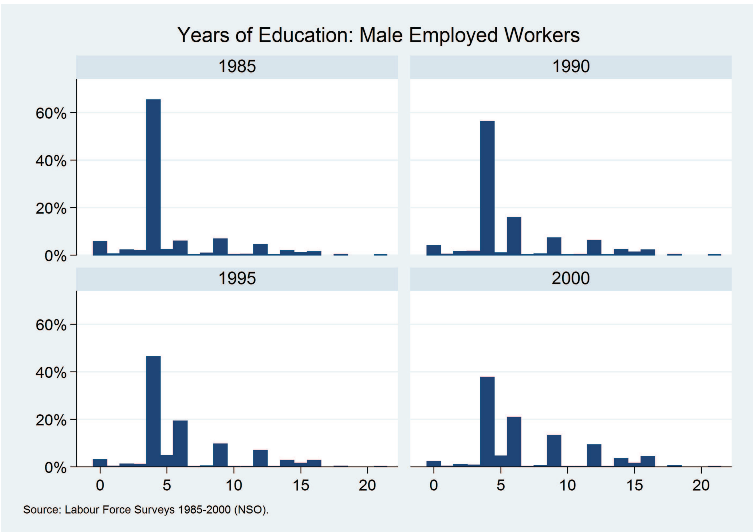
Figure 2.1: LFS Descriptive Statistics

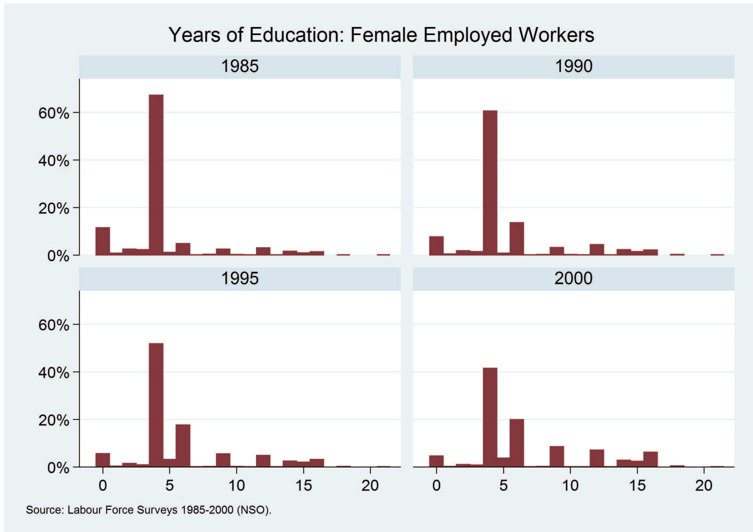
(a) Self-Employment in Thailand



(b) LFS Reported Age







(c) Years of Education

Chapter 3

Estimating Returns to Education on Earnings and Employment: A Regression Discontinuity Framework

3.1 Introduction

Education is one of the factors that play a crucial role in labour markets. The impacts of education on labour force participation, employment, occupations, and wages are found to be significant by a large number of studies. Nonetheless, the size of the impact differs greatly across countries, time periods, and variables of interest. To address a number of difficulties in estimating the causal effects of education on these labour market outcomes, various econometric methodologies have been developed in the past decades. In this chapter, I examine the impacts of education on earnings and choice of sector in Thailand, using a recently developed "reduced-form" econometric methods to estimate the causal effects of education, namely the regression discontinuity (RD) framework.

The conventional problems of estimating the effects of education on labour market outcomes comprise measurement error and the endogeneity problem arising from individual unobserved heterogeneity affecting these labour market outcomes as well as educational attainment. The instrumental variable (IV) method is often used to evaluate the causal effects of education on the outcome of interest. In the IV framework, the causal effects of education are identified by an instrument, which is related to an individual's level of education but is uncorrelated with other unobserved characteristics affecting labour market outcomes. I address these concerns about measurement error and endogeneity by using an education policy shift – the change in the compulsory schooling law – that produced exogenous variation in individuals' education. This change in compulsory schooling behaves similarly to an instrument. It induces variation in education but is unlikely to be correlated with unobserved individual factors. The only difference between the education policy shift and other instruments used in the IV literature is that the policy shift creates a discontinuous jump in years of education across birth cohorts while other instruments influence educational attainment in a smooth way. As a result, I adopt the RD framework to estimate average education effects. In this chapter, I disregard the selection into paid employment that may cause the education effects to be different from the general population. This is similar to the earlier literature and thus the estimates only apply to those in the sample. That is, the estimated impacts of education on earnings are conditional on being in the sample of paid workers.

I find that education positively influences earnings and the probability of being in the non-agricultural sector. The returns to education in Thailand during the years 1985 to 2000 are estimated to be around 8 per cent, while it increases the probability of moving into the non-agricultural sector by nearly 3 per cent. These estimated effects are slightly lower than the estimates from ordinary least squares (OLS) regressions, which could potentially reflect a positive correlation between educational attainment and unobserved individual heterogeneity affecting the outcome variables. In order to assess the importance of gender heterogeneity, the impacts of education are also estimated for men and women separately. I find that education raises earnings for women relatively more than men. On the other hand, an additional year of education helps men moving into the non-agricultural sector more than it does women. I should also emphasise that the impacts of education estimated by the RD approach only apply to sub-populations whose decisions about educational attainment change as a result of the education policy shift. As a result, the lower-than-OLS estimates could also imply relatively lower education effects among these sub-populations.[1]

Returns to education are widely studied by labour economists. This chapter makes two main contributions to the field. First, I estimate the returns to education for both wage- and self-employed workers in a developing country such as Thailand. Although there are

[1] As reviewed in Section 3.2, the existing literature finds mixed evidence. Most studies using compulsory schooling reforms in Europe and to some extent developing countries confirm the upward bias in the ordinary least squares (OLS) estimates.

a considerable number of papers studying the returns to education, most of them focus on the wage-employed population in developed countries. Some recent studies on developing countries also focus on wage-employed workers due to limited information on the earnings of the self-employed, who are a significant portion of the labour force in these developing economies. Hence, this chapter provides evidence on how education affects earnings of the majority of the labour force in a developing economy. Second, my estimates suggest how education policy could affect labour market outcomes. This is under-studied in developing economies because education policy reforms in these countries happened more recently than in developed countries. As a result, individuals who are exposed to the policy shifts have been observed only for parts of their working lives.

The structure of this chapter is as follows. Section 3.2 reviews the literature studying the effects of education on earnings and employment using an education reform as an instrument. Section 3.3 describes the identification strategy used to estimate the impacts of education on the two labour market outcomes, namely the RD framework. The data and descriptive statistics of the relevant variables from the Labour Force Survey (LFS) are discussed in Section 3.4. Section 3.5 tests the required assumptions and examines the impacts of education. It also provides robustness checks of the model specification. Section 3.6 concludes.

3.2 Using Education Reform to Identify the Effects of Education: A Review of the Literature

A number of studies examine the causal effects of education on labour market outcomes such as earnings and employment. Many of these studies measure education as years of education and identify the effects of education using the exogenous variation provided by various types of instruments. This review focuses on the literature which uses an education reform as an instrument for education and which has earnings and/or employment as outcome variables.[2]

3.2.1 Background Summary of the Returns to Education Literature

First quantified by Mincer (1974), the effects of education on earnings and other labour market outcomes at the individual level have been extensively studied. The major concern in

[2]This section does not aim to review the entire literature on private returns to education, as this has been done thoroughly by many other authors (see for example, Card (1999, 2001) and Harmon et al. (2003)).

estimating these returns involves the potential bias caused by the endogenous choice of education. That is, observed correlations between education and labour market outcomes may not reflect the causal effects of education. For earnings, this may arise if education is correlated with unobserved individual characteristics such as taste for schooling, ability in schooling or ability in the labour market.[3] The conventional hypothesis is that education-ability complementarity results in upward-biased estimates (Card, 2001). That is, an individual with high ability is assumed to find school less difficult or obtain more schooling to signal his high ability (Spence, 1973). An alternative hypothesis suggesting a downward bias is that education is compensatory for earnings capacity, meaning that highly able workers forgo more earned income when spending more years in school (Ashenfelter et al., 1999). Another source of bias is measurement error in education, which results in an attenuation bias (that is, a downward bias if the estimated coefficient is positive) (Griliches, 1977).

The earlier literature deals with the endogeneity issue by including a proxy for unobserved ability such as IQ test scores (Griliches and Mason, 1972; Griliches, 1977), or using data on siblings or twins to control for unobserved innate ability (see for example, Behrman and Taubman (1976); Chamberlain and Griliches (1975); Ashenfelter and Krueger (1994)). However, most proxies for ability are themselves influenced by or correlated to schooling, while the siblings or twins approach can only control for genetic or family factors that affect ability (Bound and Solon, 1999; Neumark, 1999). In addition, the estimated returns to education from siblings or twins data are subject to a stronger bias from measurement error (Ashenfelter and Krueger, 1994), and may not well represent the non-twin population (Blanchflower and Elias, 1993).

More recent studies use the IV method to estimate returns to education. They exploit exogenous variation in educational attainment arising from predetermined conditions such as family background or features of the education system ranging from distance to school (Card, 1993) to education reforms.[4] Imbens and Angrist (1994) argue that the results from the IV method only apply to those who change their education decisions as a result of the instrument.[5] This could be an issue if the returns to education are heterogeneous across

[3]See Griliches (1977) and Willis (1986) for surveys.

[4]For example, Ashenfelter and Zimmerman (1997) use parents' education as a family background instrumental variable. The studies using education reforms as instruments for educational attainment will be discussed in greater detail in the next sub-section.

[5]Imbens and Angrist (1994) call the instrumental variable (IV) estimates the local average treatment effects (LATE) as they represent the impacts of the treatment only among a particular sub-population.

individuals. Most studies using the IV method find a downward bias in the OLS estimates of returns to education, which is the opposite of the conventional ability bias.[6] Lang (1993) and Card (1999) refer to this as the "discount rate bias". That is, the returns to education are heterogeneous depending on individuals' tastes for schooling. The negative or downward bias in the OLS estimates is due to the fact that people who are affected by the instruments have higher costs or distastes of schooling.

3.2.2 Using Education Reforms to Identify the Effects of Education

Micro-empirical studies that use an education system or an education reform as an instrument to identify the impacts of education on earnings and employment differ greatly in their definitions of education (such as years of education, enrolment/completion levels, academic/vocational streams, public/private education), types of education reforms (including compulsory schooling years, education system, age entering school), categories of employment (employability, wage-/self-employed, sectoral employment), countries covered, and time periods. First, I divide the empirical literature on earnings returns to education into two broad categories, namely developed and developing countries. Types of education reforms also vary with stages of economic development. Second, I discuss the literature that focuses on the employment outcomes.

3.2.2.1 Returns to Education on Earnings

Education reforms in developed/advanced economies

A number of empirical studies identify the effects of education on earnings using an education reform in developed countries. While most studies of North America use an education system or a variation in education systems across states as an instrument, studies of Europe exploit various types of changes in an education system (that is an education reform) to identify the effects of education.

Angrist and Krueger (1991) are among the first studies that obtain exogenous variation in educational attainment resulting from an education system.[7] In their study on the US, quarter of birth is related to educational attainment because of school entry age and com-

[6] Although the IV coefficient estimates are often double of the OLS, they are estimated less precisely. As a result, the differences between the two are sometimes not statistically significant (Card, 1994, 1999).

[7] Unlike other more recent studies, Angrist and Krueger (1991) use the education law rather than its change (or a reform) to identify the effects of education on earnings.

pulsory school attendance laws.[8] After controlling for birth year, quarter of birth is used as an instrument for years of education in their earnings function. They find that returns to education, among employed men aged between 31 and 50 years, are around 7.5 per cent (around the compulsory schooling years), which is slightly higher than the OLS estimates. While the difference between the IV and the OLS estimates are statistically insignificant, the authors attribute the difference to bias potentially induced by measurement error. Nonetheless, Carneiro and Heckman (2003) argue against the validity of quarter of birth as an instrument due to an association between quarter of birth and some indicators of early childhood development.

More recent studies on North America report similar OLS returns but higher IV estimates than those of Angrist and Krueger (1991). By using both quarter of birth and variation in compulsory schooling laws across states as instruments, Acemoglu and Angrist (2001) find that returns to education, among men aged between 30 and 49 years in the US, range from 7.5 to 11 per cent. With the same set of instruments, Oreopoulos (2006b) estimates the returns to education in the US separately for men and all workers aged between 25 and 64 years. The IV estimates are around 7 to 13 per cent for men, and 14 to 17 per cent for all workers, both significantly higher than the OLS estimates. More recently, Clay et al. (2012) show that, by using the first compulsory school laws introduced in each of the US states during the period 1860 to 1920, the IV estimates of returns to education among men aged between 25 and 54 years are around 11 to 14 per cent, higher than the OLS estimates of around 8 per cent. When using state-specific changes in the maximum age for starting school and the minimum age for leaving schooling in Canada as an instrument, Oreopoulos (2006a) shows IV returns for all workers aged between 20 and 65 years of around 15 per cent, compared to the OLS estimates of about 10 per cent. These studies attribute higher IV estimates to heterogeneous returns to education across individuals (Card, 1999). The estimated returns refer to returns to education of specific sub-populations who are affected by variation in the instrument. In other words, the instruments identify a local average treatment effect (LATE), which exceeds the returns in the whole population, the average treatment effect (ATE).

[8] In the US, pupils are allowed to leave school when reaching a certain age, not at the end of the school year. As a result, children born in the beginning of the year (i) usually start school at an older age than those born later during the year, and therefore (ii) are allowed to drop out of school after attaining fewer years of education.

Instead of using an education system itself or a variation in education systems across regions, most studies in Europe identify the effects of education on earnings using an education reform that is either nationwide or region-specific. They use a change in the education system as an instrument. The compulsory schooling extension in Sweden, which occurred between 1948 and 1962 depending on the municipality, is used by Meghir and Palme (1999, 2005) to identify the effects of education on earnings. By adopting the IV selectivity model, they are able to treat education as a discrete variable. While the reform has a very strong effect on years of education at the lower end of the education distribution, the returns to education among wage-employed men aged between 16 and 75 years are found to be around 3 per cent with OLS and around 4 per cent with IV. These OLS and IV estimates are much lower than those in North America, and are not statistically different from each other. Meghir and Palme (1999, 2005) conclude that this is because the reform significantly changes the composition of those with higher years of education post-reform towards lower average ability. Using a similar identification strategy and specification, Aakvik et al. (2003) use the increase in compulsory schooling from 7 to 9 years in Norway during the years 1960 to 1972 to identify the effects of education on earnings among employed men aged between 38 and 47 years. They find that the IV estimates are around 10 per cent, larger than the OLS estimates of around 7.5 per cent. This supports the LATE interpretation, as in the case of the North American studies.

Meanwhile, Pischke and von Wachter (2008) estimate the impacts of education on earnings using the increase in compulsory schooling from grade 8 to 9 in West Germany that occurred between 1949 and the late 1950s depending on the state. Unlike other earlier studies with a similar research design, they find very low and close-to-zero returns to education among wage-employed workers aged between 15 and 65 years, which are also much lower than the OLS estimates (between 2 to 6 per cent). They also show that the results of close-to-zero returns are due to neither wage rigidity nor the apprenticeship training system. They suggest that the results could arise if labour market relevant skills are learnt earlier in Germany than in other countries. This conclusion is similar to the work of Oosterbeek and Webbink (2007) who study the returns to a one-year increase in vocational education caused by the vocational training reform in the Netherlands. By using a difference-in-difference method, they find that the impacts of an additional year of vocational education on earnings among employed men aged between 16 and 65 years are nearly non-existent.

Studies of developed economies that most closely relate to my analysis are those of France and the UK, using the nationwide changes in legislation on schooling ages to identify the effects of education. The changes in compulsory schooling laws in these countries concern either the school leaving age (SLA) or school leaving grade. They create exogenous variation in educational attainment by forcing individuals born after a certain year to stay in school longer, resembling increased years of compulsory schooling in Thailand.[9]

Grenet (2013) studies the returns to education in France, identified by a change in the minimum SLA from 14 to 16 years old in 1967. The paper uses the French Labour Force Survey during the years 1990 to 2002 and focuses separately on employed men and women aged between 28 and 58 years. Because the 1967 reform forces individuals who were born during and after 1953 to stay in school longer, Grenet (2013) employs the RD approach that allows the year of birth to exogenously affect educational attainment in a more flexible and discontinuous way for the cohorts before and after the reform. The paper finds the OLS returns to be around 7 per cent for men and 9 per cent for women, and very low and close-to-zero RD estimates for both men and women. This upward bias in the OLS estimates could be attributed to the fact that two additional years induced by the reform do not significantly change the educational qualifications of people at the bottom-end of the education distribution, who are affected by the reform.

For the UK, there are two important changes in the minimum SLA: from 14 to 15 years old in 1947, and from 15 to 16 years old in 1973.[10] Harmon and Walker (1995), the first paper measuring education effects resulting from both changes in the minimum SLA in the UK, use the Family Expenditure Survey (FES) between 1978 and 1986, and only focus on employed men. In this paper, the IV identification is achieved by including dummy variables for whether an individual is under the old (pre-1947 reform), middle (post-1947 and pre-1973 reform), or new (post-1973 reform) education system. Similar to earlier IV analyses especially of North America, they find a large negative bias in the OLS estimate. While the OLS estimate is around 6 per cent, the IV estimate is around 15 per cent. However, their

[9] Among the reforms used by the cited literature, the changes in minimum school-leaving age (SLA) in France and the UK are most similar to the increased years of compulsory schooling in Thailand. This is because (1) they create exogenous variation in educational attainment by forcing individuals born after a certain year to stay in school longer; (2) the reforms are nationwide; and (3) unlike other education reforms, the reforms in France and the UK do not directly affect an individual's choice of educational streams.

[10] The reforms in 1947 and 1973 affect those who were born during and after 1933 and 1958, respectively.

analysis cannot fully control for age effects due to the short duration of the data. With the FES data, it is only possible to observe older people under the old education system and younger people under the new education systems. In addition, they do not control for cohort. This is likely to play an important role in how changes in the minimum SLA affect educational attainment (Card, 1999).

Oreopoulos (2006b, 2008) corrects the cohort issue by, instead, using the RD framework.[11] He estimates the RD returns to education for employed workers using the UK's General Household Surveys (GHS) during the years 1984 to 1998. Despite focusing only on the first reform in 1947, the paper encounters the same problem of a limited age range due to the short year-span of the data. That is, the GHS data cover only older people under the pre-1947 education system and younger people under the post-1947 system. In order to make people under the different education systems comparable in terms of age, the paper therefore confines the sample to employed workers who were between 32 and 64 years old during the survey years. Oreopoulos (2006b, 2008) finds the RD returns to education to be around 7 per cent but statistically insignificant, from the OLS estimates which are around 6 per cent. While the estimated RD returns are identified only for specific sub-populations that increase their years of education in response to the reform, the paper claims that these so-called RD LATE estimates come close to the ATE estimates as these sub-populations represent the majority of the population.

Studying the same 1947 reform, Devereux and Hart (2010) use longer duration data from both the GHS during the years 1979 to 1998 and the New Earnings Survey Panel Dataset (NESPD) during the years 1975 to 2001, in order to have a wider age range (28 to 64 years) observed in the surveys, and thus to obtain more precise estimates.[12] With some additional controls, a different way of constructing the instrumental variable, and a longer period of data with better earnings information, they find the RD estimates to be around 3 per cent for the pooled sample, much lower than the OLS estimates of around 7 per cent.[13]

[11] I refer mainly to Oreopoulos (2008), which is a brief corrigendum of Oreopoulos (2006b)'s UK results. When necessary, I refer to Oreopoulos (2006b) for methodological details and background.

[12] Devereux and Hart (2010) differ from Oreopoulos (2006b, 2008), in terms of variables included as follows. First, the gender dummy variable is added to the pooled regression. Second, the reform variable is defined differently. Oreopoulos (2006b, 2008) sets the value of the reform instrument to 1 for individuals born in 1933 or after and 0 otherwise. On the other hand, in Devereux and Hart (2010), the reform variable equals 0.75 for individuals born in 1933, 1 for individuals born after 1933, and 0 otherwise. This is because the individuals affected by the law are those born after 1 April 1933, which is assumed to be 75 per cent of the 1933 cohort. Third, the main outcome variables in Devereux and Hart (2010) are (log of) hourly and weekly earnings, while those of Oreopoulos (2006b, 2008) are (log of) annual earnings.

[13] While focusing mainly on the education reform in France, Grenet (2013) also confirms that the RD

When estimating the returns to education for men and women separately, Devereux and Hart (2010) find them to be around 5 per cent for men and nearly zero (or even negative but statistically insignificant) for women. They conclude that these low returns suggest the importance of selection for those who drop out early.[14] An additional explanation is that the increased minimum SLA from 14 to 15 years old does not increase academic qualifications, and thus may not result in higher earnings. As discussed by Oreopoulos (2006b), if the RD LATE estimates converge towards the ATE estimates, these results would confirm that the ATE estimates are usually lower than the LATE estimates. By contrast, Buscha and Dickson (2012) argue that the low returns found in the UK represent an average return over a particular part of the life-cycle. They focus only on the effects of education on the earnings of individuals in their early 50s and they find slightly higher returns of around 5 per cent for the pooled sample of men and women.

Education reforms in developing economies

Unlike the aforementioned studies of developed countries which concentrate on the education systems or reforms that are related to legislation on compulsory education and schooling ages, the comparable studies of developing countries are quite diverse in terms of education reforms. The education reforms used in the latter literature range from increased public investment and/or expenditure on education to changes in compulsory schooling laws. In addition, there are substantially fewer studies on developing economies in this field. This could be because education reforms were implemented much later in these developing countries, compared to developed countries. Hence, the working life of individuals who are exposed to the reforms may have not been fully observed.

Most of the earlier studies on returns to education using an education reform focus on school construction and education subsidies. The most well known is probably the study of Duflo (2001), using school construction projects to estimate the causal return to education in Indonesia. More than 61,000 primary schools were constructed during the years 1973 to 1974 and 1978 to 1979. The school construction programme significantly increased enrolment rates among school-age children, and is expected to have a direct and exogenous impact on

returns is lower than the OLS in the UK. Using the 1973 reform, the paper finds the OLS returns to be around 10 per cent for men and 12 per cent for women, and much lower RD estimates of around 7 per cent for men and 6 per cent for women.

[14] Devereux and Hart (2010) assert that their results are in line with studies that measure the effects of education, using the 1947 reform, on other outcomes such as mortality rates, health, criminal behaviour, etc.

the educational attainment of the labour force. The intensity of the programme by region of birth and a variable indicating birth cohort are therefore used as instruments in the earnings equations. The paper separately examines the returns to education for wage- and self-employed male workers, aged between 23 and 45 years in 1995. Among wage-employed men, the IV estimates are found to be around 8 to 11 per cent, higher than the OLS estimates of around 8 per cent. On the other hand, the estimates of returns to education for self-employed men suggest an upward bias in the OLS estimates. While these IV estimates are not statistically different from the OLS estimates, Duflo (2001) argues that the difference arises from heterogeneous returns and that the IV estimates are a weighted average of the returns for individuals who are affected by the instruments differently.[15]

Similar to Duflo (2001), Simione (2012) observes that the large public investment in education and the elimination of racially separated schools in Mozambique after independence caused a remarkable increase in primary enrolment and average educational attainment. By focusing on employed individuals aged between 20 and 64 years in 2009, the paper finds high IV returns to education of around 16 per cent compared to the OLS estimates of around 4 per cent. Again, the IV estimates can be explained by heterogeneous returns to education across individuals. High returns to education are likely to occur among those who went to schools in urban areas, where the education boom was more pronounced. Exploiting a similar policy shift, Burger (2010) investigates the returns to education among formally employed South African black men using two changes in education policies in the 1990s, including the abolition of education restrictions for black people and over-aged learners. While he finds slightly higher IV estimates of around 12 per cent, compared to the OLS estimates of around 10 per cent, he also investigates the non-linearity in returns to education. The IV estimates exhibit convex returns to education for South African black men, which contradicts the conventional concavity assumption (Card, 1999). Concavity is confirmed in Tanzania, where an education reform changed the primary and middle schooling systems (Kerr and Quinn, 2010).[16]

Increases in public expenditure on education also occur in the form of education subsidies. Patrinos and Sakellariou (2011) examine returns to an additional year of education caused by

[15] Duflo (2001) highlights that the school construction project is a valid instrument only when it affects the quantity of education (educational attainment) but not the quality of education.

[16] Kerr and Quinn (2010) identify the effects of education on earnings by using an education reform in Tanzania. During the 1960s, middle school was eliminated and primary school extended to 7 years, replacing the earlier system of 4 years of primary school followed by 4 years of middle school. As a result of the reform, most children obtained 7 rather than 4 years of education.

the 1981 school voucher programme in Chile. By introducing public funding for some private schools, access to schools in Chile increased significantly, as reflected by a sharp decline in the dropout rates for basic education in 1982. The voucher reform is therefore expected to increase average educational attainment for certain birth cohorts. With the instrument indicating whether an individual was in school in 1981, and thus was affected by the reform, they find the returns to education for wage-employed men between 18 and 45 years old to be around 10 to 12 per cent for the IV estimates, compared to the OLS estimates of around 5 to 8 per cent. The higher IV returns potentially represent the returns to education of middle class students living in urban areas, as they were likely to be affected most by this reform.

More recent studies of developing economies are closer to those of developed economies. They obtain exogenous variation in educational attainment from education reforms related to compulsory schooling. The results are very diverse. Khitarishvili (2010) estimates the returns to education among wage-employed workers in Georgia between 2000 and 2004 using variation in compulsory schooling years across time and regions caused by the Soviet education reforms from 1958 to 1964. In a robustness check, he uses different reforms related to the fee abolition for upper secondary education. The returns to education are found to be very low, around 1 to 3 per cent for the OLS estimates and around 0 to 2 per cent for the IV estimates. These low returns could be explained by (i) increased educational attainment moving people to the public sector, which does not pay well, and (ii) declining quality of education over time. The latter explanation suggests that changes in the quality of education could in fact result in invalid instruments, and thus biased estimates.

In contrast to Khitarishvili (2010), some studies that use education reforms confirm high returns to education in developing economies and a downward bias in the OLS estimates. Fang et al. (2012) use the implementation of 9-year compulsory education in China in 1986 to identify the effects of education on earnings. While this law raised overall educational attainment by about 0.8 years, the effects varied considerably by gender and location due to family institutions and insufficient resources to fully implement the law in some regions. They therefore estimate the returns to education for men vs women and urban vs rural areas separately. The OLS returns are around 9 per cent for all population groups. However, the IV estimates are around 20 per cent for the overall and vary greatly across population sub-groups. The downward bias is found to be very large among men, workers in rural areas, and workers living in coastal cities. By comparing this reform to those of the US or the UK, they attribute the higher returns in China to diminishing returns to education.

With comparable instruments, similar results of high IV returns to education are also found in Turkey. The 1997 education reform in Turkey increased compulsory education from 5 to 8 years for individuals in the 1986 birth cohort and after. Under the RD framework, Aydemir and Kirdar (2013) use this reform as an instrument for educational attainment in an earnings equation estimated for men in urban areas. They conclude that the RD returns to education for men in urban areas are around 20 per cent, much higher than the OLS estimates of around 3 to 4 per cent. This is, again, attributed to the heterogeneity in returns to education and that the returns to education are higher for those at the lower end of the education distribution who were affected by the reform.[17] Note that in these studies, the changes in compulsory schooling years occurred relatively recently. This implies that the studies could not observe people who were affected by the reform at all ages and that they focus on younger people.

Closely related to my study, Leckcivilize (2013) uses the intensity of the compulsory education reform during the period 1960 to 1978 to identify the returns to education for wage earners in Thailand over the years 1994 to 2009. As discussed in the Background and Data chapter (Chapter 2), the Thai compulsory education reform raised the compulsory years of schooling from 4 years of lower primary to 6 years of primary education. The reform was firstly and partially adopted in 1960 and was subsequently implemented nationwide in 1978. More specifically, between 1960 and 1977, public schools that had the capacity to provide 6-year primary education free of charge could introduce this reform. As a result, Leckcivilize (2013) claims that, while the first cohort who are affected by the 1978 reform are those born in 1967, the earlier-than-1967 cohorts have also been exposed to the reform initiative since 1960. The paper computes the treatment intensity for these older cohorts with strong assumptions on sub-district populations and migration, and uses it as an instrument for education.[18] The paper finds zero returns to education for men and around 9 per cent

[17] An additional conclusion from this paper is that the reform also has strong spill-over effects on other levels of education.

[18] With limited historical information on the population at the sub-district level, migration, the number of schools within a district, and specific schools that advanced the reform, Leckcivilize (2013) calculates the treatment intensity by using the district population shares in 1993 to compute the province-specific probability of being exposed to the reform. In doing so, the paper implicitly assumes that (i) there is no change in the sub-district population shares between 1960 (the period of the reform initiative) and 1993; (ii) all schools in the same district advanced the reform at the same time; (iii) every wage earner works in the province he went to school (since there is no information on birth or schooling place in the LFS); and (iv)

returns for women, which are substantially lower than the OLS returns of around 11 per cent for both gender. However, this could be the result of selection into wage employment particularly among women. Furthermore, the instrument might be invalid as discussed in Footnote 18, in particular as individuals are free to choose which school to attend.

3.2.2.2 Impacts of education on employment outcomes

Studies on how schooling affects employment outcomes using an education reform as an instrument (including the above-mentioned studies) have been relatively underrepresented, in both developed and developing countries. Following Angrist and Krueger (1991), Del Bono and Galindo-Rueda (2004) show that the month of birth can be used as an instrument for education in England and Wales, as it is correlated with academic qualification due to school entry and minimum SLA. They find that additional academic qualifications have important employment consequences in England and Wales, especially for women. It increases the probabilities of participating in the labour market and of being employed by around 20 per cent for women, and around 10 per cent for men. However, the high employment impacts could be due to the fact that the paper examines the effects of additional qualifications, which imply an increase of more than one year of education. In addition, as discussed earlier, season of birth may not be a valid instrument as it may be correlated to early childhood development and some household characteristics that have direct impacts on labour market outcomes.

Using variation in compulsory schooling laws across states in the US and Canada and the increased minimum SLA in the UK as discussed earlier, Oreopoulos (2006b) finds that education lowers the likelihood of individuals being unemployed in all three countries by around 3 per cent. The IV estimates of the employment effects are not significantly different from the OLS estimates. However, a US state-specific study with a similar set of RD instruments shows no significant effects of education on employment in California and Texas (Dobkin and Ferreira, 2010). These results are also robust to specific age, race, and gender sub-populations. One possible explanation is that the effects of increased educational attainment are offset by poorer academic performance among those who are affected by the law as they are at a relatively younger age. Their results are similar to those found in the German labour market. Pischke and von Wachter (2008) examine the effects of additional years of education as a result of the increased compulsory schooling years on being employed

individuals did not self-select into schools which advanced the reform.

and being self-employed in Germany. The employment impacts of an additional year of education are found to be low: around a 0.3 per cent increase in the probability of being employed and a 0.1 per cent decrease in the probability of being self-employed.[19]

On the other hand, a few studies on employment impacts of education in developing economies find significant positive impacts of education on employability and employment in some certain sectors. While finding small impacts of education on earnings in Georgia using the Soviet education reform, Khitarishvili (2010) concludes that education is nonetheless an important factor influencing employability for both men and women, particularly in the public sector. Although not estimating the effects of education on employment directly, the study compares the standard IV estimates of returns to education to estimates that correct for sample selection bias. The higher sample selection correction estimates suggest the potential role of education in the probability of being employed.[20] Using exogenous variation in education due to the two aforementioned changes in education policy in South Africa in the 1990s, Burger (2010) also confirms the positive impacts of education on being formally employed for black men both directly by running a probit regression and indirectly by correcting for sample selection bias.

Pekkarinen and Pellicer (2013) examine the impacts of an additional year of schooling on more detailed employment outcomes, including the sector of employment and occupation, in Tunisia during the 2000s.[21] They obtain exogenous variation in educational attainment from a government policy during the 1970s, which cut costs by restricting the number of students in the final year of primary school that were promoted to secondary school. The study finds that education increases the probability of being employed, particularly in the public sector and in relatively low-skill clerical occupations. The IV estimates of the employment impacts are around 5 and 1 per cent for public administration and the private sector, respectively, compared to the OLS estimates of around 3 and 0.5 per cent. The higher IV estimates suggest relatively higher employment impacts of education for those who are affected by the policy change. Pekkarinen and Pellicer (2013) emphasise that these people are likely to have

[19] Pischke and von Wachter (2008) attribute the results to similar explanations as those for earnings outcomes (see p. 28).

[20] Khitarishvili (2010)'s further investigation on sector-specific employment is done by controlling for sector of work in the sample selection model.

[21] The sectors of employment are divided into public administration, public companies, private companies, and multi-national companies. Occupations included in the analysis range from management, scientist, technician, to secretary. The empirical analysis is conducted by running a linear probability model with a binary outcome variable indicating the status of being in each of the sectors or occupations.

been at the margin of passing to secondary school, and therefore to have relatively poorer academic performance.

To complement the existing literature, this chapter examines the impacts of educational attainment on both earnings and the sector of employment – specifically non-agricultural employment – in Thailand, providing more evidence of the effect of education in a developing economy. I identify the causal impact of education by using the change in compulsory schooling years as an instrument within the RD framework. As highlighted by some of the earlier studies, the estimates from this method represent the effects among those whose levels of educational attainment change due to the education reform. Therefore, these estimates of earnings and employment impacts of education may be potentially used to evaluate the effectiveness of the reform.

3.3 Regression Discontinuity Framework

This section describes the identification strategy used to estimate the impacts of education on the two labour market outcomes (earnings and sector of employment), namely the regression discontinuity (RD) framework. The underlying framework and the assumptions required for the RD identification strategy are described. Subsequently, the section discusses how to empirically estimate the coefficients of interest with the RD framework.

3.3.1 RD Identification Strategy

Consider the regression model characterising the causal relationship between years of education, S_i, and an outcome variable, Y_i,

$$Y_i = \beta_1 + \beta_2 \cdot S_i + \beta_3 \cdot \mathbf{X}_i + \varepsilon_i \tag{3.1}$$

where $\mathbf{X}_i$ is a vector of other observable individual controls and ε_i is an idiosyncratic error. The outcome of interest, Y_i, in this chapter includes log of hourly earnings and the sector of work. The sector of work variable is a dummy variable, taking a value of one if an individual works in the non-agricultural sector and zero otherwise.

This chapter is interested in estimating β_2, the effect of years of education on the outcome variable. Estimating Equation (3.1) by OLS could lead to a biased estimate of β_2 as a result of the correlation between years of education and any unobserved outcome-determining factors or measurement error in education. For instance, the positive correlation between unobserved ability and educational attainment (ability and signalling biases) is expected to bias the estimate upwards (Griliches, 1977; Blackburn and Neumark, 1995; Card, 2001). On the other hand, measurement error in education can bias the estimate downwards (towards zero) (Griliches, 1977; Ashenfelter and Krueger, 1994). Alternatively, a downward bias could also arise if education is compensatory for earnings capacity for some individuals (Ashenfelter et al., 1999; Walker and Zhu, 2001; Harmon et al., 2003). That is, highly able workers forgo relatively more income in the labour market when they spend more years in education.

To identify the causal effects of education on the outcomes of interest, I use the RD approach (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). The RD framework arises naturally from the discontinuity in years of education caused by the change in compulsory education that was implemented in 1978 (Education Reform Act of 1978).

As discussed in the Background and Data chapter (Chapter 2), according to this reform, pupils born before the cut-off year 1967 were subject to 4 years of compulsory lower primary education, while those born in the year 1967 and later were subject to 6 years of compulsory primary education. The change in compulsory education had two main impacts on years of education. First, it directly forced those who would have dropped out after lower-primary education had they been born before 1967 to stay in school for two more years. Second, it is expected to have indirectly raised the educational attainment of those who would have completed at least the primary level regardless of their year of birth. These relatively more educated individuals experience lower average costs as the two additional years are offered free of charge. In addition, these individuals are likely to increase their educational attainment to maintain their educational advantage over the relatively less educated (Meghir and Palme, 2005). Therefore, the difference in year of birth can be used to identify the effects of education on the outcomes of interest as it is expected to generate exogenous variation in years of education.

Let R_i be the year of birth of individual i. It is known as a running (or an assignment) variable, which is assumed to exogenously affect (increase) years of education in a discontinuous way when it exceeds a fixed threshold, $r_0 = 1967$. Given this relationship between the year of birth and potential years of education, the population can be partitioned into five sub-groups: always takers, never takers, compliers, defiers, and indefinites (Angrist et al., 1996) as displayed in the following table.[22]

Text Table 3.1: Population Classified by Potential Years of Education (S_i)

Sub-populations	Conditions
Always takers (AT)	$\lim_{r\to r_0^+} S_i(r) = \lim_{r\to r_0^-} S_i(r) \geq 6$
Never takers (NT)	$\lim_{r\to r_0^+} S_i(r) = \lim_{r\to r_0^-} S_i(r) < 6$
Compliers (C)	$\lim_{r\to r_0^+} S_i(r) > \lim_{r\to r_0^-} S_i(r)$
Defiers (DE)	$\lim_{r\to r_0^+} S_i(r) < \lim_{r\to r_0^-} S_i(r)$
Indefinite (I)	$\{\mathrm{AT} \cup \mathrm{NT} \cup \mathrm{C} \cup \mathrm{DE}\}^c$

Given the classification, the RD assumptions required for the identification of the causal average effect of education on the outcome variable for the compliers are as follows.

Assumption A (Hahn et al. (2001); Imbens and Lemieux (2008); Lee and Lemieux (2010))

A1 Fuzzy Regression Discontinuity (FRD): $\lim_{r\to r_0^+} E[S_i \mid R_i = r]$ and $\lim_{r\to r_0^-} E[S_i \mid R_i = r]$ exists and $\lim_{r\to r_0^+} E[S_i \mid R_i = r] - \lim_{r\to r_0^-} E[S_i \mid R_i = r] \neq 0$.

A2 Local Smoothness: $E[\mathbf{X}_i \mid R_i = r]$ and all other unobserved outcome-determining factors except for education are continuous in r at $r = r_0$.

A3 Monotonicity: $\lim_{r\to r_0^+} E[S_i \mid R_i = r] \geq \lim_{r\to r_0^-} E[S_i \mid R_i = r]$ for all i.

A4 Existence of Compliers at Threshold: $\lim_{r\to r_0^+} E[S_i \mid R_i = r] > \lim_{r\to r_0^-} E[S_i \mid R_i = r]$ for some i.

Note that the threshold value of year of birth, r_0, is 1967, and that around the threshold refers to those who were born around 1967. Under Assumption **A**, for a very small neighbourhood, the cut-off year acts like a local instrumental variable (Lee and Lemieux, 2010). It can be

[22] The notation in the table captures how behaviour changes around the cut-off year of birth, 1967. For example, always (never) takers do not change their years of education as a result of the reform and always attain more (less) than 6 years of schooling.

used to identify average effects for the compliers, and thus the average effects of education are referred to as the LATE (Imbens and Angrist, 1994).[23]

Assumption **A1** requires that there is a discontinuity in years of education around the cut-off year, r_0. I consider this a fuzzy RD (Hahn et al., 2001) for two reasons.[24] First, very few people born at and after 1967 were not subject to the reform due to some exceptions and delays in the law's implementation in a few remote areas (as discussed in Chapter 2, p. 10). Second, while the education reform extended compulsory education from 4 to 6 years, the average years of education among the compliers may have increased by less than 2 years. This is because the indirect effects of the reform on more educated individuals could be less than the direct effects on those who would not have completed primary education had there been no reform. However, the discontinuity could remain endogenous for those compliers who would have attained at least primary education regardless of the education regime. Because the reform does not force them to change their attainment, their decision over additional years of education may be correlated with their unobserved ability. In other words, amongst those compliers, the high-ability individuals would respond to the reform by increasing their educational attainment more. As the estimates are the weighted average treatment effects for the compliers, the degree of bias would depend on the proportion of compliers who would have completed primary education anyway.

Assumption **A2** is a smoothness condition. It assumes that individuals do not have precise control over their years of birth around the cut-off year.[25] In other words, it requires that whether the year of birth crosses the cut-off year has no discontinuous impact on the outcome variable, except by influencing years of education. This is equivalent to saying that the instrumental variable is uncorrelated with other observed and unobserved factors affecting the outcome variable.

Assumption **A3** implies that the year of birth exceeding the cut-off year can only cause individuals to weakly increase their years of education. It means there must be no defiers and indefinites in a neighbourhood around the threshold. That is, there is no instance of

[23]From the population partition, compliers refer to the sub-group of the population that raised their educational attainment as a result of the education reform. They would increase their years of education if they were born during and after the year 1967.

[24]In the sharp RD framework, $\lim_{r \to r_0^+} E[S_i \mid R_i = r] - \lim_{r \to r_0^-} E[S_i \mid R_i = r] = 2$ as individuals born in and after 1967 would be strictly forced to study for two more years. Therefore, all individuals are compliers.

[25]Imbens and Lemieux (2008) argue that it is not usually reasonable to assume no precise control only around one particular value of the running variable. In this case, they suggest the stronger assumption that the smoothness holds in all values of r.

an individual who would have reduced their educational attainment, had he been born in or after 1967. This assumption cannot be tested directly, but needs to be assessed (Imbens and Angrist, 1994). I argue that it is likely to be met for four main reasons. First, the Education Reform Act 1978 was implemented nationwide. Second, there were some exceptions made for people with extreme difficulty in attending school regularly. However, these exceptions remained the same before and after the reform. Therefore, holding other factors constant, people in the exception categories are never takers rather than defiers, as they would not have completed the primary level regardless of the education policy regime. Third, the implementation lag occurred only in very few remote areas and there is no specific reason to believe that there exist defiers in these areas. Fourth, highly educated individuals had two additional years of education free of charge after the reform. There is no reason to expect them to reduce their educational attainment.

Assumption **A4** requires that there exist compliers.

Under assumptions **A1** to **A4**, Hahn et al. (2001); Imbens and Lemieux (2008); Lee and Lemieux (2010) show that the effects of education on the outcome variables for the compliers are identified as

$$\beta_2 = \tau_{FRD} \quad = \quad \frac{\lim_{r \to r_0^+} E\left[Y_i \mid \mathbf{X}_i,\, R_i = r\right] - \lim_{r \to r_0^-} E\left[Y_i \mid \mathbf{X}_i,\, R_i = r\right]}{\lim_{r \to r_0^+} E\left[S_i \mid \mathbf{X}_i,\, R_i = r\right] - \lim_{r \to r_0^-} E\left[S_i \mid \mathbf{X}_i,\, R_i = r\right]} \tag{3.2}$$

which is equivalent to the LATE estimate in the IV approach (where the binary instrument is the indicator for $R_i \geq r_0$) (Angrist and Imbens, 1995). It is important to note that RD and IV estimates highlight the importance of heterogeneous returns to education in two ways. First, with the standard set of assumptions (**A1** to **A4**), the difference between the RD/IV estimates and the ATE reflects heterogeneous effects of education between compliers and non-compliers. Second, the RD/IV approach also allows for heterogeneity in returns among the compliers. In the presence of heterogeneous returns across levels of education, the RD/IV estimates are the weighted LATE for the compliers at each of the levels of education (Angrist and Imbens, 1995; Lee and Lemieux, 2010). Nonetheless, the RD/IV approach provides only one estimate which is the aggregated effect of education for all compliers. As a result, the RD/IV estimates remain uninformative about the shape of the education-earnings profile and the average returns for the population (that is, the ATE).[26]

3.3.2 Estimating the Effects of Education in the RD Design

A simple way of estimating the effect of education on the outcome variable in the RD design, τ_{FRD} as described in Equation (3.2), is to use polynomial regressions. This is also appropriate because my running variable, year of birth, is discrete. When the running variable is continuous, the RD approach compares the outcomes for the observations with the running variable just above and just below the cut-off point. This can be done non-parametrically by choosing the bandwidth and estimating local linear regressions on both sides of the cut-off point (Imbens and Lemieux, 2008; Lee and Lemieux, 2010; Gelman and Imbens, 2014). However, when the running variable is discrete or ordered categorical (for example, year of birth in my case, or age recorded in months or years), it is impossible to identify the parameter of interest non-parametrically as there are no observations around a very small neighbourhood of the cut-off point. In this case, Lee and Card (2008) suggest that the causal effects can only be identified parametrically with a functional form for the relationship between the running variable and the outcome of interest that is being specified. In other words, instead of choosing the bandwidth and estimating the causal effects only from the observations around the neighbourhood of the discontinuity threshold, the polynomial regressions are estimated for all observations from each side of the discontinuity threshold, and the estimates at the points very close to the cut-off can be obtained from their extrapolations.

As this is a fuzzy RD design, years of education, S_i, as a function of the year of birth can be written as

$$S_i = \gamma + \delta \cdot Z_i + \sum_{h=1}^{H} \phi_h \cdot (R_i - r_0)^h + v_i \tag{3.3}$$

where $Z_i = \mathbb{1}\{R_i \geq r_0\}$ indicates whether the running variable exceeds the cut-off value, $r_0 = 1967$ is the threshold year of birth as before, H is a polynomial order of $R_i - r_0$, and v_i is an error term which is uncorrelated with R_i.

First, consider $\lim_{r \to r_0^+} E[Y_i \mid \mathbf{X}_i, R_i = r] - \lim_{r \to r_0^-} E[Y_i \mid \mathbf{X}_i, R_i = r]$, which is the numer-

[26] In fact, the RD/IV estimates can recover the average returns for the population but with stronger assumptions. This, together with the shape of the education-earnings profile, will be discussed in the next chapter.

ator of Equation (3.2). Each term can be estimated by running a polynomial regression for the outcome variable (the reduced form equation) on each side of the discontinuity threshold

$$
\begin{aligned}
Y_i &= \alpha_r + \sum_{k=1}^{K} \pi_{r,k} \cdot (R_i - r_0)^k + \theta_{r,Y} \cdot \mathbf{X}_i \qquad \text{for } R_i \geq r_0 && (3.4) \\
Y_i &= \alpha_l + \sum_{k=1}^{K} \pi_{l,k} \cdot (R_i - r_0)^k + \theta_{l,Y} \cdot \mathbf{X}_i \qquad \text{for } R_i < r_0 && (3.5)
\end{aligned}
$$

where K is the polynomial order of $R_i - r_0$ for the outcome variable regressions. The numerator of Equation (3.2) can be estimated by $\hat{\alpha}_r - \hat{\alpha}_l$.[27]

Similarly, $\lim_{r \to r_0^+} E[S_i \mid \mathbf{X}_i, R_i = r] - \lim_{r \to r_0^-} E[S_i \mid \mathbf{X}_i, R_i = r]$ which is the denominator of Equation (3.2), can be estimated by running a polynomial regression for years of education (the first-stage equation) on each side of the discontinuity threshold

$$
\begin{aligned}
S_i &= \gamma_r + \sum_{h=1}^{H} \phi_{r,h} \cdot (R_i - r_0)^h + \theta_{r,S} \cdot \mathbf{X}_i \qquad \text{for } R_i \geq r_0 && (3.6) \\
S_i &= \gamma_l + \sum_{h=1}^{H} \phi_{l,h} \cdot (R_i - r_0)^h + \theta_{l,S} \cdot \mathbf{X}_i \qquad \text{for } R_i < r_0 && (3.7)
\end{aligned}
$$

where H is the polynomial order of $R_i - r_0$ for the years of education regressions. The estimate of the denominator of Equation (3.2) is given by $\hat{\gamma}_r - \hat{\gamma}_l$.[28] Therefore,

$$
\hat{\beta}_2 = \hat{\tau}_{FRD} = \frac{\hat{\alpha}_r - \hat{\alpha}_l}{\hat{\gamma}_r - \hat{\gamma}_l} \tag{3.8}
$$

Note that I allow for the relationships between Y_i and $(R_i - r_0)$ and between S_i and $(R_i - r_0)$ to have different polynomial coefficients on either sides of the discontinuity threshold. As long as the order of polynomial and the data window for each side of the discontinuity threshold are the same for the first- and second- stage outcomes (that is $K = H$), the presented causal effect of education on the outcome variable for the compliers is equivalent to the IV estimation of the following regression.[29]

---

[27] Alternatively, the pooled regression can be written as:
$Y_i = \alpha_l + (\alpha_r - \alpha_l) \cdot Z_i + \sum_{k=1}^{K} \pi_{l,k} \cdot (R_i - r_0)^k + \left[\sum_{k=1}^{K} (\pi_{r,k} - \pi_{l,k}) \cdot (R_i - r_0)^k\right] \cdot Z_i + (\theta_{r,Y} - \theta_{l,Y}) \cdot \mathbf{X}_i$.

[28] Alternatively, the pooled regression can be written as:
$S_i = \gamma_l + (\gamma_r - \gamma_l) \cdot Z_i + \sum_{h=1}^{H} \phi_{l,h} \cdot (R_i - r_0)^h + \left[\sum_{h=1}^{H} (\phi_{r,h} - \phi_{l,h}) \cdot (R_i - r_0)^h\right] \cdot Z_i + (\theta_{r,S} - \theta_{l,S}) \cdot \mathbf{X}_i$.

[29] Lee and Card (2008) also highlight that the parameter of interest is identified only when the order of polynomial does not exceed the number of distinct values of the running variable.

$$Y_i = \kappa_1 + \kappa_2 \cdot S_i + \sum_{k=1}^{K} \kappa_{3,k} \cdot (R_i - r_0)^k + \left[\sum_{k=1}^{K} \kappa_{4,k} \cdot (R_i - r_0)^k \right] \cdot Z_i + \kappa_4 \cdot \mathbf{X}_i \tag{3.9}$$

with $Z_i = \mathbb{1}\{R_i \geq r_0\}$ as an excluded instrument for S_i. The two-stage least squares (2SLS) estimate of κ_2 is then identical to τ_{FRD} in Equation (3.2).

3.3.3 Choice of Polynomial Order

In the case of a discrete running variable, the equivalent to bandwidth choice in the non-parametric estimation is the choice of the order of polynomial regression (Lee and Card, 2008; Lee and Lemieux, 2010). First, it is important to test whether the selected polynomial regressions are well specified. Second, it is also common practice to report a number of specifications to see to what extent the results are sensitive to the order of polynomial.[30] Lee and Card (2008) show that when the error term is normally distributed and homoskedastic, the polynomial model can be tested using a simple goodness-of-fit statistic (GOF_1) as follows.

$$\mathrm{GOF}_1 = \frac{(SSE_R - SSE_{UR})/(J-K)}{(SSE_{UR})/(N-J)} \sim F\left(J-K,\ N-J\right) \tag{3.10}$$

where SSE_R is the estimated error sum of squares of the restricted model and SSE_{UR} is the estimated error sum of squares of the unrestricted model. The unrestricted model regresses Y_i on a full set of dummy variables for the J values of the running variable, while the restricted model adopts a lower polynomial order. The statistic is distributed as $F\left(J-K,\ N-J\right)$. J is the number of values taken by the running variable, K is the number of parameters of the restricted model, and N is the number of observations. The polynomial function is too restrictive if the statistic exceeds the critical value.

In the case of a heteroskedastic error term, the goodness-of-fit test (GOF_2) can be obtained from a polynomial regression with a full set of dummy variables for the J values of the running variable, and testing whether the dummy variables in the set are jointly significant. In this case, the statistic is distributed as a Chi-Square distribution with $J-K$ degrees of freedom (that is, $\mathrm{GOF}_2 \sim \chi^2\left(J-K\right)$).[31]

[30] For the average treatment effects, a graphical presentation of the discontinuity is also helpful and informative (Lee and Lemieux, 2010). The discrete running variable provides a natural way of graphing the means of the outcome variables, Y_i, as well as the treatment variable, S_i, for each distinct value of the running variable.

[31] Lee and Card (2008) further suggest that the standard errors of the estimators in the micro-data model

3.3.4 Control Variables

I include other outcome-determining factors, $\mathbf{X}_i$, in the regression model for three reasons. First, it is important to check the validity of the local smoothness assumption (**A2**).[32] That is, the year of birth exceeding 1967 has a discontinuous impact on the outcome variable only through years of education. Second, even if the local smoothness assumption holds, observations to the left and the right of the discontinuity threshold may differ in these controls in a continuous way. In a particular year, individuals born in and after the year 1967 are likely to have less experience, be single, and have fewer children than those born before the year 1967. As these variables potentially affect the earnings process and the decision on the sector of work, including them can help to reduce the bias that may arise from their continuous correlation with year of birth. Third, accounting for additional controls may reduce the variance, and thus increase the precision of the estimates (Frölich, 2007).

For the earnings model, the control variables included are age, age squared, hours of work per week, marital status, gender, rural-urban location, province of residence, and year dummies. In addition to these variables, the number of children and elderly as well as spouse's earnings are added to the model of sector choice.

3.4 The Data and Descriptive Statistics

This section describes the nature of earnings, sectors of work, and education in Thailand from the LFS data, as well as the variables used in the empirical estimation of the causal effects of education on earnings and sector of work under the RD framework. First, I discuss the criteria for the selection of the sample used in the estimation. Second, descriptive statistics of the main variables used in this chapter are presented in order to provide an overview of the Thai labour market and the potential role of education in influencing earnings and sectors of work.

---

should be clustered according to the discrete running variable, as the polynomial regressions introduce a group structure in the standard errors.

[32] Van Der Klaauw (2008) interprets the inclusion of observed covariates as "the test for an imbalance in relevant characteristics".

3.4.1 Data Source

This chapter draws upon the LFS spanning 1985 to 2000. The sample of interest is confined to employed workers, whose ages range from 20 to 40 years old during the survey years of 1985 to 2000, for two reasons.

First, employed workers are the focus as the outcomes of interests are earnings and the decision on working in a particular sector. Note that the sample of analysis for earnings is smaller than that for the sector of work. This is because some employed workers, whose sector information is available, work for their family and are unpaid. Second, the sample includes workers whose ages were from 20 to 40 years old during the selected survey years because they were more likely to be affected by my key variable, the Education Reform Act 1978.[33] Most Thai workers have completed education by the age of 20, and this is therefore taken as the age of entering the labour market. On the other hand, capping the age at 40 helps to capture the effects of the reform. This is because persons aged above 40 years old during the selected survey years were in school much earlier than when the reform was implemented.[34]

3.4.2 Descriptive Statistics

Earnings Model

The analysis of the average effects of education on earnings uses the LFS information on paid-employed workers, aged between 20 and 40 years during the survey years 1985 to 2000. Their descriptive statistics are presented in Table 3.1. Real earnings and educational attainment moved in the same direction over time. For the pooled sample during the period from 1985 to 1996, average real earnings increased rapidly, by 4.6 per cent per annum, while average educational attainment increased by 1.4 years from 5.8 to 7.2 years. Average real hourly earnings for the pooled sample peaked in 1997 before moderately declining due to the financial crisis and remaining stable until 2000. During and after the crisis, average years of education among paid-employed workers continued to increase and reached 8 years of education by the end of the 1990s.

---

[33]This results in a sample of paid workers who were born during the years 1945 to 1980, which provides observations 22 years before 1967 (the first cohort under the 1978 reform) and 13 years after 1967.

[34]For instance, a worker whose age was 50 years old in the survey year 1985 (2000) was born in 1935 (1950), and therefore, was not subject to the new compulsory schooling law.

Considering workers by gender, the growth in average real earnings was faster for men while the increase in educational attainment was larger for women. Between 1985 and 1996, average real earnings rose by 4.8 and 4.3 per cent per annum and average educational attainment increased by 1.3 and 1.4 years, for men and women respectively. Despite having higher average educational attainment, female paid-employed workers earned slightly less per hour than male paid-employed workers throughout this period.

In addition to real hourly earnings and educational attainment, Table 3.1 also summarises other variables included in the estimation, such as age, percentage of men, percentage of paid-employed workers living in urban areas, hours of work per week, and proportion of workers in the non-agricultural sector. In 1985, around two-thirds of the paid-employed workers were men, and their relative proportion declined to around 60 per cent in 2000. Throughout the period of interest, male and female paid-employed workers also differed in characteristics other than their educational attainment. First, the average age was slightly higher among men than women. Second, among women, the shares of workers in urban areas and in non-agriculture are greater than among men, and these gaps persisted throughout the 16-year period. Lastly, paid-employed women worked fewer hours per week, although this difference narrowed over time.

Sector Model

Table 3.2 summaries the characteristics of employed workers, both paid and unpaid, aged between 20 and 40 years during the years 1985 to 2000. These observations are included in analysing the decision on sector of employment. The number of observations for this analysis is larger than that of the earnings model, as there exist workers who were employed as unpaid family workers. The reason for including unpaid family workers in the sector of work analysis is that they are considered economically active and they contribute to a household farm or business.[35]

[35] Although selection into paid- and unpaid-employment is not the focus of this chapter, it is interesting to note a few differences in the characteristics of workers in these two employment types by comparing the descriptive statistics of paid-employed workers (Table 3.1) to those of all employed workers (Table 3.2). First, a lower average age and fewer years of education among all employed workers indicate that unpaid family workers tend to be slightly younger and less educated than paid-employed workers. Second, while the proportion of non-agricultural workers among all employed workers increased significantly throughout the period, it was substantially less than that among paid-employed workers especially during the beginning of the period. This implies that most of the unpaid workers are engaged in agriculture. Third, the higher proportion of women among all employed workers, suggests that women are more likely to work for their families without pay.

The information on sectors of work shows an important structural change in the labour market in Thailand. Among all employed workers (both paid and unpaid), aged from 20 to 40 years, the proportion of non-agricultural employment increased rapidly from 36 per cent in 1985 to 56 per cent in 2000. Their average educational attainment also improved significantly from 5.2 years in 1985, which was below the 6-year primary level, to 7.7 years in 2000. During the same period, the average age of employed workers increased by around 1 year, while average hours of work per week declined marginally. Meanwhile, the proportion of employed workers in urban areas increased by 6 percentage points, from 27 per cent in 1985 to 33 per cent in 2000.

Non-agricultural and agricultural workers differed greatly in many characteristics including educational attainment, rural-urban location, marital status, and other household characteristics. First, the average educational attainment of workers improved for both sectors, with average years of education being much higher in non-agriculture compared to agriculture. The difference in years of education between workers in the two sectors remained stable at around 3 years over the entire period between 1985 and 2000. Second, while well above 60 per cent of non-agricultural workers resided in urban areas, only around 7 per cent of agricultural workers lived in urban areas. Lastly, non-agricultural workers had a smaller number of children and elderly on average, and their average spouse's earnings were substantially higher than that of agricultural workers. The difference in average spouse's earnings also increased over time.

3.4.3 Data Summary

These summary statistics show the changes in earnings and education attainment, as well as the structural change in employment between 1985 and 2000. Educational attainment improved continuously. At the same time, average real hourly earnings increased rapidly and a large share of the employed labour force was reallocated to the non-agricultural sector. These stylised facts draw attention to the role of human capital, measured by educational attainment, in explaining earnings growth and the structural change in the Thai labour market. While educational attainment, real hourly earnings, and employment in the non-agricultural sector developed in the same direction, the magnitudes of these changes differ significantly across gender. This, in addition, suggests potential heterogeneity in the effects of education on earnings and sector of work between men and women.

3.5 Empirical Results

This section estimates the effects of education on earnings and sector of employment in Thailand between 1985 to 2000. The impacts of education are estimated both for the pooled sample and for men and women separately. First, the appropriate polynomial order for the RD regressions is determined. Second, to test the discontinuity assumption, I discuss the first-stage regression estimates and present the discontinuity in educational attainment across birth cohorts in a graphical format. Third, I analyse the impacts of education on the two outcomes of interest using the RD framework, and discuss possible underlying explanations. Finally, I provide robustness checks.

3.5.1 Specification: Choice of Polynomial Order

As discussed in Sub-section 3.3.3, I use the goodness-of-fit test allowing for a heteroskedastic error term to select the polynomial order of the RD regressions. To do so, I regress my outcome variables, log of hourly earnings and a non-agricultural sector dummy, on a full set of dummy variables for all birth cohorts and then add higher-order terms of the normalised distance from the cut-off year of birth (that is, $R_i - 1967$) until all birth cohort dummies become jointly insignificant.

Table 3.3 presents the p-values from the Chi-Square tests of the (joint) statistical significance of the birth cohort dummies with the first (linear) to the fourth-order polynomials of the distance from the cut-off year of birth, 1967. That is, a p-value greater than the significance level, α, implies that the selected polynomial order explains the data well. These regressions are estimated for the pooled, male, and female samples respectively (Tables 3.3a, 3.3b, and 3.3c). The first three columns show the p-values for the earnings regressions with different sets of additional controls, while the last three columns show the corresponding results for the non-agricultural sector regressions.

From the Chi-Square test results, the second- and higher-order polynomial functions of the distance from the cut-off year of birth effectively fit the data. In particular, for all regressions with additional controls (Columns 2 and 3 for the earnings regressions, and Columns 5 and 6 for the non-agricultural sector regressions), the coefficients on the cohort dummies are statistically insignificant at the 10 per cent significance level after the second- and higher-polynomial orders are included.[36] For the regressions without any control (Column 1 for the earnings regressions and Column 4 for the non-agricultural sector regressions), the second-order polynomial sufficiently explains the data for the pooled and male samples at the 5 and 10 per cent significance levels respectively. But it may be too restrictive for the earnings regressions without any control in the female sample. Therefore, the second-order polynomial function is used as the main RD specification, while higher-order polynomial specifications are includes as robustness checks.

3.5.2 First-Stage Discontinuity: Testing Assumption A1

Effectiveness of the 1978 Education Reform

As discussed in the Background and Data chapter (Chapter 2), the Education Reform Act 1978 increased the duration of compulsory education in Thailand from 4 years of lower primary to 6 years of primary education, free of charge. As children are required to start school in the calendar year in which they became eight years old, the first birth cohort that was affected by the reform is 1967. The education reform forced less educated individuals to stay in school longer while encouraging more educated individuals to study more due to lower average costs and in an attempt to maintain their educational advantage.

To obtain an overview of the effectiveness of the 1978 Education Reform before formally testing the validity of the instrument, I follow Duflo (2001), where the effectiveness of the instrument is briefly checked by analysing its impacts on each education level. In other words, linear probability models of completing each of the education levels are estimated from the following equation.

$$Ed_{i,j} = \alpha_1 + \alpha_{2,j} \cdot \mathbb{1}\{R_i \geq 1967\} + \alpha_3 \cdot \mathbf{X}_i + \epsilon_i \qquad (3.11)$$

where $Ed_{i,j}$ is a dummy variable that indicates whether individual i born in year R_i completed j education, for j = lower than primary, primary, lower secondary, upper secondary, and tertiary. Thus, $\alpha_{2,j}$ is the impact of the compulsory schooling reform on the completion at each level of education j. $\alpha_{2,j}$ is estimated for the pooled sample of employed workers

[36]Note that a higher significance level provides a more stringent test as I look for a case of the failure to reject the null hypothesis (H_0 : The coefficients on the cohort dummies are jointly insignificant).

and for male and female employed workers separately.

Table 3.4 shows the effects of the 1978 Education Reform on years of education (Column 1) and the completion at each of the education levels, $\hat{\alpha}_{2,j}$ (Columns 2 to 6). The reform has a positive impact on years of education. It increases years of education by 0.9 years on average, and its positive impact is higher for women, compared to men. By considering each education level separately, the effects of the education reform are found to be statistically significant for all levels, in particular for primary education completion. The reform increases the probability of completing primary education by 34 per cent on average. Meanwhile, the impacts of the reform on the probabilities of completing a level higher than primary education are substantially lower, around 1 to 3 per cent. However, they remain statistically significant. While the validity of the reform as an instrument has to be formally tested, these results suggest that the assumption underlying the identification strategy is reasonable as it significantly influences educational attainment, particularly among those at the bottom and middle of the education distribution.

Testing Assumption A1

The RD framework requires that the year of birth affects years of education in a discontinuous way around the cut-off year, 1967 (assumption **A1**). This assumption can be verified by examining the effect of the instrument (whether or not an individual was born at or after 1967) in the first-stage education regressions (see Footnote 28, p. 43) or simply by looking at a plot of average years of education against birth cohort. Columns 1 to 3 of Tables 3.5 and 3.7 present the estimated coefficient on the year of birth dummy instrument in the first-stage education regression for the earnings and non-agricultural sector analyses respectively. For the sample of paid-employed workers (which will be used to estimate the earnings regressions), whether an individual was born at or after 1967 increases educational attainment by slightly more, around 0.3 years on average (Table 3.5). For the sample of employed workers (which will be used to estimate the non-agricultural sector regressions), the year of birth dummy instrument increases educational attainment by slightly higher, around 0.4 years on average (Table 3.7). In addition, the effect of the year of birth is greater for women than men in both samples. These results imply that the education reform is likely to affect women and unpaid workers slightly more than men and paid workers. These estimates of the first-stage education regressions are generally robust to the inclusion of control variables.[37] The discontinuities in educational attainment around the cut-off year are also observed when plotting the average years of education by birth cohort (Figure 3.1).

3.5.3 Effects of the 1978 Education Reform on Outcomes of Interest

The reduced form impacts of the education reform, captured by the 1967-birth cohort dummy variable, are presented in Columns 4 to 6 of Tables 3.5 and 3.7 for earnings and sector of employment models respectively. These results imply a discontinuity also in the outcome variables (log hourly earnings and the probability of working in non-agriculture) at the 1967 cut-off birth cohort. For the pooled sample of paid-employed workers in the earnings regressions (Table 3.5, Columns 4 to 6), being born during and after 1967 (and thus being under the new education regime) significantly increases earnings by 4.4 per cent on average. This effect declines to around 2 per cent when controlling for other observed characteristics. The effects are slightly higher among female workers, compared to male workers.

Considering the reform impacts on sector of employment (Table 3.7, Columns 4 to 6), being born during and after 1967 increases the probability of working in the non-agricultural sector by around 1 per cent. Without controls, these impacts appear to be larger among men. Nonetheless, with the inclusions of controls, the impacts are rather similar across gender, implying that the higher probability of being in the non-agricultural sector among male workers may be explained by different observed characteristics.

3.5.4 Effects of Education on Outcomes of Interest

Following the RD framework, the effects of education on earnings and sector of employment for compliers are calculated from the second-order polynomial regressions. That is, the education impacts are estimated from Equation (3.9) with the education reform or 1967-birth cohort dummy as an excluded instrument for years of education. The main results are shown in Tables 3.6 and 3.8. For each of the outcome variables, I first discuss the OLS estimates (Columns 1 to 3), followed by an interpretation of the RD results (Columns 4

[37]Due to the inclusion of the polynomial terms, the reform effects from the first-stage regressions (Tables 3.5 and 3.7) are much smaller than those estimated according to Equation (3.11) (Table 3.4). This highlights that the discontinuity may be overestimated when restricting the relationship between year born and education to be linear and similar on each side of the discontinuity threshold (as in Table 3.4).

to 6).[38] In both tables, different columns show the results from various specifications.[39] As discussed earlier, while the RD assumptions may be valid without conditioning, adding covariates is helpful for verifying local smoothness (assumption **A2**). Additional covariates can control for continuous relationships between other factors and the outcomes across the cut-off as well as increasing the precision of the estimates.

3.5.4.1 Earnings

I begin the analysis of returns to education by ignoring potential biases due to measurement error and/or unobserved individual heterogeneity in ability and preferences. Due to potentially systematic differences in earnings and labour market decisions, I estimate returns to education for the pooled sample, as well as men and women separately. Following Mincer (1974), I estimate a semi-logarithmic specification of (real hourly) earnings, using OLS regressions. In addition to years of education, the control variables in the earnings model include age and its square as a proxy for labour market experience; hours of work per week to control for the labour supply effect; and a rural-urban indicator. In addition, I include province and year dummies to factor out macroeconomic effects on earnings that may occur differently by location and time.[40] I also include a marital status dummy variable. The polynomial order of the distance from the cut-off year of birth is of degree two (quadratic). I should emphasise that, owing to the RD framework, the returns to education estimated in this chapter are forced to be linear, implying a constant (average) marginal return to an additional year of schooling. Also, the numbers of male and female paid employed workers are not proportionate to the male and female populations. While this chapter does not focus on the selection issues, it is worth noting that the difference in estimated returns to education between gender could be due to self-selection into paid employment.

Columns 1 to 3 in Table 3.6 display the OLS returns to education estimated from the earnings regressions. Panels (a), (b), and (c) present the results for the pooled, male, and female samples respectively. The OLS estimates without any controls suggest that an additional year of education is associated with an approximately 13 per cent increase in hourly

[38] Given that I use the same data window and polynomial function both the first-stage and the reduced form regressions, the RD estimates from the two-stage least square equations are equal to the ratio of the reduced form estimates to the first-stage estimates (as explained in p. 43).

[39] Columns 1 and 4 are the estimates without controls. Columns 2 and 5 are the estimates with controls for province and year dummies. Columns 3 and 6 are the estimates with the full set of controls.

[40] A male dummy is also included in the pooled regressions.

earnings for the pooled sample of paid employed workers (Column 1). The OLS return declines to around 9 per cent when other individual characteristics are included (Column 3), implying a possible correlation among characteristics as well as between these characteristics and years of education. When focusing on male and female paid workers separately, the OLS returns to education for women are found to be nearly 10 per cent (with a full set of controls, Column 3), slightly higher than those for men.

As the OLS returns are likely to be biased for the reasons discussed earlier, this chapter use the discontinuity in years of education at the 1967 birth cohort, as an instrumental variable. These RD results are shown in Columns 4 to 6 of Table 3.6. Column 6, which is the preferred specification as it includes the full set of controls, suggests that the OLS returns are higher than the RD returns to education. For the pooled sample, the estimated returns decline from 9 per cent when using OLS to 8 per cent with the RD. An upward bias of a similar magnitude is also observed when estimating the returns for men and women separately. Among male paid workers who are compliers, an additional year of education increases earnings by around 7 per cent, compared to the OLS estimate of 9 per cent. Meanwhile, the returns to education for women are around 8 per cent, 1 percentage point lower than the OLS estimate. From these results, I conclude that the returns to education, estimated from the RD approach, are around 8 per cent, and that women gain more from an additional year of education, compared to men.

There are two potential explanations for the upward bias in the OLS estimates. First, as the RD returns apply only to compliers, the lower RD returns could suggest the importance of selection into being a primary school dropout (Devereux and Hart, 2010). That is, the compliers have lower returns to education. Second, the upward bias in the OLS implies a positive correlation between years of education and unobserved individual heterogeneity in productivity-enhancing characteristics such as ability and preferences. This is likely to be the case when the compliers represent the majority of the population, and thus the RD returns are informative about the population (approaching the ATE) (Oreopoulos, 2008). In my case, the first explanation is more plausible because my framework is not a sharp discontinuity where everyone is a complier. In addition, compared to the literature on returns to education in Thailand discussed in the Background and Data chapter (Chapter 2), my RD estimates of returns to education are somewhat lower. As most of the earlier studies include only wage earners, my results potentially suggest lower returns to education among self-employed workers, who are also included in my analysis.

It is very interesting that for both pooled and gender-specific samples, the RD estimates of returns to education are actually higher than the OLS estimates, when including only province and year dummies (Columns 1 and 2 for the OLS, compared to Columns 4 and 5 for the RD). For example, in the simplest specification without any controls, the RD returns to education for the pooled sample are 16 per cent, compared to 13 per cent when using OLS. While this implies a potential correlation between education and these observed but excluded variables, this may be explained in two ways. First, the simple explanation is that the negative correlations between educational attainment and other observed characteristics that are included in the full set of controls are much larger than the positive correlations between education and unobserved characteristics. According to this explanation, the education reform remains an informative and valid instrument also in the specifications without the full set of controls.[41] Second, the underestimated OLS returns in these basic specifications could indicate that the education reform is not a valid instrument, when these controls are excluded. In other words, it is correlated with observed controls like age, working hours, and urban and married dummies. The second explanation is more plausible, as the probability of being under the post-reform education system is likely to be correlated with some of these excluded controls such as age (when controlling for survey year), and marital status. Of course, it could also be a combination of both explanations. Therefore, this underlines the importance of including the control variables as discussed earlier to ensure that the estimates are precise and consistent.

3.5.4.2 Non-Agricultural Sector

Next, I examine how educational attainment affects the selection into the non-agricultural sector. As before, I begin the analysis by using OLS, and thus ignoring potential endogeneity and measurement error biases. In addition to years of education, other control variables in this model include the variables in the earnings model (as individuals may choose their sector of work according to potential sectoral earnings), and other household and individual

[41] That is, it is correlated to years of education and uncorrelated with both other observed but excluded controls and unobserved characteristics. If the negative correlations between education and these excluded controls, such as age, working hours per week, and rural-urban indicator, are much larger than the positive correlation between education and ability or preferences, the OLS estimates of the without-control specifications would be downwardly biased.

characteristics that could affect preferences to work in a specific sector. These household and individual characteristics are the number of children and elderly, and spouse's earnings.[42] The polynomial order of the distance from the cut-off year of birth is of degree two (quadratic). Note that, the total number of observations, especially for women, is higher than in the earnings model because of unpaid workers, who are likely to work in the agricultural sector on family farms.

While the linear probability model can generate predicted outcomes outside of a sensible range (zero to one), it remains a popular modelling framework as some econometric problems including endogeneity are relatively straightforward to address within this model as opposed to probit and logit models (Miguel et al., 2004; Wooldridge, 2002, 2010).[43] In addition, it is important to recognise that the linear probability framework tends to provide "good estimates of the partial effects on the response probability near the center of the distribution of X [an explanatory variable]", compared to around its extreme values (Wooldridge, 2002, p. 455). This implies that the linear probability model can provide a good estimate of the average effect. Following Miguel et al. (2004), I also compare the results using OLS and probit (Table 3.A.1 in the Appendix) and they are rather similar.[44]

The OLS effects of education on the sector of work are shown in Columns 1 to 3 of Table 3.8. Education seems to be positively correlated with working in non-agriculture. Without control variables, an additional year of education is associated with a nearly 6 per cent increase in the probability of working in the non-agricultural sector. This estimate declines to around 3 per cent when all controls are included. In the OLS estimation, the impacts of education on the probability of being in the non-agricultural sector are above 3 per cent for women, slightly higher than for men.

However, the OLS impacts may be biased because of measurement error and/or other endogeneity problems. For example, individuals with poor non-agricultural ability may study less and also choose to be in the agricultural sector. As in the earnings model, I use the RD framework to identify the education effects on sector of employment for the compliers.

[42]For the number of children, I include two variables: the number of children aged below 6 years and the number of children aged between 6 and 13 years.

[43]Following the econometric literature, I refer to an OLS regression with a binary outcome as a linear probability model.

[44]Miguel et al. (2004) study the impacts of economic conditions on the likelihood of civil conflict in Africa using a linear probability model (following Wooldridge (2002)). They first show that the basic results are identical, whether estimated with a linear OLS or probit, and then conduct their further econometric analysis with the linear OLS model. Sergenti (2006) replicates this study using a probit model and concludes that the findings are very similar to those of Miguel et al. (2004).

The RD results are shown in Columns 4 to 6 of Table 3.8. As argue above, I mainly focus on the results in Column 6, which includes the full set of controls. Overall, the positive impacts of education on working in the non-agricultural sector decline but remain significant. The upward bias in the OLS implies a positive relationship between the unobserved factors for selection into the non-agricultural sector and educational attainment. Alternatively, it could suggest a smaller education effect among the compliers. For the pooled sample, the estimated impact of education on the probability of being in the non-agricultural sector declines from 3 per cent to around 2.5 per cent. When considering male and female workers separately, the results change to some extent, reflecting heterogeneity in impacts by gender. While the OLS shows comparable impacts of education on being in the non-agricultural sector for men and women, the RD impacts for women decline substantially to below 2 per cent, while those for men remain around 3 per cent. The results suggest that education helps men move out of agriculture towards non-agriculture relatively more than it does women.

3.5.4.3 Robustness Checks: Higher Polynomial Orders

Although I used the optimal polynomial order (as identified in Sub-section 3.5.1), it is important to additionally test whether the chosen polynomial order is misspecified. This could lead to biased estimates of the discontinuity and erroneous conclusion about statistical significance (Lee and Lemieux, 2010). Specifically, a misspecified polynomial order could identify a discontinuity, when in fact it does not exist in the data, and vice versa. Therefore, the main test in this robustness check is the statistical significance of the estimates rather than their magnitudes. I report results for alternative polynomial orders which are higher than the second order, as they are believed to be relatively less restrictive.

Table 3.A.2 in the Appendix presents the returns to education using polynomial RD regressions with third- and fourth-order polynomials of the distance from the 1967 cut-off year of birth. Overall, the RD results in terms of statistical significance are generally robust to the choice of polynomial order. In the specifications with the full set of controls, the impacts of an additional year of education on earnings remain significantly positive and are higher for women, compared to men. The magnitude of the returns is similar between the second- (Table 3.6) and third-order polynomials at around 8 per cent for the pooled sample, and slightly lower (higher) for men (women).

However, the coefficient estimates change more substantially in the RD regressions with a quartic polynomial. In particular, among male paid employed workers, the returns to education decline from 7 per cent to 4 per cent, while the returns for women remain stable at around 8 per cent. If the quadratic and cubic functions of birth cohorts are believed to be too restrictive for educational attainment and earnings, the quartic RD regressions should be preferred which would lead to much lower returns to education for men. On the other hand, the quartic function may be too flexible and may not describe the data well. As most of the aforementioned literature on RD returns to education uses linear or quadratic functions, a quartic polynomial would be hard to justify. Furthermore, it is important to bear in mind that the appropriate test concerns the statistical significance which is robust to all polynomial orders considered.

Table 3.A.3 in the Appendix reports the impacts of education on the probability of being in the non-agricultural sector using the RD regressions with cubic and quartic polynomials. The results are robust to the choice of polynomial order both in terms of statistical significance and magnitude. For the pooled sample, the impact of an additional year of education on the probability of working in the non-agricultural sector is around 3 per cent. These impacts remain higher for men, at around 3 to 4 per cent, compared to around 2 per cent for women.

3.5.5 Empirical Results Summary

In summary, the increase in educational attainment, induced by a change in compulsory years of schooling, had significant positive impacts on both earnings and the probability of working in the non-agricultural sector. In addition, the effects of education on these two outcomes of interest differ across gender. For the earnings regressions, an additional year of education is expected to increase hourly earnings by 8 per cent on average, around 7 per cent for men and above 8 per cent for women. Meanwhile, education increases the probability of being in the non-agricultural sector by 3 per cent on average, and slightly above 3 per cent for men and 2 per cent for women. These RD estimates on both earnings and the probability of working in non-agriculture are lower than the OLS effects. The conventional explanation relates this upward bias in the OLS estimates to a positive correlation between education and unobserved individual ability. However, I argue that it is more plausible that the lower RD estimates are due to heterogeneous effects of education across individuals. As a result, the RD estimates imply lower education effects among the compliers, who increase their years of education as a result of the education reform.

It is important to note that the empirical results should be interpreted with caution especially when comparing the RD returns across gender. A significantly smaller proportion of female paid employed workers raises the question of whether women who are unemployed or unpaid are systematically different from those who are in paid employment. If this is true, the estimated female returns are likely to be subject to a sample selection bias when the population of interest is all employed workers.

3.6 Conclusion

In this chapter, I have shown, using Thai labour force data, that an additional year of education, induced by the 1978 Education Reform, plays an important role in raising earnings and the probability of working in the non-agricultural sector in Thailand. My sample includes both wage- and self-employed workers to assess the impacts of education on earnings, and both paid and unpaid workers to assess the impacts on sector of employment. As the education reform generated an exogenous discontinuity in years of education across birth cohorts, I apply the RD approach to identify the causal impacts of education. This method is similar to the IV approach, and thus provides estimates that are applicable for the compliers, whose educational attainment changes as a result of the reform.

I first highlight that the 1978 Education Reform, which changes compulsory education from 4 years of lower primary to 6 years of primary, significantly increases the educational attainment of the Thai labour force. Subsequently, with the RD framework, I find the returns to education to be around 8 per cent on average, during the years 1985 to 2000 for the compliers, while its positive impacts on the probability of working in the non-agricultural sector are around 3 per cent. More importantly, when estimating the education impacts for men and women separately, the heterogeneity in impacts across gender is confirmed. It is also important to stress that the RD education effects are found to be lower than the OLS estimates, which could be explained by both a positive correlation between education and unobserved individual heterogeneity, and the fact that the RD estimates only apply to sub-populations. I argue that the latter is likely to be more important as the compliers are not necessarily representative of the population. Furthermore, the RD estimates may still suffer from an upward bias as the increase in education could be endogenous for those compliers who would have attained at least primary education regardless of the education regime (as I discussed in the context of the RD assumptions). However, I do not expect this to be a major problem given that around 60 per cent of the labour force attained less than 6 years of primary education (Chapter 2, Figure 2.1c).

I should emphasise that the estimated education effects in this chapter do not take into account sample selection issues. In other words, the returns to education are estimates conditional on being a paid employed workers. Similarly, the education impacts on the probability of working in the non-agricultural sector are conditional on being employed. Given very low unemployment and high labour force participation in Thailand, the estimates of the education impacts on the sector of employment may be affected relatively little by sample selection. However, due to the large amount of unpaid family workers, particularly in the agricultural sector and among women, the estimated returns to education are unlikely to be similar to the unconditional returns for the entire population. Especially for women, if their decisions about being in paid employment are based on individual returns, the conditional returns are likely to be higher than the unconditional returns. Similarly, if there exists an unobserved factor which influences both earnings and the decision about paid employment in the same direction, the conditional returns could underestimate the unconditional returns.[45]

To conclude, the RD empirical results in this chapter suggest that the education policy shift is crucial in changing the level of human capital, measured by years of education, and thus in changing labour market outcomes including earnings and sector of employment. Nonetheless, its effects differ across gender. Given drastically different numbers of male and female paid employed workers, it is important to further understand, in particular for women, how education influences the probability of having a paid job and what the unconditional returns to education would be, including for those who are not able to or decide not to get a paid job.

[45]For instance, motivation and/or labour market ability are expected to positively affect both earnings and the decision on paid employment. Even if they were uncorrelated with education at the population level, this could still lead to a downward (sample selection) bias in the OLS estimates. More specifically, those with low education need to have very high motivation and/or labour market ability to be in the paid employment sample. This results in a negative correlation between education and unobserved individual heterogeneity among paid employed workers, and thus a downward bias in the OLS estimates.

Table 3.1: LFS Descriptive Statistics of Paid-Employed Workers (Earnings Model)

	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
LFS Paid-Employed Workers																
# of observations	13,050	12,446	12,100	13,344	18,609	17,851	20,033	19,575	18,991	37,522	37,638	36,376	36,114	33,680	33,958	33,013
Avg. age	30.2	30.1	30.0	30.2	30.1	30.1	30.1	30.3	30.2	30.3	30.6	30.5	30.7	30.8	30.8	30.8
1{Male}	68%	68%	65%	65%	66%	65%	65%	64%	64%	62%	62%	63%	62%	61%	61%	60%
1{Urban}	38%	39%	36%	38%	37%	37%	37%	37%	37%	38%	39%	39%	41%	40%	39%	40%
Avg. years of educ.	5.8	6.1	6.2	6.4	6.4	6.5	6.7	6.8	6.9	7.1	7.1	7.2	7.4	7.8	8.0	8.1
Avg. hourly earnings (THB)	22.3	22.4	22.0	25.1	23.8	26.1	27.1	30.3	32.1	33.4	36.8	36.6	42.5	34.9	34.6	34.0
Avg. hours of work per week	51.8	51.9	51.2	51.6	52.8	52.0	51.7	52.0	51.4	51.6	51.8	51.5	48.8	50.4	49.9	50.4
1{Non-agriculture}	54%	56%	57%	57%	57%	59%	63%	63%	65%	66%	70%	70%	72%	70%	70%	71%
LFS Male Paid-Employed Workers																
Avg. age	30.5	30.4	30.3	30.5	30.4	30.4	30.3	30.5	30.4	30.5	30.8	30.7	31.0	31.1	31.1	31.1
1{Urban}	33%	35%	32%	34%	33%	33%	33%	33%	33%	34%	36%	35%	37%	36%	36%	36%
Avg. years of educ.	5.6	5.9	6.0	6.2	6.2	6.3	6.5	6.6	6.7	6.9	6.9	6.9	7.1	7.6	7.8	7.8
Avg. hourly earnings (THB)	22.3	22.4	21.9	25.4	24.2	26.0	27.3	30.5	32.3	34.2	38.1	37.3	42.5	34.9	35.4	34.3
Avg. hours of work per week	53.1	53.0	52.4	52.8	54.0	53.2	52.9	53.2	52.3	52.8	52.9	52.5	50.1	51.4	50.6	51.1
1{Non-Agriculture}	49%	50%	51%	52%	51%	53%	57%	57%	59%	61%	65%	66%	67%	64%	64%	65%
LFS Female Paid-Employed Workers																
Avg. age	29.6	29.6	29.4	29.6	29.6	29.6	29.6	29.8	29.8	29.8	30.1	30.1	30.2	30.3	30.3	30.4
1{Urban}	46%	48%	44%	45%	44%	44%	44%	43%	43%	43%	44%	44%	47%	46%	44%	45%
Avg. years of educ.	6.1	6.5	6.6	6.8	6.7	6.9	7.0	7.1	7.3	7.4	7.5	7.5	7.9	8.3	8.4	8.5
Avg. hourly earnings (THB)	22.2	22.4	22.1	24.5	23.1	26.2	26.6	30.0	31.6	32.1	34.8	35.3	42.4	35.0	33.5	33.6
Avg. hours of work per week	48.8	49.6	48.9	49.4	50.5	49.6	49.5	50.0	49.7	49.5	50.0	49.7	46.8	48.8	48.8	49.4
1{Non-Agriculture}	66%	69%	69%	67%	69%	71%	74%	73%	76%	75%	77%	78%	80%	80%	80%	79%

Source: Author's calculation from the Labour Force Surveys 1985-2000 (NSO).
Notes:
* This is a sample of the paid-employed labour force aged from 20 to 40 years, which will be used in the earnings model.
** Earnings are adjusted for inflation and regional price differences using the regional headline CPI.

Table 3.2: LFS Descriptive Statistics of Employed Workers (Sector Model)

	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
LFS Employed Workers (both paid and unpaid)																
# of observations	19,599	18,744	17,868	20,148	28,100	26,227	28,024	26,982	25,977	51,702	51,054	48,485	48,928	45,818	44,901	43,565
Avg. age	29.0	29.1	29.0	29.0	29.0	29.2	29.2	29.3	29.3	29.5	29.9	30.0	30.1	30.1	30.3	30.3
1{Male}	55%	54%	54%	53%	53%	54%	54%	54%	54%	54%	54%	55%	54%	54%	54%	54%
1{Urban}	27%	28%	28%	27%	26%	27%	27%	26%	28%	29%	31%	31%	32%	32%	32%	33%
Avg. years of educ.	5.2	5.5	5.7	5.8	5.8	6.0	6.1	6.2	6.4	6.5	6.6	6.7	6.9	7.3	7.5	7.7
Avg. hours of work per week	52.9	53.2	52.1	52.5	54.0	51.8	52.8	52.9	51.9	52.7	52.5	51.9	50.4	51.3	50.4	51.0
1{Non-agriculture}	36%	37%	40%	38%	37%	40%	44%	42%	47%	47%	53%	54%	54%	53%	55%	56%
LFS Non-agricultural Workers (both paid and unpaid)																
Avg. age	29.5	29.7	29.4	29.5	29.3	29.5	29.5	29.9	29.7	29.7	29.9	29.9	30.0	30.2	30.2	30.2
1{Male}	57%	57%	54%	55%	55%	55%	55%	56%	56%	55%	55%	56%	55%	53%	53%	53%
1{Urban}	62%	63%	59%	60%	59%	58%	54%	54%	53%	53%	53%	52%	54%	54%	53%	53%
Avg. years of educ.	7.1	7.5	7.6	7.9	7.8	7.8	7.9	7.9	8.0	8.1	8.0	8.0	8.3	8.8	9.0	9.1
Avg. hours of work per week	50.2	50.3	50.1	50.4	51.6	51.2	51.0	51.5	51.4	51.1	51.1	51.6	47.9	50.2	50.1	50.7
1{Married}	66%	65%	64%	63%	62%	62%	65%	66%	65%	65%	67%	67%	64%	65%	65%	64%
Avg. # of children aged <6	0.4	0.4	0.4	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Avg. # of children aged 6-13	0.5	0.5	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.3	0.3	0.3	0.3	0.3	0.3
Avg. # of elderly	0.05	0.04	0.05	0.05	0.05	0.06	0.05	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.04
Avg. spouse's hourly earnings	11.2	11.2	11.5	13.1	11.8	12.7	14.0	16.1	17.5	17.2	17.6	17.9	21.9	17.3	16.3	16.3
LFS Agricultural Workers (both paid and unpaid)																
Avg. age	28.8	28.7	28.7	28.6	28.8	28.9	28.9	28.9	29.1	29.3	29.9	30.0	30.1	30.0	30.3	30.4
1{Male}	53%	53%	54%	53%	52%	53%	54%	52%	53%	53%	53%	53%	54%	56%	56%	56%
1{Urban}	7%	6%	7%	7%	7%	7%	6%	6%	5%	7%	6%	7%	6%	7%	7%	7%
Avg. years of educ.	4.1	4.3	4.5	4.5	4.6	4.7	4.7	5.0	5.0	5.1	5.1	5.2	5.3	5.6	5.7	5.9
Avg. hours of work per week	54.4	55.0	53.4	53.7	55.4	52.2	54.2	53.9	52.3	54.1	54.1	52.1	53.3	52.5	50.7	51.3
1{Married}	84%	83%	83%	84%	84%	83%	84%	85%	84%	83%	85%	85%	84%	82%	83%	82%
Avg. # of children aged <6	0.6	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.4	0.4	0.4	0.4	0.4
Avg. # of children aged 6-13	0.9	0.9	0.9	0.8	0.8	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Avg. # of elderly	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.07	0.07	0.07	0.07
Avg. spouse's hourly earnings	3.0	2.9	3.3	4.0	3.7	3.8	3.9	4.6	4.8	5.7	5.9	5.1	5.8	5.0	5.2	5.9

Source: Author's calculation from the Labour Force Surveys (1985-2000).
Notes:
* This is a sample of the employed labour force (both paid and unpaid) aged from 20 to 40 years, which will be used in the sector model.

Table 3.3: P-Values for the Tests of Polynomial Specifications

(a) Pooled Sample: 1985-2000

Dependent variable:	Earnings Regressions Log of hourly earnings			Non-Agri. Sector Regressions 1{Non-Agriculture}		
	(1)	(2)	(3)	(4)	(5)	(6)
Linear	0.0000	0.0039	0.0445	0.0000	0.0057	0.0632
Quadratic	0.0856	0.1630	0.4025	0.0660	0.1759	0.3419
Cubic	0.1053	0.6944	0.9666	0.1589	0.3196	0.4756
Quartic	0.6944	0.9853	1.0000	0.7514	0.9968	1.0000
Year and province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y

(b) Male Sample: 1985-2000

Dependent variable:	Earnings Regressions Log of hourly earnings			Non-Agri. Sector Regressions 1{Non-Agriculture}		
	(1)	(2)	(3)	(4)	(5)	(6)
Linear	0.0065	0.0346	0.0589	0.0000	0.0277	0.0360
Quadratic	0.1132	0.2253	0.4458	0.1139	0.2429	0.3755
Cubic	0.2467	0.8463	0.9933	0.2896	0.4997	0.6428
Quartic	0.8433	1.0000	0.9997	0.7902	0.9854	0.9891
Year and province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y

(c) Female Sample: 1985-2000

Dependent variable:	Earnings Regressions Log of hourly earnings			Non-Agri. Sector Regressions 1{Non-Agriculture}		
	(1)	(2)	(3)	(4)	(5)	(6)
Linear	0.0000	0.0017	0.0041	0.0000	0.0015	0.0411
Quadratic	0.0069	0.1441	0.1609	0.0647	0.1898	0.3571
Cubic	0.0573	0.1579	0.3131	0.1167	0.3119	0.4466
Quartic	0.1441	0.7348	0.7735	0.6314	0.7467	0.9203
Year and province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y

Notes:
* I use two different samples of the labour force aged from 20 to 40 years. First, the sample for earnings regressions included all the paid employed. Second, in the non-agricultural sector regressions, both the paid and unpaid are included.
** Other controls include age, age-squared, gender dummy (for pooled regressions), hours of work per week, a rural-urban indicator, and marital status. The number of children and elderly, and the spouse's earnings are additionally included in the non-agricultural sector regressions.
*** The p-value refers to the following hypothesis test: H_0: The coefficients on the year of birth dummies are jointly insignificant, when regressing the dependent variable on the year of birth dummies and the higher order polynomials in the distance from the cut-off year of birth 1967. Therefore, failing to reject the null (p-value greater than the significance level, α) implies that the chosen polynomial order explains the data well.

Table 3.4: Effectiveness of the 1978 Education Reform

Dependent variable:	Years of educ.	1{Less than prim.}	1{Prim.}	1{Lower second.}	1{Upper second.}	1{Tertiary}
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient estimates on 1{Born at and after 1967}: $\alpha_{2,j}$ for Columns (2) to (6)						
Pooled sample	0.881***	-0.383***	0.342***	0.0186**	0.0163***	0.00615***
	(0.0951)	(0.0282)	(0.0215)	(0.00770)	(0.00401)	(0.00206)
Male sample	0.750***	-0.367***	0.340***	0.00681***	0.0183***	0.00189*
	(0.0828)	(0.0263)	(0.0206)	(0.00193)	(0.00371)	(0.00110)
Female sample	1.049***	-0.402***	0.344***	0.0296***	0.0142***	0.0141***
	(0.116)	(0.0311)	(0.0226)	(0.00872)	(0.00501)	(0.00328)

Notes:
* This uses the sample of the employed labour force (both paid and unpaid) aged from 20 to 40 years.
** Controls include age, age-squared, gender dummy (for pooled regressions), hours of work per week, a rural-urban indicator, marital status, province dummies, and year dummies.
*** The standard errors are clustered at the year of birth level. * p<0.1, ** p<0.05, *** p<0.01.

Table 3.5: First-Stage Discontinuity and Reduced Form Earnings Regressions

(a) Pooled Sample: 1985-2000

Dependent variable:	First-Stage Discontinuity Education (years)			Reduced Form Log of hourly earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
1{Born at and after 1967}	0.281***	0.330***	0.301***	0.0440**	0.0464***	0.0242***
	(0.0748)	(0.0682)	(0.0753)	(0.0176)	(0.0121)	(0.00781)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.044	0.112	0.228	0.012	0.259	0.416
# of observations	394,300	394,300	394,300	394,300	394,300	394,300

(b) Male Sample: 1985-2000

Dependent variable:	First-Stage Discontinuity Education (years)			Reduced Form Log of hourly earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
1{Born at and after 1967}	0.290***	0.314***	0.268***	0.0516***	0.0458***	0.0203**
	(0.0739)	(0.0660)	(0.0638)	(0.0135)	(0.0109)	(0.00787)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.043	0.121	0.227	0.013	0.295	0.438
# of observations	223,800	223,800	223,800	223,800	223,800	223,800

(c) Female Sample: 1985-2000

Dependent variable:	First-Stage Discontinuity Education (years)			Reduced Form Log of hourly earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
1{Born at and after 1967}	0.284***	0.366***	0.370***	0.0354	0.0479**	0.0317*
	(0.0970)	(0.0911)	(0.106)	(0.0309)	(0.0223)	(0.0156)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.042	0.095	0.228	0.010	0.201	0.379
# of observations	170,500	170,500	170,500	170,500	170,500	170,500

Notes:
* This is a sample of the paid employed labour force aged from 20 to 40 years.
** Other controls include age, age-squared, gender dummy (for pooled regressions), hours of work per week, a rural-urban indicator, and marital status.
*** The standard errors are clustered at the year of birth level. * p<0.1, ** p<0.05, *** p<0.01.

Table 3.6: OLS and RD Earnings Regressions

(a) Pooled Sample: 1985-2000

	OLS			RD-2SLS		
Dependent variable:	Log of hourly earnings			Log of hourly earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.134***	0.111***	0.0943***	0.157**	0.141***	0.0804***
	(0.00307)	(0.00293)	(0.00225)	(0.0769)	(0.0440)	(0.0296)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.292	0.443	0.522	0.286	0.442	0.523
# of observations	394,300	394,300	394,300	394,300	394,300	394,300

(b) Male Sample: 1985-2000

	OLS			RD-2SLS		
Dependent variable:	Log of hourly earnings			Log of hourly earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.140***	0.111***	0.0913***	0.178***	0.146***	0.0756**
	(0.00347)	(0.00307)	(0.00219)	(0.0685)	(0.0425)	(0.0364)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.263	0.443	0.519	0.246	0.437	0.519
# of observations	223,800	223,800	223,800	223,800	223,800	223,800

(c) Female Sample: 1985-2000

	OLS			RD-2SLS		
Dependent variable:	Log of hourly earnings			Log of hourly earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.125***	0.112***	0.0984***	0.125	0.131**	0.0856**
	(0.00237)	(0.00268)	(0.00247)	(0.0989)	(0.0532)	(0.0352)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.359	0.464	0.542	0.362	0.474	0.542
# of observations	170,500	170,500	170,500	170,500	170,500	170,500

Notes:
* This is a sample of the paid employed labour force aged from 20 to 40 years.
** Other controls include age, age-squared, gender dummy (for pooled regressions), hours of work per week, a rural-urban indicator, and marital status.
*** The standard errors are clustered at the year of birth level. * p<0.1, ** p<0.05, *** p<0.01.

Table 3.7: First-Stage Discontinuity and Reduced Form Sectoral Regressions

(a) Pooled Sample: 1985-2000

Dependent variable:	First-Stage Discontinuity Education (years)			Reduced Form 1{Non-Agricultural Sector}		
	(1)	(2)	(3)	(4)	(5)	(6)
1{Born at and after 1967}	0.411***	0.455***	0.422***	0.0111	0.0175***	0.0107***
	(0.0581)	(0.0592)	(0.0517)	(0.00867)	(0.00506)	(0.00352)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of ($R_i - 1967$)	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.048	0.137	0.246	0.004	0.265	0.383
# of observations	546,122	546,122	546,122	546,122	546,122	546,122

(b) Male Sample: 1985-2000

Dependent variable:	First-Stage Discontinuity Education (years)			Reduced Form 1{Non-Agricultural Sector}		
	(1)	(2)	(3)	(4)	(5)	(6)
1{Born at and after 1967}	0.390***	0.407***	0.350***	0.0209*	0.0200***	0.0107**
	(0.0526)	(0.0568)	(0.0506)	(0.0103)	(0.00709)	(0.00498)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of ($R_i - 1967$)	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.037	0.123	0.239	0.002	0.247	0.362
# of observations	271,865	271,865	271,865	271,865	271,865	271,865

(c) Female Sample: 1985-2000

Dependent variable:	First-Stage Discontinuity Education (years)			Reduced Form 1{Non-Agricultural Sector}		
	(1)	(2)	(3)	(4)	(5)	(6)
1{Born at and after 1967}	0.440***	0.519***	0.499***	-0.000370	0.0146***	0.00930**
	(0.0775)	(0.0766)	(0.0718)	(0.00756)	(0.00425)	(0.00378)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of ($R_i - 1967$)	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.062	0.155	0.272	0.008	0.288	0.419
# of observations	274,257	274,257	274,257	274,257	274,257	274,257

Notes:
* This is a sample of the employed labour force (both paid and unpaid) aged from 20 to 40 years.
** Other controls include age, age-squared, gender dummy (for pooled regressions), hours of work per week, a rural-urban indicator, marital status, the number of children and elderly, and spouse's earnings.
*** The standard errors are clustered at the year of birth level. * p<0.1, ** p<0.05, *** p<0.01.

Table 3.8: OLS and RD Sectoral Regressions

(a) Pooled Sample: 1985-2000

	OLS			RD-2SLS		
Dependent variable:	1{Non-Agricultural Sector}			1{Non-Agricultural Sector}		
	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.0582***	0.0420***	0.0319***	0.0269	0.0384***	0.0253***
	(0.000313)	(0.000413)	(0.000446)	(0.0212)	(0.0108)	(0.00899)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.189	0.352	0.423	0.135	0.353	0.423
# of observations	546,122	546,122	546,122	546,122	546,122	546,122

(b) Male Sample: 1985-2000

	OLS			RD-2SLS		
Dependent variable:	1{Non-Agricultural Sector}			1{Non-Agricultural Sector}		
	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.0566***	0.0408***	0.0306***	0.0535**	0.0492***	0.0306**
	(0.000491)	(0.000602)	(0.000626)	(0.0269)	(0.0165)	(0.0147)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.169	0.326	0.398	0.173	0.327	0.399
# of observations	271,865	271,865	271,865	271,865	271,865	271,865

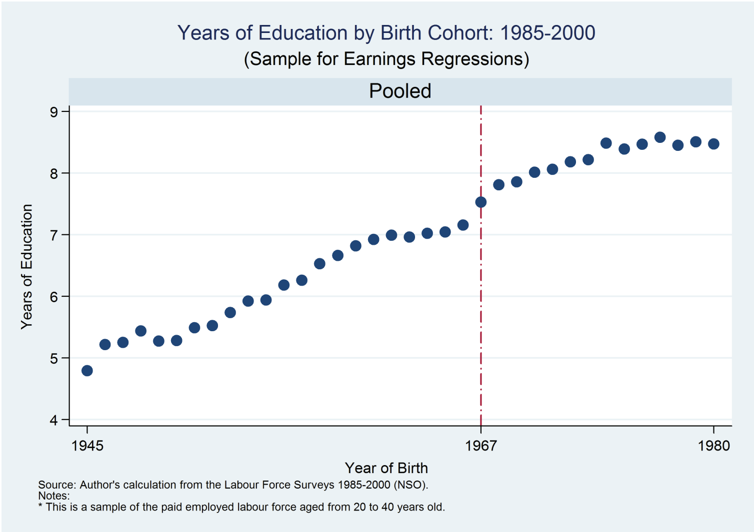
(c) Female Sample: 1985-2000

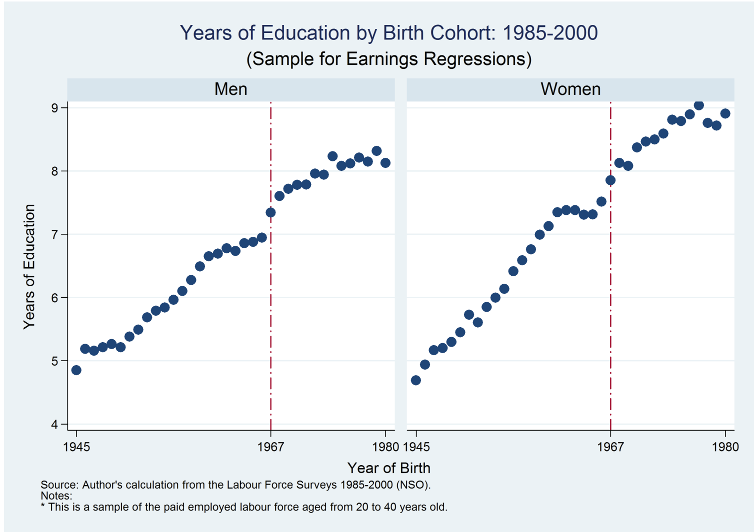
	OLS			RD-2SLS		
Dependent variable:	1{Non-Agricultural Sector}			1{Non-Agricultural Sector}		
	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.0600***	0.0432***	0.0317***	-0.000840	0.0282***	0.0186**
	(0.000362)	(0.000408)	(0.000473)	(0.0169)	(0.00907)	(0.00839)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.211	0.383	0.459	0.003	0.372	0.454
# of observations	274,257	274,257	274,257	274,257	274,257	274,257

Notes:
* This is a sample of the employed labour force (both paid and unpaid) aged from 20 to 40 years.
** Other controls include age, age-squared, gender dummy (for pooled regressions), hours of work per week, a rural-urban indicator, marital status, the number of children and elderly, and spouse's earnings.
*** The standard errors are clustered at the year of birth level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

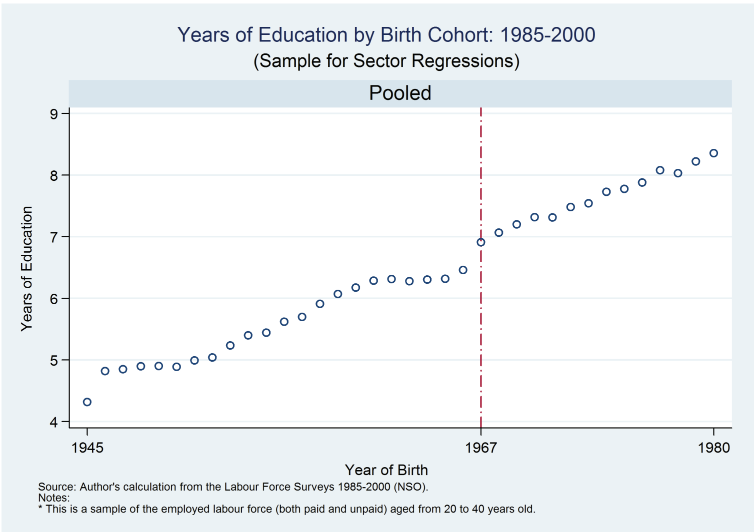
Figure 3.1: Educational Attainment by Birth Cohort

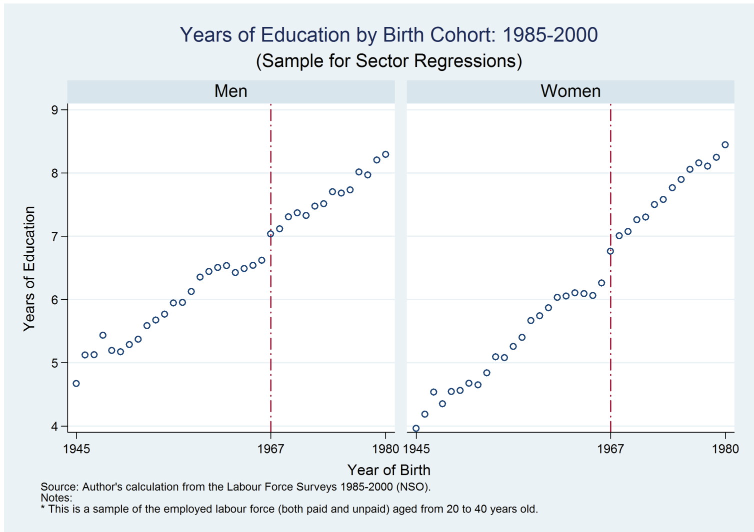
(a) Sample for Earnings Regressions





(b) Sample for Sector Regressions





3.A Chapter Appendix

Table 3.A.1: OLS and Probit Sectoral Regressions

(a) Pooled Sample: 1985-2000

Dependent variable:	OLS			Probit		
1{Non-Agri. Sector}	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.0582***	0.0420***	0.0319***	0.0714***	0.0645***	0.0597***
	(0.000313)	(0.000413)	(0.000446)	(0.000881)	(0.000956)	(0.00131)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.189	0.352	0.423			
# of observations	546,122	546,122	546,122	546,122	546,122	546,122

(b) Male Sample: 1985-2000

Dependent variable:	OLS			Probit		
1{Non-Agri. Sector}	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.0566***	0.0408***	0.0306***	0.0671***	0.0596***	0.0550***
	(0.000491)	(0.000602)	(0.000626)	(0.00116)	(0.00124)	(0.00157)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.169	0.326	0.398			
# of observations	271,865	271,865	271,865	271,865	271,865	271,865

(c) Female Sample: 1985-2000

Dependent variable:	OLS			Probit		
1{Non-Agri. Sector}	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.0600***	0.0432***	0.0317***	0.0777***	0.0716***	0.0662***
	(0.000362)	(0.000408)	(0.000473)	(0.000941)	(0.000984)	(0.00143)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
R-squared	0.211	0.383	0.459			
# of observations	274,257	274,257	274,257	274,257	274,257	274,257

Notes:
* This is a sample of the employed labour force (both paid and unpaid) aged from 20 to 40 years.
* The probit estimates presented are marginal probit effects, evaluated at the means of the explanatory variables.
*** Other controls include age, age-squared, gender dummy (for pooled regressions), hours of work per week, a rural-urban indicator, marital status, the number of children and elderly, and spouse's earnings.
**** The standard errors are clustered at the year of birth level. * p<0.1, ** p<0.05, *** p<0.01.

Table 3.A.2: Robustness Checks for RD Earnings Regressions: Higher Polynomial Orders

(a) Pooled Sample: 1985-2000

	RD-2SLS			RD-2SLS		
Dependent variable:	Log of hourly earnings			Log of hourly earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.155**	0.0765***	0.0757***	0.149***	0.0625***	0.0599***
	(0.0645)	(0.0280)	(0.0275)	(0.0370)	(0.0158)	(0.0171)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of	Cubic	Cubic	Cubic	Quartic	Quartic	Quartic
$(R_i - 1967)$						
R-squared	0.287	0.438	0.521	0.291	0.420	0.511
# of observations	394,300	394,300	394,300	394,300	394,300	394,300

(b) Male Sample: 1985-2000

	RD-2SLS			RD-2SLS		
Dependent variable:	Log of hourly earnings			Log of hourly earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.176***	0.0850***	0.0701**	0.161***	0.0603***	0.0411**
	(0.0571)	(0.0274)	(0.0329)	(0.0290)	(0.0170)	(0.0204)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of	Cubic	Cubic	Cubic	Quartic	Quartic	Quartic
$(R_i - 1967)$						
R-squared	0.248	0.446	0.517	0.258	0.422	0.497
# of observations	223,800	223,800	223,800	223,800	223,800	223,800

(c) Female Sample: 1985-2000

	RD-2SLS			RD-2SLS		
Dependent variable:	Log of hourly earnings			Log of hourly earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.123	0.0633	0.0821**	0.126**	0.0606	0.0799***
	(0.0874)	(0.0467)	(0.0344)	(0.0598)	(0.0373)	(0.0279)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of	Cubic	Cubic	Cubic	Quartic	Quartic	Quartic
$(R_i - 1967)$						
R-squared	0.362	0.432	0.541	0.362	0.426	0.540
# of observations	170,500	170,500	170,500	170,500	170,500	170,500

Notes:
* This is a sample of the paid employed labour force aged from 20 to 40 years.
** Other controls include age, age-squared, gender dummy (for pooled regressions), hours of work per week, a rural-urban indicator, and marital status.
*** The standard errors are clustered at the year of birth level. * p<0.1, ** p<0.05, *** p<0.01.

Table 3.A.3: Robustness Checks for RD Sectoral Regressions: Higher Polynomial Orders

(a) Pooled Sample: 1985-2000

Dependent variable:	RD-2SLS 1{Non-Agricultural Sector}			RD-2SLS 1{Non-Agricultural Sector}		
	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.0575***	0.0344***	0.0343***	0.0604***	0.0312***	0.0306***
	(0.0120)	(0.00888)	(0.00842)	(0.0211)	(0.00891)	(0.00814)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Cubic	Cubic	Cubic	Quartic	Quartic	Quartic
R-squared	0.191	0.352	0.424	0.192	0.349	0.424
# of observations	546,122	546,122	546,122	546,122	546,122	546,122

(b) Male Sample: 1985-2000

Dependent variable:	RD-2SLS 1{Non-Agricultural Sector}			RD-2SLS 1{Non-Agricultural Sector}		
	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.0879***	0.0478***	0.0447***	0.0892***	0.0442***	0.0422***
	(0.0142)	(0.0150)	(0.0145)	(0.0249)	(0.0147)	(0.0133)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Cubic	Cubic	Cubic	Quartic	Quartic	Quartic
R-squared	0.129	0.329	0.391	0.125	0.331	0.393
# of observations	271,865	271,865	271,865	271,865	271,865	271,865

(c) Female Sample: 1985-2000

Dependent variable:	RD-2SLS 1{Non-Agricultural Sector}			RD-2SLS 1{Non-Agricultural Sector}		
	(1)	(2)	(3)	(4)	(5)	(6)
Education (years)	0.0268*	0.0225**	0.0242***	0.0330*	0.0209**	0.0201**
	(0.0137)	(0.00931)	(0.00863)	(0.0184)	(0.00888)	(0.00876)
Year & province dummies	-	Y	Y	-	Y	Y
Other controls	-	-	Y	-	-	Y
Polynomial order of $(R_i - 1967)$	Cubic	Cubic	Cubic	Quartic	Quartic	Quartic
R-squared	0.149	0.364	0.459	0.170	0.360	0.456
# of observations	274,257	274,257	274,257	274,257	274,257	274,257

Notes:
* This is a sample of the employed labour force (both paid and unpaid) aged from 20 to 40 years.
** Other controls include age, age-squared, gender dummy (for pooled regressions), hours of work per week, a rural-urban indicator, marital status, the number of children and elderly, and spouse's earnings.
*** The standard errors are clustered at the year of birth level. * p<0.1, ** p<0.05, *** p<0.01.

Chapter 4

Testing the Importance of Human Capital for Structural Change: Sectoral Returns to Education

4.1 Introduction

There is widespread evidence that labour reallocation, structural transformation, and economic growth are interrelated in developed and developing economies. While traditional growth models explaining structural transformation are largely silent about human capital, more recent studies highlight that human capital development is crucial for economic growth driven by structural transformation. In this chapter, I use micro-level labour data to investigate the relative importance of human capital to the Thai economy during the period of high economic growth and structural transformation.

Structural transformation refers to the process of change in the fundamental structure of an economy (see for example, Lewis (1954); Kuznets (1955); Krueger (2008); McMillan and Rodrik (2011)). It is also likely to have implications for economic growth. In this chapter, structural transformation is defined as a significant shift from an agricultural to a non-agricultural (manufacturing and service) based economy. In Thailand, the government initially formulated the policies for industrialisation in the 1961 five-year National Economic and Social Development plan. Under the Act for the Promotion of Industrial Investment 1962, economic policy focused on import-substituting industrialisation (ISI). During the 1960s and 1970s, ISI policies mainly supported companies that produced products for the domestic market using income tax incentives and tariff protection of local production (Kuchiki, 2007; National Economic and Social Development Board, 2010). The second phase of industrialisation in Thailand was introduced in 1980, focusing on export promotion especially in the labour-intensive manufacturing sectors. Various export incentives such as a reduction in business taxes and the establishment of export processing zones were used (Hernandez, 2004; Kuchiki, 2007). Privileged access to subsidised credit and licensing was granted to exporters (Rock, 1995; Hernandez, 2004). The high export growth occurred during a conducive macroeconomic environment such as low fiscal deficits and inflation as well as a competitive exchange rate (Siriprachai, 1996). In addition, increased public investment in roads, railroad networks, electricity generation, and irrigation supported these industrialisation policies (World Bank, 1993). Government spending on human capital has also been substantial with the objective of achieving universal primary education. Consequently, structural transformation towards the non-agricultural sector was observed both in terms of production and employment.

By focusing on the labour market aspects, I attempt to assess the role of education, as a measure of human capital, in facilitating structural transformation in Thailand. First, I estimate the effects of education on the sectoral sorting process. Second, I evaluate how the effects of education on earnings differ across the two sectors of work – namely agriculture and non-agriculture. If human capital plays an important role in structural transformation driven growth, the effects of education on earnings (that is, the returns to education) should be larger in the more productive sector.

Compared to Chapter 3, which estimates the effects of education on earnings and sector of work using the regression discontinuity (RD) framework, I adopt a more flexible approach in this chapter. First, the sectoral sorting process is allowed to be correlated with the selection into paid or unpaid employment. The correlation in the decisions about the sector of work and the type of employment has an important implication on the sectoral sorting process due to a significant proportion of unpaid family workers in the agricultural sector. Second, the effects of education on earnings are allowed to be heterogeneous across sectors, levels of education, and individuals. This allows me to analyse how education enhances structural transformation. Third, I attempt to estimate these effects of education for the entire population. To estimate the two sectoral earnings processes separately, I adopt a double selection correction technique to control for potential sample selection bias arising from non-random selection into a specific sector or paid employment. At the same time, I use a control function (CF) approach to correct for potential endogeneity of education. The CF approach corrects for this endogeneity while allowing for heterogeneity in returns across levels of education and individuals. Furthermore, with additional parametric assumptions on the correlation between the unobservables in the sectoral earnings and educational choice processes, the CF approach recovers the population-level effects of education. Similar to Chapter 3, I use the change in the compulsory schooling law in 1978 to produce exogenous variation in individuals' education.

My empirical results confirm the importance of human capital in facilitating the structural transformation in Thailand during the high growth period between 1985 and 2000. I find that education has significant positive effects on reallocating workers from the agricultural sector to the non-agricultural sector for all workers, and from unpaid to paid employment for the female labour force. The effects of education on earnings are found to be significantly non-linear and different across sectors. The returns to education at a level below primary school are higher in the agricultural sector. However, the agricultural returns decline, as educational attainment increases, and eventually become negative at around and after 7 years of school. That is, the education-earnings profile is concave in the agricultural sector. On the other hand, the returns to education in the non-agricultural sector are very low at low levels of education, but increase with years of education. In other words, the non-agricultural sector has a convex education-earnings profile. The non-agricultural returns become higher than the agricultural returns at around 6 years of primary school and beyond that the gap in the returns is increasing with years of education. Compared to men, women's education-earnings profile is more concave in the agricultural sector and less convex in the non-agricultural sector.

This chapter makes two main contributions to the study of returns to education and structural transformation. First, I provide a framework that links the macroeconomic growth models related to structural transformation to labour market outcomes. I test the relevance of these models using micro-level labour data. Second, I attempt to estimate returns to education which are heterogeneous across levels of education and individuals, and apply to the entire population. Earlier studies of returns to education used the conventional instrumental variable (IV) approach together with a set of standard assumptions. These studies are able to recover only a linear effect of education on earnings which applies to those sub-populations whose education decisions change in response to the instrument.

The structure of this chapter is as follows. Section 4.2 reviews the two competing growth models related to structural transformation and their labour market implications. Section 4.3 outlines the empirical framework and discusses the identification strategy used to estimate the effects of education on sector of work and employment type, and sectoral returns to education. Sector 4.4 describes the data and presents descriptive statistics of the relevant sample and variables from the Labour Force Survey (LFS). Sector 4.5 discusses the empirical results which consist of two main parts. First, I discuss how education affects the sorting into paid employment and the sector of work. Second, I estimate the sectoral returns to education. I also provide a discussion of how the results are related to the two structural transformation growth models. Section 4.6 concludes.

4.2 Modelling Structural Transformation and Economic Growth: A Review of the Two Main Structural Change Growth Models

Given the evidence of structural transformation in many developed and developing countries, there has been a growing concern about labour reallocation towards the high-productivity sector. This section reviews the two major economic growth models related to structural transformation. They make different assumptions about labour markets, and thus propose different driving forces of economic growth. The first of these is the dual economy growth model formulated by Lewis (1954) which attributes the process of structural transformation to labour reallocation despite holding the level of human capital constant. In other words, the dual economy growth model assumes a segmented labour market which results in different levels of labour productivity and wages across sectors. The second model of economic growth to be reviewed is the human capital augmented neoclassical growth model pioneered by Lucas (1988), Romer (1989), and Mankiw et al. (1992). It focuses on the role of human capital in supporting structural transformation by reallocating workers towards the high-productivity sector. The subsequent section discusses how to empirically test these two growth models using micro-level labour data.

4.2.1 The Dual Economy: Real Structural Change

The dual economy theory describes a labour market as being divided into two segments, the traditional sector and the modern sector (Lewis, 1954, 1979). The modern sector is often referred to as the manufacturing or advanced sector, in which firms compete freely and workers get paid according to their marginal productivity of labour. On the other hand, there is underemployment in the traditional sector as the marginal productivity of labour is very low and possibly close to zero. This is usually consistent with subsistence agriculture. Labour income in the traditional sector is assumed to exceed the marginal productivity of labour. It is likely to be close to the subsistence level or the average productivity of labour due to institutional constraints, such as income-sharing within households. Because of diminishing returns to factors of production and a limited supply of land, the marginal labour productivity in the traditional sector is likely to be zero.

The dual economy exists as long as mobility between the two sectors is limited. Lewis (1954) claims that the unlimited labour supply from the traditional sector to the modern sector can enhance total productivity growth. In other words, moving the labour force from the traditional sector to the modern sector increases the modern sector's productivity without losses in agricultural productivity. The income of workers who move to the modern sector also increases. Later, Ranis and Fei (1961, 1964, 1997) include the additional assumption that withdrawing the labour force from the traditional sector will at some point raise the marginal productivity of labour in the traditional sector above zero. As a result, the period of unlimited labour supply ends. Reallocating additional workers to the modern sector reduces productivity in the traditional sector. Eventually, labour incomes equal to the marginal productivity of labour in each sector. Higher marginal labour productivity in the modern sector will continue to be a source of earnings differentials. Hence, there could still be room for economic growth by further structural transformation through labour reallocation.

Another variant of the dual economy growth model assumes a non-competitive modern sector as labour incomes are set above market-clearing levels (Harris and Todaro, 1970). This type of economy is partitioned into the rural sector and the urban sector, where formal jobs are located. The urban formal sector pays above the rural informal sector. For example, this could arise because of a minimum wage or because formal sector firms are more likely to pay efficiency wages (Shapiro and Stiglitz, 1984). Unlike the Lewis model, the labour force is freely mobile across sectors. In equilibrium, the uncertainty of obtaining a formal sector job and thus the possibility of urban unemployment imply that the expected wage in the urban sector equals the rural wage.

It is important to emphasise that these dual economy models are largely silent about the role of human capital. By claiming that productivity gains could be obtained simply through labour reallocation, the model implicitly assumes sectoral productivity differentials at all human capital levels. In other words, all workers, regardless of their levels of human capital, are more productive in the modern sector. In this chapter, I therefore refer to this process of economic growth as real structural transformation since it is a genuine change from predominantly subsistence agriculture to an economy with increasing manufacturing and other highly-productive economic activities.

4.2.2 The Competitive Labour Market: Human Capital Augmented Structural Change

The human capital augmented growth model (Lucas, 1988; Romer, 1989; Mankiw et al., 1992; Hall and Jones, 1999) is developed from the neoclassical Solow growth model (1956), which concludes that the long-run source of per worker output growth is exogenous technological progress. In the neoclassical growth model, factors of production include capital, labour, and technological progress.[1] The difference between the original neoclassical growth model and the human capital augmented one is the definition of the labour input. While the Solow growth model defines the labour input as the employed workforce, the human capital augmented growth model expresses it as efficiency units of the employed workforce, measured by the total amount of productive services supplied by workers. In other words, the human capital augmented growth model assumes that the productivity of workers varies according to their human capital or skill levels. Hence, in addition to the exogenous technological

[1] This model rests on the following fundamental neo-classical conditions. First, the production function exhibits constant returns to scale in both of the two inputs, capital and labour. Second, all inputs are essential for production. The marginal productivity of each input is positive, diminishing, and satisfies Inada (1963) conditions. Third, the change in the capital stock in each period is an exogenous constant fraction of output-saved net capital depreciation, whilst population growth and the rate of technological progress are assumed to be exogenously constant.

progress, development in human capital could enhance economic growth as it makes workers more productive. While not being explicit about structural transformation, the human capital augmented neoclassical economy could be viewed as comprising two sectors, the low-productivity and the high-productivity sectors. Human capital plays a role in structural transformation, and thus economic growth by reallocating workers from the low-productivity sector towards the high-productivity sector.

More recent growth models emphasise the role of human capital in shaping skill levels and enhancing structural transformation by moving labour towards the high-productivity sector, which is growing faster as a result of skill-biased technical change (SBTC) (see for example, Bound and Johnson (1992); Acemoglu (1998); Autor et al. (1998)). First, human capital is positively correlated with skills. Second, the high-productivity sector favours skilled over unskilled workers because high human capital is complementary with the capital and technology used in the high-productivity sector. In these models, increasing the number of skilled workers does not necessarily result in lower marginal productivity, and thus lower skilled labour income. This could be because new technologies used in the advanced sector are complementary to skills (Bound and Johnson, 1992; Acemoglu, 1998). Some studies link these new technologies to rapid investment in computers and R&D (Krueger, 1993; Berman et al., 1994; Autor et al., 1998). As a result, the increase in human capital raises the number of skilled workers who are more productive in the advanced sector.

Unlike the dual economy models, structural transformation explained by the neoclassical growth and the SBTC models requires an increase in human capital. That is, human capital plays an important role in shifting the labour force from the traditional to the advanced sector, in which workers with high levels of human capital are more productive. Hence, I refer to this as a human capital augmented structural transformation.

4.2.3 Identifying the Relevance of the Two Structural Change Growth Models from Micro Data

The two growth models have important implications for the labour market, in particular for labour earnings and the sectoral sorting process, which can be tested using micro data. By using educational attainment as a measure of human capital and defining the advanced sector as non-agricultural activities, the two structural change models differ in the following predictions about the labour market.

First, in human capital augmented structural change education is expected to play an important role in sorting workers from the agricultural sector into the non-agricultural sector. This can be tested using micro data by estimating the effects of education on the sectoral sorting process. That is, I estimate whether an increase in educational attainment raises the probability of being in the non-agricultural sector.

Second, the role of human capital in explaining earnings differentials across sectors is expected to be larger in the human capital augmented structural change model. In this model, higher labour productivity and earnings in the non-agricultural sector are attributable to a higher level of human capital. On the other hand, in the real structural change model, higher earnings in the non-agricultural sector may not fully be explained by human capital differentials because higher productivity could be achieved by solely reallocating workers across sectors. One way to test the importance of human capital for structural transformation is to estimate the returns to education, separately in the two sectors. In the real structural change model, the earnings gap between the agricultural and non-agricultural sectors is constant for all levels of education. That is, the returns to education are similar in the two sectors. Meanwhile, the human capital augmented structural change model suggests that human capital is valued relatively more in the non-agricultural sector, which would thus have higher return to education. If the returns to education are non-linear, they are expected to be higher in the non-agricultural sector, in particular at a middle and/or high level of education.

These testable predictions of the two structural change models are summarised in the table below.[2] In the next section, I outline the econometric framework used to estimate the sectoral sorting and sector-specific earnings processes. I also clarify the endogeneity and sample selection concerns that arise in estimating these models.

[2]Note that the formulation of the earnings process draws upon the micro-based human capital theories. They stress the importance of education, training, and other productive ability in shaping individual productivity, and thus earnings (see for example, Becker (1994); Mincer (1974); Card (1999)). In these micro-based models, there is an incentive to invest in human capital, if individuals anticipate human capital to increase their future earnings. Workers, who are paid their marginal product, therefore choose the amount of human capital that maximises their expected life-time earnings. This observation results in the formulation of the human capital earnings functions, in which wages paid are a function of productivity-enhancing factors such as education, work experience, and training.

Text Table 4.1: Labour Market Implications for the Two Structural Change Growth Models

	Real Structural Change	**Human Capital Augmented Structural Change**
1. Sectoral sorting process	• Unexplained by education.	• Education helps to reallocate workers from the agricultural sector to the non-agricultural sector.
2. Non-agricultural earnings premium	• Unexplained by the different levels of educations in the two sectors.	• Explained by the different levels of education in the two sectors. It is likely that the premium is larger for workers with a high level of education.
3. Sectoral returns to education	• $\mathrm{R}_{\mathrm{Agric.}} \simeq \mathrm{R}_{\mathrm{Non-agric.}}$; for all levels of education.	• $\mathrm{R}_{\mathrm{Agric.}} < \mathrm{R}_{\mathrm{Non-agric.}}$; particularly at a high level of education (if allowing for non-linear effects of education on earnings).

Notes: * $\mathrm{R}_{\mathrm{Agric.}}$ and $\mathrm{R}_{\mathrm{Non-agric.}}$ are returns to education in the agricultural and non-agricultural sectors respectively.
** In this chapter, I test Points 1 and 3 using micro-level labour force data.

4.3 Econometric Framework and Identification Strategy

This section describes the econometric framework used to estimate the effects of education on (i) labour earnings in the two sectors of employment; and (ii) on the sectoral sorting process. First, I specify a general model of the earnings process and discuss potential sources of bias in the estimation of the sectoral returns to education. The two major econometric concerns are the bias arising from an endogenous explanatory variable and a non-randomly selected sample used to estimate the sector-specific earnings function. Second, I propose an identification strategy to correct for these concerns. The sample-selection correction also provides an answer to whether the impacts of education on the sectoral sorting process are substantial. Subsequently, I discuss the key variables in the model.

4.3.1 The Endogenous Switching Model of Sectoral Earnings

Following the literature on wage determination and returns to education, I estimate an earnings equation to assess the impacts of education on earnings in the agricultural and non-agricultural sectors.

$$\ln w_i = \alpha_0 + \beta_1 \cdot S_i + \beta_2 \cdot S_i^2 + \boldsymbol{\gamma} \cdot \mathbf{X}_i + \varepsilon_i \tag{4.1}$$

where w_i is hourly earnings of individual i, S_i is years of education, $\mathbf{X}_i$ is a vector of other observable individual productivity-enhancing characteristics, and ε_i is an idiosyncratic error.[3] Generally, the main productive characteristics influencing labour earnings are edu-

[3] The time period subscript, t, is dropped for notational convenience. In the estimation, I use repeated

cation and labour market experience, which are both measures of human capital.

This earnings function can be derived from a simple educational choice model, in which individuals maximise the discounted present value of earnings (net of the cost of schooling) by choosing their level of education (Becker, 1994; Mincer, 1974; Willis, 1986) (also see Footnote 2, p. 80). Given that the only cost of schooling in these models is foregone earnings, the maximisation problem results in a linear relationship between education and earnings.[4] More recently, models of endogenous schooling choice take into account that individuals have different tastes for and returns to schooling, as well as facing different financial constraints (Card, 1999, 2001).[5] Their models also assume a concave relationship between education and earnings. Equation (4.1) follows this non-linear specification.

Human capital can also be acquired through other individual characteristics, $\mathbf{X}_i$, such as labour market experience and/or training on the job. It is important to include these other productivity-enhancing characteristics, particularly when they are likely to be correlated with education.[6]

To analyse the sectoral earnings process, I adopt an endogenous switching model. That

---

cross-sectional data.

[4]That is, at a given discount rate, r, an individual maximises his discounted net present value of earnings, $\mathrm{PV}(S_i) = \int_{S_i}^{\infty} w(S_i) \cdot e^{-rt} dt$, with respect to years of education, S_i. As a result, $\frac{w'(S_i)}{w(S_i)} = \frac{d \ln w(S_i)}{dS_i} = r$ or $\ln w_i = \alpha + r \cdot S_i$.

[5]In Card's model, at a given discount rate, r, an individual maximises his net utility,

$$\mathrm{U}(S_i,\, c(t)) = \int_{S_i}^{\infty} u(c(t)) \cdot e^{-rt} dt + \int_0^{S_i} u(c(t) - \phi(t)) \cdot e^{-rt} dt \qquad \text{with respect to } S_i$$

where $c(t)$ is consumption in period t, and $\phi(t)$ is a convex function, reflecting the disutility of schooling compared to work. Hence, $u(c(t))$ is out-of-school utility and $u(c(t) - \phi(t))$ is in-school utility. Letting $w(S_i,\, t)$ denote real earnings, the intertemporal budget constraint can be written as

$$\int_0^{\infty} c(t) \cdot e^{-rt} dt = \int_0^{S_i} [P(t) - T(t)] \cdot e^{-rt} dt + \int_{S_i}^{\infty} w(S_i,\, t) \cdot e^{-rt} dt$$

where $P(t)$ are part-time earnings while at school and $T(t)$ are tuition costs. Letting $w(S_i,\, t) = f(S_i) \cdot h(t - S_i)$ and assuming that $T(S_i) \simeq P(S_i)$, the optimal schooling choice equates the marginal utility, $\frac{f'(S_i)}{f(S_i)}$, and marginal cost of schooling, $d(S)$. With the assumptions on the functional forms of the marginal return and marginal cost, this condition can be written as

$$\frac{f'(S_i)}{f(S_i)} = b_i - k_1 \cdot S_i \equiv d(s) = r_i + k_2 \cdot S_i$$

where b_i and r_i represent individual heterogeneity in marginal returns and costs of schooling respectively. By assuming that the marginal return, $\frac{f'(S_i)}{f(S_i)}$, is linear in education, the relationship between education and earning is quadratic. That is, $\ln w_i = \alpha + b_i \cdot S_i - \frac{1}{2} \cdot k_1 \cdot S_i^2$.

[6]An experience-earnings profile is often specified as being non-linear (see for example, Mincer (1974) and Card (1999, 2001)). In particular, Card (2001) assumes that (i) log earnings are additively separable in education and labour market experience, and (ii) the relationship between earnings and labour market experience is concave.

is, the two earnings equations for the agricultural and non-agricultural sectors are estimated separately.

$$\ln w_{i,A} = \alpha_{0,A} + \beta_{1,A} \cdot S_i + \beta_{2,A} \cdot S_i^2 + \boldsymbol{\gamma_A} \cdot \mathbf{X}_i + \varepsilon_{i,A} \tag{4.2}$$

$$\ln w_{i,NA} = \alpha_{0,NA} + \beta_{1,NA} \cdot S_i + \beta_{2,NA} \cdot S_i^2 + \boldsymbol{\gamma_{NA}} \cdot \mathbf{X}_i + \varepsilon_{i,NA} \tag{4.3}$$

where subscripts A and NA refer to the agricultural and non-agricultural sectors respectively.

4.3.2 Sample Selection Problem

In addition to the problems of endogenous regressors (which arises from the correlation between the explanatory variables, S_i and $\mathbf{X}_i$, and ε_i), a simple ordinary least squares (OLS) estimation of Equations (4.2) and (4.3) could lead to a bias in the returns to education for two reasons related to sample selection.

First, a sample selection bias could arise from the non-random exclusion of non-participants in the labour force (both voluntary and involuntary) and unpaid workers in the working population. The model is estimated for the sample of paid workers only, who may differ systematically from the rest of the population. This problem is closely related to sample selection bias from labour force participation addressed by Gronau (1974) and Heckman (1974, 1979). While Thailand's labour force participation has been high and the unemployment rate has remained low and stable over time, the non-random participation in paid and unpaid jobs is of particular concern.[7] For example, those who have relatively high returns to education may be more likely to secure a paid job in the labour market, while those with relatively low returns may prefer not to participate in paid employment and only work as an unpaid family worker. As a result, the returns to education estimated for the sample of paid workers, would overestimate the returns to education for the population. On the other hand, given a likely positive effect of education on the probability of paid employment, poorly educated workers would need to have a relatively high motivation to secure a paid job. If motivation is also one of the unobserved variables positively affecting earnings, the correlation between educa-

[7] This is because there are a substantial number of workers who are engaged in unpaid family employment, and thus excluded from the estimation of the earnings equation. The Thai labour force is described in detail in the Background and Data chapter (Chapter 2).

tion and the error terms of the earnings function would be negative for paid workers. This implies that the estimated returns to education obtained from the sample of paid workers would underestimate the returns for the population. In sum, if individuals select themselves into paid employment on the basis of some attributes that also affect their earnings, this process should be taken into account when estimating the earnings process and returns to education for the population.

Second, as the earnings function is estimated separately for the agricultural and non-agricultural sectors, the OLS estimates of the returns to education are exposed to another sample selection bias related to the non-random selection into the sector of work. Given the decision to work, either for a paid or unpaid job, individuals have to decide in which sector they want to work. In fact, agricultural and non-agricultural workers could differ from each other in many ways as they select themselves into sectors based on their labour market characteristics, preferences, and sector-specific ability. If these sector-determining factors are correlated with education and unobserved factors affecting earnings, the estimates of the sectoral returns to education can be biased. This selection issue is closely related to the model of occupational choice formulated by Roy (1951), in which individuals self-select into a sector depending on their comparative advantage. It is important to note that the direction of the bias is uncertain even if there is positive sectoral sorting based on comparative advantage. For example, when workers have an absolute advantage only in their selected sector over those who work in the other sector, the direction of bias is likely to be positive in both sectors. However, if some workers have an absolute advantage in both sectors, the sample selection bias would be positive in the sector (chosen by these workers) and negative in the other.

Empirical Framework for Sample Selection Correction

To correct for the two potential selectivity biases, I formulate the two sample selection problems as follows. First, individuals choose to work in paid employment as long as the utility of being paid and employed is greater than the utility of working as an unpaid family worker. Let P_i^* be a latent variable indicating the additional utility of being a paid worker compared with working as an unpaid family worker. The process of selection into paid employment can be written as

$$P_i = \begin{cases} 1 & \text{if } P_i^* > 0 \\ 0 & \text{if } P_i^* \leq 0 \end{cases} \qquad \text{and } P_i^* = \boldsymbol{\pi} \cdot \mathbf{W}_i + u_i \tag{4.4}$$

where P_i is a binary variable indicating whether an individual is a paid worker, $\mathbf{W}_i$ is a vector of all characteristics affecting the decision to work in paid employment and u_i is the error term for this decision making process. This reduced-form equation of the selection into paid employment is quite similar to the structural equation of labour force participation specified by Wooldridge (2010). In that model, individuals maximise their utility by choosing whether to participate in the labour market taking into account their productivity enhancing characteristics, disutility of working, and other individual and household characteristics that may influence their participation preferences.[8]

Second, individuals sort themselves into sectors depending on their expected utility in each sector. This expected utility usually depends on expected sectoral earnings as well as individual preferences for sectors, which could be influenced by other individual and household characteristics. Let A_i^* be a latent variable indicating the net utility of working in the agricultural sector (including the foregone utility from working in non-agriculture). The process of sorting into agricultural employment can be written as

$$A_i = \begin{cases} 1 & \text{if } A_i^* > 0 \\ 0 & \text{if } A_i^* \leq 0 \end{cases} \qquad \text{and } A_i^* = \boldsymbol{\tau} \cdot \mathbf{G}_i + v_i \tag{4.5}$$

where A_i is a dummy variable, taking the value of 1 when an individual is an agricultural worker and 0 when he is in non-agriculture, $\mathbf{G}_i$ is a vector of all characteristics affecting the decision about the sector of employment and v_i is the respective error term.

The two above-mentioned selection processes divide the total sample of the labour force into four categories as follows.

[8]While Wooldridge (2010) provides a simple model of labour force participation, I focus on the selection into paid employment. His basic model of labour force participation is motivated by the earlier model of wage offers and observed wage distributions by Gronau (1974).

Text Table 4.2: Potential Outcomes for the Selection into Paid and Agricultural Employment

		Agriculture (A_i)	
		$A_i = 1$	$A_i = 0$
Paid (P_i)	$P_i = 1$	$H_1 : P_i^* > 0$ and $A_i^* > 0$	$H_2 : P_i^* > 0$ and $A_i^* \leq 0$
	$P_i = 0$	$H_3 : P_i^* \leq 0$ and $A_i^* > 0$	$H_4 : P_i^* \leq 0$ and $A_i^* \leq 0$

As only individuals from sub-samples H_1 and H_2 are included in the estimation of the earnings equations, Equations (4.2) and (4.3) can be written as follows.

$$E(\ln w_{i,A} \mid S_i,\, \mathbf{X}_i,\, \varphi_1) = \alpha_{0,A} + \beta_{1,A} \cdot S_i + \beta_{2,A} \cdot S_i^2 + \boldsymbol{\gamma_A} \cdot \mathbf{X}_i + E(\varepsilon_{i,A} \mid S_i,\, \mathbf{X}_i,\, \varphi_1) \quad (4.6)$$

$$E(\ln w_{i,NA} \mid S_i,\, \mathbf{X}_i,\, \varphi_2) = \alpha_{0,NA} + \beta_{1,NA} \cdot S_i + \beta_{2,NA} \cdot S_i^2 + \boldsymbol{\gamma_{NA}} \cdot \mathbf{X}_i + E(\varepsilon_{i,NA} \mid S_i,\, \mathbf{X}_i,\, \varphi_2) \quad (4.7)$$

where the conditional arguments φ_1 and φ_2 denote the joint outcome of the two selection processes for sub-samples H_1 and H_2 respectively. Abstracting from the potential endogeneity of education at the population level, $E(\varepsilon_{i,A} \mid S_i,\, \mathbf{X}_i,\, \varphi_1) \neq 0$ and $E(\varepsilon_{i,NA} \mid S_i,\, \mathbf{X}_i,\, \varphi_2) \neq 0$ imply non-random sample selection. In other words, the correlation between the error terms in the sectoral earnings equations and the two selection processes could lead to a bias in the coefficient estimates from the OLS method.[9]

Identification Strategy for Sample Selection Correction

To obtain unbiased coefficient estimates, a number of assumptions for the conditional means, variances, and the distribution of the error terms in all processes are required. Different methods to correct for non-random selection can be distinguished according to the correlation between the error terms of the two selection processes, u_i and v_i. When decisions on the type and sector of employment are made independently (that is, the correlation between u_i and v_i is zero), the two processes can be modelled in a similar way to the single selection case (Heckman, 1979).[10] However, the two selection processes are likely to be correlated.

[9] The size of selection bias depends on the correlations between the error terms in the selection processes and sectoral earnings equations.

[10] That is, with the assumption of joint normality between each of the selection error terms and the error terms from the earning process, the three correction terms (known as the inverse Mill's ratio (IMR)) can be computed from the estimates of the two probit selection models ($\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\tau}}$) :

Some unobserved characteristics such as motivation, sector-specific ability, or preferences on hours of work may affect both whether individuals choose to be in paid employment and in the non-agricultural sector. Hence, I allow for this correlation. To do so, let $R_{i,j}$ be a vector of the three error terms $(\varepsilon_{i,j},\ u_i,\ v_i)^{'}$ and assume that these error terms have zero means and are jointly normally distributed. That is, $R_{i,j} \sim N(\mathbf{0},\ \Sigma_j)$, where Σ_j is the covariance matrix of the three error terms as follows.[11]

$$\Sigma_j = \begin{bmatrix} \sigma_j^2 & \sigma_j \cdot \rho_{ju} & \sigma_j \cdot \rho_{jv} \\ & 1 & \rho_{uv} \\ & & 1 \end{bmatrix} \qquad \text{for} \quad j = A \,\text{and}\, NA$$

Given the above structure, coefficient estimates of the sectoral earnings equations and the two selection processes can be obtained by maximum likelihood estimation (MLE). The likelihood function to be maximised is

$$\begin{aligned} L \;=\; & \prod_{i=1}^{N} \Big\{ l\left(P_i^* \leq 0,\ A_i^* \leq 0\right)^{(1-P_i)\cdot(1-A_i)} \cdot l\left(P_i^* \leq 0,\ A_i^* > 0\right)^{(1-P_i)\cdot(A_i)} \cdot \\ & l\left(P_i^* > 0,\ A_i^* \leq 0\right)^{(P_i)\cdot(1-A_i)} \cdot l\left(P_i^* > 0,\ A_i^* > 0\right)^{(P_i)\cdot(A_i)} \Big\} \end{aligned} \qquad (4.8)$$

where the parts of the likelihood function for the unpaid workers in the two sectors can be expressed as follows.

$$\begin{aligned} l\left(P_i^* \leq 0,\ A_i^* \leq 0\right) &= \Phi_2\left(-\boldsymbol{\pi} \cdot \mathbf{W}_i,\ -\boldsymbol{\tau} \cdot \mathbf{G}_i;\ \rho_{uv}\right) \\ l\left(P_i^* \leq 0,\ A_i^* > 0\right) &= \Phi_2\left(-\boldsymbol{\pi} \cdot \mathbf{W}_i,\ \boldsymbol{\tau} \cdot \mathbf{G}_i;\ -\rho_{uv}\right) \end{aligned}$$

The parts of the likelihood function for the paid workers in the two sectors can be expressed accordingly.

(i) the IMR for paid employment, $\lambda_i^P = \phi\left(\hat{\boldsymbol{\pi}} \cdot \mathbf{W}_i\right) / \Phi\left(\hat{\boldsymbol{\pi}} \cdot \mathbf{W}_i\right)$;
(ii) the IMR for agricultural employment, $\lambda_i^A = \phi\left(\hat{\boldsymbol{\tau}} \cdot \mathbf{G}_i\right) / \Phi\left(\hat{\boldsymbol{\tau}} \cdot \mathbf{G}_i\right)$; and
(iii) the IMR for non-agricultural employment, $\lambda_i^{NA} = -\phi\left(\hat{\boldsymbol{\tau}} \cdot \mathbf{G}_i\right) / \Phi\left(-\hat{\boldsymbol{\tau}} \cdot \mathbf{G}_i\right)$.

where $\phi\,()$ and $\Phi\,()$ are the univariate standard normal density and distribution functions respectively. The IMR for paid employment is added to both sectoral earnings equations, while the IMRs for agricultural and non-agricultural employment are included separately in their respective earnings equations. Alternatively, all the process can be estimated jointly using the maximum likelihood estimation (MLE).

[11] The covariance between $\varepsilon_{i,A}$ and $\varepsilon_{i,NA}$ is not directly identifiable in this kind of model. Hence, the covariance matrix is split into two matrices for the agricultural and non-agricultural sectors (see Vijverberg (1993a,b) and Koop and Poirier (1997) for more details). As only the binary variables, P_i and A_i, are observed, σ_u^2 and σ_v^2 are normalised to one for identification purposes (Wooldridge, 2010).

$$\begin{aligned} l\left(P_i^* > 0,\ A_i^* \leq 0\right) &= \Phi_2\left(\boldsymbol{\pi} \cdot \mathbf{W}_i,\ -\boldsymbol{\tau} \cdot \mathbf{G}_i;\ -\rho_{uv} \mid \ln w_{i,NA}\right) \cdot f\left(\ln w_{i,NA}\right) \\ l\left(P_i^* > 0,\ A_i^* > 0\right) &= \Phi_2\left(\boldsymbol{\pi} \cdot \mathbf{W}_i,\ \boldsymbol{\tau} \cdot \mathbf{G}_i;\ \rho_{uv} \mid \ln w_{i,A}\right) \cdot f\left(\ln w_{i,A}\right) \end{aligned}$$

where $\Phi_2\left(\cdot,\ \cdot;\ \rho\right)$ is a standard bivariate normal cumulative distribution function with a specified correlation coefficient, ρ, and $f\left(\ln w_{i,j}\right)$ is the log earnings density function for sector j.[12]

The estimates obtained from maximising Equation (4.8) are known as the full-information maximum likelihood (FIML) estimators. While the FIML method provides asymptotically efficient estimates, it is computationally intensive, especially when there are a large number of parameters to be estimated. In addition, the FIML method relies on the joint normality of the error terms. An alternative to the FIML is a simple Heckman (1979) two-step approach with extended correction terms. Fishe et al. (1981) and Tunali (1986) are among the first studies that allow for multiple sample selection processes and correlation in the error terms using the Heckman two-step approach.[13] The two-step estimation is more robust to misspecified distributional assumptions (see for example, Ham (1982); Tunali (1986); Co et al. (1999); Wetzels and Zorlu (2003); Ghinetti (2014)).[14] For the correction of the double selection process, the joint distributional assumptions could be reduced to

S1. $\left(u_i,\ v_i\right) \sim N\left(\mathbf{0},\ \begin{bmatrix} 1 & \rho_{uv} \\ & 1 \end{bmatrix}\right)$

S2. $E\left(\varepsilon_{i,j} \mid u_i,\ v_i\right) = \rho_{ju} \cdot E\left(u_i \mid v_i\right) + \rho_{jv} \cdot E\left(v_i \mid u_i\right) = \rho_{ju} \cdot \rho_{uv} \cdot v_i + \rho_{jv} \cdot \rho_{uv} \cdot u_i$

Then, the conditional expectation of the error terms in Equations (4.6) and (4.7) can be re-written as

[12]That is, $f\left(\ln w_{i,j}\right) = \frac{1}{\sigma_j} \cdot \phi\left(\frac{\ln w_{i,j} - \alpha_{0,j} - \beta_{1,j} \cdot S_i - \beta_{2,j} \cdot S_i^2 - \boldsymbol{\gamma_j} \cdot \mathbf{X}_i}{\sigma_j}\right)$.

[13]It is important to also emphasise that there is a substantial debate on whether the correlation between the error terms in the outcome equation and the selection process(es) have an effect on the performance of the two-step Heckman estimation (see Puhani (2000) for a comparison of full-information maximum likelihood (FIML) and two-step Heckman estimations). Most studies reviewed by Puhani (2000) are inconclusive about whether the FIML or the two-step Heckman estimation is superior. In addition to efficiency gains, FIML may be preferred in the absence of exclusion restrictions (that is, a simple Tobit model). However, it is computationally difficult to find an optimal set of parameters. As a result, the literature usually adopts the two-step Heckman estimation. It also performs relatively better when the number of observations is large.

[14]In the case of a single selection process, the two-step Heckman estimation usually assumes a bivariate normal distribution between the two error terms (one from the earnings function and the other from the selection process). But this assumption could be reduced to marginal normality and linearity in the conditional expectation of the error terms (Heckman, 1979).

$$
\begin{aligned}
E\left(\varepsilon_{i,A} \mid S_i,\, \mathbf{X}_i,\, \varphi_1\right) &= \sigma_A \cdot E\left(e_{i,A} \mid u_i > -\boldsymbol{\pi}\cdot\mathbf{W}_i,\, v_i > -\boldsymbol{\tau}\cdot\mathbf{G}_i\right) \\
&= \sigma_A \cdot \rho_{Au} \cdot \lambda_{i,P-A} + \sigma_A \cdot \rho_{Av} \cdot \lambda_{i,A-A}
\end{aligned} \tag{4.9}
$$

$$
\begin{aligned}
E\left(\varepsilon_{i,NA} \mid S_i,\, \mathbf{X}_i,\, \varphi_1\right) &= \sigma_{NA} \cdot E\left(e_{i,NA} \mid u_i > -\boldsymbol{\pi}\cdot\mathbf{W}_i,\, v_i \leq -\boldsymbol{\tau}\cdot\mathbf{G}_i\right) \\
&= \sigma_{NA} \cdot \rho_{NAu} \cdot \lambda_{i,P-NA} + \sigma_{NA} \cdot \rho_{NAv} \cdot \lambda_{i,A-NA}
\end{aligned} \tag{4.10}
$$

where $e_{i,j} \sim N(0,\,1)$ and $E(\varepsilon_{i,j}) = \sigma_j \cdot E(e_{i,j})$. The correction terms for the two non-random sample selection processes, $\lambda_{i,P-j}$ and $\lambda_{i,A-j}$ for $j = A,\ NA$ are as follows.

$$
\begin{aligned}
\lambda_{i,P-A} &= \phi\left(\boldsymbol{\pi}\cdot\mathbf{W}_i\right)\cdot\Phi\left[\frac{\boldsymbol{\tau}\cdot\mathbf{G}_i - \rho_{uv}\cdot\boldsymbol{\pi}\cdot\mathbf{W}_i}{\sqrt{1-\rho_{uv}^2}}\right]\cdot\Phi_2\left(\boldsymbol{\tau}\cdot\mathbf{G}_i,\, \boldsymbol{\pi}\cdot\mathbf{W}_i;\, \rho_{uv}\right)^{-1} \\
\lambda_{i,A-A} &= \phi\left(\boldsymbol{\tau}\cdot\mathbf{G}_i\right)\cdot\Phi\left[\frac{\boldsymbol{\pi}\cdot\mathbf{W}_i - \rho_{uv}\cdot\boldsymbol{\tau}\cdot\mathbf{G}_i}{\sqrt{1-\rho_{uv}^2}}\right]\cdot\Phi_2\left(\boldsymbol{\tau}\cdot\mathbf{G}_i,\, \boldsymbol{\pi}\cdot\mathbf{W}_i;\, \rho_{uv}\right)^{-1} \\
\lambda_{i,P-NA} &= \phi\left(\boldsymbol{\pi}\cdot\mathbf{W}_i\right)\cdot\Phi\left[-\frac{\boldsymbol{\tau}\cdot\mathbf{G}_i - \rho_{uv}\cdot\boldsymbol{\pi}\cdot\mathbf{W}_i}{\sqrt{1-\rho_{uv}^2}}\right]\cdot\Phi_2\left(-\boldsymbol{\tau}\cdot\mathbf{G}_i,\, \boldsymbol{\pi}\cdot\mathbf{W}_i;\, -\rho_{uv}\right)^{-1} \\
\lambda_{i,A-NA} &= -\phi\left(\boldsymbol{\tau}\cdot\mathbf{G}_i\right)\cdot\Phi\left[\frac{\boldsymbol{\pi}\cdot\mathbf{W}_i - \rho_{uv}\cdot\boldsymbol{\tau}\cdot\mathbf{G}_i}{\sqrt{1-\rho_{uv}^2}}\right]\cdot\Phi_2\left(-\boldsymbol{\tau}\cdot\mathbf{G}_i,\, \boldsymbol{\pi}\cdot\mathbf{W}_i;\, -\rho_{uv}\right)^{-1}
\end{aligned} \tag{4.11}
$$

where $\phi\,()$ and $\Phi\,()$ are the univariate standard normal density and distribution functions respectively. $\Phi_2\,(\cdot,\ \cdot\,;\ \rho)$ is a standard bivariate normal cumulative distribution function with a specified correlation coefficient, ρ.

The two-step estimation for the sectoral earnings model with a double selection process can be implemented by: (1) jointly estimating the coefficients of the two selection processes, $\hat{\boldsymbol{\tau}}$, $\hat{\boldsymbol{\pi}}$, and $\hat{\rho}_{uv}$ using a bivariate probit model (Equations (4.4) and (4.5)), and using those estimates to compute the correction terms, $\hat{\lambda}_{i,P-j}$ and $\hat{\lambda}_{i,A-j}$; and (2) estimating the two sectoral earnings function including the estimated correction terms. The sectoral earnings functions (Equations (4.6) and (4.7)) can be rewritten with the correction terms as follows.

$$E\left(\ln w_{i,A} \mid S_i,\, \mathbf{X}_i,\, P_i = 1,\, A_i = 1\right) = \alpha_{0,A} + \beta_{1,A} \cdot S_i + \beta_{2,A} \cdot S_i^2 + \boldsymbol{\gamma_A} \cdot \mathbf{X}_i + \delta_{Au} \cdot \hat{\lambda}_{i,P-A} + \delta_{Av} \cdot \hat{\lambda}_{i,A-A} \tag{4.12}$$

$$E\left(\ln w_{i,NA} \mid S_i,\, \mathbf{X}_i,\, P_i = 1,\, A_i = 0\right) = \alpha_{0,NA} + \beta_{1,NA} \cdot S_i + \beta_{2,NA} \cdot S_i^2 + \boldsymbol{\gamma_{NA}} \cdot \mathbf{X}_i + \delta_{NAu} \cdot \hat{\lambda}_{i,P-NA} + \delta_{NAv} \cdot \hat{\lambda}_{i,A-NA} \tag{4.13}$$

It is important to note that when the selection processes matter, the asymptotic variance of the coefficient estimates in the earnings and selection equations needs to be corrected. This happens because of heteroskedastic errors and the fact that $\hat{\lambda}_{i,P-j}$ and $\hat{\lambda}_{i,A-j}$ are generated regressors.

Technically, an exclusion restriction assumption is not required for the model to be identified. However, when $(S_i,\, \mathbf{X}_i) = \mathbf{W}_i = \mathbf{G}_i$, the coefficient estimates in the sectoral earnings equation are identified only by the non-linearity of the correction terms. When there is a little variation in $\boldsymbol{\pi} \cdot \mathbf{W}_i$ and $\boldsymbol{\tau} \cdot \mathbf{G}_i$, the correction terms can be approximated by a linear function of $(S_i,\, \mathbf{X}_i)$. This implies a severe collinearity among the regressors which potentially increases the standard errors of the coefficient estimates (Puhani, 2000; Wooldridge, 2010). Exclusion restrictions are therefore helpful to improve the precision of the estimates. These are variables which directly affect the decision on the employment type, but do not directly influence sector choice and sectoral earnings. Similarly for the sectoral sorting process, it requires factors that affect the decision about the sector of work but can be excluded from the decision on the employment type and the sectoral earnings model.

4.3.3 Endogeneity Problem

The main variable of interest, educational attainment, S_i and S_i^2, is likely to be endogenous. That is, it is likely to be correlated with the unobserved individual productivity-enhancing characteristics, for the reasons that have already been discussed in the Chapter 3. In addition, as shown by Card (1999), individual heterogeneity can arise from differences in both marginal costs and returns of education. The effect of this heterogeneity together with the standard endogeneity problem can be added to the sectoral earnings equations as follows.

$$\ln w_{i,j} = \bar{\alpha}_{0,j} + \bar{\beta}_{1,j} \cdot S_i + \beta_{2,j} \cdot S_i^2 + \boldsymbol{\gamma_j} \cdot \mathbf{X}_i + \alpha_{i,0,j} + \left(\beta_{i,1,j} - \bar{\beta}_{1,j}\right) \cdot S_i + \varepsilon_{i,j} \quad (4.14)$$

for $j = A$ and NA, and where $\alpha_{i,0,j}$, $\left(\beta_{i,1,j} - \bar{\beta}_{1,j}\right)$, and $\varepsilon_{i,j}$ are unobserved. The presence of individual heterogeneity in both the intercept, $\alpha_{i,0,j}$, and the slope of the sectoral earnings function, $\beta_{i,1,j}$, causes a correlation between the unobservables and educational attainment.[15] As a result the OLS estimate of the return to education, $E\left(\beta_{1,j}\right) + 2 \cdot E\left(\beta_{2,j}\right) \cdot S_i$, would be biased.

Even if valid and informative instruments existed, the IV technique would only allow for unobserved individual characteristics, but not for heterogeneity in returns across individuals (Card, 1999, 2001). In the presence of this heterogeneity, the IV estimates would only be informative for sub-populations, whose decisions on education change as a result of the instrument (compliers). These IV estimates are known as the local average treatment effects (LATE) (Imbens and Angrist, 1994; Angrist and Imbens, 1995). Furthermore, with more than one endogenous variable (for instance, the non-linear effect of education), it is much harder to distinguish the compliers. As a result, the LATE literature typically does not allow for these non-linearities (as reviewed in Chapter 3).

In addition, to move to an estimate for the population (that is, the average treatment effect (ATE)), the IV approach requires stronger assumptions: Imbens and Wooldridge (2007a,b) make an additional assumption on a conditional covariance of unobserved heterogeneous returns and years of education.[16] Alternatively, Wooldridge (1997) assumes homoskedastcity of returns to education as well as linearity in years of education.[17]

Due to a limited number of instruments and the assumptions of heterogeneous and non-linear returns to education, I use the control function (CF) approach to estimate the average returns to education for the population. The CF approach was initially applied to the analysis of education by Garen (1984) and recently reviewed by Card (1999, 2001); Imbens and Wooldridge (2007a) and Imbens and Lemieux (2008). By treating endogeneity as an omitted variable, the endogeneity problem can be solved by including a control for the correlation

[15]The individual heterogeneity in the slope is also known as a random coefficient of educational attainment.
[16]However, Card (2001) suggests that this restrictive assumption is not suitable for a discrete endogenous variable and can be violated even when the endogenous variable is roughly continuous.
[17]Wooldridge (1997) also further requires that the individual heterogeneity terms are mean independent of the instruments, which is stronger than the uncorrelatedness that is usually required.

between omitted unobserved heterogeneity and years of education. Compared to the IV, the CF approach enables the population estimation (that is, the ATE) even when there exists a non-linear effect of education and/or an education choice is influenced by heterogeneous returns across individuals (Blundell et al., 2004; Imbens and Wooldridge, 2007a). It also provides more precise estimates when allowing for a non-linear effect and more robust estimates when the effect is heterogeneous across individuals (Card, 2001; Söderbom et al., 2006; Rijkers et al., 2010). This however requires parametric assumptions on the correlation between the unobservables in the sectoral earnings and schooling equations. Schooling is endogenous and can be modelled as a function of its marginal costs and benefits, which are also affected by some instrumental variables that do not affect earnings. Then, the optimal schooling decision can be written as

$$S_i = \boldsymbol{\theta} \cdot \mathbf{Z}_i + \zeta_i \tag{4.15}$$

where $\mathbf{Z}_i$ is the set of variables that influence schooling decision and ζ_i is the error term which also contains the unobserved heterogeneity in returns to education, $\left(\beta_{i,1,j} - \bar{\beta}_{1,j}\right)$, and unobserved individual characteristics, $\alpha_{i,0,j}$. This is also known as a first-stage regression similarly to that of the IV method.

Following Imbens and Wooldridge (2007a), the CF approach then assumes that $\mathbf{Z}_i$ is mean independent of $\alpha_{i,0,j}$, $\left(\beta_{i,1,j} - \bar{\beta}_{1,j}\right)$, and other factors in ζ_i, and that the conditional expectation of the unobserved heterogeneity components are linear in the residual, ζ_i, from the first-stage regression (Equation (4.15)). These can be written as

$$E\left(\beta_{i,1,j} - \bar{\beta}_{1,j} \mid S_i,\, \mathbf{X}_i,\, \mathbf{Z}_i\right) = \psi_1 \cdot \zeta_i \tag{4.16}$$

$$E\left(\alpha_{i,0,j} \mid S_i,\, \mathbf{X}_i,\, \mathbf{Z}_i\right) = \psi_0 \cdot \zeta_i \tag{4.17}$$

These assumptions imply

$$
\begin{aligned}
\ln w_{i,j} &= \bar{\alpha}_{0,j} + \bar{\beta}_{1,j} \cdot S_i + \beta_{2,j} \cdot S_i^2 + \boldsymbol{\gamma_j} \cdot \mathbf{X}_i + \\
&\quad \psi_0 \cdot \zeta_i + \psi_1 \cdot \zeta_i \cdot S_i + \varepsilon_{i,j} \qquad (4.18)
\end{aligned}
$$

Or,

$$
\begin{aligned}
E\left(\ln w_{i,j} \mid S_i,\, \mathbf{X}_i,\, \mathbf{Z}_i\right) &= \bar{\alpha}_{0,j} + \bar{\beta}_{1,j} \cdot S_i + \beta_{2,j} \cdot S_i^2 + \boldsymbol{\gamma_j} \cdot \mathbf{X}_i + \\
&\quad \psi_0 \cdot \zeta_i + \psi_1 \cdot \zeta_i \cdot S_i \qquad (4.19)
\end{aligned}
$$

which can be estimated using the predicted residual, $\hat{\zeta}_i$, from the first-stage regression (Equation (4.15)). The restriction that the conditional expectation is linear in the schooling residual can also be relaxed by adding a non-linear term in the schooling residual.

In addition, as my instrument, Z_i, is a binary variable indicating whether individuals are in the pre-education reform regime, a simple extension is to allow the conditional expectation to have different coefficients for each period (Card, 2001). Then, Equations (4.16) and (4.17) become

$$
E\left(\beta_{i,1,j} - \bar{\beta}_{1,j} \mid S_i,\, \mathbf{X}_i,\, Z_i\right) = \psi_{1,Z_0} \cdot (1 - Z_i) \cdot \zeta_i + \psi_{1,Z_1} \cdot Z_i \cdot \zeta_i \qquad (4.20)
$$

$$
E\left(\alpha_{i,0,j} \mid S_i,\, \mathbf{X}_i,\, Z_i\right) = \psi_{0,Z_0} \cdot (1 - Z_i) \cdot \zeta_i + \psi_{0,Z_1} \cdot Z_i \cdot \zeta_i \qquad (4.21)
$$

Therefore,

$$
\begin{aligned}
E\left(\ln w_{i,j} \mid S_i,\, \mathbf{X}_i,\, \mathbf{Z}_i\right) &= \bar{\alpha}_{0,j} + \bar{\beta}_{1,j} \cdot S_i + \beta_{2,j} \cdot S_i^2 + \boldsymbol{\gamma_j} \cdot \mathbf{X}_i + \\
&\quad \psi_{0,Z_0} \cdot \zeta_i + (\psi_{0,Z_0} + \psi_{0,Z_1}) \cdot Z_i \cdot \zeta_i + \\
&\quad \psi_{01,Z_0} \cdot \zeta_i \cdot S_i + (\psi_{1,Z_0} + \psi_{1,Z_1}) \cdot Z_i \cdot \zeta_i \cdot S_i \qquad (4.22)
\end{aligned}
$$

This sectoral earnings equation can be estimated by firstly running the first-stage regression (Equation (4.15)), and secondly estimating Equation (4.22), which includes the predicted residual from the first-stage regression, $\hat{\zeta}_i$, and its interactions with other variables as specified. Hence, in addition to the assumptions imposed in the CF technique, the identification strategy requires a valid instrument, which does not directly affect sectoral earnings.

It is important to discuss how strict assumptions required by the CF technique are compared to the IV approach. The mean independence exogeneity conditions (between Z_i and $\alpha_{i,0,j}$, $\left(\beta_{i,1,j} - \bar{\beta}_{1,j}\right)$, and ζ_i) impose a non-trivial restriction on both structural and first-stage reduced form error terms (Imbens and Wooldridge, 2007a,b). First, the mean independence of Z_i from the unobserved heterogeneity in individual productivity-enhancing characteristics and returns to education, $\alpha_{i,0,j}$ and $\left(\beta_{i,1,j} - \bar{\beta}_{1,j}\right)$, is stronger than the uncorrelatedness condition required in the general IV method. Second, assuming that Z_i is mean independent of the first-stage error term, ζ_i, is not an appropriate assumption when the endogenous variable is discrete (Card, 2001; Imbens and Wooldridge, 2007a,b).[18] As discussed earlier, a non-discrete endogenous variable is also required by the IV approach. Given that schooling is a count variable, most studies assume that it is sufficiently continuous to satisfy this condition (see for example Card (2001); Blundell et al. (2004); Söderbom et al. (2006); Kerr and Quinn (2010); Burger (2010); Falco et al. (2014)).[19]

4.3.4 Control Function Approach with a Double Selection Process

Numerous studies in the past estimate the LATE returns to education taking into account only the endogenous choice of education. Some more recent studies additionally allow for returns to education to be heterogeneous across different levels of education and individuals but attempt to provide estimates only for the employed population. On the other hand, most related studies that control for sample selection ignore the endogeneity problem.[20] Only a few studies attempt to address the issues of endogeneity and sample selection simultaneously.[21] Among these, most of them exclude the education variable from the selection process, but use all other exogenous variables from both the IV and sample selection regressions (Wooldridge,

[18] This independence condition also requires more knowledge about the process that generates the schooling outcome. That is, the first-stage process is unlikely to be a reduced form equation as it is in the IV approach.

[19] If the education variable is too discrete, one alternative would be to explicitly model the process that produced this discreteness, for instance by using an ordered logistic model. However, such techniques would typically require even stronger distributional assumptions.

[20] Buchinsky (2001) estimates the returns to education by quantile with a correlation for selection into wage-employment. Hence, while he allows returns to education to be heterogeneous across wage levels, he ignores the endogenous choice of education.

[21] In any case, these studies still do not allow for heterogeneous returns to education across individual and/or levels of education.

2002; Arabsheibani and Mussurov, 2007; Chen and Hamori, 2009). The education variable is dropped as they assume that it does not influence employment (Wooldridge, 2002), and therefore, they ignore the endogeneity problem in the selection equation. Other authors include education in the selection process but treat it as exogenous (García et al., 2001; Arrazola et al., 2003; Arrazola and de Hevia, 2008). To my knowledge, only Das et al. (2003) and Burger (2010) attempt to control for the endogeneity of education in both the earnings and selection processes, as well as correcting for the sample selection problem. Burger (2010) is similar to my analysis as he (i) allows for heterogeneous returns across levels of education and individuals; (ii) uses the CF approach to correct for the endogeneity in both the selection and earnings processes; and (iii) assumes homogeneous impact of education in the selection process for identification purposes. However, he considers only the selection into employment, while I model the selection into paid employment and sectors. My estimation proceeds in three main steps. First, I obtain the first-stage residuals for the entire sample, include them in the bivariate probit estimation of the two selection processes, and compute the selection correction terms (as discussed in Sub-section 4.3.2, Equation (4.11)). Second, I estimate the first-stage residuals for the earnings sample, which will be used to correct for the endogeneity problem in the sectoral earnings functions. Third, I include the sample selection correction terms together with the CF endogeneity and heterogeneity correction terms in the estimation of the sectoral earnings. Therefore, for sector $j = A$ and NA, the earnings equation to be estimated can be written as

$$\begin{aligned} E\left(\ln w_{i,j} \mid S_i,\, \mathbf{X}_i,\, \mathbf{Z}_i\right) &= \bar{\alpha}_{0,j} + \bar{\beta}_{1,j} \cdot S_i + \beta_{2,j} \cdot S_i^2 + \boldsymbol{\gamma_j} \cdot \mathbf{X}_i + \\ & \quad \psi_{0,Z_0} \cdot \hat{\zeta}_i + \left(\psi_{0,Z_0} + \psi_{0,Z_1}\right) \cdot Z_i \cdot \hat{\zeta}_i + \\ & \quad \psi_{01,Z_0} \cdot \hat{\zeta}_i \cdot S_i + \left(\psi_{1,Z_0} + \psi_{1,Z_1}\right) \cdot Z_i \cdot \hat{\zeta}_i \cdot S_i + \\ & \quad \delta_{ju} \cdot \hat{\lambda}_{i,P-j} + \delta_{Av} \cdot \hat{\lambda}_{i,A-j} \end{aligned} \tag{4.23}$$

4.3.5 Instruments and Exclusion Restriction

As discussed in the identification strategy, an instrumental variable is required to control for the endogeneity problem and an exclusion restriction is necessary for the sample selection corrections. The instrumental variable for the endogenous choice of education and heterogeneous returns needs to produce exogenous variation in workers' education. In other words, the instrumental variable must be associated with educational attainment while being uncorrelated with any unobserved factors influencing earnings. Similar to the other chapters in this thesis, exogenous variation in educational attainment arises from the change in compulsory education that was implemented in 1978 (Thailand's Education Reform Act of 1978). The 1978 Education Reform increased compulsory education from 4 years of lower primary to 6 years of primary school. Regardless of their observed and unobserved labour market characteristics and ability, individuals who are born in and after 1967 are forced to stay in school at least for two more years, compared to the older cohorts.[22] Hence, I use a binary variable indicating whether an individual is in the old or new education regime as an instrument for educational attainment.

The exclusion restriction in the paid employment selection equation are the variables that influence the utility of being an unpaid family worker and/or the reservation wage, but do not directly affect labour earnings. These variables are quite similar to those influencing labour force participation. In this chapter, I use household background characteristics which would affect the non-pecuniary utility of working at home, and thus being unpaid. They include the number of children and elderly in a family as well as the spouse's earnings. It is important to emphasise that the effect of some of these variables on the decision about paid employment may differ across gender. For example, in a country where most household heads are men, the number of young children is expected to be positively correlated with the probability of paid employment for men but negatively correlated for women.

The exclusion restriction used to identify the process of selection into sectors are the provincial labour market characteristics. Specifically, I use the lagged changes in the sectoral employment shares at a province level. The data do not contain useable information on parent background characteristics which are likely to be linked to sector of work.[23] Changes in the sectoral employment share reflect the relative labour demand for each sector, and hence, affect on the probability of being in that sector. It is possible that the exclusion restriction fails for the contemporaneous change in the sector share because earnings are

[22]In Thailand, children were required to start school in the calendar year in which they became eight years old. As a result, the change in compulsory schooling from 4 to 6 years implemented in 1978 affects individuals who are born in and after 1967.

[23]The literature has used various parent characteristics such as father's and mother's education or occupation, as well as some household characteristics such as siblings' occupation as exclusion restrictions the choice of sector. The Labour Force Survey (LFS) collects information on parents' education and occupation only for respondents who are still living with their parents. The LFS also asks about the sector of work and occupation in the previous year, which could be a good predictor of the current sector of work. However, there are too many missing values in these variables.

related to changes in the relative demand across sectors. Hence, I use the first and second lags of province-level changes in sectoral employment shares as exclusion restrictions in the process of selection into the sector of employment.

4.4 The Data and Descriptive Statistics

This section presents the key characteristics of the Thai labour market, including labour force participation, and types and sectors of employment, as well as characteristics of the Thai labour force such as educational attainment, age, gender, and residential location. First, I describe how the sample used in the empirical analysis is constructed. Second, I discuss the descriptive statistics of the main variables.

4.4.1 Data Source

In this chapter, I use labour force data from the LFS during the years 1985 to 2000.[24] The sample is restricted to employed workers, who aged between 20 and 65 years old during the survey years for two main reasons.

First, employed workers are the focus of this chapter as the research question involves returns to education and the roles of education in sorting workers into different sectors of work. Second, the age range of the sample is restricted to 20 to 65 years old. As discussed in the earlier chapter, while the working age population is typically defined as between 16 and 65 years old, I choose a lower bound of 20 years old because most Thai workers have completed education and enter the labour force by this age level. The retirement age for wage-employment in Thailand was around 60 during the period of study. However, self-employed workers tend to be active in the labour market beyond that age.

4.4.2 Descriptive Statistics

Labour Market Characteristics

As discussed in the Background and Data chapter (Chapter 2), the LFS information on labour force participation, employment, and unemployment shows high labour utilisation in

[24] The sector selection, employment selection, and earnings equations are estimated over observations from 1988 to 2000. This is because lagged province-level changes in sectoral employment shares, which are exclusion restrictions, are also computed from the LFS. The second lag of these variables is available first in 1988.

Thailand. First, the labour force participation rates for the population aged between 20 to 65 years old were well above 90 and 70 per cent for men and women respectively, which are much higher than in other developing economies. Second, unemployment rates were relatively low and stable even during the financial crisis. Third, low unemployment was not offset by high underemployment as observed in some other agricultural-based developing countries. On the other hand, the Thai labour market has experienced a high proportion of unpaid family workers, particularly in the agricultural sector and among women. This evidence suggests possible non-random selection into paid employment which supports my empirical approach of estimating the returns to education across sectors as well as the effects of education on sorting individuals into different sectors.

Earnings, Sector of Employment, and Educational Attainment

Tables 4.1 to 4.3 describe the characteristics of the sample used in the analysis by gender, sector of work, and type of employment. Table 4.1 shows that the proportion of employed workers engaged in agriculture has been substantial but gradually and continuously decreasing over time. This decline was slightly faster for women at around 20 percentage points, compared to a decline of 15 percentage points observed among men. The proportion of men is slightly higher in the sample, while paid workers are predominantly male, reflecting a substantial number of employed women working for their family's farm or business and being unpaid. In 1985, around 58 per cent of female employed workers were unpaid, decreasing over time to 40 per cent by the end of 2000. With relatively stable rates of labour force participation, unemployment, and underemployment, this suggests significant movement of female unpaid workers towards paid jobs. Furthermore, this decline is roughly in line with the share of female workers moving from agriculture to non-agriculture. Meanwhile, the average level of education increased over time for both male and female workers. A faster increase in years of education attained by female workers narrowed the gender education gap. Average hourly earnings grew faster for women, at about 3.6 per cent annually during the years 1985 and 2000, compared to the annual growth rate of 3.1 per cent for men's average hourly earnings.[25]

Table 4.2 summarises the labour force characteristics by the sector of work. Both male

[25] The median hourly earnings grew slightly faster than the average hourly earnings over the same period. The annual growth rates of median hourly earnings were about 3.7 and 3.3 per cent for women and men, respectively.

and female non-agricultural workers have attained significantly more years of education and have been paid substantially higher than those working in agriculture. The average level of education in non-agriculture was well above completed primary (6 years) and increased by around 2 years over the 16-year period. At the same time, agricultural workers attained less than primary education on average and their educational attainment increased by around 1.2 years over the same period. The sectoral hourly earnings gap increased over time for both men and women, reflecting faster earnings growth in the non-agricultural sector.[26] It is important to note that while hourly earnings in the agricultural sector have been relatively similar for men and women, the non-agricultural hourly earnings of men have been substantially above those of women.

Table 4.2 also presents the shares of urban dwellers as well as paid workers in each of the sectors of work. Most agricultural workers resided in rural areas, while more than half of non-agricultural workers lived in urban areas. The proportion of paid employment for men has been high and quite stable, at around 70 per cent for the agricultural sector and above 90 per cent for the non-agricultural sector. On the other hand, less than a third of female agricultural workers were paid while around 80 per cent of those in the non-agricultural sector were paid. Paid employment increased substantially over time for women in both sectors.

The cross-tabulations of these labour force characteristics by employment type are shown in Table 4.3. The levels of education of paid and unpaid men were quite similar, while the average years of education among female paid workers were significantly higher than female unpaid workers. In addition, men who were in unpaid employment were much younger than those working in a paid job. Meanwhile, the average age between the two types of employment were more similar for women. Most male and female unpaid workers were engaged in agriculture, while female unpaid workers were less likely to be in agriculture, compared to men.

4.4.3 Data Summary

To summarise, the LFS data highlight the structural transformation in employment and earnings during the period between 1985 and 2000. First, there has been a continuous move-

[26] For employed men, average agricultural and non-agricultural earnings grew 1.9 and 2.4 per cent annually, respectively. On the other hand, women's agricultural and non-agricultural earnings grew 2.7 and 3.0 per cent annually.

ment of employment out of the agricultural sector. Second, earnings differentials across sectors increased over time. These two stylised facts potentially reflect higher productivity in the non-agricultural sector. This together with a significant increase in educational attainment, especially in the non-agricultural sector, also suggests a role for human capital in facilitating this structural change. In addition, the descriptive statistics confirm systematic differences between male and female workers. While unemployment has remained low, the data suggest a possible non-random selection into paid versus unpaid employment, which is also related to the level of education, particularly among women.

4.5 Empirical Results

This section investigates the effects of education on the type and sector of employment as well as the sectoral returns to education in Thailand during the period 1985 to 2000. The impacts of education on these outcomes of interest are estimated for men and women separately. First, the impacts of education on the type and sector of employment are estimated jointly using a bivariate probit model as describes in the earlier section. I also discuss the exclusion restriction for the double selection model in more detail. Second, I estimate the returns to education, separately for the agricultural and non-agricultural sectors, taking into account the endogeneity and sample selection concerns as well as allowing for heterogeneous returns across levels of education and individuals. The sample selection correction uses the results from the bivariate probit estimation in the first part. Finally, the relevance of the results to the two aforementioned structural change growth models is discussed.

4.5.1 Measuring the Effects of Education on Type and Sector of Employment

I begin the empirical analysis by exploring the impacts of education on the selection into paid employment and the agricultural sector. As potential earnings could affect the choices of sector and employment type, I include the explanatory variables from the earnings regressions both selection processes. They are years of education and its square (to capture the non-linear effects of education on the outcomes of interest), age and its square, hours of work per week, a rural-urban indicator, province and year dummies, and marital status. I also specify a number of exclusion restrictions which are important for the identification strategy of the subsequent sectoral earnings model. These variables are included only in the selection regressions and not the sectoral earnings model. For the selection into paid employment, these variables are the number of children below 6 years old and between 6 and 13 years old, the number of elderly in the household, and the earnings of the spouse. The variables included only in the sector choice process are first and second lags of the province-level changes in manufacturing and services employment shares.

I allow for the two selection processes to be correlated to each other. They are estimated jointly using a bivariate probit model. Tables 4.4 and 4.5 show the results for male female workers respectively.[27] The second and third columns of each table present the initial results which ignore the potential bias due to unobserved heterogeneity that affects both education and selection. The fourth and fifth columns of each table display the results which correct for the endogeneity problems by adding the CF variables obtained from estimating the residuals, ς_i, of the first-stage regression of educational attainment for all employed workers (Column 1 of each table).[28] The discussion of the impacts of education on the two selection processes is mainly based on the graphical results in Figures 4.1 and 4.2 (for men and women, respectively). These graphs present the predicted probability of being in paid employment and the predicted probability of working in the agricultural sector as educational attainment varies, holding other regressors at their mean values. The predicted conditional probability of working in the agricultural sector given that a worker is paid is also displayed in the same figures. The coefficient estimates of the bivariate probit model are easier to interpret from the graphs, especially when the non-linear terms are included.[29]

For men, Figure 4.1 highlights the role of education in reallocating workers out of agriculture and into non-agriculture, while its impact on sorting workers into paid employment is quite minimal.[30] The shape of the education-probability profile does not change much for the selection into paid employment after correcting for the endogeneity of education. For the selection into agriculture and with controlling for endogeneity, the decline in the probability

[27] The empirical results, when assuming that the two selection processes are independent are presented in Table 4.A.1 and Figure 4.A.1 in the Appendix. This is not the preferred specification as the correlation between the unobservables of the two processes is found to be statistically significant (to be discussed in the following paragraphs).

[28] The first-stage regression is estimated for men and women separately.

[29] Without a no non-linear term, the bivariate probit or probit results could simply be presented as average marginal effects. However, the marginal effects are allowed to change in a more flexible way in this chapter due to the non-linear effects of education.

[30] The results in Figure 4.1a are based on the coefficient estimates in Columns 2 and 3 of Table 4.4, while those in Figure 4.1b are the coefficient estimates in Columns 4 and 5 of Table 4.4.

of working in the agricultural sector as education increases is particularly large for those who attained 6 years of primary education and below. On the other hand, this negative impact of education on agricultural employment becomes somewhat smaller for men with higher-than-primary education in the specification with the endogeneity correction. This suggests that the correlation between education and the unobserved factors encouraging individuals to work in agriculture is positive for below-average educated workers, while it is negative for those who have at least primary education.[31]

Holding other factors constant and after correcting for endogeneity, male workers with 6 years of primary education have a 30 percentage point lower probability of working in agriculture compared to those without education (Figure 4.1b). An additional year of education at the primary level would reduce the probability of working in agriculture by around 5 percentage points from 50 to 45 per cent. For male workers who completed lower secondary school (9 years) and high school (12 years), the probability of working in agriculture would decline further to 37 and 28 per cent, respectively. On the other hand, the probability of being in paid employment is well above 80 per cent for men at all levels of education. The likelihood of being a paid employee declines only slightly from 90 to 84 per cent as educational attainment increases from 0 to 4.5 years. This likelihood is then increasing gradually with education, and reaches nearly 100 per cent for male workers who completed at least high school (12 years). Considering the conditional probability, the probability of working in agriculture given that an individual is being paid is lower than the marginal probability of working in agriculture, especially for male workers with high school education and below. This implies a significant negative correlation between the two selection processes as also shown in Table 4.4 by the estimates for ρ_{uv}.

Table 4.4 also presents the estimates of other coefficients in the male workers' bivariate probit model. An additional year of age, as a proxy for labour market experience, increases the probability of being in paid employment, but its positive effect decreases over time and become zero at around 53 years old. At the same time, higher age reduce the probability of working in agriculture until around 41 years old on average. Taken together, this suggests that the youngest workers are more likely to be unpaid and work in agriculture. The exclusion restriction variables are informative about the two selection processes. The number

[31] This unobserved factor could be viewed as agriculture-specific ability or individual preferences towards agricultural work.

of children positively affects men's probability of working for a paid job while the number of elderly in the household and spouse's earnings tend to have very small negative effects. For the selection into the agricultural sector, faster growth in the share of manufacturing employment decreases the probability of working in agriculture, while faster growth in the share of services employment has the opposite effect. As most of the manufacturing workers are wage-employed and a substantial proportion of services workers are self-employed, the empirical results additionally suggest that faster growth in wage-employment and slower growth in self-employment would help moving male workers out of the agricultural sector.[32]

For women, Figure 4.2 emphasises the role of education in both reallocating workers across sectors and sorting them into paid employment.[33] It also suggests the importance of the endogeneity problem as the shape of the education-predicted probability profiles changes substantially after using the control function correction. When treating educational attainment as exogenous, an additional year of education increases the probability of being in paid employment for all levels of education and monotonically decreases the probability of working in the agricultural sector for female workers who attained an undergraduate degree (16 years) and below. However, the impact of education on the probability of being in paid employment becomes much lower after correcting for the endogeneity of education, suggesting a positive relationship between the unobserved factors influencing paid employment and education. The smaller negative effect of education on the probability of being in agriculture implies a negative relationship between educational attainment and the unobserved factors influencing the selection into agriculture. Similarly to the results for male workers, this negative association emphasises the selection into sectors based on sector-specific ability, which in turn is significantly correlated with educational attainment and/or the ability that is useful for studying.

More specifically, for female workers after controlling for endogeneity (Figure 4.2b), the effect of education on the probability of being in paid employment is positively concave, and eventually becomes negative after 12 years of education. The probability of working for a paid job increases by around 15 percentage points, from 40 per cent for women with no education to around 55 per cent for women with high school education (12 years). The role of

[32] The positive effect of the increasing share of services employment on the probability of working in agriculture may additionally be explained by the fact that a substantial proportion of services workers in Thailand are self-employed and are concentrated in provinces which are also agricultural base.

[33] The results in Figure 4.2a are consistent with the coefficient estimates in Columns 2 and 3 of Table 4.5, while those in Figure 4.2b are consistent with the coefficient estimates in Columns 4 and 5 of Table 4.5.

education in sorting workers into paid employment is larger for women than for men. At the same time, the probability of working in agriculture declines as years of education increase, in particular for female workers who completed at least 4 years of lower primary education. The probability of being an agricultural worker increases slightly from 52 per cent when having no education to 54 per cent when attaining lower primary education (4 years), and then declines gradually and monotonically to around 45 per cent when completing high school education (12 years). In addition, conditioning on being in paid employment, the probability of working in the agricultural sector declines substantially for all levels of education, reflecting a negative correlation between the two selection processes. This negative correlation is larger for female than male workers (Table 4.5, the estimates for ρ_{uv}).

The other coefficient estimates of the female workers' bivariate probit model are presented in Table 4.5. Most of the results are similar to men, in terms of the direction of the effects, with the exception of the number of the elderly in the household that positively affects the probability of working for a paid job for women. Similarly to male workers, the variables used in the exclusion restriction are informative about the two selection processes. It is important to also note that the number of children aged between 6 to 13 years old encourages female workers to be in paid employment, but its effect is much lower than that of male workers.[34]

4.5.2 Estimating Sectoral Returns to Education

Next, I examine how educational attainment affects earnings. The returns to education estimated in this chapter differ from the RD returns in Chapter 3 in a number of ways. First, education can affect earnings in a non-linear fashion as the square of years of education is included in the earnings function. Second, the returns to education are allowed to be heterogeneous across individuals. By allowing for heterogeneous returns across levels of education and individuals, the CF method is used to correct for the endogeneity of education. Third, to test the role of human capital on structural transformation, the returns to education are estimated separately for the agricultural and non-agricultural sectors. By non-randomly dividing the sample of workers into two groups, a correction for self-selection into each sector

[34] The positive effects of the number of children on paid employment may be somewhat counter-intuitive as mothers with young children may be more likely to stay at home, taking care of their children, and thus only work as an unpaid family worker. This actually holds in the descriptive statistics. That is, the proportion of female paid employed workers declines as the number of children increases. However, once controlling for other explanatory variables, these effects disappear, reflecting a correlation between the number of children and other regressors. For example, the number of elderly, who might also be involved with child care, has a positive effect on the probability of paid employment.

of work is required so that the estimates of the sectoral returns to education are comparable across sectors and informative about the overall labour market. Note that the sectoral returns are estimated for men and women separately as their decisions on the employment type and sector of work are shown to be different.

It is important to note that the relationship between education and earnings is assumed to be time invariant, although year dummies are included to capture the time trend (for example, in overall productivity). Without this assumption, one would need a separate instrument for each of the interaction terms between year dummies and education (and education squared). This is a strong assumption because structural transformation could change the human capital augmented productivity, and hence the returns to education. Therefore, my estimates are the average returns over this period. In other words, if these returns increase over time, I would overestimate the returns for the earlier years and underestimate for the later years.

The empirical results on sectoral returns to education are presented in Table 4.6 for the agricultural sector and Table 4.7 for the non-agricultural sector. As the returns are non-linear in years of education, I include and refer most of the result discussion to a graphical presentation of the estimated marginal returns in Figures 4.3 and 4.4 for the two sectors. Increasing (upward sloping) marginal returns imply a convex education-earnings profile, while decreasing (downward sloping) marginal returns imply a concave education-earnings profile.

4.5.2.1 OLS Sectoral Returns to Education

I begin the analysis of sectoral returns to education by ignoring the potential biases in the estimates due to non-random selection into paid employment and sector of work as well as unobserved heterogeneity in productivity-enhancing characteristics and in returns to education. Similar to Chapter 3, I estimate a semi-logarithmic specification of the real hourly earnings equation (Mincer, 1974) using OLS. Years of education and its square are the explanatory variables of interests as their coefficient estimates are the returns to education.[35] The OLS estimates thus only control for observed characteristics.

[35] Following Chapter 3, the other control variables are age and its square as a proxy of labour market experience; hours of work per week to control for the labour supply effect; and a rural-urban indicator. In addition, I include province and year dummies to factor out macroeconomic effects on earnings that may occur differently by location and time. A marital status dummy is also included.

Table 4.6, Columns 2 and 7 show the OLS results for men and women respectively, in the agricultural sector. The left panels of Figure 4.3 present how these OLS estimated agricultural returns vary with education. The OLS returns to education in the agricultural sector are positive but decline as years of education increase for men (Figure 4.3a). At 4.4 years of education, which is around the average level of education for male agricultural paid workers, the return to education is 6.2 per cent. The returns remain positive but decline to 5.9 per cent for those who attained 6 years of primary education and to 4.8 per cent for those who completed 12 years of education. Meanwhile, the OLS returns to education for female agricultural workers increase with years of education (Figure 4.3b). Despite having a convex education-earnings profile, their returns are lower than those of male agricultural workers for most levels of education (until lower secondary school, that is, 9 years of education). The return to education at their average level of education (4 years) is around 3.5 percent. It increases to 4.2 per cent at 6 years of primary education and further increases to 6.1 per cent for those who completed 12 years of education.

On the other hand, Table 4.7, Columns 2 and 7 suggest that the OLS education-earnings profile is convex for both men and women in the non-agricultural sector. The left panels of Figure 4.4 present how these OLS estimated non-agricultural returns vary with education. The OLS non-agricultural returns to education are higher for women but the gap declines with years of education. For male workers in the non-agricultural sector (Figure 4.4a), the OLS return to education at their average years of schooling (7.8 years) is 7.4 per cent, and it increases to 9.9 and 12.3 per cent for those with high school (12 years) and undergraduate education (16 years) respectively. The average years of education for female non-agricultural workers is also 7.8 years, similar to that of men, while women's return at this level of education is higher at 9.2 per cent (Figure 4.4b). The returns to education for women working in the non-agricultural sector increase to 11.2 and 13.1 per cent when they completed high school (12 years) and undergraduate education (16 years).

It is important to note that, for men, the observed relationship between years of education and earnings is concave in the agricultural sector and convex in the non-agricultural sector. That is, the returns to education are higher in the agricultural sector for a low level of education. But as years of education increase, the agricultural returns decline while the non-agricultural returns increase. The OLS returns to education in the two sectors are roughly equal at 5.5 years of education. This suggests that, based on observables characteristics, education is valued relatively more in the non-agricultural sector. However, the higher returns in the non-agricultural sector could be misleading if male workers select into the sector of work based on their unobserved characteristics that are correlated with the unobserved heterogeneity in productivity-enhancing characteristics. On the other hand, the OLS education-earnings profiles in both sectors are convex for women. Returns to education in the non-agricultural sector are higher than those in the agricultural sector for all levels of education. In addition, a stronger convexity in the non-agricultural sector implies that the non-agricultural returns to education increase at a faster rate than those in the agricultural sector, so the gap in returns widens as years of education increase. Of course, the OLS returns for women, in both sectors, may be subject to various biases. In particular, the substantial proportion of female unpaid family workers, especially in the agricultural sector, causes a substantial amount of non-random selection in the sectoral earnings equation. If female unpaid family workers differ systematically from female paid workers, the returns to education would be subject to a self-selection bias. Furthermore, the sectoral returns for men and women are likely to be biased due to the fact that educational attainment is likely to be correlated with unobserved factors that affect productivity and thus earnings, and that the returns to education may be heterogeneous across individuals.

4.5.2.2 Sectoral Returns to Education with a Correction for Sample Selection Bias

As described in the empirical framework (Section 4.3), the OLS estimates of the sectoral returns could be biased if the observations included in the analysis differ systematically from the excluded observations. I correct for sample selection bias for the selection into paid employment and sector of work simultaneously. First, the two selection processes are estimated jointly using the bivariate probit model (Sub-section 4.3.2). Second, the correction terms, $\lambda_{\cdot,j}$, are generated from the estimates of the double selection model (in Tables 4.4 and 4.5). As found in Sub-section 4.5.1, the two selection processes are highly correlated, so I correct for them jointly.

Table 4.6, Columns 3 and 8 (and also the left panels of Figure 4.3) show the earnings function results with the correction for sample selection bias for men and women in the agricultural sector. Compared with the simple OLS, the effect of education on earnings has a lower linear effect and becomes more concave for men (Figure 4.3a). At the average level of education of 4.4 years for male agricultural workers, the return to education declines significantly to around 2.5 per cent. Moving along the education distribution, it decreases at a faster rate than the OLS return, reaching zero at around 8 years of education and then becoming negative. Meanwhile, the effect of education on earnings for women in the agricultural sector has a lower linear effect and is more convex after controlling for sample selection bias (Figure 4.3b). That is, the returns to education for female agricultural workers increase faster as education increases. The more convex education-earnings profile together with the lower linear effect indicates that in the OLS the returns to education are overestimated for those with low education (3 years and less), and underestimated for those with medium and high education. At the average level of education of 4 years for female agricultural workers, the return to education increases slightly to 3.9 per cent, compared to the OLS return of 3.5 per cent.

It is important to highlight that, in the agricultural sector, the direction of the selection bias for men is opposite to that for women. Hence, the selection processes jointly affect the returns in a different way across gender.[36] First, the upward bias in the OLS returns to education for male agricultural workers suggests that male workers who are unpaid or work in the non-agricultural sector would have obtained less agricultural returns had they worked for a paid agricultural job. As the majority of excluded workers are non-agricultural workers, this could imply that male non-agricultural workers have less skills needed for agriculture and lower agricultural returns to education, compared to male agricultural workers. Second, the downward bias in the OLS returns for female agricultural workers show that those female workers who are unpaid or work in the non-agricultural sector would have higher agricultural returns to education, compared to female paid agricultural workers. As the sample of female workers is rather mixed, it is more difficult to identify which of the selection effects dominates. One explanation for the downward bias in the OLS returns would be a negative correlation between education and motivation in the sample of female paid agricultural workers. An alternative and complementary explanation could be that female non-agricultural workers have an absolute advantage in the agricultural sector over those in the agricultural sector but still choose to be in non-agriculture based on their individual comparative advantage.

[36] As the two selection processes are highly correlated, it is unclear how to estimate them separately.

Table 4.7, Columns 3 and 8 (and also the left panels of Figure 4.4) present the earnings function results with sample selection correction for the non-agricultural sector. The results for men and women are quite similar. The effects of education on non-agricultural earnings are more convex with a smaller linear effect after controlling for sample selection. The smaller linear effect implies that the OLS non-agricultural returns to education are upwardly biased, but the positive bias declines as years of education increase and eventually becomes negative (downward) at a high level of education (Figure 4.4). For men in the non-agricultural sector, their OLS returns are larger than the selection-corrected returns when they attain less than high school education (12 years). In contrast, for men who attain more than 12 years of education, the non-agricultural returns are higher than the OLS. For women in the non-agricultural sector, the positive selection bias becomes negative at around 11 years of education. At the average levels of education of 7.8 years in the non-agricultural sector, the selection-corrected returns decline to around 6.6 and 8.2 per cent for men and women respectively.

These results suggest that the majority of workers who are not in the non-agricultural sector or work for an unpaid job would have had lower non-agricultural returns had they obtained a paid non-agricultural job.[37] These lower returns reflect a strong positive sorting into the non-agricultural sector based on both comparative and absolute advantage.

4.5.2.3 Sectoral Returns to Education with a Correction for Endogeneity of Education

The other potential bias comes from the endogenous choice of education, due to unobserved productivity-enhancing characteristics (which are correlated to education) and heterogeneity in returns to education across individuals. I adopt the CF approach to obtain consistent estimates of the sectoral returns to education. First, the first-stage regression of education is estimated by using the 1978 Education Reform to provide exogenous variation in educational attainment. Second, the CF method generates additional variables, from the first-stage regression of education, to control for the endogeneity bias.

Table 4.6, Columns 1 and 6, for men and women respectively, show the results of the first-stage regression for paid workers in the agricultural sector. The results of the first-stage

[37] More than 95 per cent of the employed workers (both paid and unpaid) attained less than or equal to 12 years of high school.

regression for paid workers in the non-agricultural sector are shown in Table 4.7, Columns 1 and 6. In addition to the dummy indicating whether an individual was in school before or after the education reform, all other variables from the earnings equation are included. The change in compulsory education in 1978 has significant positive effects on years of education for paid employed workers in both sectors. Its effects are significantly higher for the paid agricultural workers, who have lower educational attainment on average. The education reform increases educational attainment by around 0.78 and 0.76 years for male and female agricultural paid workers respectively. At the same time, it raises educational attainment of male and female workers in the non-agricultural sector by around 0.10 and 0.13 years.

Abstracting from the sample selection bias, the correction for the endogeneity of education shows that the OLS returns to education are downwardly biased in all cases, except for the returns for female agricultural workers.[38] Table 4.6, Columns 4 and 9 (and also the right panels of Figure 4.3) display the agricultural earnings results after the endogeneity correction for men and women respectively. Compared to the OLS returns, the effects of education on agricultural earnings for men become more concave but the linear effect is also larger, meaning that there exists a downward bias in the OLS estimates but the negative bias is decreasing in years of education (Figure 4.3a). At the average level of education of 4.4 years for male paid agricultural workers, the return to education increases to around 10.4 per cent, compared to 6.2 per cent found in OLS. The endogeneity-corrected agricultural returns for male workers decline at a faster rate as education increases, being around 9.7 and 7.3 when educational attainment is primary (6 years) and high school (12 years) respectively. More interestingly, the effects of education on agricultural earnings, for women, turn from convex in the OLS to be concave when taking into account the endogeneity problem (Figure 4.3b). The linear effect also becomes much larger than in the OLS, so the endogeneity-controlled returns are higher than the OLS at low levels of education suggesting a downward bias in the OLS. On the other hand, they are lower than the OLS at medium and high levels of education. That is, there is an upward bias in the OLS at these education levels. At the average level of education of 4 years for female agricultural workers, the returns after correcting for the endogeneity problem are around 6.4 per cent, nearly double those found in the OLS estimation. The two estimated returns are equal at around 5.5 years of education.

[38] To be precise, for women, the downward bias in the OLS agricultural returns still exists for those with less than 6 years of education.

The CF agricultural returns decline to around 3 per cent for female workers with 6 years of primary education and to around -7.2 per cent for those who completed high school (12 years).

The endogeneity correction results for non-agricultural earnings are presented in Table 4.7, Column 4 and 9 (and also the right panels of Figure 4.4) for men and women respectively. The education-earnings profile of male non-agricultural workers becomes more convex, and the linear effect of education also increases (Figure 4.4a), suggesting a downward bias in the OLS. For women working in the non-agricultural sector (Figure 4.4b), the education-earnings profile also becomes more convex, while the linear effect of education decreases. Hence, the OLS non-agricultural returns for women are upwardly biased at low levels of education, but this positive bias decreases as years of education increase. The bias is zero at around 4 years of education and the OLS return becomes downwardly biased at high levels of education. At most levels of education, the downward bias is larger for men than women. For both men and women in the non-agricultural sector, at the average level of education of 7.8 years, the non-agricultural returns are around 12.8 and 11.7 per cent respectively.

There are a few potential explanations for the downward bias in the returns to education observed in male workers in both sectors, female non-agricultural workers, and female agricultural workers with low education. First, it could be that education and unobserved productivity-enhancing characteristics are compensatory. That is, highly able workers attain less education as their foregone earnings due to more education are substantial. This explanation is the opposite to the conventional ability bias which results in an upward bias in the OLS estimates. Second, measurement error may explain the downward bias when the returns are positive (Griliches, 1977).[39] I would expect that measurement error affects every sub-sample of workers in a similar way. However, the OLS returns for female agricultural workers are positively biased.[40] Third, while the CF method tries to recover the average treatment effect, it may not be able to fully control for the effects of heterogeneous returns to education across individuals.[41] In this case, the estimated returns to education are only informative for some high returns individuals who are affected by the education reform.[42]

[39] Measurement error in educational attainment result in an attenuation bias in the OLS returns to education when ability bias is relatively small (Griliches, 1977).

[40] One may argue that measurement error is likely to be higher for older cohorts (known as recall bias). However, the female agricultural workers are not substantially different from other workers, and they are even slightly older.

[41] For example, if the heterogeneity in returns also occurs for the non-linear effects of education.

[42] These individuals may not choose to study without the reform despite high returns because of high costs

However, comparing these results to Chapter 3 which provides the LATE estimates, this does not seem to be the case.

In sum, these potential explanations are not fully conclusive. It is also important to remember that the CF approach rests on strong assumptions of the correlation between the unobservables in the selection and earnings processes. Furthermore, these results have not taken into account the sample selection bias, which is shown to be statistically significant in the OLS estimation where education is exogenous. There could also be potential interaction effects between the endogeneity bias and the non-random selection.

4.5.2.4 Sectoral Returns to Education Controlling for Both Non-random Selection and the Endogeneity of Education

Both the selection into sector and paid employment and the endogenous choice of education have significant impacts on the returns to education. Therefore, the final estimation method, which is the preferred specifications, controls for both at the same time. The selection correction terms generated from the bivariate probit model of paid and agricultural employment (Columns 4 and 5 of Tables 4.4 and 4.5) and the CF terms (and its interaction with years of education) from the first-stage education regression (Columns 1 and 6 of Tables 4.6 and 4.7) are included in the sectoral earnings model. Table 4.6, Columns 5 and 10 (and also the right panels of Figure 4.3) present these results for men and women in the agricultural sector. The corresponding results for the non-agricultural sector are in Table 4.7, Columns 5 and 10 (and also the right panels of Figure 4.4). In all cases except for women working in agriculture, the effects of the selection correction on the returns to education are quite similar whether they are applied to the OLS or the CF specifications.

In the agricultural sector, male workers have a more concave education-earnings profile compared to the OLS (Figure 4.3a). As the linear effect is greater, the agricultural returns at a low level of education (less than 4 years of lower primary education) are higher than the OLS returns. At and after their average level of education (4 years), a stronger concavity results in the downward adjustment of the returns, compared to the OLS. Due to the concavity of the education-earnings profile, the returns to education become negative after 7.5 years of education. The returns to education for female agricultural workers after controlling for both potential biases look quite similar (albeit being more concave) to those of men in the

of education. In this case, the control function (CF) estimates may give the LATE returns to education.

same sector. The returns to education for female agricultural workers are positive, but this positive effect declines as years of education increase and eventually becomes negative at around 6.5 years of education (Figure 4.3b). These negative agricultural returns are the results from extrapolating the estimated relationship between earnings and education beyond the observed range of educational attainment. Over the period of interest, the proportion of agricultural paid workers with primary education (6 years) or less ranges from 90 to 97 per cent for men and from 94 to 99 per cent for women.

For non-agricultural earnings (both men and women), the education-earnings profile is more convex after correcting for the selection and endogeneity bias (Table 4.7, Columns 5 and 10), compared to the OLS. The linear effects of education change significantly from positive to negative, which decreases the non-agricultural returns at a low level of education substantially. As a result, the returns to education for men with 4 years of lower primary education and less, and for women with less than 2 years of education, are negative (Figure 4.4). Because of the stronger convexity, the returns are higher at medium and high levels of education compared to the OLS. At the average years of education (7.8 years), the non-agricultural returns are slightly higher than the OLS, being at around 8.4 and 10 per cent for men and women respectively.

4.5.3 The Implications of the Results for the Structural Change Growth Models

The empirical results show that education has significant positive effects on both selection into non-agriculture and on earnings, which support the underlying assumptions of the human capital augmented structural change growth model.

First, an additional year of education significantly increases the probability of working in the non-agricultural sector for both men and women (Sub-section 4.5.1). The role of education in reallocating workers towards the non-agricultural sector is larger for men. In addition, education helps sorting female workers into paid employment.

Second, returns to education are found to be higher in the non-agricultural sector for workers with at least primary education (Sub-section 4.5.2). This emphasises the role of education in moving workers from agriculture to non-agriculture, and suggests that this reallocation is likely to happen only after a certain level of schooling. Figure 4.5 summarises the estimated returns to education in the agricultural and non-agricultural sectors, for men and women separately, after correcting for both the sample selection bias and the endogeneity of education. The returns to education differ substantially across the two sectors of work, and they vary with the level of education. Different returns to education across sectors and levels of education suggest that the two sectors value education at various levels differently. Education below the primary level (6 years) is more valuable in agriculture for both male and female workers as the agricultural return is higher than the non-agricultural returns at that level of education. For men, the returns to the first four years of education are negative in the non-agricultural sector. However, as returns are decreasing in education for agriculture and increasing for non-agriculture, the returns in the non-agricultural sector are higher when educational attainment is equal to the primary level and higher. Hence, higher levels of education are valued more in the non-agricultural sector. Due to decreasing returns and extrapolating the estimates beyond the observed range of educational attainment, education has a negative impact on agricultural earnings after 7 years for both men and women.

4.6 Conclusion

In this chapter, I have outlined a framework that links the macroeconomic structural transformation growth models to the micro-level labour market outcomes. I investigated the importance of human capital to the structural transformation in Thailand between 1985 and 2000, using labour force data. To do so, I examine (i) how education, as a measure of human capital, helps reallocating workers from the agricultural to the non-agricultural sectors; and (ii) how returns to education differ across the two sectors. The latter results show how human capital, at different levels, is valued in the two sectors of production. I use a CF approach to recover the population estimates, allowing for heterogeneous returns to education across sectors of work, levels of education, gender, as well as individuals. The 1978 Education Reform, which increased compulsory education by two years, is used to provide exogenous variation in individuals' education. In order to be able to compare the two sectoral returns to education, I also correct for sample selection bias, which may arise from non-random selection into paid employment and sectors, using a double selection model.

I find that education significantly increases the probability of being in the non-agricultural sector for all workers, while its positive impact on the likelihood of being in paid employment is only significant for female workers. This could be because most men work for a paid job in any case, while there is a substantial proportion of women working for their families and being unpaid. The effects of education on earnings are found to be significantly non-linear and heterogeneous across the two sectors, reflecting that different levels of human capital are relevant to each of the sectors. The education-earnings profile is concave and convex in the agricultural and non-agricultural sectors respectively. In other words, the returns to education in the agricultural sector are decreasing in education while those in the non-agricultural sector are increasing in education. More specifically, the agricultural returns are higher than the non-agricultural returns for education below the primary level, while the opposite holds for education at the primary level and above. The empirical results support the human capital augmented structural change model, which emphasises the role of human capital in transferring workers towards the higher productivity sector.

It is important to highlight the role of unobserved productivity-enhancing characteristics, individual heterogeneity in returns, and non-random selection, as they all significantly affect the estimated returns in the earnings model. It is difficult to disentangle how each of these factors affects the bias in the OLS estimates, as they are correlated to each other. Nevertheless, the results suggest that, for the non-agricultural sector, the endogeneity of education is likely to cause a downward bias in the OLS, while non-random self-selection leads to an upward (downward) bias at low (high) levels of education. On the other hand, for the agricultural sector, the endogenous choice of education causes a downward bias among men, while the direction of the bias is unclear for women. I should also emphasise that the CF estimates, while being applicable to the population, rest on relatively strong assumptions on the correlations between the unobservables in the sectoral earnings and educational choice processes. Without these assumptions, it is only possible to recover the estimates for sub-populations.

To conclude, my results suggest that human capital is crucial for economic growth in Thailand between 1985 and 2000, which is driven by structural transformation. Human capital facilitates the reallocation of labour towards paid employment and the more productive sector, whose expansion spurs economic growth. Human capital at various levels is valued differently in these sectors. Medium and high levels of human capital are required for jobs in the more productive sector.

Table 4.1: LFS Descriptive Statistics of Employed Workers (Age 20-65)

Male workers	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Avg. hourly earnings	23.46	23.26	23.47	27.85	26.11	28.82	28.92	32.58	34.58	36.71	40.74	40.55	46.48	38.40	39.20	37.39
Median hourly earnings	14.08	13.67	13.47	16.37	16.40	17.83	17.61	19.47	20.60	22.39	24.71	26.08	26.11	23.82	23.91	22.97
	(34.18)	(34.74)	(34.76)	(46.94)	(39.18)	(44.44)	(43.49)	(48.51)	(49.00)	(56.16)	(60.67)	(63.68)	(73.73)	(51.28)	(56.30)	(53.37)
Avg. years of educ.	5.07	5.31	5.46	5.56	5.59	5.66	5.79	5.86	6.05	6.13	6.14	6.16	6.32	6.60	6.80	6.87
	(3.21)	(3.32)	(3.43)	(3.51)	(3.40)	(3.44)	(3.49)	(3.52)	(3.59)	(3.58)	(3.57)	(3.56)	(3.69)	(3.78)	(3.91)	(3.87)
Avg. age	35.54	35.55	35.85	35.67	35.56	35.86	35.76	35.86	35.95	36.12	37.10	37.24	37.40	37.70	37.86	38.10
	(11.54)	(11.46)	(11.64)	(11.68)	(11.68)	(11.68)	(11.61)	(11.66)	(11.53)	(11.51)	(11.65)	(11.63)	(11.59)	(11.52)	(11.44)	(11.52)
1{Urban}	27%	28%	27%	27%	27%	27%	27%	27%	28%	29%	30%	30%	31%	30%	31%	31%
	(0.44)	(0.45)	(0.44)	(0.45)	(0.44)	(0.45)	(0.44)	(0.44)	(0.45)	(0.45)	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)
1{Agriculture}	64%	63%	61%	62%	64%	61%	58%	58%	55%	54%	50%	48%	49%	51%	49%	49%
	(0.48)	(0.48)	(0.49)	(0.48)	(0.48)	(0.49)	(0.49)	(0.49)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
1{Paid empl.}	79%	79%	80%	78%	77%	79%	80%	80%	83%	81%	84%	85%	83%	83%	85%	85%
	(0.41)	(0.41)	(0.40)	(0.41)	(0.42)	(0.41)	(0.40)	(0.40)	(0.38)	(0.39)	(0.37)	(0.36)	(0.37)	(0.37)	(0.36)	(0.36)
Frequency	15,496	14,752	14,163	15,155	20,840	19,959	22,212	22,007	20,980	43,568	43,064	41,625	42,721	40,525	39,955	39,798

Female workers	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Avg. hourly earnings	21.78	21.29	21.17	24.34	23.63	27.52	26.42	30.23	32.42	32.74	35.46	36.17	44.00	36.96	35.75	37.07
Median hourly earnings	13.02	12.73	13.47	14.85	14.35	16.29	15.98	17.98	20.18	20.48	22.09	22.86	24.70	23.42	22.81	22.50
	(31.61)	(28.49)	(24.77)	(33.21)	(32.69)	(46.73)	(39.38)	(38.19)	(42.15)	(41.91)	(45.07)	(47.38)	(58.66)	(44.13)	(48.57)	(121.30)
Avg. years of educ.	4.38	4.64	4.83	4.94	4.98	5.11	5.24	5.36	5.54	5.66	5.68	5.73	5.97	6.28	6.40	6.59
	(3.11)	(3.21)	(3.38)	(3.48)	(3.34)	(3.44)	(3.51)	(3.55)	(3.63)	(3.66)	(3.67)	(3.71)	(3.85)	(4.00)	(4.15)	(4.16)
Avg. age	35.39	35.48	35.51	35.49	35.42	35.68	35.56	35.72	35.79	35.76	36.88	37.02	37.17	37.38	37.50	37.72
	(11.32)	(11.41)	(11.46)	(11.49)	(11.46)	(11.55)	(11.32)	(11.44)	(11.26)	(11.15)	(11.32)	(11.26)	(11.26)	(11.20)	(11.09)	(11.06)
1{Urban}	25%	25%	26%	26%	25%	26%	26%	25%	27%	28%	29%	29%	30%	31%	32%	32%
	(0.44)	(0.44)	(0.44)	(0.44)	(0.43)	(0.44)	(0.44)	(0.44)	(0.44)	(0.45)	(0.45)	(0.46)	(0.46)	(0.46)	(0.47)	(0.47)
1{Agriculture}	68%	66%	63%	66%	67%	64%	60%	62%	57%	57%	54%	52%	52%	50%	47%	48%
	(0.47)	(0.47)	(0.48)	(0.47)	(0.47)	(0.48)	(0.49)	(0.49)	(0.49)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
1{Paid empl.}	42%	43%	48%	45%	42%	45%	49%	48%	51%	52%	54%	55%	55%	56%	60%	60%
	(0.49)	(0.49)	(0.50)	(0.50)	(0.49)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)	(0.49)
Frequency	14,691	13,949	13,718	14,680	20,254	19,389	21,147	21,256	20,094	41,254	41,324	39,378	40,834	38,774	38,046	38,129

Source: Author's calculation from the National Labour Force Surveys 1985-2000 (NSO), using the surveys' population weight variable. Standard deviations in parentheses.
Notes: * The average and median hourly earnings are computed from the pooled sample of paid employed workers. They are adjusted for inflation and regional price differences using the regional headline CPI.

Table 4.2: LFS Descriptive Statistics of Employed Workers by Sector of Work (Age 20-65)

Male workers	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Male workers in agriculture																
Avg. hourly earnings	13.64	11.37	12.04	16.31	14.70	16.26	14.18	15.97	15.90	18.24	21.25	22.00	21.68	20.47	20.84	18.22
Median hourly earnings	7.58	7.40	7.18	8.50	8.97	9.73	8.75	9.77	9.46	10.18	10.81	12.36	11.91	13.42	12.54	12.45
Avg. years of educ.	4.07	4.17	4.37	4.37	4.50	4.52	4.51	4.66	4.70	4.81	4.73	4.80	4.85	5.09	5.15	5.28
Avg. age	35.91	35.79	36.09	35.94	35.94	36.23	36.38	36.29	36.59	36.70	38.29	38.53	38.58	38.61	38.99	39.33
1{Urban}	8%	7%	7%	8%	8%	8%	7%	7%	6%	8%	7%	7%	7%	8%	8%	8%
1{Paid empl.}	71%	70%	72%	69%	68%	69%	69%	69%	72%	69%	73%	74%	72%	73%	76%	75%
Male workers in non-agriculture																
Avg. hourly earnings	36.68	38.14	37.32	41.72	40.59	43.18	43.59	49.37	51.61	52.28	55.58	54.13	64.32	52.93	53.45	52.20
Median hourly earnings	26.46	27.28	26.57	27.34	27.24	27.91	28.28	31.73	31.39	32.01	35.59	34.29	36.14	33.02	32.66	31.58
Avg. years of educ.	6.85	7.22	7.17	7.50	7.49	7.43	7.51	7.53	7.67	7.67	7.54	7.45	7.72	8.19	8.39	8.39
Avg. age	34.90	35.15	35.47	35.25	34.91	35.28	34.91	35.27	35.18	35.44	35.92	36.02	36.27	36.75	36.77	36.93
1{Urban}	61%	63%	58%	60%	61%	58%	54%	54%	54%	53%	52%	51%	53%	54%	53%	53%
1{Paid empl.}	93%	93%	93%	94%	94%	94%	94%	95%	95%	95%	95%	95%	95%	94%	94%	94%

Female workers	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Female workers in agriculture																
Avg. hourly earnings	12.29	10.55	10.62	12.92	13.62	16.30	13.01	14.38	14.22	16.74	17.70	18.93	19.33	18.32	17.52	18.35
Median hourly earnings	8.45	7.74	7.96	8.54	9.11	9.31	9.18	10.58	10.23	11.68	12.80	13.91	14.64	13.57	13.52	14.97
Avg. years of educ.	3.53	3.73	3.79	3.80	4.00	4.04	4.04	4.19	4.21	4.33	4.24	4.30	4.39	4.52	4.53	4.67
Avg. age	35.81	35.75	36.16	36.04	35.94	36.28	36.31	36.21	36.71	36.70	38.42	38.43	38.56	38.87	38.99	39.57
1{Urban}	7%	7%	7%	7%	7%	7%	6%	6%	5%	7%	7%	7%	7%	7%	7%	7%
1{Paid empl.}	25%	24%	29%	27%	24%	25%	26%	26%	27%	28%	29%	29%	28%	30%	32%	34%
Female workers in non-agriculture																
Avg. hourly earnings	28.21	27.71	27.59	31.64	29.62	33.68	32.84	38.21	40.33	39.72	42.77	42.67	52.84	43.67	42.10	43.99
Median hourly earnings	19.44	19.36	19.13	20.79	19.45	20.95	21.71	24.32	25.57	25.60	27.61	27.44	28.88	26.76	26.25	26.58
Avg. years of educ.	6.16	6.44	6.60	7.13	6.94	7.04	7.00	7.25	7.33	7.39	7.34	7.29	7.65	8.03	8.08	8.32
Avg. age	34.50	34.95	34.41	34.43	34.39	34.61	34.46	34.94	34.55	34.53	35.09	35.47	35.68	35.88	36.16	36.04
1{Urban}	63%	62%	59%	63%	60%	59%	56%	56%	56%	56%	55%	54%	56%	56%	54%	55%
1{Paid empl.}	78%	79%	80%	79%	79%	81%	81%	83%	83%	84%	82%	84%	83%	83%	84%	84%

Source: Author's calculation from the National Labour Force Surveys 1985-2000 (NSO), using the surveys' population weight variable.
Notes: * The average and median hourly earnings are computed from the pooled sample of paid employed workers. They are adjusted for inflation and regional price differences using the regional headline CPI.

Table 4.3: LFS Descriptive Statistics of Employed Workers by Type of Employment (Age 20-65)

Male workers	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Male workers in paid employment																
Avg. years of educ.	5.12	5.38	5.44	5.57	5.57	5.66	5.83	5.90	6.08	6.18	6.17	6.16	6.34	6.62	6.80	6.83
Avg. age	37.68	37.43	37.79	37.98	37.98	37.99	37.56	37.98	37.67	37.91	38.64	38.56	38.91	39.27	39.19	39.42
1{Urban}	31%	32%	30%	31%	31%	31%	31%	31%	31%	32%	32%	33%	34%	33%	33%	34%
1{Agriculture}	57%	56%	55%	55%	56%	53%	50%	50%	48%	46%	43%	42%	42%	45%	44%	44%
Male workers in unpaid employment																
Avg. years of educ.	4.89	5.06	5.55	5.52	5.63	5.65	5.60	5.71	5.94	5.94	5.99	6.20	6.21	6.53	6.80	7.05
Avg. age	27.53	28.61	28.03	27.43	27.43	27.74	28.60	27.38	27.82	28.40	29.26	29.69	29.82	29.95	30.26	30.74
1{Urban}	14%	13%	15%	13%	13%	13%	13%	12%	12%	14%	15%	15%	14%	15%	17%	17%
1{Agriculture}	88%	88%	86%	89%	90%	90%	88%	89%	87%	88%	83%	84%	84%	83%	79%	79%

Female workers	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Female workers in paid employment																
Avg. years of educ.	5.24	5.56	5.70	5.95	5.90	6.08	6.16	6.28	6.49	6.57	6.58	6.59	6.92	7.28	7.30	7.48
Avg. age	36.16	36.36	36.08	36.14	36.11	36.15	36.01	36.31	35.98	35.76	36.49	36.66	36.84	37.03	37.19	37.15
1{Urban}	43%	44%	40%	43%	42%	41%	40%	40%	41%	41%	41%	41%	43%	43%	42%	43%
1{Agriculture}	40%	37%	38%	39%	37%	35%	32%	33%	30%	30%	29%	27%	26%	26%	26%	27%
Female workers in unpaid employment																
Avg. years of educ.	3.75	3.96	4.04	4.12	4.31	4.33	4.37	4.53	4.56	4.66	4.64	4.67	4.81	4.98	5.06	5.23
Avg. age	34.83	34.83	34.99	34.96	34.92	35.29	35.14	35.19	35.58	35.76	37.33	37.45	37.56	37.83	37.95	38.59
1{Urban}	13%	12%	13%	12%	13%	13%	13%	12%	12%	13%	15%	15%	15%	16%	17%	17%
1{Agriculture}	88%	88%	86%	87%	88%	88%	85%	88%	86%	86%	82%	83%	82%	80%	79%	79%

Source: Author's calculation from the National Labour Force Surveys 1985-2000 (NSO), using the surveys' population weight variable.

Table 4.4: Estimates of the Double Selection Process: Male Sample

	FIRST_M	BIPROBIT_M		$\text{BIPROBIT}_\text{M} + \text{CF}$	
Dependent variable:	Educ (years)	$\mathbf{1}\{\text{Paid}\}$	$\mathbf{1}\{\text{Agri}\}$	$\mathbf{1}\{\text{Paid}\}$	$\mathbf{1}\{\text{Agri}\}$
	(1)	(2)	(3)	(4)	(5)
$\mathbf{1}\{\text{Reform}\}$	0.358***				
	(0.0769)				
Educ		-0.0485***	-0.0719***	-0.153***	-0.157***
		(0.0107)	(0.0120)	(0.0294)	(0.0279)
Educ-sq/100		0.569***	-0.449***	1.601***	0.332***
		(0.0798)	(0.0803)	(0.0460)	(0.0574)
Age	0.0294	0.147***	-0.0656***	0.151***	-0.0631***
	(0.0278)	(0.00886)	(0.00527)	(0.00203)	(0.00175)
Age-sq/100	-0.125***	-0.138***	0.0810***	-0.142***	0.0778***
	(0.0329)	(0.00752)	(0.00560)	(0.00258)	(0.00231)
#children (<6years)	0.0208	0.0557***		0.0578***	
	(0.0699)	(0.0151)		(0.00596)	
#children (6-13years)	-0.308***	0.172***		0.167***	
	(0.0289)	(0.0135)		(0.00483)	
#elderly	0.0322	-0.215***		-0.213***	
	(0.0760)	(0.0225)		(0.00959)	
Spouse's earnings	0.000200***	-0.0000384***		-0.0000368***	
	(0.0000220)	(0.00000336)		(0.000000887)	
L1.D1.Manufac.	-0.0773		-0.301*		-0.303***
	(0.669)		(0.181)		(0.0721)
L1.D1.Services	0.542**		0.352**		0.347***
	(0.230)		(0.164)		(0.0493)
L2.D1.Manufac.	-0.352		-0.629***		-0.627***
	(0.530)		(0.209)		(0.0712)
L2.D1.Services	0.389*		0.290**		0.285***
	(0.209)		(0.137)		(0.0490)
$\hat{\zeta}_i$				0.0481*	-0.157***
				(0.0287)	(0.0279)
$\hat{\zeta}_i \cdot$Educ				-0.0120***	0.332***
				(0.000464)	(0.0574)
$\hat{\rho}_{uv}$		-0.674***		-0.678***	
		(0.0174)		(0.00465)	
R-squared	0.294				
# of observations	412,409	412,409		412,409	
Log likelihood		-14,8594,356.3		-148,393,666.5	

Notes:
* This is a sample of the employed labour force aged from 20 to 65 years old, including unpaid workers.
** Dependent variables are the years of education in the first-stage regression (Column 1) and the dummies of paid employment and agricultural employment in the other columns.
*** Other controls include hours of work per week, rural-urban indicator, marital status, and year and province dummies.
**** The standard errors for the first-stage and exogenous bivariate probit regression (Columns 1, 2, and 3) are clustered at the province level. Other regressions use bootstrapped standard errors with 250 replications to account for generated regressors.
*p<0.1, **p<0.05, ***p<0.01.

Table 4.5: Estimates of the Double Selection Process: Female Sample

	FIRST_F	BIPROBIT_F		$\text{BIPROBIT}_\text{F} + \text{CF}$	
Dependent variable:	Educ (years)	**1**{Paid}	**1**{Agri}	**1**{Paid}	**1**{Agri}
	(1)	(2)	(3)	(4)	(5)
1{Reform}	0.492***				
	(0.0682)				
Educ		-0.0225*	-0.0173	0.0672***	0.0275
		(0.0120)	(0.0115)	(0.0174)	(0.0206)
Educ-sq/100		0.686***	-0.970***	-0.322***	-0.344***
		(0.109)	(0.0868)	(0.0393)	(0.0593)
Age	-0.0137	0.0692***	-0.0381***	0.0605***	-0.0239***
	(0.0211)	(0.00800)	(0.00616)	(0.00193)	(0.00240)
Age-sq/100	-0.108***	-0.0728***	0.0455***	-0.0637***	0.0418***
	(0.0289)	(0.00757)	(0.00658)	(0.00167)	(0.00196)
#children (<6years)	-0.127**	0.0791***		0.0739***	
	(0.0536)	(0.0218)		(0.00444)	
#children (6-13years)	-0.394***	0.0611***		0.0620***	
	(0.0232)	(0.0125)		(0.00329)	
#elderly	0.181**	0.0736***		0.0719***	
	(0.0814)	(0.0202)		(0.00692)	
Spouse's earnings	0.0000878***	-0.0000208***		-0.0000214***	
	(0.00000346)	(0.00000121)		(0.000000393)	
L1.D1.Manufac.	0.0510		-0.117		-0.119
	(0.591)		(0.191)		(0.0764)
L1.D1.Services	0.916***		0.292*		0.293***
	(0.301)		(0.168)		(0.0442)
L2.D1.Manufac.	0.274		-0.554***		-0.556***
	(0.709)		(0.189)		(0.0784)
L2.D1.Services	0.384		0.220		0.221***
	(0.232)		(0.141)		(0.0475)
$\hat{\zeta}_i$				-0.0419**	0.0275
				(0.0167)	(0.0206)
$\hat{\zeta}_i$·Educ				0.0121***	-0.344***
				(0.000446)	(0.0593)
$\hat{\rho}_{uv}$		-0.841***		-0.840***	
		(0.0137)		(0.00399)	
R-squared	0.315				
# of observations	394,559	394,559		394,559	
		-143,464,427.2		-143,300,419.1	

Notes:
* This is a sample of the employed labour force aged from 20 to 65 years old, including unpaid workers.
** Dependent variables are the years of education in the first-stage regression (Column 1) and the dummies of paid employment and agricultural employment in the other columns.
*** Other controls include hours of work per week, rural-urban indicator, marital status, and year and province dummies.
**** The standard errors for the first-stage and exogenous bivariate regression (Columns 1, 2, and 3) are clustered at the province level. Other regressions use bootstrapped standard errors with 250 replications to account for generated regressors.
*p<0.1, **p<0.05, ***p<0.01.

Table 4.6: Estimates of Returns to Education in the Agricultural Sector

	Male Sample					**Female Sample**				
	FIRST_M	OLS_M	$\text{OLS}_\text{M} + \lambda_A$	CF_M	$\text{CF}_\text{M} + \lambda_A$	FIRST_F	OLS_F	$\text{OLS}_\text{F} + \lambda_A$	CF_F	$\text{CF}_\text{F} + \lambda_A$
Dependent variable:	Educ (years)	Log of hourly earnings				Educ (years)	Log of hourly earnings			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1{Reform}	0.781***					0.759***				
	(0.0581)					(0.0838)				
Educ		0.0703***	0.0502***	0.122***	0.107***		0.0223***	0.0104**	0.133***	0.179***
		(0.00310)	(0.00334)	(0.0342)	(0.0360)		(0.00401)	(0.00425)	(0.0494)	(0.0498)
Educ-sq/100		-0.0910***	-0.291***	-0.204	-0.710**		0.162***	0.367***	-0.856*	-1.295**
		(0.0235)	(0.0343)	(0.310)	(0.338)		(0.0367)	(0.0822)	(0.512)	(0.528)
Age	-0.0983***	0.0299***	0.0772***	0.0385***	0.0797***	-0.0806***	0.0157***	0.0464***	0.0183***	0.0509***
	(0.00752)	(0.00155)	(0.00412)	(0.00193)	(0.00417)	(0.0130)	(0.00234)	(0.00420)	(0.00322)	(0.00451)
Age-sq/100	0.0624***	-0.0292***	-0.0731***	-0.0379***	-0.0756***	0.0316**	-0.0194***	-0.0527***	-0.0220***	-0.0575***
	(0.00797)	(0.00176)	(0.00432)	(0.00211)	(0.00437)	(0.0133)	(0.00272)	(0.00481)	(0.00355)	(0.00509)
$\hat{\lambda}_{P,A}$			0.577***		0.561***			0.621***		0.685***
			(0.0466)		(0.0478)			(0.0723)		(0.0784)
$\hat{\lambda}_{A,A}$			0.387***		0.398***			-0.190**		-0.262***
			(0.0464)		(0.0489)			(0.0815)		(0.0885)
$\hat{\zeta}_i$				-0.0458**	-0.0367				-0.0757**	-0.107***
				(0.0215)	(0.0227)				(0.0315)	(0.0319)
$\hat{\zeta}_i$·Educ				0.00112	0.00398				-0.0552**	-0.0862***
				(0.00315)	(0.00333)				(0.0237)	(0.0238)
1{Reform}·$\hat{\zeta}_i$				-0.0657***	-0.0544***				0.0113**	0.0187***
				(0.0164)	(0.0173)				(0.00522)	(0.00531)
1{Reform}·$\hat{\zeta}_i$·Educ				0.00245	0.00619*				0.00854	0.0163***
				(0.00322)	(0.00341)				(0.00534)	(0.00542)
Other controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.162	0.385	0.389	0.385	0.389	0.202	0.286	0.291	0.287	0.291
# of observations	124,541	124,541	124,541	124,541	124,541	46,769	46,769	46,769	46,769	46,769

Notes:
* This is a sample of the paid employed labour force aged from 20 to 65 years old.
** Dependent variables are the years of education in the first-stage regressions (Columns 1 and 6) and log of hourly earnings in the other columns.
*** Other controls include hours of work per week, rural-urban indicator, marital status, and year and province dummies.
**** The standard errors for the first-stage and OLS regressions (Columns 1, 2, 6, and 7) are clustered at the province level. Other regressions use bootstrapped standard errors with 250 replications to account for generated regressors. *p<0.1, **p<0.05, ***p<0.01.

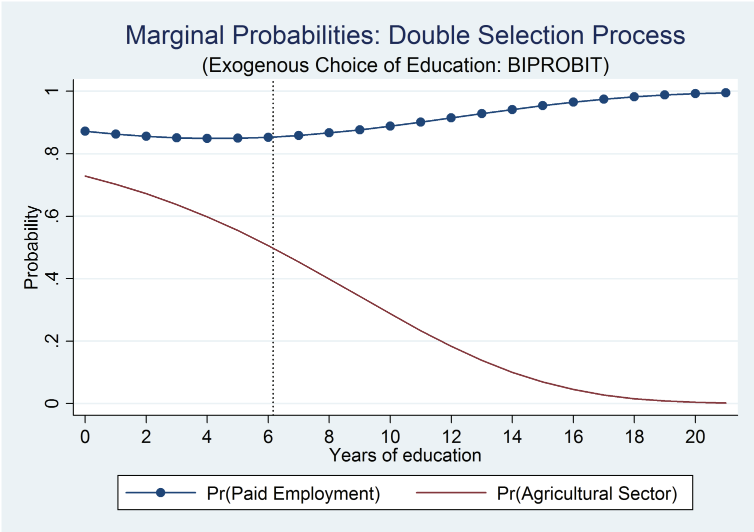
Table 4.7: Estimates of Returns to Education in the Non-Agricultural Sector

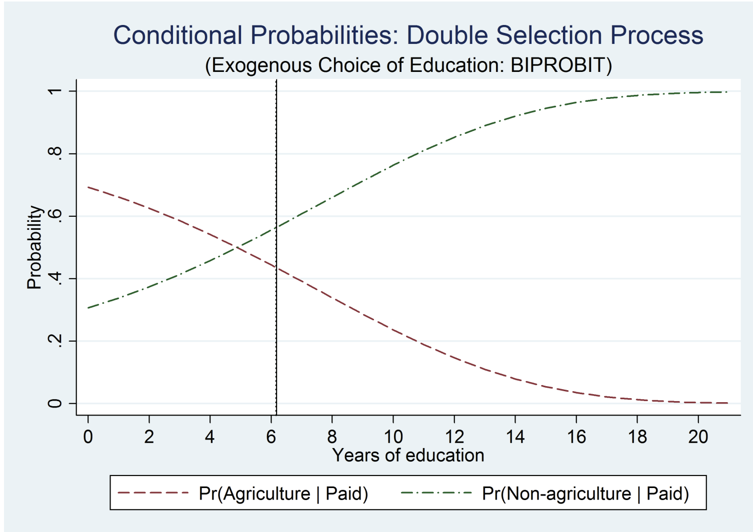
	Male Sample					Female Sample				
	FIRST_M	OLS_M	$\text{OLS}_\text{M} + \lambda_{NA}$	CF_M	$\text{CF}_\text{M} + \lambda_{NA}$	FIRST_F	OLS_F	$\text{OLS}_\text{F} + \lambda_{NA}$	CF_F	$\text{CF}_\text{F} + \lambda_{NA}$
Dependent variable:	Educ (years)	Log of hourly earnings				Educ (years)	Log of hourly earnings			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1{Reform}	0.103*					0.133**				
	(0.0584)					(0.0617)				
Educ		0.0274***	0.00637***	0.0541***	-0.0909***		0.0550***	0.0183***	0.0298***	-0.0209***
		(0.00151)	(0.00175)	(0.00774)	(0.00809)		(0.00148)	(0.00172)	(0.00497)	(0.00509)
Educ-sq/100		0.298***	0.383***	0.476***	1.123***		0.239***	0.406***	0.558***	0.773***
		(0.00796)	(0.00828)	(0.0235)	(0.0245)		(0.00776)	(0.00892)	(0.0191)	(0.0197)
Age	0.147***	0.0681***	0.0927***	0.0746***	0.0950***	0.121***	0.0679***	0.0829***	0.0719***	0.0849***
	(0.0304)	(0.000920)	(0.00155)	(0.00117)	(0.00167)	(0.0268)	(0.00105)	(0.00114)	(0.00136)	(0.00142)
Age-sq/100	-0.278***	-0.0647***	-0.0900***	-0.0711***	-0.0912***	-0.309***	-0.0694***	-0.0836***	-0.0726***	-0.0844***
	(0.0330)	(0.00119)	(0.00179)	(0.00142)	(0.00192)	(0.0279)	(0.00136)	(0.00144)	(0.00163)	(0.00168)
$\hat{\lambda}_{P,A}$			0.580***		0.585***			0.705***		0.633***
			(0.0234)		(0.0249)			(0.0177)		(0.0185)
$\hat{\lambda}_{A,A}$			0.128***		0.260***			0.354***		0.411***
			(0.0113)		(0.0116)			(0.0102)		(0.0102)
$\hat{\zeta}_i$				-0.0340***	0.0441***				0.00692	0.0167***
				(0.00744)	(0.00760)				(0.00430)	(0.00428)
$\hat{\zeta}_i$·Educ				-0.00201***	-0.00830***				-0.0232***	-0.0226***
				(0.000262)	(0.000268)				(0.00376)	(0.00376)
1{Reform}·$\hat{\zeta}_i$				-0.0630***	0.00759				-0.00317***	-0.00390***
				(0.00632)	(0.00649)				(0.000208)	(0.000220)
1{Reform}·$\hat{\zeta}_i$·Educ				-0.00155***	-0.00696***				-0.00378***	-0.00390***
				(0.000298)	(0.000297)				(0.000272)	(0.000285)
Other controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.174	0.492	0.498	0.494	0.502	0.236	0.553	0.569	0.558	0.575
# of observations	236,226	236,226	236,226	236,226	236,226	194,341	194,341	194,341	194,341	194,341

Notes:
* This is a sample of the paid employed labour force aged from 20 to 65 years old.
** Dependent variables are the years of education in the first-stage regressions (Columns 1 and 6) and log of hourly earnings in the other columns.
*** Other controls include hours of work per week, rural-urban indicator, marital status, and year and province dummies.
**** The standard errors for the first-stage and OLS regressions (Columns 1, 2, 6, and 7) are clustered at the province level. Other regressions use bootstrapped standard errors with 250 replications to account for generated regressors. *p<0.1, **p<0.05, ***p<0.01.

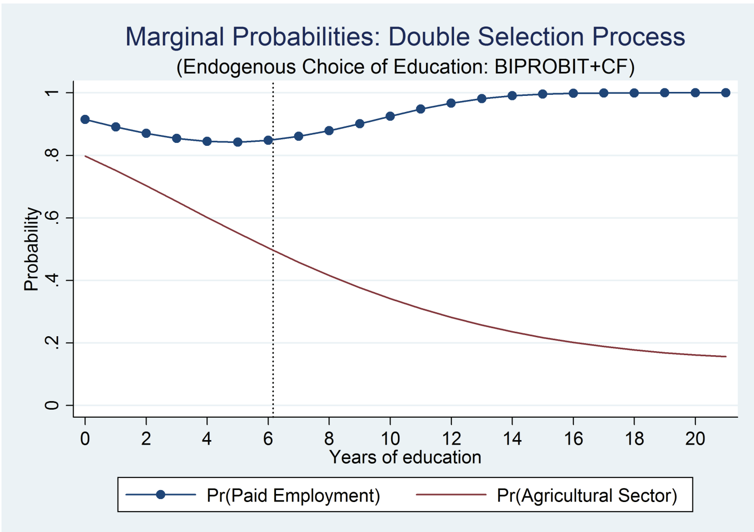
Figure 4.1: The Role of Education in the Double Selection Process: Male Sample

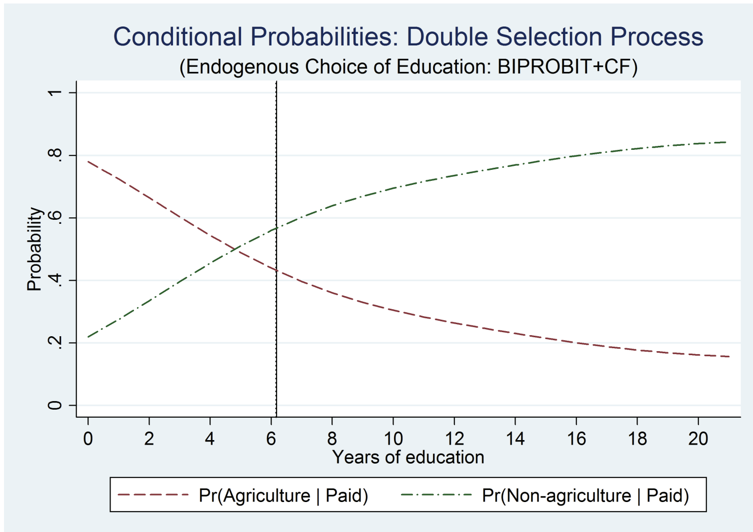
(a) Exogenous Choice of Education





(b) Endogenous Choice of Education

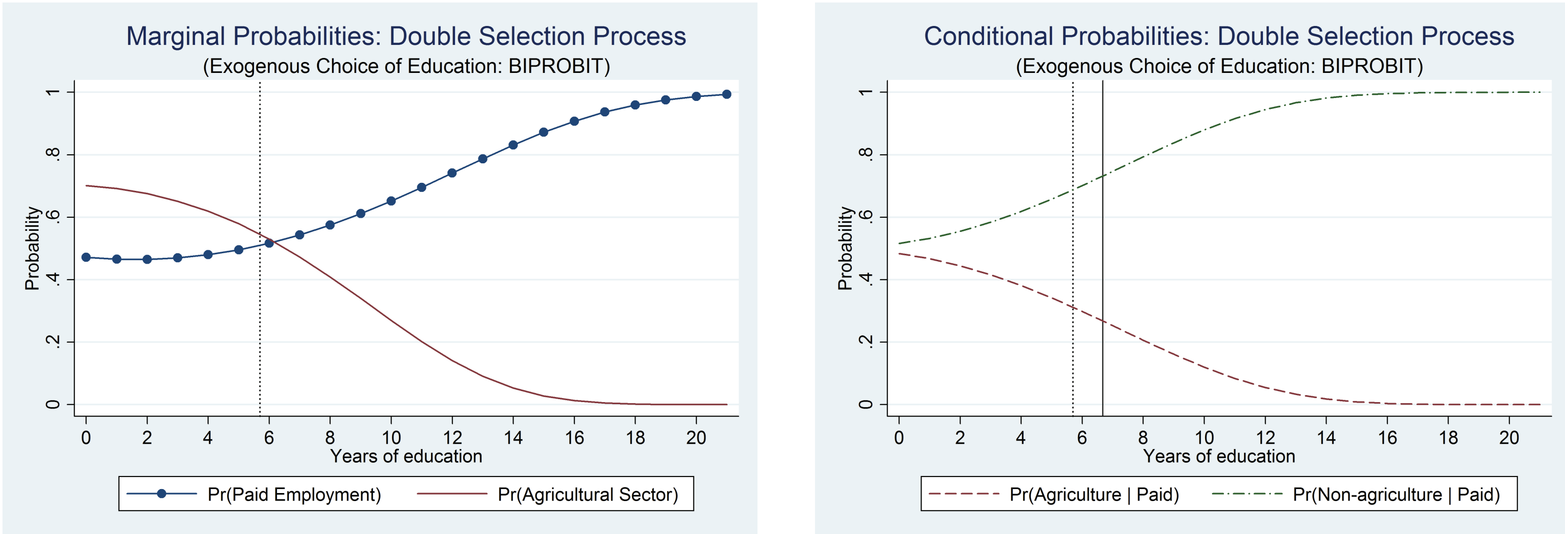




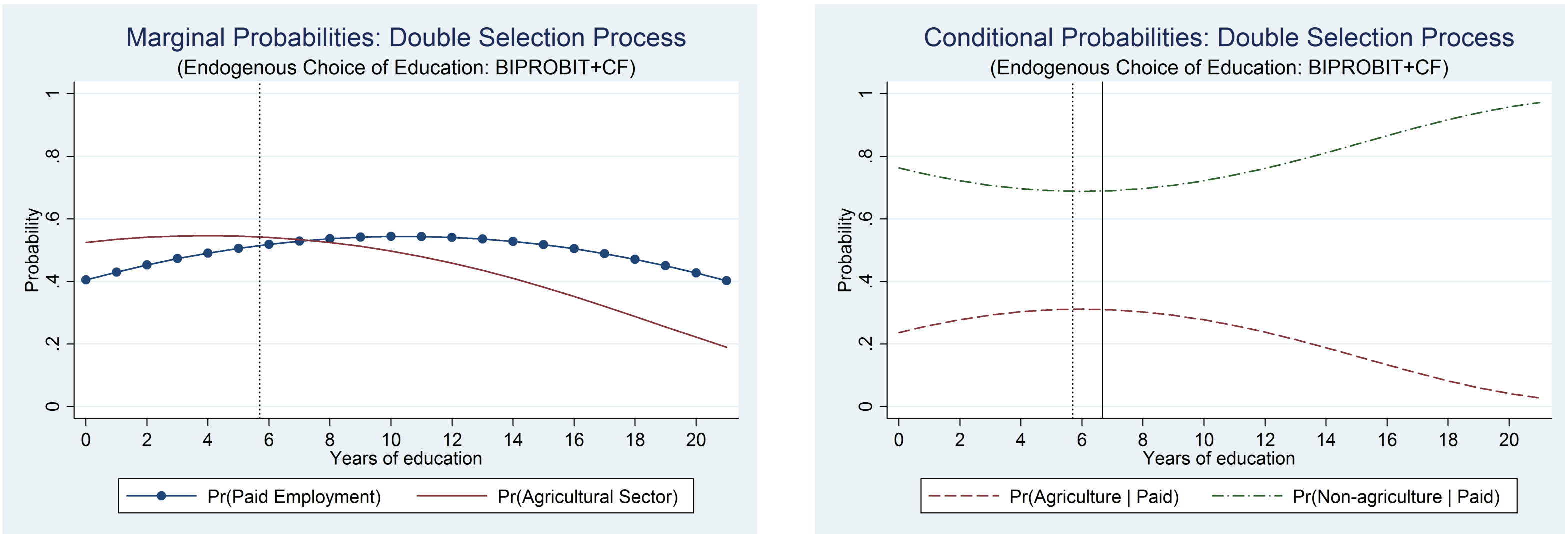
Notes: * The graphs present the predicted marginal and conditional probabilities by varying the years of education (holding other regressors at their mean values) from the bivariate probit regressions (Table 4.4, Columns 2 to 5).
** The dotted and solid vertical lines refer to the average years of education of the employed labour force and the paid employed labour force (aged between 20 and 65 years old), respectively.

Figure 4.2: Roles of Education in the Estimated Double Selection Process: Female Sample

(a) Exogenous Choice of Education



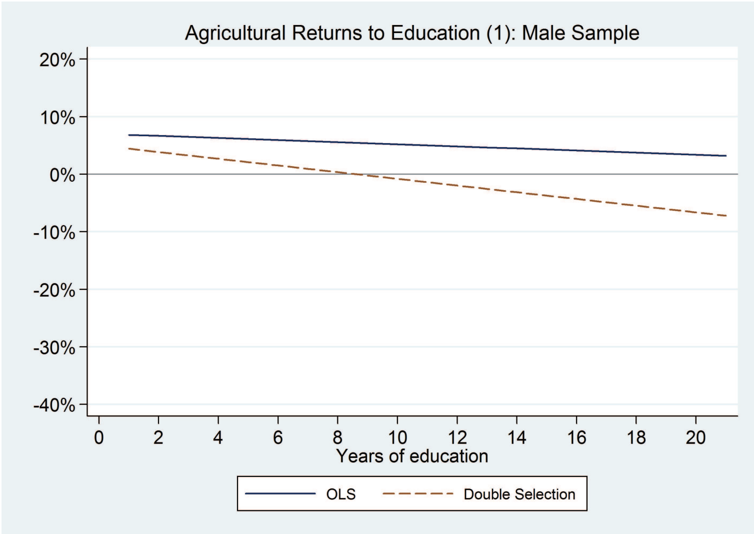
(b) Endogenous Choice of Education

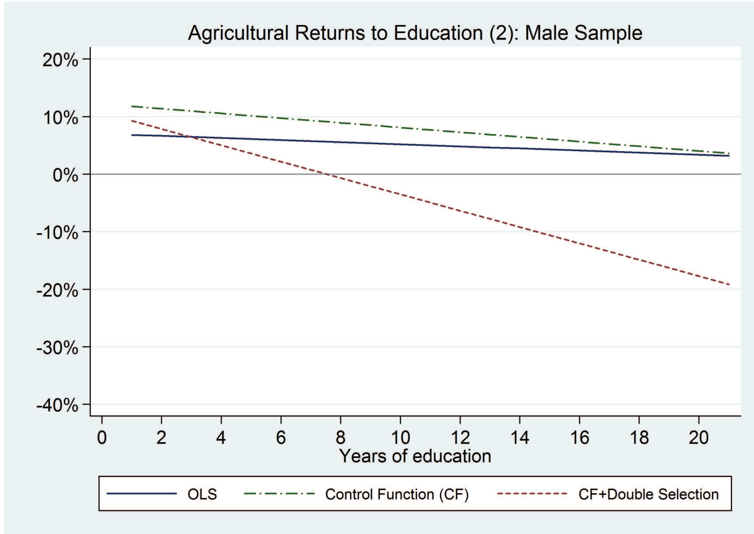


Notes: * The graphs present the predicted marginal and conditional probabilities by varying the years of education (holding other regressors at their mean values) from the bivariate probit regressions (Table 4.5, Columns 2 to 5).
** The dotted and solid vertical lines refer to the average years of education of the employed labour force and the paid employed labour force (aged between 20 and 65 years old), respectively.

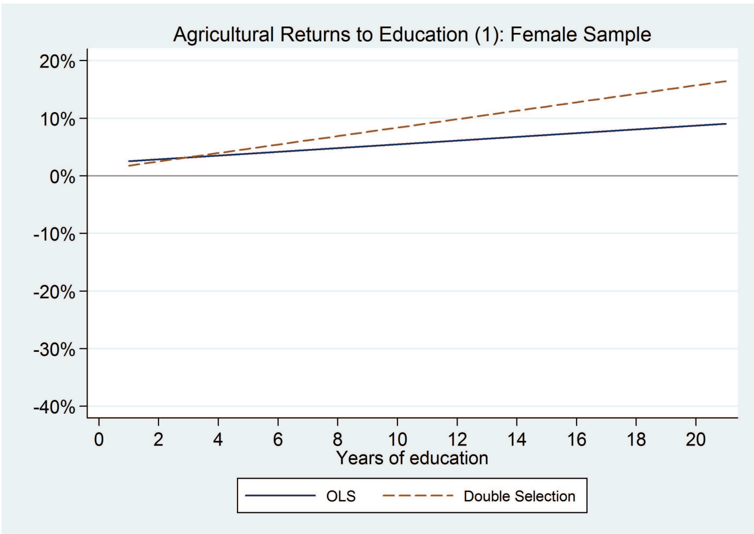
Figure 4.3: Agricultural Returns to Education

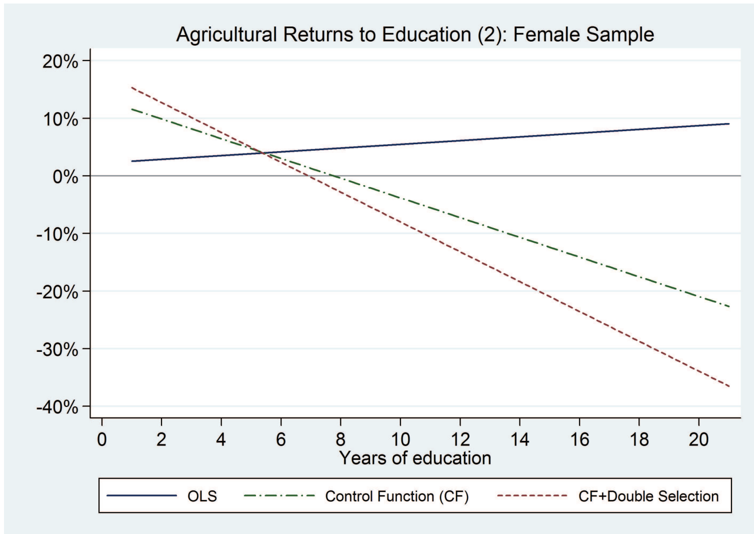
(a) Male Sample





(b) Female Sample

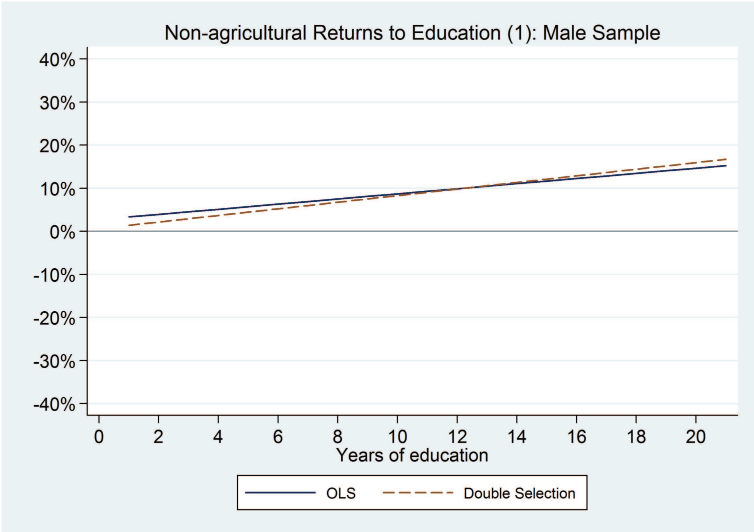


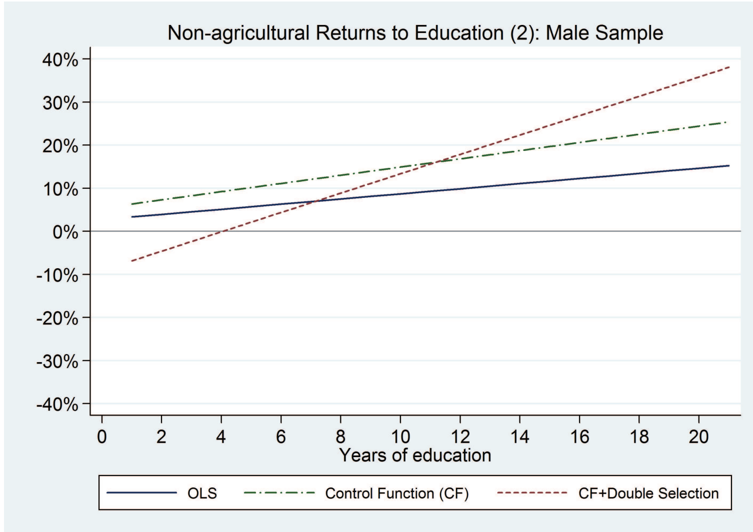


Notes: * The graphs present the estimated sectoral returns to education from the four main specifications: OLS, double selection controls (OLS+ λ_A), control function (CF), and control function+double selection controls ($CF + \lambda_A$) as shown in in Table 4.6.
** Increasing returns to education imply a convex education-earnings profile, while decreasing returns to education imply a concave education-earnings profile.

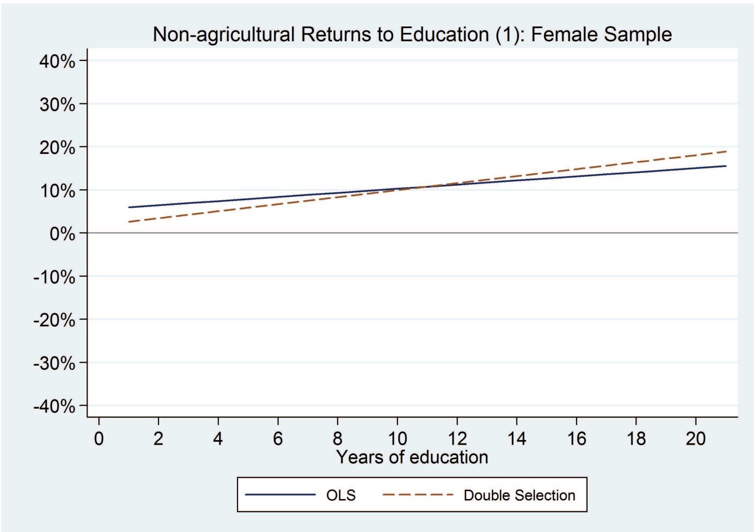
Figure 4.4: Non-Agricultural Returns to Education

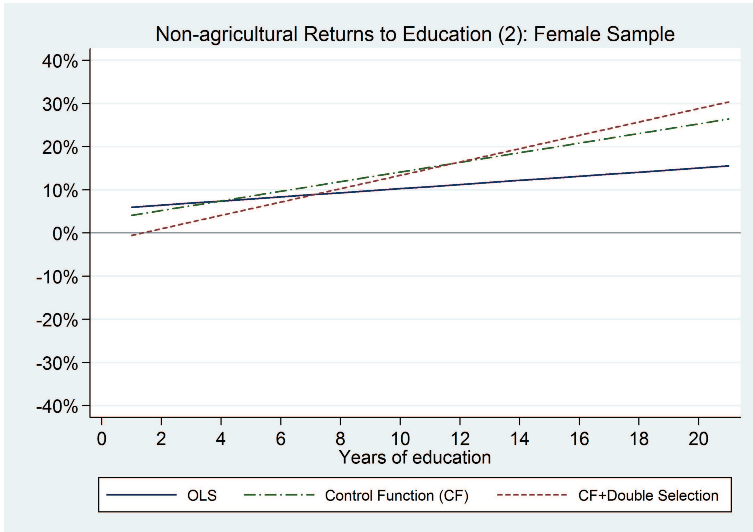
(a) Male Sample





(b) Female Sample

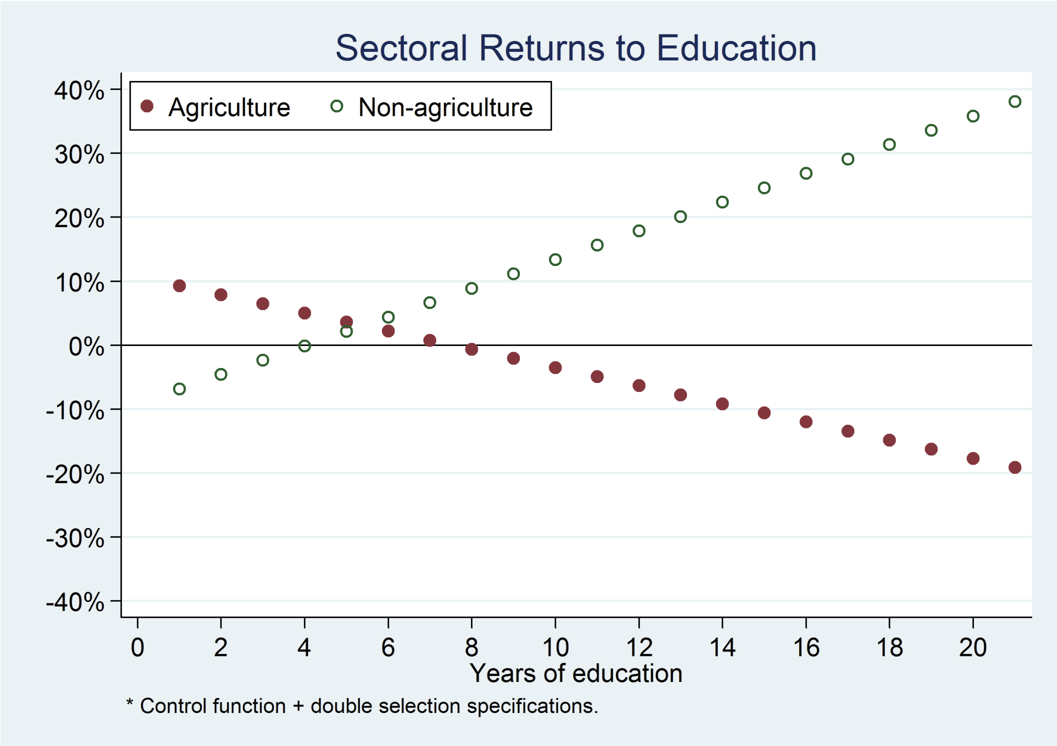




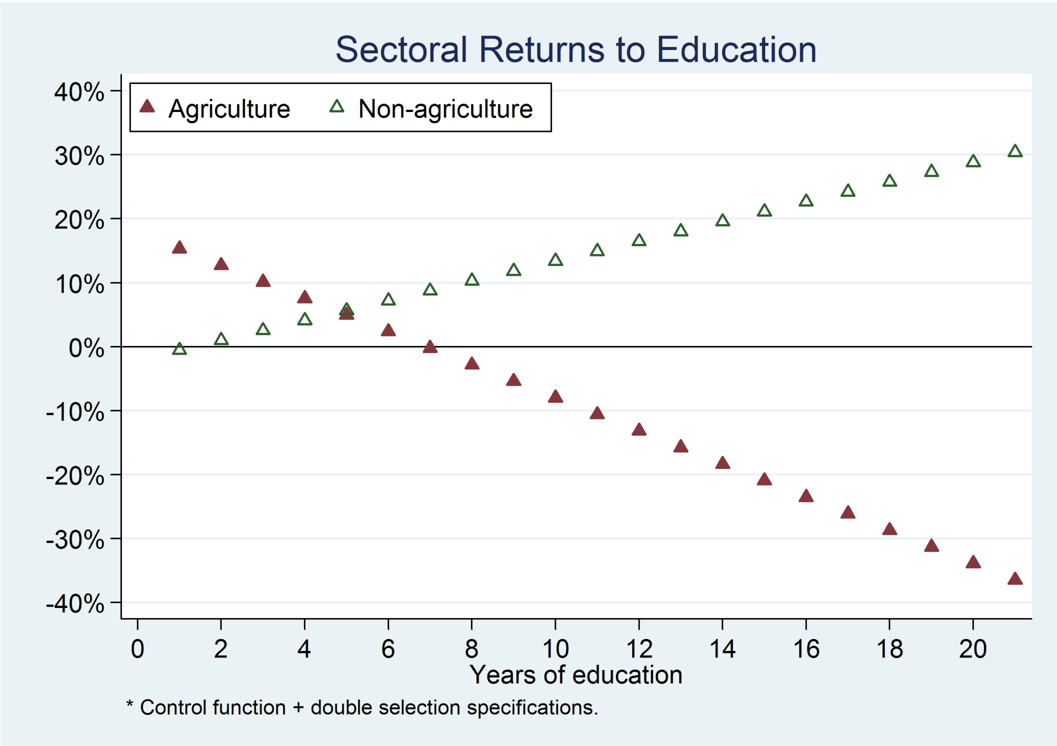
Notes: * The graphs present the estimated sectoral returns to education from the four main specifications: OLS, double selection controls (OLS+ λ_{NA}), control function (CF), and control function+double selection controls ($CF + \lambda_{NA}$) as shown in Table 4.7.
** Increasing returns to education imply a convex education-earnings profile, while decreasing returns to education imply a concave education-earnings profile.

Figure 4.5: Summary of the Estimated Sectoral Returns to Education

(a) Male Sample



(b) Female Sample



Notes:
* The graphs present the estimated sectoral returns to education from the regressions which simultaneously take into account the sample selection and the endogenous choice of education (Tables 4.6 and 4.7, Columns 5 and 10).
** Increasing returns to education imply a convex education-earnings profile, while decreasing returns to education imply a concave education-earnings profile.

4.A Chapter Appendix

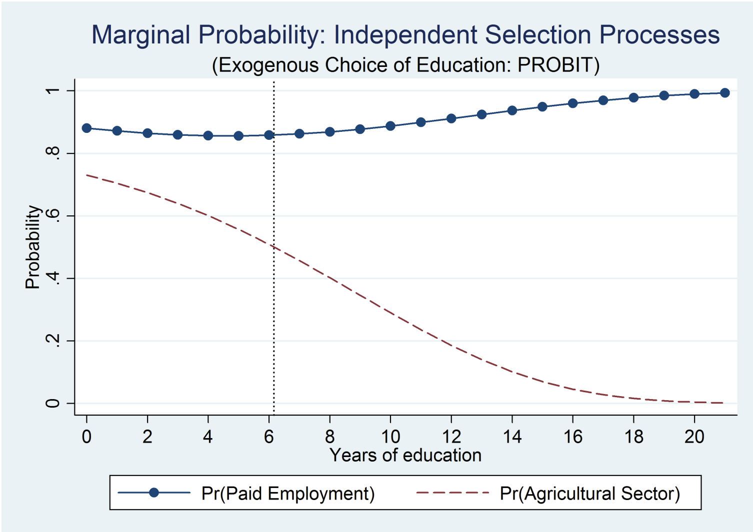
Table 4.A.1: Estimates of the Two Independent Selection Processes

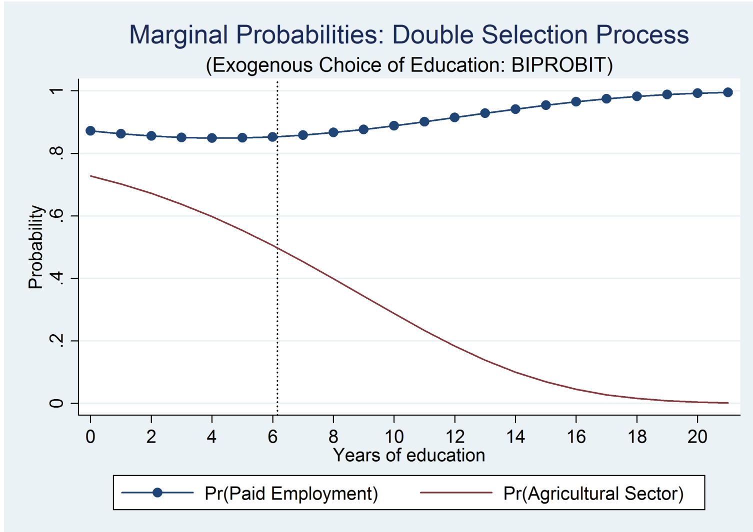
	Male Sample		Female Sample	
	PROBIT$_{\mathrm{M}}$		PROBIT$_{\mathrm{F}}$	
Dependent variable:	**1**{Paid}	**1**{Agri}	**1**{Paid}	**1**{Agri}
	(1)	(2)	(3)	(4)
Educ	-0.0502***	-0.0716***	-0.0237**	-0.0108
	(0.00993)	(0.0120)	(0.0121)	(0.0113)
Educ-sq/100	0.537***	-0.451***	0.664***	-1.068***
	(0.0693)	(0.0777)	(0.106)	(0.0742)
Age	0.147***	-0.0647***	0.0684***	-0.0415***
	(0.00846)	(0.00402)	(0.00718)	(0.00516)
Age-sq/100	-0.136***	0.0811***	-0.0728***	0.0504***
	(0.00714)	(0.00442)	(0.00660)	(0.00580)
#children (<6years)	0.0394**		0.0512**	
	(0.0157)		(0.0251)	
#children (6-13years)	0.169***		0.0423***	
	(0.0142)		(0.0145)	
#elderly	-0.261***		0.0378	
	(0.0232)		(0.0235)	
Spouse's earnings	-0.0000283***		-0.0000147***	
	(0.00000152)		(0.00000239)	
L1.D1.Manufac.		-0.416**		-0.106
		(0.204)		(0.241)
L1.D1.Services		0.407**		0.433*
		(0.186)		(0.228)
L2.D1.Manufac.		-0.687***		-0.645***
		(0.250)		(0.245)
L2.D1.Services		0.328**		0.309*
		(0.156)		(0.187)
# of observations	412,409	412,409	394,559	394,559
Log likelihood	-69,278,645.9	-85,699,333.2	-89,676,268.9	-64,960,238.0

Notes:
* This is a sample of the employed labour force aged from 20 to 65 years old, including unpaid workers.
** Dependent variables are the dummy of paid employment and agricultural employment.
*** Other controls include hours of work per week, rural-urban indicator, marital status, and year and province dummies.
**** The standard errors are clustered at the province level. *p<0.1, **p<0.05, ***p<0.01.

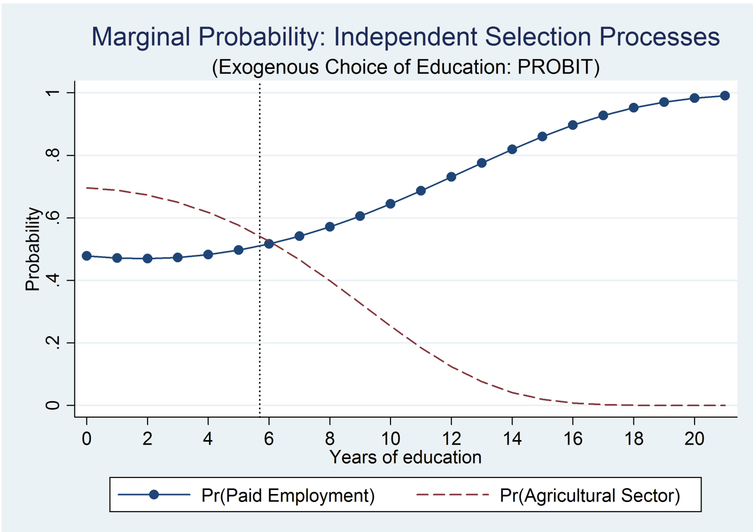
Figure 4.A.1: The Role of Education in the Independent Selection Processes

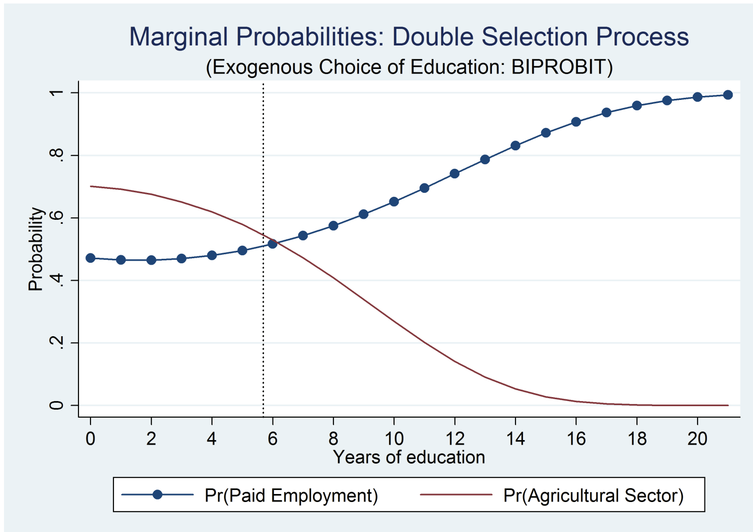
(a) Male Sample





(b) Female Sample





Notes: * The graphs present the predicted probabilities by varying the years of education (holding other regressors at their mean values) from the probit and bivariate probit regressions prior to taking into account the endogenous choice of education (Tables 4.4 and 4.5, Columns 2 and 3, and Table 4.A.1).
** The dotted vertical line refers to the average years of education of the employed labour force (aged between 20 and 65 years old).
*** For comparison, the graphs on the righ-hand-side are reproduced from Figures 4.1 and 4.2.

Chapter 5

Estimating the Inequality Treatment Effect of a Change in Compulsory Schooling

5.1 Introduction

There exists a burgeoning literature on the effects of education on earnings, as reviewed by Card (2001). However, most of the empirical literature has concentrated on identifying the average returns to education and the shape of education-earnings profiles for the entire population and/or particular welfare subgroups. Less attention has been devoted to estimating the effects of education on the distribution of earnings. As education is considered a crucial factor for economic development, understanding both its average and distributional effects are important policy questions. In this chapter, I attempt to explore the distributional impacts of education on earnings in Thailand.

Thailand is an interesting country for studying the distributional impacts of education on earnings during the structural change period. In the 1990s, rapid economic growth in Thailand was accompanied by rising average earnings and a very gradual decline in earnings inequality in the labour market. At the same time, the average level of educational attainment increased significantly due to a sharp fall in the primary school dropout rate especially among the poor and in rural areas. While the other two chapters of the thesis have analysed how education affects earnings and the sector of employment, this chapter investigates how education on average, particularly at the primary level, can improve or worsen earnings inequality. Using micro-level data from the Labour Force Survey (LFS), the main questions posed are whether the returns to primary education completion are heterogeneous across the earnings distribution, and if so, whether they result in higher or lower earnings inequality in Thailand. In a later part of the chapter, I analyse these distributional effects of education on earnings for the rural and urban areas separately.

I seek to identify the causal effects of primary education completion, using an education policy shift – the change in the compulsory schooling law – that produced exogenous variation in individuals' education. Particularly, the increase in compulsory education from 4 years of lower primary school to 6 years of primary school in the late 1970s makes an individual's probability of completing at least 6 years of primary education a discontinuous function of his year of birth, with a substantive positive jump in 1967. This enables me to use a regression discontinuity (RD) approach to identify the distributional effects of primary education completion as proposed by the most recent micro-econometrics literature (Frölich and Melly, 2010b; Frandsen et al., 2012). First, the two counterfactual earnings distributions related to the binary treatment (primary school completion) are identified in the RD framework. Second, the earnings quantiles and other measures of earnings inequality are estimated from the counterfactual distributions of earnings. Finally, the differences in these earnings quantiles and inequality measures are used to describe the distributional treatment effects of primary education completion.

I find that the increased primary education completion rate results in lower earnings inequality as the returns to primary education completion are larger for the poor, compared to the rich. The causal distributional effects of primary education completion on earnings are more pronounced in the first half of the 1990s and among workers in rural Thailand. In addition, by considering various measures of earnings inequality, I find that the change in the primary education completion rate has a larger effect on the bottom and the middle of the distribution than the top.

This chapter offers four main contributions to the study of education and earnings inequality. First, I examine how education, at the basic level of primary schooling, affects earnings inequality. In contrast, most of the literature on earnings inequality focuses on the impacts of education at the secondary and higher levels of education. Basic education is of particular policy relevance in developing countries, especially in the early stages of economic development when the average level of educational attainment is low. Second, I use micro-level data to demonstrate the link between education and earnings inequality. This, together with the RD framework, enables me to control for individual unobservable heterogeneity that may affect both an individual's decision on education investment and his earnings. As a result, my estimated parameters have a causal interpretation. Third, while most studies using micro-level data focus only on quantile treatment effects, I adopt the recently proposed identification strategy in the RD framework that can identify the entire counterfactual distributions of earnings. Therefore, the distributional effects estimated in this chapter can be summarised by a single number, which is the difference in an inequality measure between the two counterfactual distributions. A single number may be more easily conveyed to policy makers and to the public. In addition, unlike other related studies using quantile regressions, the estimated effect of primary education completion at each quantile in this chapter is an unconditional effect, which is more relevant for overall earnings inequality. Finally, given that this is one of the first empirical applications of this method, this chapter requires substantial work on programming which will facilitate future applications of this technique.

The structure of the chapter is as follows. Section 5.2 reviews different models, which try to explain earnings inequality by changes in educational attainment, and their empirical evidence. Section 5.3 briefly recaps the general RD framework that can be used to identify the average effects of primary education completion on earnings. Section 5.4 and Section 5.5 focus on the identification strategy for the distributional impacts of primary education completion developed from the RD framework, and subsequently outline the empirical estimation method. The descriptive statistics of the relevant variables from the LFS are presented in Section 5.6. Section 5.7 examines the causal impacts of primary education completion on the overall distribution of earnings, and Section 5.8 tests these impacts on the rural and urban earnings distributions separately. The robustness of the results to alternative model specifications is considered in Section 5.9 and Section 5.10 concludes.

5.2 Education and Earnings Inequality: A Review of Theoretical Explanations and Empirical Evidence

This chapter is related to a number of different strands of literature. This section summarises the literature on education and earnings inequality in two parts and discusses how the chapter contributes to each of them. First, I review the literature on the theoretical links between education and earnings inequality and define what this relationship means for this chapter.[1] Second, I discuss the relevant empirical studies. These empirical studies differ not only in their definitions of education and earnings, but also in data types, countries and time periods covered, and estimation techniques.[2] The discussion of the empirical literature is divided into macro- and micro-data studies.

5.2.1 Education and Earnings Inequality: Theoretical Explanations

The literature on the links between education and the distribution of earnings can be divided into two substantive groups according to the education level of interest, namely basic and higher education. First among those focusing on basic education are Becker and Chiswick (1966), who explain how the distributions and shapes of marginal returns and marginal costs of education investment – as well as the correlations between the two – affect the earnings distribution. In their model, the dispersion in earnings would be less than that in education investment when the marginal returns are decreasing, and would exceed that in education investment when the marginal costs are increasing.[3] In addition, the negative correlations between factors influencing the marginal returns and marginal costs result in lower earnings inequality. They also highlight briefly that equal opportunities to invest in education would potentially reduce the inequalities in investment, and thus in earnings. Free public schools and compulsory schooling laws are, albeit imperfect, examples of the elimination of unequal investment opportunities as they reduce the costs of education.[4]

[1] This chapter focuses on the inequality in the labour market. Therefore, the variable of interest is earnings, which comprise income from labour activity. Some of the reviewed literature also uses labour income. Similarly, I refer to this as "earnings". Other studies use total income, which includes labour and capital incomes, as well as transfers and remittances. This is referred to, in this section, as "income".

[2] The details of the relevant estimation techniques used in the reviewed literature will be discussed in Section 5.4.

[3] This is because: (1) decreasing marginal returns imply large investors would receive lower returns; and (2) increasing marginal costs mean large investors would have higher costs.

[4] Under free public school provisions or compulsory schooling laws, some costs of schooling – such as living expenses and foregone earnings – may not be subsidised.

Under different model set-ups, Chiswick (1969) and Eckstein and Zilcha (1994) explain, in more detail, how compulsory education affects the distributions of education, returns to education, and earnings. Given decreasing (positive) marginal returns, increasing marginal costs, and a negative relationship between marginal returns and marginal costs, compulsory education is likely to reduce the inequalities in education and in earnings (Chiswick, 1969).[5] This is because compulsory schooling laws are expected to positively affect individuals at the bottom of both the earnings and education distributions, more than those at the top of these distributions.[6] First, compulsory schooling laws would reduce the gap between the marginal costs and the marginal returns at the compulsory schooling level for those who would not have studied up to this level, had there been no law enforcement. Second, better educated individuals are likely to raise their own attainment in response to increases in compulsory schooling, due to lower average costs and an effort to maintain their educational advantage over the less educated (Meghir and Palme, 2005). For those at the bottom of the distributions, while their marginal and average returns decrease, they will later on experience an increase in their post-education earnings. Chiswick (1969) also notes that, nonetheless, economic growth associated with an increased demand for skilled or educated workers would change the earnings structure, and therefore, reduce the effects of compulsory education in improving earnings inequality. Under the overlapping generations model, Eckstein and Zilcha (1994) further exhibit the long-run effects of compulsory schooling on the intra-generational distribution of earnings. In their model, compulsory education reduces earnings inequality as education for the poor is financed by transfers from the rich.

Similarly to free public school provision and compulsory schooling laws, Schultz (1963) contends that public education is a potential factor to increase human capital and reduce earnings inequality. Glomm and Ravikumar (1992) develop an overlapping generations model of endogenous economic growth, in which the choice of education regime is endogenous. They conclude that public schools, where investment in the school quality is made through majority voting, reduce earnings inequality. This conclusion is also supported by other theoretical studies (Saint-Paul and Verdier, 1993; Zhang, 1996; Sylwester, 2002a,b). However,

[5] That is, individuals with high marginal returns are more likely to have low marginal costs. For instance, this could be due to the positive correlation between genetic ability and parental wealth, and the positive correlation between ability and the likelihood of obtaining merit-based scholarships.

[6] For example, a relatively higher degree of risk aversion, limited sources of funds, and relatively less knowledge about their own abilities may deter the poor from obtaining basic education (Stiglitz, 1973; Deininger, 2003).

the association between earnings inequality and public expenditures on education remains ambiguous (Stiglitz, 1973; Fields, 1980; Jimenez, 1986; Ram, 1989; Barro, 2000; Sylwester, 2002a,b). For instance, the effect might not show up when the poor are too poor and do not have sufficient resources to attend school (Sylwester, 2002a,b) or when public expenditures on education are highly concentrated in higher-level education (Stiglitz, 1973).

More recently, however, education has been criticised as one of the major contributors to increased earnings inequality. I should emphasise that there is one important difference between the more recent work and most of the earlier studies. While the earlier work on education and earnings inequality focuses on a basic level of education (which is often provided free of charge and/or compulsory), the recent studies tend to concentrate on the effects of education at a higher level (for instance, vocational, upper-secondary, or university education). Their theoretical predictions are developed from the basic human capital model and the Mincerian earnings equation. That is, increased earnings inequality is associated closely with rising returns to education and experience. The increasing returns to education are often explained by a rise in the demand for highly-skilled or educated workers due to new technology, which is assumed to grow faster than the increase in human capital or education (for example, see Bound and Johnson (1992); Katz and Murphy (1992); Krueger (1993); Card and DiNardo (2002)).[7] As a result of higher demand for skilled workers, the earnings gains due to an additional year of schooling are disproportionately higher at the top end of the education and earnings distributions. Thus, the earnings inequality between groups with different levels of education increases. Note that in these models, the effects of education on earnings inequality are expected to be less significant and subsequently negative when the change in the supply catches up with the increased demand for highly-skilled or educated workers.[8]

The two types of models, focusing on different levels of education and relying on different underlying assumptions, reach contradictory conclusions on the relationships between education and earnings inequality. On the one hand, the studies focusing on basic education argue that education reduces inequality. On the other hand, the literature on skill-biased technical change (SBTC) suggests that education increases inequality (although only during the stage

[7]This is known as the skill-biased technical change (SBTC) hypothesis. In these studies, skilled workers refer to, for example, those who use computers at work, or university graduates.

[8]Knight and Sabot (1983) call the initial increase in inequality the "composition effect", and the subsequent decrease the "compression effect".

when demand for educated workers outstrips the supply). The focus of this chapter is to investigate the distributional impacts of primary education, as a result of the change in the compulsory schooling law, on earnings in Thailand. This is closely related to the traditional models, which emphasise the roles of basic education and equal access to basic education in reducing earnings inequality.

5.2.2 Education and Earnings Inequality: Empirical Evidence

There are a number of empirical studies, which examine the aforementioned theoretical relationships between education and earnings inequality. They differ widely in terms of definitions of education (years of education, enrolment/completion levels, or compulsory/public education), income types (total income, earnings, consumption, or per-capita unit), data types (macro- or micro-level datasets), countries covered (developed or developing countries, and single- or cross-country analyses), and time periods, as well as empirical specifications and identification strategies. Here, I divide the empirical literature into two broad categories according to data type, namely macro- and micro-level evidence.

Macro-level Analysis

First, the macro-level analysis involves an investigation at the country or regional level, of which the unit of analysis is larger than the individual or the household. The relationship between education and earnings inequality is commonly captured by regressing various inequality indices as dependent variables on measures of education, such as the average years of education, education completion rates, and their variations.[9] Earlier work finds a negative relationship between education and earnings inequality.[10] Becker and Chiswick (1966) show that earnings inequality is negatively correlated with the average level of schooling at the state level in the US. Using a combination of country-level data from multiple sources, Ahluwalia (1976) finds a negative cross-country relationship between income inequality and human capital, measured by the literacy rate and secondary school enrolment, among 62 countries (14 of which are developed countries and 6 of which are socialist countries). Most

[9] These dependence variables are, for example, the Gini coefficient (and its change), the variance of log, and the earnings share of a particular decile.

[10] Some of the studies, in fact, are interested in the relationship between economic development and inequality (known as the Kuznets inverted-U curve (Kuznets, 1955)). While economic development and inequality are usually measured by per-capita income and earnings/income inequality respectively, education is often one of the control variables in their model specifications.

of the other cross-country literature also confirms this negative association between education and earnings inequality (for example, Marin and Psacharopoulos (1976); Psacharopoulos (1977); Winegarden (1979); Ram (1981); Park (1996)). Furthermore, Ram (1990) suggests that the relationship between the average level of schooling and earnings inequality depends on the degree of economic development and the relationship between the level and inequality of education.[11]

The more recent studies address problems with data quality, including measurement error, inconsistent measurement and differences in variable definitions across countries (for example, see Anand and Kanbur (1993); Deininger and Squire (1996, 1998); De Gregorio and Lee (2002)). The internationally comparable panel datasets used in the more recent work are, for instance, the income and inequality dataset from Deininger and Squire (1996), the World Bank's World Development Indicators (WDI), the human capital statistics from Barro and Lee (1993, 1996), and the Penn World Tables. However, the empirical evidence remains inconclusive and dependent upon other country-specific factors. Education is found to be roughly negatively correlated with income inequality in Deininger and Squire (1998). With the same results for the cross-country regressions, Checchi (2000) and De Gregorio and Lee (2002) further examine differences in the relationship between education and income inequality across regions and sets of countries. The negative relationship remains significant for most of the developed economies, but only for some of the developing economies. Given the positive relationship between income inequality and educational inequality, De Gregorio and Lee (2002) argue that the correlation between average education and income inequality depends on how increased education affects the inequality in educational attainment.[12] In addition, the relationship between education and income inequality could also differ across levels of education. Barro (1999) finds that income inequality is negatively correlated with primary education but positively correlated with higher-than-primary education. On the other hand, Rodriguez-Pose and Tselios (2009) attempt to give a causal interpretation to this relationship by using dynamic panel estimation techniques and find the opposite empirical results. That is, increased education, especially at the secondary level and higher, leads to

[11] In Ram (1990), the relationship between education and earnings inequality operates through the positive relationship between earnings inequality and education inequality. That is, the average level of education affects the inequality in education, which is, in turn, positively correlated with earnings inequality.

[12] In other words, it could be that, increased education reduces (increases) the inequality in education in developed (developing) countries, and hence, is negatively (positively) correlated with income inequality. This is similar to the assumption made by Ram (1990).

higher income inequality.

While the cross-country studies using macro-level data are interesting, they are intensely criticised for their choice of countries included and international data comparability. In addition, they fail to explain any within-country differences at the individual level, both observable and unobservable, which was the original motivation by the theoretical literature. The duration of the data is often too short to allow for the cross-country heterogeneity. The empirical results using macro-level data should, therefore, be considered with caution.

Micro-level Analysis

Given the potential problems in cross-country analyses, the country-specific interests, and the increasing attention to important heterogeneity at both the country and individual levels, more studies in the past decade utilise micro-level data to examine the relationship between education and earnings inequality. These micro-level data contain information on earnings and other sources of income, educational attainment, and other socio-economic characteristics at the individual or household level.

Most of the micro-level studies employ quantile regression techniques to investigate earnings inequality and its relation to education. In other words, they use individual-level information to study how the correlation between education and earnings differs across the distribution of earnings. The pioneering work on education and the conditional earnings distribution using quantile regressions is done by Chamberlain (1991) and Buchinsky (1994) to explain the continuous rise in US earnings inequality prior to the 1990s.[13] The papers find that returns to education, especially at the high school and university levels, are higher and grow faster at the upper quantiles of the conditional distribution of earnings. That is, secondary and higher education is likely to be positively correlated with earnings inequality as outlined in the SBTC literature. The empirical results hold in most studies on developed countries.[14] However, Abadie (1997) finds the opposite results using Spanish earnings data. The correlation between education at all levels and earnings is more strongly positive among poor, compared to the rest of the distribution. This implies the opposite result, that educa-

[13] Most of the earnings distributions in the cited literature are the conditional ones, which are conditional on age, experience, education level, or unobserved ability (measured by the residuals of the earnings regressions). This implies an interest in the relationship between education and within-group earnings inequality.

[14] See for example, Fitzenberger et al. (2001); Weber and Ammermüller (2003) and Prasad (2004) for Germany, Harmon et al. (2003) for the UK, Martins and Pereira (2004) for 15 European countries and the US, Mata and Machado (2005) for Portugal, and Lemieux (2006) for the US.

tion is negatively correlated with earnings inequality.

On the other hand, there are substantially fewer studies on developing economies.[15] The results are rather mixed and depend, to some extent, on levels of education. Mwabu and Schultz (1996) find that the returns to secondary and higher level education are higher for the upper quantiles only among white South African males, whereas the returns to primary education are higher for the lower tail of the distribution of all South African males. That is, in South Africa, primary education reduces earnings inequality of every population group, while secondary and higher education increases earnings inequality among white male workers. Blom et al. (2001) also suggest that tertiary education and earnings inequality are positively related due to demand for skilled labour outstripping the supply in Brazil. Extending the quantile analysis to several countries in East Asia and Latin America, Patrinos et al. (2009) show that increased educational attainment, measured by years of schooling, is associated with higher earnings inequality in Latin America, but lower earnings inequality in East Asia.

The above-mentioned micro-level literature does not address the potential problem of omitted variable bias. Therefore, the results from these studies can only be interpreted as a correlation between the two variables of interest. To my knowledge, there are very few studies that address the endogeneity of education and attempt to estimate the causal impacts of education on earnings inequality, using micro-level datasets.[16] In addition, accounting for the endogeneity of education changes the results substantially. Examples include Girma and Kedir (2005) for Ethiopia and Brunello et al. (2009) for European countries.[17] Both studies conclude that education, measured by years of schooling, reduces earnings inequality because its contribution to earnings is higher at the lower end of the earnings distribution. Note that the reversal of the results after accounting for endogeneity of education could also arise due to different schooling levels considered. The earlier literature concentrates more on secondary and higher education. By contrast, the mean years of schooling are relatively low in Ethiopia, compared to other countries, and increasing education potentially implies a change around the mean. Similarly, Brunello et al. (2009)'s study focuses on the impact

[15] This could be due to the more limited data availability in developing countries.

[16] I exclude a number of studies on China because these are confined to specific sub-populations, for instance, urban areas or provinces (for example, Knight and Song (2003); Wang (2011); Appleton et al. (2012); Messinis (2013)).

[17] They adopt the instrumental variables (IV) quantile regression framework, which will be discussed in Section 5.4.

of increased years of education, in response to changes in compulsory schooling laws, on the distribution of earnings. Changes in compulsory schooling laws directly affect individuals at the lower end of the education distribution.

To add to the limited evidence on developing economies and to fill a gap in literature, this chapter investigates the causal distributional impacts of education on earnings in Thailand. I focus on education at a basic and compulsory level as it plays a crucial role in achieving sustainable and equitable economic growth in the initial development process. In addition, the chapter applies the methodology proposed by Frölich and Melly (2010b) and Frandsen et al. (2012), which enables me to estimate not only the causal impacts at each quantile but also the causal impacts of education on various earnings inequality measures.[18]

5.3 Regression Discontinuity Framework

This section recaps the general framework for the RD design with a binary treatment. The RD framework will be used to identify the distributional effect of the primary education treatment on hourly earnings, which is the focus of this chapter.

In the basic setting for the Rubin Causal Model (RCM), researchers are interested in the causal effect of a binary intervention or a treatment (Holland, 1986). The so-called treatment effect is usually heterogeneous across individuals. In my context, the treatment variable is the completion of primary education and the outcome variables are hourly earnings and their distribution.

Following the common practice in the literature on RD (for example, Hahn et al. (2001); Imbens and Lemieux (2008); Lee and Lemieux (2010)) and programme evaluation (for example, Angrist and Krueger (1991); Imbens and Angrist (1994); Angrist et al. (1996); Heckman et al. (2001)), let Y_i^0 and Y_i^1 denote the pair of potential earnings for an individual i. Y_i^0 are the potential earnings when individual i left school before completing primary education, and Y_i^1 is the potential outcome when individual i completed at least primary school. For an individual i, only the earnings corresponding to his educational completion are observed. Let $D_i \in \{0, 1\}$ denote the status of primary education completion, with $D_i = 1$ if an individual i achieved at least primary education, and $D_i = 0$ otherwise. The observed earnings, Y_i, can therefore be written as

[18] The method is called the regression discontinuity distributional treatment effects (RD-DTE), which will be discussed in detail in Section 5.4.

$$Y_i = Y_i^0 \cdot (1 - D_i) + Y_i^1 \cdot D_i = \begin{cases} Y_i^0 & \text{if } D_i = 0 \\ Y_i^1 & \text{if } D_i = 1 \end{cases} \tag{5.1}$$

As the pair Y_i^0 and Y_i^1 are never observed together, the treatment effect or programme evaluation literature focuses on average effects of the treatment. That is, the literature uses the average of $Y_i^1 - Y_i^0$ over the population, rather than the individual-level effects. However, an econometric problem arises from estimating the effect of primary education completion on earnings by averaging $Y_i^1 - Y_i^0$ because of the potential correlation between primary education completion and unobserved productivity-enhancing attributes. For instance, low-ability workers may decide to leave school early and to work for a low-paid job. In this case, education is a signal of workers' ability, and the average of $Y_i^1 - Y_i^0$ is likely to over-estimate the effect of the primary education on earnings (Griliches, 1977; Card, 1999). This problem is similar to the conventional endogeneity problem in the returns to education literature.

In addition to earnings, Y_i, and the dummy for primary education completion, D_i, a vector of pre-treatment variables or covariates (R_i, X_i) may be observed. Let R_i be the year of birth of an individual i, and X_i be a vector of other individual characteristics. R_i and X_i are assumed not to have been affected by the education received. In my RD framework, the year of birth variable, R_i, is known as a running (or an assignment) variable. It determines the treatment status, which is primary education completion in this case, either completely or partially. Thailand's Education Reform Act 1978 increased compulsory schooling by two years from 4 years of lower primary to 6 years of primary education, free of charge. As children were required by law to start school in the calendar year in which they became eight years old, pupils who were eleven years old or younger at the time the reform was introduced, were compelled to attend two additional years of schooling. This suggests that the first cohort potentially affected by the reform comprises those born in 1967, whereas those born before 1967 might have dropped out of school before completing the primary level. Therefore, primary education completion is determined partially by the year of birth being on either side of the fixed value, 1967.[19] That is, the probability of completing primary education, as a function of the year of birth, is expected to be discontinuous at the cut-off

[19]This is known as a fuzzy RD (Hahn et al., 2001). This is because there were exceptions and some delay in the law implementation in very few remote areas.

year of 1967. Therefore, any discontinuity in earnings as a function of the year of birth at the 1967 cut-off can be interpreted as evidence of a causal average treatment effect of primary education completion (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010).[20] The RD average treatment effects are identified only for the sub-populations that change their treatment status in response to the running variable exceeding the discontinuity point. The RD literature refers to such sub-populations as the compliers and to the RD average treatment effects as the local average treatment effects.[21]

5.4 Distributional Treatment Effects in the RD Framework

This section focuses on the distributional impacts of primary education completion on earnings, and their identification strategy used in this chapter. First, I outline the definition of the distributional treatment effects, particularly of primary education completion on earnings, and their measures. Second, the identification strategy in the RD set-up proposed by Frölich and Melly (2010b) and Frandsen et al. (2012) is presented. It will be used to address the endogenous relation between primary education completion and any individual unobserved attributes.[22] Subsequently, I discuss the interpretation of the estimated distributional treatment effects and its relevance to my context. Lastly, the section briefly assesses how the selected RD identification strategy performs, compared to other related approaches.

5.4.1 Distributional Treatment Effects: Quantile and Inequality Measures

For the evaluation of primary education completion, especially as a result of a change in the compulsory schooling law, it is reasonable to assume that the policy-maker is interested in both the average and the distributional effects on earnings. Let W be a social welfare function that depends on both the mean earnings and the distribution of earnings. It can be written as

$$W(F) = \Omega(\mu(F), v(F))$$

[20]Note that the year of birth may itself be associated with potential earnings. However, the RD framework only requires this association, if exists, to be smooth.

[21]Hence, the RD estimand can be interpreted equivalently to that of the IV estimator of the average treatment effect (Imbens and Angrist, 1994).

[22]This is similar to the endogeneity problem in the returns to education literature.

where F is a distribution of earnings, μ is the mean earnings, and v is a distributional measure of earnings.[23] Let $F_{Y^0}(y)$ and $F_{Y^1}(y)$ denote the distributions of the potential earnings under the situations in which no one completed primary education and everyone completed primary education, respectively. Using the RCM set up with potential earnings, the distributional treatment effect of primary education completion on earnings is the difference in a given distributional measure, v, between these two hypothetical cases. The distributional treatment effect is, therefore, defined as

$$\triangle^{v}_{DTE} = v\left(F_{Y^1}(y)\right) - v\left(F_{Y^0}(y)\right) \tag{5.2}$$

According to the various distributional measures of earnings, the distributional treatment effects can be broadly divided into two groups. The first are the quantile treatment effects, which measure the treatment effects along different points of the earnings distribution (Abadie et al., 2002; Chernozhukov and Hansen, 2005; Guiteras, 2008; Frandsen et al., 2012). The second comprises the inequality treatment effects, which measure the overall distributional changes using other inequality indices (Frölich and Melly, 2010b; Firpo and Pinto, 2011). They can be written as follows.

$$\triangle^{\tau}_{QTE} = Q^{\tau}\left(F_{Y^1}(y)\right) - Q^{\tau}\left(F_{Y^0}(y)\right) \tag{5.3}$$

$$\triangle^{I}_{ITE} = I\left(F_{Y^1}(y)\right) - I\left(F_{Y^0}(y)\right) \tag{5.4}$$

where $Q^{\tau}\left(F_{Y^d}\right) = \inf\left\{F_{Y^d}(u) \geq \tau\right\}$ is the τ-th quantile of potential earnings and $I\left(F_{Y^d}\right)$ is a summary inequality index of the potential earnings, for $d \in \{0, 1\}$. While the two hypothetical distributions, $F_{Y^0}(y)$ and $F_{Y^1}(y)$, are unobserved, the observed difference in a given measure of earnings inequality between the primary school graduates and primary school dropouts is likely to be a biased estimate of the distributional treatment effects of primary education completion. This is because of the aforementioned potential correlation between education received and unobserved individual characteristics affecting earnings.

[23] The distributional measures are functionals of the distribution function. They include earnings quantiles and other inequality indices, such as the Gini coefficient, the inter-quartile range, the variance of log, the coefficient of variation, or other inequality indices that belong to the Generalised Entropy (GE) class.

5.4.2 Identification of Distributional Treatment Effects in the RD Design

The identification strategy of this chapter mainly follows that of Frölich and Melly (2010b) and Frandsen et al. (2012). It involves two steps. In the first step, the distributions of earnings, under the two counterfactual circumstances related to the binary treatment (primary education completion), are identified in the RD framework. In the second step, the quantiles and other inequality measures are estimated from the counterfactual distributions of earnings. Finally, the distributional treatment effects are given by the differences in the inequality measures of the two counterfactual distributions.

Similarly to the RD average treatment effects identification, the running variable (year of birth), R_i, plays a special role in identifying the distributional treatment effects in the RD framework. An individual i's dummy variable for primary education completion, D_i, is a function of the year of birth. That is, $D_i = D_i(R_i)$. Due to the Education Reform Act 1978, the year of birth influences the probability of completing primary education in a discontinuous way when it exceeds a fixed threshold, $r_0 = 1967$.

Let Z_i be a binary indicator of the year of birth exceeding the threshold, that is, $Z_i = \mathbb{1}\{R_i \geq r_0\}$, and let D_i^0 and D_i^1 denote the limits of $D_i(r)$ as r approaches r_0 from below and above, respectively. That is, $D_i^0 \equiv lim_{r \to r_0^-} D_i(r)$ and $D_i^1 \equiv lim_{r \to r_0^+} D_i(r)$. Then, the population can be partitioned into five sub-groups: always takers, never takers, compliers, defiers, and indefinite (Angrist et al., 1996) as displayed in the following table.

Text Table 5.1: Population Classified by Primary Education Completion Status (D_i)

Sub-populations	Conditions
Always takers (AT)	$D_i^0 = 1$ and $D_i^1 = 1$
Never takers (NT)	$D_i^0 = 0$ and $D_i^1 = 0$
Compliers (C)	$D_i^0 < D_i^1$
Defiers (DE)	$D_i^0 > D_i^1$
Indefinite (I)	$\{\text{AT} \cup \text{NT} \cup \text{C} \cup \text{DE}\}^c$

With these population partitions, Frölich and Melly (2010b) and Frandsen et al. (2012) discuss the assumptions required for the identification of the two counterfactual distributions for the compliers, $F_{Y^0|C}(y)$ and $F_{Y^1|C}(y)$, which will be subsequently used to identify the distributional effects of primary education completion for the compliers, as follows.[24]

Assumption I (Frölich and Melly, 2010b and Frandsen et al., 2012)

I1 Fuzzy Regression Discontinuity (FRD): $\lim_{r \to r_0^+} \Pr(D = 1 \mid R = r) > \lim_{r \to r_0^-} \Pr(D = 1 \mid R = r)$.

I2 Local Smoothness: $F_{Y^d|D^0, D^1, R}\left(y \mid d^0, d^1, r\right)$ is continuous in r at r_0 for $d^0, d^1 \in \{0, 1\}$.

$E\left[D^j \mid R = r\right]$ is continuous at r_0 for $j \in \{0, 1\}$.

I3 Monotonicity: $\lim_{r \to r_0} \Pr\left(D^1 \geq D^0 \mid R = r\right) = 1$ and $\lim_{r \to r_0} \Pr(\text{Indefinite} \mid R = r) = 0$.

I4 Density at Threshold: $F_R(r)$ is differentiable at r_0 and $\lim_{r \to r_0} f_R(r) > 0$.

Note that the threshold value of year of birth, r_0, is 1967, and that around the discontinuity threshold refers to those who were born around 1967 (before, and in or after).

Assumption **I1** is the key feature of the fuzzy RD design (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). That is, the probability of completing primary education changes discontinuously at the threshold value of the year of birth. Unlike the sharp RD design, the change in the probability of completing primary education does not have to be from 0 to 1 at the threshold.[25] This is similar to the application of the change in compulsory education in Thailand for two reasons. First, some individuals born before 1967 continued their studies to and beyond the primary level. Second, a very few people born at and after 1967 were not subject to the reform due to some exceptions and the delay in the law's implementation in very few remote areas (as discussed in Chapter 2, p. 10).

Assumption **I2** is a smoothness condition. It assumes that the conditional distribution functions are smooth in the year of birth around the threshold year of 1967.[26] The sufficient smoothness on both sides of the threshold implies that the difference in the earnings distribution on either side of the threshold is due to the discontinuous change in the probability of completing primary education.

[24] Frölich and Melly (2010b) and Frandsen et al. (2012) study the local quantile treatment effects in the RD framework. Their identification strategy is developed from the IV outcome distribution estimates of Imbens and Rubin (1997) and the IV quantile regression estimator of Abadie et al. (2002). They first estimate the counterfactual distributions, and then they, obtain the quantile treatment effects from the difference at each quantile of the two distributions. Given that some inequality measures are functionals of a distribution function, similar results can be obtained (see Frölich and Melly (2010b) for the Gini coefficient and the Lorenz curve, and Firpo and Pinto (2011) for other inequality indicators).

[25] In the sharp RD framework, $\lim_{r \to r_0^+} \Pr(D = 1 \mid R = r) - \lim_{r \to r_0^-} \Pr(D = 1 \mid R = r)$ is equal to one, and therefore, all individuals are compliers. That is, Assumption **I3** is satisfied automatically.

[26] Imbens and Lemieux (2008) argue that it is not usually reasonable to assume smoothness only for one particular value of the running variable. In this case, they suggest to a stronger assumption that $F_{Y^d|D^0, D^1, R}\left(y \mid d^0, d^1, r\right)$ and $E\left[D^j \mid R = r\right]$ are continuous in r for all Y, for $d^0, d^1, j \in \{0, 1\}$.

Assumption **I3** is known as a monotonicity assumption. It means that the decision on primary education completion can only be affected by the year of birth crossing the threshold in one direction. It also implies that there exist no defiers and indefinites in the neighbourhood around the threshold. That is, there does not exist anybody who would have left school before the primary level under the Education Reform Act 1978, had he decided to complete the 6 years of primary education when only 4 years of lower primary education was compulsory. This assumption cannot be directly tested (Imbens and Angrist, 1994), but needs to be assessed. I argue that it is likely met for three main reasons. First, the Education Reform Act 1978 was implemented nationwide. Second, some exceptions were made for people with extreme difficulty in attending school regularly. However, the change in the law did not affect these exceptions. Therefore, holding other factors constant, people in the exception categories are never takers rather than defiers, as they would not have completed the primary level regardless of the education policy regime. Third, the implementation lag occurred only in very few remote areas and there is no specific reason to believe that there exist defiers in these areas.

Assumption **I4** requires that observations close to the threshold exist.

Given these assumptions (**I1** to **I4**), the two counterfactual distributions for the compliers can be identified (Frandsen et al., 2012) as follows.[27]

$$F_{Y^1|C}(y) = \frac{\lim_{r \to r_0^+} E\left[\mathbb{1}\{Y \leq y\} \cdot D \mid R = r\right] - \lim_{r \to r_0^-} E\left[\mathbb{1}\{Y \leq y\} \cdot D \mid R = r\right]}{\lim_{r \to r_0^+} E\left[D \mid R = r\right] - \lim_{r \to r_0^-} E\left[D \mid R = r\right]} \tag{5.5}$$

$$F_{Y^0|C}(y) = \frac{\lim_{r \to r_0^+} E\left[\mathbb{1}\{Y \leq y\} \cdot (1-D) \mid R = r\right] - \lim_{r \to r_0^-} E\left[\mathbb{1}\{Y \leq y\} \cdot (1-D) \mid R = r\right]}{\lim_{r \to r_0^+} E\left[1-D \mid R = r\right] - \lim_{r \to r_0^-} E\left[1-D \mid R = r\right]} \tag{5.6}$$

Subsequently, the distributional impacts of primary education completion on earnings for the compliers are simply (1) the difference between the two estimated marginal distributions of potential earnings for the compliers at a particular quantile; and (2) the difference between the inequality measures from the two estimated distributions of potential earnings for the compliers. They can be written as

[27]The proof taken from Frandsen et al. (2012) is provided in detail in the Appendix, p. 188.

$$\triangle^{\tau}_{LQTE|C} = Q^{\tau}\left(F_{Y^1|C}(y)\right) - Q^{\tau}\left(F_{Y^0|C}(y)\right) \quad (5.7)$$

$$\triangle^{I}_{LITE|C} = I\left(F_{Y^1|C}(y)\right) - I\left(F_{Y^0|C}(y)\right) \quad (5.8)$$

where $Q^{\tau}(\cdot)$ and $I(\cdot)$ are defined as before (Equations (5.3) and (5.4)). Similarly to the RD average treatment effects, the RD distributional treatment effects are identified only for the compliers. Hence, they can be called the local distributional treatment effects.[28]

5.4.3 Interpretation of Local Distributional Treatment Effects

In this chapter, the local quantile treatment effects obtained from the RD design (Equation (5.7)) are the causal impacts of primary education completion on hourly earnings at each of the earnings distribution quantiles, for the compliers. In other words, the RD quantile treatment effects reflect the average earnings premiums of primary school graduates relative to primary school dropouts among the compliers, at each of the quantiles. Therefore, they provide useful insights into the distributional effects of primary education completion on earnings.

Alternatively, the RD inequality treatment effects (Equation (5.8)) summarise the distributional effects of primary education completion on hourly earnings for the compliers into a single number. This describes how primary education completion impacts on a specific inequality measure. I should emphasise that the estimated RD inequality treatment effects are the changes in earnings inequality, measured by a particular inequality index, as a result of moving from the situation in which none of the compliers completed primary education to one in which every complier did. In other words, the inequality treatment effects are calculated from the two extreme RD counterfactual earnings distributions, neither of which is observed in reality. Therefore, the estimated inequality treatment effects are likely to be the upper bound estimates.

In order to obtain the relevant distributional impacts of primary education completion due to the Education Reform Act 1978, one should look at the inequality treatment effects scaled by the actual change in the rate of primary education completion for the compliers.

[28]Or the local quantile treatment effects (LQTE) and the local inequality treatment effects (LITE).

Assuming a linear relationship between earnings inequality and the rate of primary education completion, this is the estimated RD inequality treatment effect multiplied by the percentage point change in the compliers' primary education completion rate due to the reform.[29] Although the change in the primary education completion rate among the compliers is unobserved, it could be roughly proxied by the percentage change in the primary education completion rates between the cohorts right before and after the reform.[30]

Unconditional vs Conditional Distributional Treatment Effects

When the parameters of interest are distributional treatment effects, the interpretation of the conditional and unconditional treatment effects can be quite different (Frölich and Melly, 2010a; Fort, 2012; Frölich and Melly, 2012). For example, consider the differences between rural and urban areas, the unconditional effect at the 80th percentile refers to the absolute high earners in the national sample. On the other hand, the effect at the 80th percentile conditional on rural-urban residential location refers to the high earners within each area, who may not be the high earners overall. Presuming a strong positive correlation between earnings and urban location, it is possible that the majority of those at the unconditional 80th percentile reside in urban areas. It may also be that the earnings of the 80th percentile within rural areas are below those of the median urban workers. Therefore, unlike in a mean regression, the interpretations of the effects at each percentile are different for the conditional and unconditional cases. Similarly, the conditional inequality treatment effects capture the impacts on earnings inequality in a particular area, whereas the unconditional inequality treatment effects refer to the changes in overall earnings inequality.

In this chapter, I investigate both unconditional and conditional distributional impacts of primary education completion on hourly earnings. First, it is necessary to know the unconditional distributional treatment effects, when researchers are interested in the effects of the treatment on the entire distribution. Especially in the case of earnings, the welfare change of the unconditional poor may impact on the overall and between-group inequality more than that of the urban residents with relatively low earnings. Second, one may also be interested the conditional distributional treatment effects as they are informative and multi-dimensional. By allowing the distributional treatment effects to be heterogeneous across the

[29] This assumption is natural, given that only the change in earnings inequality per a 100 percentage point change in the primary education completion rate is estimated.

[30] It is unobserved because an individual's type is not always identifiable from the observed variables.

conditioning variables, the conditional distributional treatment effects help to explain what drive the overall change in the earnings distribution. Although they are not decompositions of the total effects, they would still be very informative if researchers are interested in particular welfare groups and/or when the distributional treatment effects differ greatly across groups. In my case, the conditional distributional treatment effects are calculated for the rural and urban areas. That is, in addition to its effects on the entire distribution, I am interested in how primary education completion affects the earnings distributions of the rural and urban areas separately. The distributional treatment effects are expected to differ greatly with the area of residence, due to a different sectoral composition, and thus required educational qualifications in each area.

5.4.4 Identifying Local Distributional Treatment Effects: RD vs IV

As mentioned earlier, my approach to identify the quantile and inequality treatment effects follows the RD framework developed by Frölich and Melly (2010b) and Frandsen et al. (2012). This sub-section briefly discusses how this method is developed, and how the selected RD estimators compare with other related approaches.

The distributional treatment effects in the RD design are closely related to studies using a (binary) instrumental variable (IV) to deal with possible self-selection into a (binary) treatment. The study of Imbens and Rubin (1997) is among the first to estimate the entire distribution of the outcome variable for the compliers under different treatment statuses, using the IV local average treatment effect (LATE) assumptions.[31] Subsequently and more specifically, Abadie et al. (2002) estimate the effects of a treatment on conditional quantiles for compliers in the IV models, also in the LATE framework. Their estimated parameter is the difference in the τ-th (conditional) quantile of the treated and untreated outcomes, not the τ-th (conditional) quantile of the difference in the treated and the untreated outcomes, for the compliers.[32] Under the same LATE framework, Frölich and Melly (2008, 2012) recently developed the IV estimators of the unconditional quantile treatment effects for compliers. They also show how to estimate the unconditional quantile treatment effects when it is

[31]The assumptions, under which an IV estimator can be interpreted as a local average treatment effect, are: (1) stable unit treatment value assumption (SUTVA); (2) exclusion restriction; (3) strict monotonicity; and (4) random assignment of the instrument (Imbens and Angrist, 1994; Angrist et al., 1996).

[32]The latter requires stronger assumptions to identify the joint distribution (of treated and untreated outcomes). For example, Heckman and Smith (1997) discuss the models where features of the distribution of the difference in the treated and the untreated outcomes are identified. They argue that this may be of interest for questions regarding the political economy of social programme.

necessary to include control variables.[33] In this case, the unconditional effects are weighted averages of the conditional effects.[34] The main difference between Abadie et al. (2002) and Frölich and Melly (2008, 2012) apart from the conditional-unconditional interpretations, is that the latter is fully non-parametric while the former imposes a linear specification in the control variables.

Subsequently, Frölich and Melly (2010b) and Frandsen et al. (2012) focus on the LATE framework and develop interpretations of the quantile and inequality treatment effects, both conditional and unconditional, under the RD design.[35] They point out that the RD setting violates the "non-trivial assignment" assumption required in Abadie et al. (2002) and Frölich and Melly (2008; 2012).[36] This is because the RD instrument, Z_i, is determined completely by the running variable, R_i. As a result, the IV quantile treatment effects methods by Abadie et al. (2002) and Frölich and Melly (2008; 2012) cannot always be used for the RD setting.[37]

As formalised by Hahn et al. (2001), RD designs for a framework with a binary treatment and a binary instrumental variable potentially produce more credible causal inference than an IV approach, because they can justify a (local) Wald estimator with milder assumptions.[38] However, when comparing the RD design to the IV approach, it is important to note the following points. First, there are no tests for the validity of neither the RD design nor the IV approach (Lee and Lemieux, 2010).[39] Second, the RD estimates are applicable to the sub-populations of individuals around the discontinuity threshold (Hahn et al., 2001), while the IV-LATE estimates are applicable to the sub-populations affected by the instrument (Imbens and Angrist, 1994). In my case, the identification assumptions are likely to be more plausible under the RD framework. This is due to the fact that the nature of the education

[33]For example, when the instrument is independent of the outcome variable only conditionally on covariates. In addition, Frölich and Melly (2008) show that even if the IV assumptions are valid without conditioning, adding covariates is still helpful to reduce the standard errors of the estimators.

[34]Given the control variable, X, the weight is the density of $dF_{X|C}$ of X, among compliers. See Frölich and Melly (2012) for more details.

[35]The inequality treatment effects measured in Frölich and Melly (2010b) are changes in the Lorenz curve and the Gini coefficient.

[36]The non-trivial assignment condition requires that the instruments are not perfectly determined by other covariates.

[37]In particular, when the running variable is one of the control variables in the conditional IV quantile treatment effects.

[38]When using IV to identify a causal effect, the instrument must be exogenous to all other factors that affect the outcome variable. On the other hand, RD designs only require individuals to have imprecise control over the assignment variables (Lee and Lemieux, 2010).

[39]These untestable assumptions are (1) For the RD design: the continuity in all factors, both observed and unobserved affecting the outcome variable (I2 Local smoothness); and (2) For the IV approach: the instrument must be uncorrelated with the unobserved factors influencing the outcome variable.

reform creates a discontinuity in educational attainment (owing to the year of birth, which cannot be precisely controlled). In addition, the local smoothness assumption (I2) is unlikely to be controversial.

Note that, recently developed alternative approaches to estimate quantile treatment effects with endogenous treatment are from Chernozhukov and Hansen (2005) and Guiteras (2008), for the IV and RD estimates respectively. The main differences between this alternative literature and the LATE literature are in terms of identifying assumptions. First, the rank invariance assumption is required in Chernozhukov and Hansen (2005) and Guiteras (2008). This is a very strong assumption, implying that a treatment must not alter the rank ordering of individuals. Meanwhile, the authors do not require that exceeding the discontinuity threshold have a monotonic effect on treatment status. As a result, their estimates are applicable to the entire population, compared to only a sub-population under the LATE framework. Second, Guiteras (2008)'s approach is only applicable for a continuous outcome variable, while those of the LATE framework (Abadie et al., 2002; Frölich and Melly, 2010b; Frandsen et al., 2012) allow for the outcome variable to be discrete, continuous, or hybrid (such as earnings with a mass point at zero or top coding). Although the LFS earnings information is not top-coded and my sample of interest is employed workers with positive earnings, the distributional treatment effects from the LATE framework would be generally more applicable to the analysis of earnings due to the nature of the outcome variable.

5.5 Estimation of Local Distributional Treatment Effects in the RD Design

This section describes the empirical estimation method used to estimate the counterfactual distributions of potential earnings and the distributional treatment effects for the compliers in the RD framework as specified in Equations (5.5), (5.6), (5.7), and (5.8).

A simple way of estimating the distribution of potential earnings for compliers, $F_{Y^1|C}(y)$ and $F_{Y^0|C}(y)$ as described in Equations (5.5) and (5.6), in the RD design is to use polynomial regressions. Also, the RD polynomial regressions are appropriate as my running variable, year of birth, is discrete (see the discussion in Chapter 3, Sub-section 3.3.2).

As this is a fuzzy RD design, the probability of completing primary education, D_i, as a function of the year of birth, R_i can be written as

$$
\begin{aligned}
E\left(D_i \mid R_i = r\right) = & \quad \Pr\left(D_i = 1 \mid R_i = r\right) = & \quad \gamma + \delta \cdot Z_i + g\left(R_i - r_0\right) \\
D_i = & \quad \Pr\left(D_i = 1 \mid R_i = r\right) + v_i = & \quad \gamma + \delta \cdot Z_i + g\left(R_i - r_0\right) + v_i
\end{aligned}
$$

where $Z_i = \mathbb{1}\left\{R_i \geq r_0\right\}$ indicates whether the running variable exceeds the cut-off value, $r_0 = 1967$ is the threshold year of birth as before, $g\left(\cdot\right)$ is a polynomial function of $R_i - r_0$, and v_i is an error term which is independent of R_i.[40]

First, consider the regression for the cumulative distribution of earnings, on both sides of the threshold year of birth.

$$
\begin{aligned}
\left(\hat{\alpha}_{1l},\, \underline{\hat{\beta}}_{1l}\right) &= \arg\min_{\alpha_{1l},\, \underline{\beta}_{1l}} \sum_{R_i < r_0} \left(\mathbb{1}\left\{Y_i \leq y\right\} \cdot D_i - \alpha_{1l} - f_{1l}\left(R_i - r_0\right)\right)^2 \\
\left(\hat{\alpha}_{1r},\, \underline{\hat{\beta}}_{1r}\right) &= \arg\min_{\alpha_{1r},\, \underline{\beta}_{1r}} \sum_{R_i \geq r_0} \left(\mathbb{1}\left\{Y_i \leq y\right\} \cdot D_i - \alpha_{1r} - f_{1r}\left(R_i - r_0\right)\right)^2
\end{aligned}
$$

and

$$
\begin{aligned}
\left(\hat{\alpha}_{0l},\, \underline{\hat{\beta}}_{0l}\right) &= \arg\min_{\alpha_{0l},\, \underline{\beta}_{0l}} \sum_{R_i < r_0} \left(\mathbb{1}\left\{Y_i \leq y\right\} \cdot \left(1 - D_i\right) - \alpha_{0l} - f_{0l}\left(R_i - r_0\right)\right)^2 \\
\left(\hat{\alpha}_{0r},\, \underline{\hat{\beta}}_{0r}\right) &= \arg\min_{\alpha_{0r},\, \underline{\beta}_{0r}} \sum_{R_i \geq r_0} \left(\mathbb{1}\left\{Y_i \leq y\right\} \cdot \left(1 - D_i\right) - \alpha_{0r} - f_{0r}\left(R_i - r_0\right)\right)^2
\end{aligned}
$$

where $\underline{\beta}_{jl}$ and $\underline{\beta}_{jr}$ are vectors of polynomial coefficients from the corresponding polynomial functions, $f_{jl}\left(\cdot\right)$ and $f_{jr}\left(\cdot\right)$, for $j \in \{0, 1\}$.

Second, consider the regression for the primary education completion indicator.

$$
\begin{aligned}
\left(\hat{\gamma}_{1l},\, \underline{\hat{\phi}}_{1l}\right) &= \arg\min_{\pi_{1l},\, \underline{\phi}_{1l}} \sum_{R_i < r_0} \left(D_i - \gamma_{1l} - g_{1l}\left(R_i - r_0\right)\right)^2 \\
\left(\hat{\gamma}_{1r},\, \underline{\hat{\phi}}_{1r}\right) &= \arg\min_{\pi_{1r},\, \underline{\phi}_{1r}} \sum_{R_i \geq r_0} \left(D_i - \gamma_{1r} - g_{1r}\left(R_i - r_0\right)\right)^2
\end{aligned}
$$

and

[40] The sharp RD is a special case when $\gamma = 0$, $g\left(\cdot\right) = 0$, and $\delta = 1$.

$$\left(\hat{\gamma}_{0l}, \underline{\hat{\phi}}_{0l}\right) = \arg\min_{\pi_{0l}, \underline{\phi}_{0l}} \sum_{R_i < r_0} \left((1 - D_i) - \gamma_{0l} - g_{0l}(R_i - r_0)\right)^2$$
$$\left(\hat{\gamma}_{0r}, \underline{\hat{\phi}}_{0r}\right) = \arg\min_{\pi_{0r}, \underline{\phi}_{0r}} \sum_{R_i \geq r_0} \left((1 - D_i) - \gamma_{0r} - g_{0r}(R_i - r_0)\right)^2$$

where $\underline{\phi}_{jl}$ and $\underline{\phi}_{jr}$ are vectors of polynomial coefficients from the corresponding functions, $g_{jl}(\cdot)$ and $g_{jr}(\cdot)$, for $j \in \{0, 1\}$.

With the same orders of polynomial used in $f(\cdot)$ and $g(\cdot)$ on each side of the threshold year of birth, the counterfactual cumulative distributions for the compliers (Equations (5.5) and (5.6)) can be estimated as the ratios of the discontinuities in the cumulative distribution of earnings and the primary education completion indicator (Imbens and Lemieux, 2008) as follows.

$$\hat{F}_{Y^1|C}(y) = \frac{\hat{\alpha}_{1r} - \hat{\alpha}_{1l}}{\hat{\gamma}_{1r} - \hat{\gamma}_{1l}}$$
$$\hat{F}_{Y^0|C}(y) = \frac{\hat{\alpha}_{0r} - \hat{\alpha}_{0l}}{\hat{\gamma}_{0l} - \hat{\gamma}_{0r}}$$

Alternatively, the cumulative distributions of the potential earnings for the compliers can be estimated by polynomial regressions with an instrumental variable (Hahn et al., 2001; Lee and Lemieux, 2010). Define

$$P_i = \begin{bmatrix} \pi_1 \\ \mathbb{1}\{R_i < r_0\} \cdot f_l(R_i - r_0) \\ \mathbb{1}\{R_i \geq r_0\} \cdot f_r(R_i - r_0) \end{bmatrix}$$

where $f_l(\cdot)$ and $f_r(\cdot)$ are polynomial functional forms. Then, the estimated equation can be written as

$$\mathbb{1}\{Y_i \leq y\} \cdot D_i = \alpha \cdot D_i + P_i^{'} + \varepsilon_i \tag{5.9}$$

By estimating Equation (5.9), using the two-stage least squares (2SLS) method with $Z_i = \mathbb{1}\{R_i \geq r_0\}$ as the excluded instrument for D_i, the estimate of α is identical to $\hat{F}_{Y^1|C}(y)$. Similarly, the estimator $\hat{F}_{Y^0|C}(y)$ can be obtained as the 2SLS estimator of θ from the following equation.

$$\mathbb{1}\{Y_i \leq y\} \cdot (1 - D_i) \quad = \quad \theta \cdot (1 - D_i) + P_i^{'} + \varepsilon_i \tag{5.10}$$

Note that I allow for the slopes and curvatures of the regression lines to be different on each side of the discontinuity threshold. For example, in the quadratic case where $f_l(R_i - r_0) = \pi_{l2} \cdot (R_i - r_0) + \pi_{l3} \cdot (R_i - r_0)^2$ and $f_r(R_i - r_0) = \pi_{r2} \cdot (R_i - r_0) + \pi_{r3} \cdot (R_i - r_0)^2$, the estimated equation for the distribution of the potential earnings of the treated group (Equation (5.9)) will be

$$\begin{aligned}\mathbb{1}\{Y_i \leq y\} \cdot D_i \quad = \quad & \alpha \cdot D_i + \pi_1 + \mathbb{1}\{R_i < r_0\} \cdot \left[\pi_{l2} \cdot (R_i - r_0) + \pi_{l3} \cdot (R_i - r_0)^2\right] \\ & + \mathbb{1}\{R_i \geq r_0\} \cdot \left[\pi_{r2} \cdot (R_i - r_0) + \pi_{r3} \cdot (R_i - r_0)^2\right] + \varepsilon_i\end{aligned}$$

5.5.1 Choice of Polynomial Order

As discussed in the RD returns to education chapter (Chapter 3), it is important to test whether the selected polynomial regressions are well specified and lead to correct estimates of discontinuity (Lee and Lemieux, 2010).[41] The polynomial model can be tested using a goodness-of-fit statistic, as specified in Chapter 3 (Sub-section 3.3.3).

It is common to use the year-of-birth clustered standard errors as the polynomial regressions introduce a group structure to the standard errors (Lee and Card, 2008). However, in this chapter, I will present bootstrapped standards errors of the estimators, as my main variables of interest are earnings quantiles and inequality measures, which are regenerated from the estimated cumulative distributions of earnings.

5.5.2 Unconditional Distributional Treatment Effects: Adding Covariates

The unconditional inequality treatment effects under the RD design can be obtained from the polynomial regression model as discussed above. Additionally, Frölich and Melly (2010b) show how to incorporate additional covariates in the RD estimation of the unconditional cumulative distributions of the outcome. There are two main reasons for checking the robustness of the results when covariates are included. First, the inclusion of observed covariates

[41] The choice of the order of polynomial regression is equivalent to the bandwidth choice in the non-parametric estimation when a running variable is continuous.

as controls can be interpreted as "the test for an imbalance in relevant characteristics" (Van Der Klaauw, 2008). If RD estimates are sensitive to the inclusion of individual covariates, the local smoothness assumption (**I2**) is likely to be violated.[42] Second, including observed covariates increases the precision of the estimates in the case of average effects (Frölich, 2007).

Following Frölich and Melly (2010b), when Assumptions **I1** to **I4** hold conditionally on the covariates, X, and the common support restriction holds, the identification results above now apply immediately to the counterfactual distributions as well as the distributional treatment effects conditionally on X.[43] They further demonstrate that the unconditional effects, which are the effects for all compliers irrespective of their values of X, can be obtained by first calculating the counterfactual distributions conditional on $X = x$ using the propensity score match method, and thereafter integrating these conditional counterfactual distributions with respect to X.[44]

Following Frölich and Melly (2010b), let $p_\varepsilon(x) = \Pr(R \geq r_0 \mid X = x,\, R \in (r_0 - \varepsilon, r_0 + \varepsilon))$. The unconditional counterfactual cumulative distribution with covariates, X, can be written as[45]

$$F_{Y^1|C}(y) = \lim_{\varepsilon \to 0} \frac{E\left[\mathbb{1}\{Y \leq y\} \cdot \frac{\mathbb{1}\{R \geq r_0\} - p_\varepsilon(x)}{p_\varepsilon(x)\cdot(1-p_\varepsilon(x))}(2D-1) \mid R \in (r_0 - \varepsilon, r_0 + \varepsilon),\, D = 1\right]}{E\left[\frac{\mathbb{1}\{R \geq r_0\} - p_\varepsilon(x)}{p_\varepsilon(x)\cdot(1-p_\varepsilon(x))}(2D-1) \mid R \in (r_0 - \varepsilon, r_0 + \varepsilon),\, D = 1\right]} \tag{5.11}$$

$$F_{Y^0|C}(y) = \lim_{\varepsilon \to 0} \frac{E\left[\mathbb{1}\{Y \leq y\} \cdot \frac{\mathbb{1}\{R \geq r_0\} - p_\varepsilon(x)}{p_\varepsilon(x)\cdot(1-p_\varepsilon(x))}(2D-1) \mid R \in (r_0 - \varepsilon, r_0 + \varepsilon),\, D = 0\right]}{E\left[\frac{\mathbb{1}\{R \geq r_0\} - p_\varepsilon(x)}{p_\varepsilon(x)\cdot(1-p_\varepsilon(x))}(2D-1) \mid R \in (r_0 - \varepsilon, r_0 + \varepsilon),\, D = 0\right]} \tag{5.12}$$

Subsequently, the unconditional quantile and inequality treatment effect can be obtained as defined in Equations (5.7) and (5.8). It is important to note that, the unconditional effects with covariates should not depend on the variables included in X. Different sets of control variables X can be used to estimate the same object, which is useful for examining robustness of the results to the set of control variables.

[42] In general, this is equivalent to, but more comprehensive than, testing whether the relationship between the running variable and other observed covariates is smooth around the neighbourhood of the discontinuity.

[43] The common support restriction is defined as $\lim_{r \to r_0^+} \operatorname{supp}(X \mid R = r) = \lim_{r \to r_0^-} \operatorname{supp}(X \mid R = r)$.

[44] Under the condition that the rate of convergence does not depend on the number of covariates (Frölich, 2007; Firpo and Pinto, 2011).

[45] The proof taken from Frölich and Melly (2010b) is provided in detail in the Appendix, p. 189.

5.5.3 Conditional Distributional Treatment Effects: Different Effects in Rural and Urban Areas

As mentioned briefly in Sub-section 5.4.3, the conditional distributional treatment effects allow the impacts of primary education completion on earnings distribution to be heterogeneous across control observables. Given the impacts of primary education completion on the entire distribution of earnings, the conditional effects may help to explain the underlying mechanisms that drive the overall changes.

In this chapter, I am interested in how primary education completion affects the overall earnings inequality as well as that of the rural and urban areas separately. The conditional counterfactual distributions of earnings can be obtained by estimating the counterfactual cumulative distributions of earnings for the urban and rural sub-populations separately.

$$F_{Y^1|C,\,X=u}(y) = \frac{\lim_{r\to r_0^+} E[\mathbb{1}\{Y\le y\}\cdot D|R=r,\,X=ur]-\lim_{r\to r_0^-} E[\mathbb{1}\{Y\le y\}\cdot D|R=r,\,X=ur]}{\lim_{r\to r_0^+} E[D|R=r,\,X=ur]-\lim_{r\to r_0^-} E[D|R=r,\,X=ur]} \tag{5.13}$$

$$F_{Y^0|C,\,X=u}(y) = \frac{\lim_{r\to r_0^+} E[\mathbb{1}\{Y\le y\}\cdot(1-D)|R=r,\,X=ur]-\lim_{r\to r_0^-} E[\mathbb{1}\{Y\le y\}\cdot(1-D)|R=r,\,X=ur]}{\lim_{r\to r_0^+} E[1-D|R=r,\,X=ur]-\lim_{r\to r_0^-} E[1-D|R=r,\,X=ur]} \tag{5.14}$$

where u is the dummy for residing in an urban area.

5.5.4 Choice of Inequality Measures

The economic literature on inequality generally agrees that, there are four properties that inequality measures should satisfy (see for example, Anand (1983); Heshmati (2004); Cowell (2011)). First, mean or scale independence requires that the inequality measure is invariant to a change in everyone's income by the same proportion. Second, population size independence means that the inequality measure remains unchanged when the number of people at each earnings level changes by the same proportion. Third, the Pigou-Dalton condition indicates that the inequality measure decreases when there is any transfer of a positive amount from a richer to a poor person that does not change their relative ranks. Fourth, sub-group decomposability means that the inequality measures of the entire population can be additively decomposed into the inequality within sub-group as well as the inequality between these sub-groups (Shorrocks, 1984).

The first three properties are generally required for studying inequality in the overall distribution or that of a specific sub-group, while the decomposability property is required for understanding of sub-group factors determining overall inequality.

In this chapter, I focus on the following inequality measures: the Gini coefficient (G), the standard deviation of logarithms (SDL), Theil's entropy index L (Theil-L, L), and Theil's entropy index T (Theil-T, T). They are defined as follows.

$$\begin{aligned}
G &= 1 - \frac{1}{\mu} \int_{0}^{\infty} (1 - F(y))^2 \, dy \\
SDL &= \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (\ln Y_i - \ln \tilde{\mu})^2} \\
L &= \sum_{i=1}^{N} \frac{1}{N} \cdot ln \left(\frac{\mu}{Y_i} \right) \\
T &= \sum_{i=1}^{N} \frac{Y_i}{Y} \cdot ln \left(\frac{Y_i}{\mu} \right)
\end{aligned}$$

where Y_i is earnings of an individual i, $F(y)$ is the cumulative distribution function of earnings, Y is the total earnings of the entire population, N is the number of observations, μ is the arithmetic mean earnings, and $\tilde{\mu}$ is the geometric mean earnings.

All of these inequality measures, except for the standard deviation of logarithms, satisfy the mean independence, the population size independence, and the Pigou-Dalton properties.[46] The Gini coefficient and the standard deviation of logarithms are more traditional measures of inequality. The Gini coefficient measures the ratio of the area between the Lorenz curve, which plots the income share of the cumulative population, and the equality line to the area of the triangle under the equality line. In a general case of non-negative earnings, it ranges from 0 (perfect equality) and 1 (perfect inequality). The standard deviation of logarithms is the more straight-forward inequality measure, which is a useful summary measure when earnings are approximately log-normal and are less sensitive to extreme outliers at the top.[47] On the other hand, the Theil-L and Thiel-T indices measure the divergence between earnings shares and population shares (Theil, 1967; Anand, 1983). When there is perfect

[46] The standard deviation of logarithms does not satisfy the Pigou-Dalton condition for earnings above a certain level of income (Anand, 1983)

[47] It is not defined if there is a person in the distribution with zero or negative earnings, which does not happen in my case.

equality, both the Theil-L and Theil-T indices are zero. Among these inequality measures, the Theil-L index is most sensitive to the changes at the bottom of the distribution, while the Gini coefficient is sensitive to the changes around the middle of the distribution. The Gini coefficient is not additively decomposable by sub-groups, whereas the other three selected inequality measures differ in their static and dynamic sub-group decomposability (Anand, 1983; Cowell, 2011).

5.6 The Data and Descriptive Statistics

This section describes the nature of earnings inequality in Thailand from the LFS data, as well as the variables used in the estimation of the distributional effects of primary education completion under the RD framework. First, I discuss the criteria for selecting the sample used in the empirical estimation. Subsequently, descriptive statistics of the main variables for the estimation sample are presented in order to provide an overview of earnings inequality and the potential role of primary education in influencing the earnings distribution in Thailand.

5.6.1 Data Source

The empirical analysis in this chapter draws upon labour force data from the Thai LFS during the years 1991 to 2000. The sample of interest is confined to paid workers, whose ages range from 20 to 40 years during the survey years 1991 to 2000, for three major reasons. First, this chapter focuses on how primary education completion impacts on earnings inequality and the earnings information is only available for those who are employed and paid.[48]

Second, the sample includes workers aged between 20 and 40 years old during the selected survey years (1991 to 2000) because they were more likely to be affected by my key variable, the Education Reform Act 1978.[49] Most Thai workers have completed education by the age of 20, and so, this is taken as the age of entering the labour market. On the other hand,

[48]This analysis, therefore, only answers questions about inequality in the labour market, and does not say anything about poverty and income inequality. It is important to emphasise that inequality in the labour market does not always have implications on the overall welfare for two main reasons. First, earnings are labour income, which is one of the many sources of income influencing a consumption pattern, and thus utility. The other sources of income are, for instance, capital income, remittance and transfers, and in-kind benefits. In addition, welfare is often measured by consumption expenditures, in particular in developing countries in which consumption maybe better reflects long-term resources available to individuals. Second, earnings are a measure of income at the individual level. Due to resource sharing and economies of scale in a household, welfare, if indicated by income, is generally measured at a per-capita or per-equivalent-size unit.

[49]This results in the sample of paid workers who were born during the years 1951 to 1980, which provides the observations 16 years before 1967 (the first cohort under the 1978 reform) and 13 years subsequent to 1967.

capping the age at 40 years helps to capture the effects of the reform. This is because persons aged above 40 years during the selected survey years were in school much earlier than when the reform was implemented.[50]

Third, although the LFS data have been available since 1985, the analysis in this chapter uses only the information from 1991 to 2000 because 1985 is too early as a starting year for the analysis. In 1985, individuals who were part of the first cohort under the 1978 reform were excluded from my sample of analysis as they were only 18 years old in that year.[51] Hence, these early years do not add much useful information for the RD framework, especially for a separate analysis of each survey year which is considered in this chapter.

5.6.2 Descriptive Statistics

Earnings, Earnings Inequality, and Educational Attainment

Figures 5.1 summarises the two main variables of interest, real hourly earnings and educational attainment of paid workers, aged from 20 to 40 years old, from the LFS 1991 to 2000. Real hourly earnings take into account the over-time changes in price-levels as well as the regional price differentials. The pooled sample data shows a marked increase in the mean and a decline in inequality (Figure 5.1a). Between 1991 and 1996, the average (median) real hourly earnings increased by 5.1 (6.5) per cent per annum, while the Gini coefficient decreased by 4 percentage points from 0.50 to 0.46. The average real hourly earnings peaked in 1997 before sharply declining as a result of the financial crisis and remained stable until the end of the 1990s. Earnings inequality, measured by the Gini coefficient, also rose back to the initial level in the financial crisis year of 1997 before dropping to 0.45 by the end of the 1990s. Other earnings inequality measures, including the standard deviation of logarithms (SD of log) and the two Theil indices, exhibit similar patterns of a very gradual decline in inequality between 1991 and 1996 and an upward jump during the financial crisis year.[52] Figure 5.1b shows that the gradual decline in earnings inequality during the 1990s may be explained by faster growth rates in earnings of workers at the bottom quintiles, compared to those at the top, in most years except for the bubble period before the financial crisis.

[50]For instance, a worker whose age was 50 years old in the survey year 1991 (2000) was born in 1941 (1950), and therefore, was not subject to the new compulsory schooling law as he was in the fourth year of school (lower primary education) in 1952 (1961).

[51]These are individuals who were born in 1967.

[52]The relative change seems larger in the two Theil measures, which are both bottom-sensitive.

Over the same period, the average educational attainment of paid workers increased from 6.7 years in 1991 to 8.1 years in 2000 (Figure 5.1c). The proportion of workers with less than primary education has dropped remarkably and continuously, while those of other levels of education have increased accordingly. Figure 5.1d suggests that the composition of those who completed at least the primary level differ greatly across the earnings distribution. A clear majority of workers in the lower earnings quintiles, who attained primary school education and higher, quit school right after their primary education completion. On the other hand, most of those at the top earnings quintile continued their studies until they finished at least high school. This implies that the increase in the primary completion rate is driven by different mechanisms in different earnings groups.

Figure 5.2 presents the same descriptive statistics of earnings and education by area of residence (rural and urban). Albeit with similar changes over time, average earnings were substantially higher in urban areas, than rural areas (Figure 5.2a). On the other hand, earnings inequality within the rural and urban areas declined slightly and was of comparable levels. Similarly to the pooled observations, the poor attained higher annual earnings growth in most years, compared to the rich (Figure 5.2c). More interestingly, Figure 5.2b summarises the average educational attainment which differed greatly between the rural and urban areas. The average years of education increased from 5.6 years in 1991 to 7 years in 2000 for the rural areas, and from 8.6 years in 1991 to 9.7 years in 2000 for the urban areas. The increase in average years of education in both the rural and urban areas was mainly driven by sharp declines in the proportion of primary school dropouts. This coincided with the increasing proportion of primary school graduates in the rural areas, while in the urban areas a rising proportion of high school graduates is observed.

Other Labour Force Characteristics

Table 5.1 summarises other important characteristics of paid workers aged from 20 to 40 years old during the survey years 1991 to 2000. The average age of these observations has remained stable at around 30 to 31 years old. The average age of the rural sample has been slightly higher than that of the urban sample but this gap has narrowed over time. Additionally, the proportion of male paid workers has been higher in rural areas. The average hours of work for these paid workers have been around 50 hours per week, and have been slightly higher for the rural workers, compared to the urban workers. This suggests no evidence of underemployment in the LFS.

The 1978 Education Reform

From the aforementioned descriptive statistics, the increase in average years of education coincided with the rapid decline in the proportion of workers with less than primary education. This could be potentially driven by the Education Reform Act 1978, which increased compulsory education from 4 years of lower primary to 6 years of primary education, free of change. In other words, holding other factors constant, a person who was born in 1966 and only studied for 4 years (of lower primary education) would have been forced to stay in school for two more years until he finished the primary level, had he been born a year later.[53] This resulted in a sharp increase in the proportion of primary school graduates for the cohorts born in 1967 and after, as demonstrated in Figure 5.3.[54]

5.6.3 Data Summary

To summarise, the LFS data exhibit the evolution of earnings, earnings inequality, and educational attainment during the period of 1991 to 2000. While educational attainment and average real hourly earnings increased consecutively, earnings inequality remained rather stagnant and declined only slightly over the same period. These patterns hold when examining the rural and urban areas separately. Note that changes in education completion rates were most pronounced at the primary level due to a strong decline in the proportion of workers dropping out of primary school. This is potentially driven by the Educational Reform Act 1978 which increased compulsory education from 4 years of lower primary to 6 years of primary education. As studied in the other two chapters, education has an important role in increasing labour earnings (at the average level), and shifting workers towards the more productive sectors during the structural transformation process of Thailand. It would be very interesting to investigate how the observed increase in primary education completion, in response to the change in the compulsory schooling law, interacts with earnings inequality. That is, how education, measured by primary education completion, affects labour earnings at different points of the distribution.

[53] The birth year of the first cohort that was affected by the reform is 1967.

[54] Primary school graduates refer to persons who completed at least primary education.

5.7 Empirical Evidence: Overall Distributional Impacts of Primary Education Completion on Earnings

This section investigates the unconditional impacts of primary education completion on the distribution on hourly earnings in Thailand during the period 1991 to 2000. The distributional effects are estimated for the pooled year sample and each of the survey year samples. For ease of interpretation, the empirical results are presented in a graphical format.[55] First, the two counterfactual distributions of earnings are estimated, using the RD method as proposed in Sections 5.4 and 5.5. Second, to assess the overall distributional impacts of primary education completion, I analyse the quantile treatment effects, which are the effects of primary education completion on earnings at each of the earnings quantiles. Third, to obtain a statistic of the inequality impacts, the changes in the selected inequality measures are estimated and discussed. I also discuss how these estimates are relevant to the observed inequality in Thailand.

5.7.1 Specification: Choice of Polynomial Order

I use the goodness-of-fit test allowing for a heteroskedastic error term as described in Chapter 3 (Sub-section 3.3.3) to select an appropriate polynomial order of the RD polynomial regressions. To choose a polynomial order for the estimated cumulative earnings distribution when every complier completed at least primary education, I regress $\mathbb{1}\{Y_i \leq y\} \cdot D_i$ on a full set of dummy variables for all birth years and then add higher-order terms of the normalised distance from the cut-off year of birth until those year of birth dummies are no longer jointly significant.[56] Subsequently, a polynomial order for the estimated cumulative earnings distribution when none of the compliers completed primary education can be obtained by repeating this procedure with $\mathbb{1}\{Y_i \leq y\} \cdot (1 - D_i)$ as a dependent variable.

Table 5.2 presents the p-values from the Chi-Square tests of the (joint) statistical significance of year of birth dummies with the first (linear) to the fifth-order polynomials of the distance from the cut-off year of birth. These regressions are estimated for the pooled year (1991-2000) sample at 4 different values of y, corresponding to the earnings at the 20th, 40th, 60th, and 80th percentiles.[57] A p-value greater than the significance level, α, implies that

[55] This is also due to the large number of parameter estimates from quantiles and inequality measures.
[56] That is the higher order terms of $R_i - r_0$ or year of birth $- 1967$.
[57] The Chi-Square test results (not shown) are robust to the ten separate samples for each of the survey

the selected polynomial order explains the data well. From the Chi-Square test results, the second- and higher-order polynomial functions of the distance from the cut-off year of birth effectively fit the data. This is because once the second- and higher-polynomial orders are added, the coefficients on the years of birth dummies become statistically insignificant at the 10 per cent significance level.[58] I therefore use the quadratic function of the distance from the cut-off year of birth in my main RD specification. Additionally, I provide the estimation results for the alternative higher polynomial orders as a robustness check in Section 5.9.

5.7.2 First-Stage Discontinuity: Testing Assumption I1

The first-stage discontinuity shows how the effects of primary education completion are identified by the year of birth. It can be naturally and transparently presented, using a simple graph of the average value of the primary education completion dummy for each of the birth cohorts.[59] For the estimation of the unconditional distributional treatment effects, the first-stage discontinuity is shown in Figure 5.3a. For the analysis of the pooled year sample, the primary education completion rate jumps from 44 per cent for the 1966 cohort to 74 per cent for the 1967 cohort, and to 95 per cent for the 1980 cohort (top left panel in Figure 5.3a). The discontinuities in the rate of primary education completion are also significant and around 22 to 32 per cent when considering each of the survey year samples separately.

Figure 5.3b presents the first-stage discontinuities for the estimation of the distributional treatment effects conditional on the residential location. These jumps in the primary education completion rate are drawn for the rural and urban samples separately. On average, the discontinuities in the rate of primary education completion are 35 and 21 per cent for the rural and urban areas, respectively. The smaller discontinuity in the urban areas could be explained by the fact that the proportion of primary school graduates was already higher in the urban areas prior to the reform. This suggests a lower proportion of compliers in urban areas, compared to rural areas.

---

years.

[58]Note that a higher significance level provides a more stringent test as I look for the case of failure to reject the null hypothesis.

[59]This is because my running variable is year of birth, which is discrete. For the case of a continuous running variable, the first-stage discontinuity must be checked by graphing the means of the treatment variable for each of the bins of the running variable from a formal bandwidth selection procedure, such as the cross-validation criterion (Lee and Lemieux, 2010).

The year of birth discontinuity as an instrument clearly predicts primary school completion, as required by assumption **I1**.

5.7.3 Unconditional RD Counterfactual Earnings Density

Following the method proposed by Frölich and Melly (2010b) and Frandsen et al. (2012), the two counterfactual earnings distributions for the compliers are calculated from the polynomial regressions as discussed in Section 5.5. The two counterfactual earnings distributions are estimated from Equations (5.9) and (5.10) with 150 different values of y.[60]

Figure 5.4a plots the estimated density functions of potential log hourly earnings, obtained from the cumulative earnings distributions estimated for the pooled sample and for each survey year separately. The non-treated outcome line refers to the estimated density function of log hourly earnings for the compliers when everyone drops out before completing primary school, whereas the treated outcome line refers to the case when all compliers are primary school graduates. For the pooled year sample, the graphs show very clearly that, for the population of compliers who were born close to the cut-off year 1967, primary education completion leads to a rightward shift in earnings, especially for the poor. As a result, the earnings density becomes more concentrated around the mode.

Considering the estimated density functions from each of the survey years separately, primary education completion significantly narrows the earnings density. However, the distributional effect declines over time. During the early 1990s, primary education completion had a relatively larger impact on increasing the earnings of those in the bottom half of the earnings distribution. This suggests an increase in the share of total earnings going to those in the bottom half, which potentially reduces earnings inequality, as a result of primary education completion. The impact of primary education completion on narrowing the earnings density is significant, but less pronounced in the later years before the financial crisis in July 1997. On the other hand, for the post-financial crisis years, primary education completion shifts nearly every compliers' earnings to the right. Therefore, the shape of the earnings density stays relatively unchanged. These results are also robust to the inclusion of covariates such as gender, province, and area of residence (Figure 5.A.1a in the Appendix), suggesting that the local smoothness assumption (**I2**) is satisfied.[61]

[60] That is, the counterfactual distributions are evaluated at 150 points. The estimation at each point involves a separate regression.

[61] For the results from the RD counterfactual distribution with covariates, the year dummies are also

Additionally, the density (and thus the distribution) of log hourly earnings can be considered reasonably smooth. This allows me to invert the distribution functions to obtain the marginal distributions of log hourly earnings at a particular quantile, which will be subsequently used to for the calculation of the quantile treatment effects.

5.7.4 Unconditional Quantile Treatment Effects

Quantile regressions provide the overall picture of the distributional impacts of primary education completion, as they allow for the impacts of primary education completion to vary along the entire earnings distribution. First, I estimate the unconditional Mincerian earnings function using the ordinary least squares approach (OLS) and the unconditional quantile regression approach by assuming that the primary education completion variable is exogenous.[62] The exogenous quantile regression estimates are informative and potentially indicate the bias that might be introduced by ignoring the endogeneity issues. For instance, an upward bias in the inequality impacts may suggest a positive correlation between education and ability for high-earnings groups, and between returns to primary education completion and earnings.[63] Subsequently, the unconditional quantile treatment effects of primary education completion, allowing for endogenous education choice, are calculated from the RD counterfactual distributions of log hourly earnings obtained in the previous sub-section. Note that these coefficient estimates are expected to be high. This is due to the fact that primary education completers could be either those who studied just up to the primary level or those who attained a higher level of education. In other words, the average and the quantile effects can be interpreted as the unconditional earnings premiums for all primary school graduates, relative to the primary school dropouts.

OLS and Quantile Regressions

Figure 5.4b reports the unconditional average effects from the OLS and the quantile effects from the quantile regressions for each of the ventiles. These coefficient estimates are obtained

included in the pooled year estimations. These counterfactual distributions are calculated from integrating the conditional effects with respect to each respective category of control variables as discussed in Sub-section 5.5.2.

[62]To make the estimates comparable to those from the unconditional quantile treatment effects, there are no other control variables in the OLS and the quantile regression estimations. For the pooled year regressions, the year dummies are included.

[63]That is, the exogenous estimates are higher than the endogenous estimate, in particular for individuals at the top of the earnings distribution.

by assuming that the completion of primary education is uncorrelated with other earnings-enhancing attributes. According to the pooled year sample regressions, the unconditional average return to primary education completion is around 71 per cent. In other words, the hourly earnings of those completing at least primary education are, on average, 71 per cent higher than those of the primary school dropouts. These average earnings premiums between the primary school graduates and the dropouts decreased over time from 85 per cent in 1991 to 58 per cent in 2000.

In the same figure, the unconditional exogenous quantile treatment effects, however, show that the average estimates conceal important heterogeneity in the impacts of primary education completion across the earnings distribution. Taking primary education completion as exogenous, primary education completion is expected to reduce earnings inequality as it yields the highest increase in hourly earnings for the bottom 20 per cent of the earnings distribution. Primary education completion raises hourly earnings for this earnings group by 93 per cent on average, compared to the group without primary education. The impacts among the richest 20 per cent are, on average, 73 per cent, around the unconditional average premium from the OLS regression. On the other hand, the impact of having at least primary education for the workers in the middle of the earnings distribution (40th to 60th percentiles) is around 57 per cent, which is relatively low and is lower than the average impacts.

When estimating the quantile impacts for each year separately, the patterns of negative inequality impacts remain quite similar over time, with the exception of the post-financial crisis years, when the impacts for the poorest and the richest 20 per cent are comparable and relatively close to the average impacts. Note that the high levels of earnings premiums for primary school graduates at the lower end of the earnings distribution are potentially consistent with the received wisdom about higher returns for primary schooling (Psacharopoulos, 1994).[64] If this is the case, the high returns to primary completion among the poor may directly suggest the importance of primary schooling and compulsory education.

Nonetheless, the unconditional OLS and exogenous quantile regression estimations fail to allow factors such as unobserved ability and other unobserved productivity-enhancing characteristics to be correlated with primary education completion. The average and quantile treatment effects estimators from the proposed RD counterfactual distribution of earnings

[64]This interpretation is only correct if most of the low earners with at least primary education studied only up to the primary level.

can be used to control for this unobserved individual heterogeneity. It is then possible to estimate the causal difference in earnings between those who completed at least primary education and those who did not, both on average and for each of the earnings deciles.

RD Quantile Treatment Effects

Figure 5.4c displays the results from the average and quantile estimates of the RD counterfactual earnings distributions without controlling for any covariates, with point-wise bootstrapped 95 per cent confidence bands.[65] The results suggest that the overall effects of primary education completion on hourly earnings are much larger among the poor. Therefore, holding other factors constant, primary education completion is likely to have reduced earnings inequality in Thailand between 1991 and 2000. For the pooled year sample, the estimated average effect of primary education completion on hourly earnings is 16 per cent, which is much lower than the OLS estimates when assuming that primary education completion is exogenous. The large positive bias in the OLS estimation implies a strong positive correlation between unobserved factors and whether or not an individual completes primary level education. This is sometimes called ability bias (Griliches, 1977; Card, 1999). The effect of primary education completion on earnings is negatively associated with earnings as shown by the downward sloping line in the pooled sample. Primary education completion has larger effects in increasing hourly earnings for the poorest 20 per cent (around 43 per cent), whereas these effects become negative for the richest 20 per cent (around minus 8 per cent). At the same time, it increases hourly earnings for the second, third, and fourth quintiles by 25, 14, and 5 per cent, respectively.[66] When allowing for heterogeneous effects over time, the patterns remain similar and of greater magnitudes in the first half of the 1990s. For the second half of the 1990s, the heterogeneity in the impacts of primary education completion is smaller (that is, the line is flatter). The effects of primary education completion on hourly earnings are relatively homogeneous across the earnings distribution, around the average effects, during the two years following the financial crisis.

The results of the RD quantile treatment effects are generally robust to the inclusion of control variables, such as gender, province, and rural-urban area of residence (Figure 5.A.1b

[65] Note that the 95 per cent confidence intervals are wider for the early years of the 1990s due to the smaller sample sizes.

[66] Alternatively, when separating the observations into two halves, the hourly earnings premiums of primary school graduates relative to primary school dropouts are around 30 per cent for the bottom half and only around 1 per cent for the top half.

in the Appendix). The shapes of the quantile treatment effects-earnings profiles remain broadly unchanged. For the pooled year sample, primary education completion increases earnings of the poorest 20 per cent by around 40 to 45 per cent, and reduces earnings of the richest 20 per cent by around 7 to 9 per cent. The estimation results for each year with various control variables also remain robust. That is, primary education completion benefits the bottom half more than the top half and this pattern was much stronger in the first half of the 1990s. Its effects on hourly earnings were roughly similar across earnings groups during the years 1998 and 1999.

5.7.5 Unconditional Inequality Treatment Effects

From the estimated unconditional quantile treatment effects, I can conclude that the effects of primary education completion on hourly earnings are higher for the poor, compared to the median and rich workers, especially in the first half of the 1990s. This suggests that primary education might reduce earnings inequality. In other words, ceteris paribus, a higher proportion of individuals who completed at least primary education would lead to lower earnings inequality. As the quantile treatment effects are calculated from the RD counterfactual earnings distributions, the inequality treatment effects can also be computed accordingly. That is, the inequality treatment effects are equal to the difference in inequality measures from the two counterfactual earnings distributions.

As mentioned earlier, I focus on four inequality measures, namely the SD of log, the Gini coefficient, and the Theil-L and the Theil-T indices of hourly earnings. Figure 5.5a presents the predicted effects of primary education completion on the SD of log and the Gini coefficient for each year during the period from 1991 to 2000, and for different sets of control variables.[67] The predicted effects are calculated from the estimated inequality treatment effects, multiplied by the expected percentage point changes in the primary education completion rates as discussed in Sub-section 5.4.3.[68] The results are generally robust and more precise when the additional control variables are included. On average, the increase in the primary education completion rate is predicted to reduce the SD of log and the Gini coefficient by 0.05 and 0.02 respectively. Considering each survey year separately, the pre-

[67]Negative values mean inequality reductions.

[68]The estimated inequality treatment effects, which are the change in inequality measures when the rate of primary education completion increases by 100 percentage points, together with the bootstrapped 95 per cent confidence intervals, are shown in Figure 5.A.2a in the Appendix.

dicted impacts of primary education completion on the SD of log and the Gini coefficient are significantly negative for most years and never significantly positive. They are statistically insignificant during the post-crisis years (1998 to 1999 for both the SD of log and the Gini coefficient) and in addition during the period 1992 to 1993 for the Gini coefficient. For the years that the predicted impacts are statistically significant, the increased rate of primary education completion is expected to reduce the SD of log by 0.01 to 0.05 and the Gini coefficient by 0.01 to 0.03.

Figure 5.5b shows the predicted effects of primary education completion on the Theil-L and Theil-T indices for each year during the same period. The results are quite similar to those of the SD of log and the Gini coefficient, in terms of the directions and over-time patterns, but less significant statistically in particular for the Theil-T index.[69] They are also robust to the inclusion of control variables and more precise as the bootstrapped 95 per cent confidence intervals shrink, with the exception of 1991, when the predicted effects on the two Theil indices become much larger. Holding other factors constant, the increase in the number of primary school graduates leads to lower earnings inequality, measured by the two Theil indices. The predicted impacts of primary education completion on the Theil-L index range from (minus) 0.01 to (minus) 0.05, except for the post-crisis years (1998 to 1999), whereas those on the Theil-T index are statistically insignificant for most years even after including covariates.

As the predicted inequality treatment effects are more pronounced for the Gini coefficient and the Theil-L index compared to the Theil-T index, the change in the primary education completion rate has a larger effect on the bottom and the middle of the distribution than the top.

5.7.6 Summary of the Overall Distributional Impacts

To summarise, the impacts of primary education completion on earnings are significant, and more importantly are heterogeneous along the earnings distribution. Primary education completion has larger positive impacts on the earnings of the poor and the vulnerable, compared to those of the top earners, especially in the first half of the 1990s. This implies that primary education is one of the fundamental policy tools for tackling overall earnings

[69] The bootstrapped 95 per cent confidence intervals of the estimates are presented together with the estimated inequality treatment effects in Figure 5.A.2b in the Appendix.

inequality in Thailand.

Nonetheless, its impacts on earnings seem to be homogeneous in the late 1990s, after the financial crisis. Although the sources of the changes in impact heterogeneity cannot be concluded in this chapter, this could be explained by the two following reasons. First, the financial crisis had larger negative impacts on the poor (Sussangkarn et al., 1999; Fallon and Lucas, 2002) and on the primary school graduates at the lower end of the relative education distribution (Krongkaew et al., 2006).[70] Presuming a positive correlation between educational attainment and earnings, this implies that the crisis substantially reduced the earnings premiums between the primary school graduates and dropouts for the poor.[71] Second, the rapid increase, and thus potential over-supply, of workers with primary education only may lead to lower returns to primary education, which may be consistent with lower primary school premiums for the lower quantiles.

5.8 Empirical Evidence: Distributional Impacts of Primary Education Completion on Earnings in Rural and Urban Thailand

In this section, I further allow for heterogeneous impacts of primary education completion across rural and urban areas. Policy-makers often relate studies on inequality to inequality within and between rural and urban areas. While primary education is likely to be a basic foundation and compulsory in developing countries, it is not considered a sufficient driver for developed or fast growing economies. Its impacts on earnings and earnings inequality are, therefore, expected to differ with the level of development. This could be directly applied to a specific country study by conducting the analysis of the rural and urban areas separately.

[70] Krongkaew et al. (2006) find that, during the post crisis years, the earnings of the workers with primary and junior secondary education decline while those of the workers with vocational or university education remain stable.

[71] This could be the case when the majority of the primary school graduates in the lower quantiles only studied up to the primary level, while most primary school graduates in the upper quantiles completed other education levels higher than that. This is also supported by the descriptive statistics from the LFS (Figure 5.1d).

5.8.1 Rural-Urban Conditional Quantile Treatment Effects

Rural-Urban OLS and Quantile Regressions

With the assumption of exogenous educational attainment, the upper panels of Figure 5.6 display the average effects of primary education completion on earnings from the OLS and the quantile effects at each of the ventiles from the quantile regressions, for the rural and urban areas separately. These estimates reveal that the effects of primary education completion may not only differ across earnings groups and time, but also across the levels of development of an area.

First, the quantile patterns of the exogenous hourly earnings premiums between the primary school graduates and dropouts for rural workers (Figure 5.6a) are reasonably similar to those combining rural and urban observations. This is not too surprising given that majority of population is in rural areas. According to the pooled year regression for the rural areas, hourly earnings of the primary school graduates are, on average, 59 per cent higher than those of the dropouts. The hourly earnings premiums in the rural area increased slightly in the first few years of the 1990s before declining over time. Considering the premiums in the rural area at each ventile, the pooled year regression indicates that the primary school graduate premiums are around 75, 47, and 58 per cent for the bottom, middle, and top 20 per cent of the rural earnings distribution, respectively.[72] If primary education completion was not correlated to other productivity-enhancing factors, these premiums would suggest potential negative impacts of primary education completion on earnings inequality. Although the impacts on earnings inequality are slightly weaker for 1998, the patterns of the negative impacts generally hold over time when looking at each survey year separately.

Second, the profiles of the primary school graduate premiums and earnings quantiles for the urban areas differ greatly from the pooled and the rural samples (Figure 5.6b). The urban average premiums from the OLS regression for the pooled year sample are around 56 per cent, slightly lower than those of the rural areas. However, the premiums of primary education completion are relatively higher for the upper quantiles of the urban earnings distribution, both on average and especially in the second half of the 1990s. This implies the opposite inequality impacts of primary education completion for the urban areas.

[72] The premiums are around 64 and 52 per cent for the bottom and the upper halves of the rural earnings distribution respectively.

The different primary school premiums-quantile patterns between the rural and urban samples are striking. However, as discussed earlier, these exogenous regression estimations are potentially biased due to the correlation between education and other factors influencing earnings. In addition, this correlation may be significantly different between the rural and urban samples.

Rural-Urban RD Quantile Treatment Effects

The lower panels of Figure 5.6 present the results from the mean and quantile estimates of the RD counterfactual earnings distributions for the rural and urban observations separately.[73] Similar to the format of the pooled sample estimation, these are the results without controlling for any covariates. For the robustness check of an alternative specification with additional covariates, the estimated conditional RD quantile treatment effects are shown in Figure 5.A.3 in the Appendix.

For the rural areas, the patterns of primary school premiums and earnings quantiles are similar to those of the national sample (Figure 5.6c). The effects of primary education completion on hourly earnings are much larger among the bottom half of the rural earnings distribution. These effects are also higher than the mean effects. From the pooled year sample, the earnings of the primary school graduates are on average around 17 per cent higher than those of the dropouts. Considering the quantile treatment effects, primary education completion results in a 38 per cent increase in hourly earnings for the poorest 20 per cent, whereas it has a negative effect for the richest 20 per cent. The impact of primary education completion on hourly earnings is around the average impact for the middle 20 per cent of the earnings distribution. This implies that, ceteris paribus, primary education completion has a negative impact on earnings inequality by increasing hourly earnings of the poor proportionately more than the rich. Primary education completion also impacts on the earnings distribution differently over time. During the years 1991 to 1997, its positive impacts on earning were relatively higher among the bottom half of the earnings distribution. Subsequently, the impacts became more homogeneous across earnings groups for two years after the financial crisis before resuming the pattern again in 2000. The shapes of the quantile treatment effect-earnings profiles are also fairly robust to the inclusion of control variables

[73] These conditional quantile treatment effects are derived from separate RD counterfactual earnings densities for rural and urban areas (as discussed in Sub-section 5.5.3), similarly to the unconditional case.

(Figure 5.A.3a in the Appendix).[74]

The quantile effects of primary education completion on urban hourly earnings are significantly different (Figure 5.6d). For the pooled year sample, the estimated quantile effects are positive for the bottom half of the distribution and negative for the top half of the distribution. However, most of the quantile effects are not statistically different from the mean, and thus rather homogeneous. This implies that, the overall impact of primary education completion on earnings inequality is negative, but likely to be statistically insignificant. Considering the estimates from each survey year separately, the quantile patterns change over time. From 1991 to 1994, primary education completion increased earnings of the bottom half by less than those of the top half. This is likely to have increased earnings inequality. By contrast, from 1995 to 1997, the effects were larger for the bottom half of the earnings distribution. Similarly to the rural and national samples, the impacts after the financial crisis are rather homogeneous across the quantiles. These relationships are also robust to the inclusion of covariates (Figure 5.A.3b in the Appendix).

5.8.2 Rural-Urban Conditional Inequality Treatment Effects

Next, I focus on the changes in the selected earnings inequality measures due to primary education completion for the rural and urban samples separately. These inequality measures are computed from the RD counterfactual earnings distribution, as mentioned earlier, to summarise the overall inequality treatment effects. The predicted effects of primary education completion are presented in Figure 5.7.[75]

First, Figure 5.7a summarises the predicted effects of primary education completion on earnings inequality in the rural areas of Thailand between 1991 and 2000. The results are as expected from the quantile treatment effects in most years. That is, primary education completion reduces earnings inequality, especially during the years before the financial crisis. The results are roughly robust to the inclusion of covariates, except for 1991. The left panel of Figure 5.7a shows that the impacts of the increased primary education completion rate due to the education reform are expected to be significantly negative, with the exception of the year 1993 and the post-crisis years for both the SD of log and the Gini coefficient when the impacts are statistically insignificant. The predicted impacts on the SD of log also fluctuate

[74] The control variables included are a gender dummy and region dummies.

[75] The estimated inequality treatment effects and their bootstrapped 95 per cent confidence bands are shown in Figure 5.A.4 in the Appendix.

more and are of greater magnitudes than those on the Gini coefficient. Primary education completion is expected to reduce the SD of log by 0.02 to 0.08 and the Gini coefficient by 2 to 6 percentage points. The right panel of Figure 5.7a displays the predicted changes in earnings inequality measured by the Theil indices. The increased primary education completion has negative impacts on the two Theil indices; however, the impacts are statistically insignificant in most years especially for the Theil-T index. The Theil-L index is expected to decrease statistically significantly by 0.05 during the pre-crisis years.

Second, the impacts of primary education completion on earnings inequality in the urban areas are in mixed directions, nonetheless statistically insignificant for most years (Figure 5.7b). For the SD of log and the Gini coefficient, primary education completion had positive impacts on earnings inequality in the first half of the 1990s, although the effects are not statistically significant for most years and most specifications with the exception of the period 1992 to 1993. For the late 1990s, the inequality treatment effects were negative, but relatively small and statistically insignificant in some years. The results are similar in terms of directions, patterns, and statistical significance for the two Theil indices. This suggests that the increased primary education completion rate is expected to have increased earnings inequality in the urban labour market during the period 1992 to 1993, but with no statistically significant impacts in other years.

5.8.3 Summary of the Rural-Urban Distributional Treatment Effects

In summary, increased primary education completion as a result of the education reform affects the rural and urban earnings distributions differently. Primary education completion is expected to reduce earnings inequality among rural workers as it raises hourly earnings of the bottom of the distribution proportionately more than the top. By contrast, primary education completion does not reduce earnings inequality in urban areas, and even increase inequality between 1992 and 1993. Although the rural-urban distributional treatment effects are not complete decompositions of the total effects, these upward impacts of primary education completion on earnings inequality in the urban areas potentially offset the downward inequality impacts in the rural areas, and thus explain the statistically insignificant impacts on the overall earnings distribution over the years 1992 and 1993 for some inequality measures.

This implies that while basic and compulsory formal education plays an important role in reducing inequality in the initial stage of economic development (that is, the rural areas), it may not be adequate to deal with inequality in the fast-growing urban areas. This is potentially explained by the difference in the composition of primary school graduates in the rural and urban areas, as well as the rising returns to education, as discussed in the literature review. From the descriptive statistics, it can be seen that most of the rural primary school graduates are those who quit school right after their primary education (Figure 5.2b). On the other hand, the majority of urban primary school graduates are those who studied up to at least the high school level. This implies that the increase in the primary education completion rate as a result of the education reform is likely to lead to a more equal distribution of education among rural workers, compared to that among urban workers. This is because the education reform is expected to directly affect individuals who would have quit school before the primary level had there been no change in the compulsory schooling law. The more equal distribution of education in the rural areas can lead to a decline in earnings inequality.[76] However, among urban workers, compulsory education at the primary level may not sufficiently encourage equality in education, due to a bigger proportion of high school and university graduates. In addition, increasing returns to education at upper levels of schooling owing to a rise in demand for highly-skilled workers can influence the overall impacts of primary education completion on urban earnings inequality to eventually be positive.[77] Therefore, the impacts of education on earnings are heterogeneous across not only the earnings levels but also the rural-urban area of residence.

5.9 Additional Checks

5.9.1 Higher Polynomial Orders

In addition to checking for the robustness to the inclusion of covariates described in the main empirical result sections, it is important to test whether the chosen polynomial order is misspecified as this can lead to biased estimates of the discontinuity as well as erroneous interpretations of statistical significance (Lee and Lemieux, 2010). This section reports results for alternative polynomial orders. As the second polynomial order is justifiable by

[76] Ram (1990); De Gregorio and Lee (2002).
[77] Bound and Johnson (1992); Katz and Murphy (1992); Krueger (1993); Card and DiNardo (2002).

the goodness-of-fit test, the results should be robust to higher polynomial orders which are less restrictive (Lee and Card, 2008).

Figure 5.A.5 in the Appendix displays the quantile treatment effects of primary education completion on hourly earnings using the polynomial RD regressions with the first (linear) to the fifth-order polynomials of the running variable (year of birth) when no covariates are included. The bootstrapped 95 per cent confidence bands are from the fifth-order polynomial specification. First, the empirical results for the second- to the fifth-order polynomial specifications are generally similar to each other. That is, for the pooled and the rural samples primary education completion increases earnings of the bottom half of the distribution proportionately more than those of the top half. By contrast, the impacts of primary education completion on earnings in the urban areas fluctuate over time and are rather homogeneous across earnings levels. Note that the estimates from the first-order polynomial specification are inconsistent with the rest, which suggests that a linear specification is too restrictive and does not fit the data well. Second, while the estimates of quantile treatment effects are very similar for polynomials of orders two to five in terms of the impacts on earnings inequality, the difference among polynomials of orders three to five is even smaller. Lastly, despite not being shown here, the empirical results are also robust to the higher polynomial orders when other control variables are included.

5.9.2 Compliers along the Distributions

One may cast doubts on how the effects of primary education at a very top of the earnings distribution are. While the distributional treatment effects estimated in this chapter apply to the compliers — who change their primary education decision due to the 1978 Education Reform, there may be no compliers at all among the top earners (particularly in urban areas). As a result, the quantile treatment effects for these people may not be identified, and thus the conclusion of overall inequality treatment effects could be misleading.

It is not possible to identify compliers as only one outcome is actually observed. Nonetheless, a group of potential compliers may be established. They are likely to be those who were born before 1967 and were primary school dropouts, and/or some of those who were born at and after 1967 and were primary school graduates. The RD approach, by construction, assumes that individuals in the discontinuity neighbourhood are similar to each other and only differ in their running variable (in a discontinuous way). As a result, an approximate proportion of the potential compliers could be obtained by plotting the proportion of primary school graduates before and after the cut-off year of birth. The difference in the primary completion rates between those who were born before and at the cut-off year is expected to be informative about the importance of the compliers. Note that this is similar to testing the significance of the discontinuity along the earnings distribution.

To obtain a rough picture of how substantial the compliers are, Figure 5.A.6 in the Appendix shows the primary completion rates by birth cohort and earnings decile, for the pooled and rural-urban samples. For the pooled sample (Figure 5.A.6a), the discontinuity exists but declines substantially as moving up the earnings distribution (in particular for the 9th and 10th deciles). Meanwhile, Figure 5.A.6b indicates large heterogeneity in a discontinuity, and thus the significance of compliers across urban and rural areas. When separating the samples into urban and rural residents, the discontinuity remains substantial across decile groups in rural areas. However, this discontinuity becomes very small and is likely to be insignificant for those urban dwellers in the 9th and 10th earnings deciles.

These results suggest that the quantile treatment effects for the top earners in urban areas may not be identified because the education reform had a very little impact on these people. In other words, this suggests that returns to primary education induced by the reform are insignificant for the rich. If this is the case, the positive returns to primary education among the poor should still imply the equalising effects of the primary education among the compliers.

5.10 Conclusion

In this chapter, I have shown, using Thai labour force data, that increased primary education completion played an important role in shaping earnings inequality in Thailand. During the period of high economic growth in the 1990s, observed earnings inequality in the Thai labour market reduced gradually while average educational attainment increased with a remarkable shift out of lower-than-primary completion especially among low-paid workers and in rural areas. The quantile regression approach suggests potential heterogeneity in returns to primary education completion across earnings levels. The observed earnings premiums between primary school graduates and dropouts at different points of the earnings distribution exhibit a U-shape. That is, they are highest for the bottom 20 per cent of the earnings distribution, lowest for the middle 20 per cent, and around the average level for the top 20 per cent.

The crucial question that I am able to answer using the RD approach is whether this heterogeneous relationship between primary education completion and earnings across earnings levels is due to unobservable heterogeneity between workers. That is, the question is whether primary education completion increases the earnings of the lower quantiles relatively more than those of the upper quantiles, and thus causes a reduction in earnings inequality. To identify the impacts of primary education completion at different points of the earnings distribution, I make use of a natural experiment, where depending on birth cohorts, individuals' school attendance was governed by different compulsory schooling laws. Once I control for individual unobservable characteristics, the earnings premiums between primary school graduates and dropouts drop substantially while the heterogeneous premiums across earnings levels remain, and in fact show a clearer distributional impact of primary education completion on earnings. The impact of primary education completion on earnings declines as one moves up the earnings distribution. In other words, primary education completion is found to reduce earnings inequality. In addition, I find that overall national impacts are driven mainly by rural workers, and that in contrast, primary education completion actually increases inequality among urban workers. This could be because there are increasing returns to medium and higher-level education in urban areas, and the increased primary education completion rate in response to a change in the compulsory education affects education and earnings inequalities differently between rural and urban areas. I also find that primary education completion had no significant impact on earnings inequality during the post-financial crisis years. Although I am unable to explain whether this is caused by the financial crisis using the LFS data and the proposed RD framework, I have discussed the relevant literature that examines this issue directly. The descriptive statistics also showed that the decline in the premiums among the poor could be explained by the over-supply of workers with only primary education.

I should emphasise that the distributional treatment effects estimated in the RD framework are identified only for the compliers, who are those who would not have completed primary education had there been no change in the compulsory schooling law. However, as the number of compliers is relatively large and increases over time due to the law enforcement, the RD estimates are expected to be informative about the population.

To conclude, my results suggest that basic education like primary schooling is an important policy tool for dealing with earnings inequality in developing countries, as it benefits the poor relatively more than the rich. Nonetheless, basic education may not be sufficient to foster earnings equality in a fast-growing economy where returns to education are increasing as the level of education rises. The results also point towards a need for further empirical studies on education and earnings inequality which may differ greatly across various welfare groups and over time within the same economy. For instance, to make progress in this regard, it is crucial to understand how education affects inequality between rural and urban groups, and to explain the changes in the distributional impact over time.

Table 5.1: LFS Descriptive Statistics

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Pooled LFS Sample										
Avg. age	30.1	30.3	30.2	30.3	30.6	30.5	30.7	30.8	30.8	30.8
1{Male}	65%	64%	64%	62%	62%	63%	62%	61%	61%	60%
1{Urban}	37%	37%	37%	38%	39%	39%	41%	40%	39%	40%
Avg. years of educ.	6.7	6.8	6.9	7.1	7.1	7.2	7.4	7.8	8.0	8.1
Avg. hours of work per week	51.7	52.0	51.4	51.6	51.8	51.5	48.8	50.4	49.9	50.4
Rural LFS Sample										
Avg. age	30.14	30.47	30.20	30.29	30.73	30.65	30.96	30.90	30.88	30.89
1{Male}	68%	68%	67%	66%	66%	66%	66%	65%	64%	63%
Avg. years of educ.	5.6	5.7	5.8	6.0	6.2	6.2	6.4	6.8	7.0	7.1
Avg. hours of work per week	52.8	53.0	52.0	52.2	52.7	51.7	50.9	51.0	50.1	50.7
Urban LFS Sample										
Avg. Age	29.93	29.95	30.13	30.21	30.26	30.26	30.33	30.55	30.65	30.65
1{Male}	58%	58%	58%	56%	57%	58%	57%	55%	56%	55%
Avg. years of educ.	8.6	8.6	8.8	8.8	8.6	8.6	8.9	9.5	9.7	9.7
Avg. hours of work per week	49.7	50.4	50.3	50.5	50.4	51.0	45.9	49.5	49.6	50.0

Source: Author's calculation from the Labour Force Surveys 1991-2000 (NSO).
Notes: * This is a sample of the paid employed labour force aged from 20 to 40 years old.

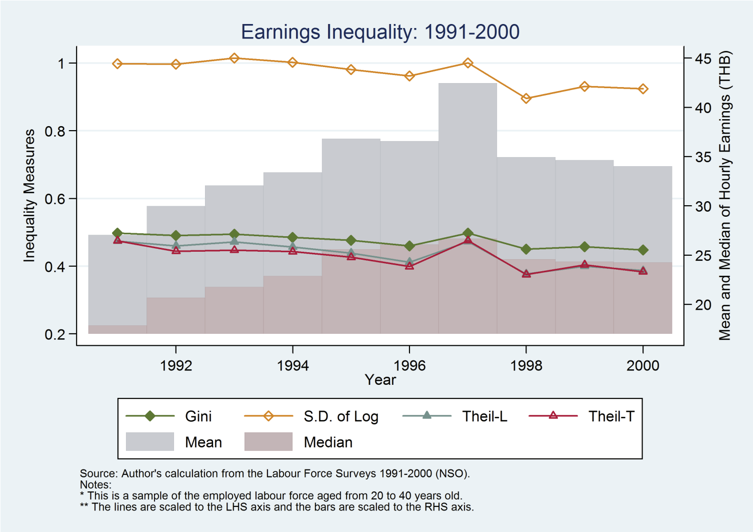
Table 5.2: P-Values for the Tests of Polynomial Specifications

For the RD estimation of $F_{Y^0\|C}(y)$: Dependent variable $= \mathbb{1}\{Y_i \leq y\} \cdot (1 - D_i)$				
Polynomial order	$y = P(20)$	$y = P(40)$	$y = P(60)$	$y = P(80)$
Linear	0.8945	0.1424	0.0009	0.0000
Quadratic	0.9979	0.9990	0.4339	0.1904
Cubic	0.9992	0.9999	0.9880	0.8545
Quartic	1.0000	1.0000	0.9993	0.9777
Quintic	1.0000	0.9960	0.9731	1.0000
For the RD estimation of $F_{Y^1\|C}(y)$: Dependent variable $= \mathbb{1}\{Y_i \leq y\} \cdot D_i$				
Polynomial order	$y = P(20)$	$y = P(40)$	$y = P(60)$	$y = P(80)$
Linear	0.109	0.007	0.000	0.000
Quadratic	0.320	0.100	0.096	0.262
Cubic	0.797	0.466	0.697	0.434
Quartic	0.884	0.869	0.975	0.989
Quintic	0.991	0.995	0.980	1.000

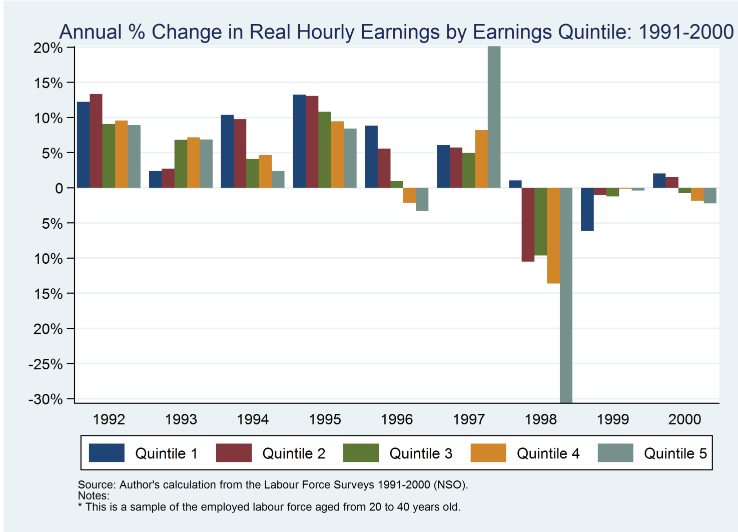
Notes:
* P-value concerns the following hypothesis test, H_0: The coefficients on the year of birth dummies are jointly insignificant, when regressing the dependent variable on the year of birth dummies and the higher order terms in the distance from the cut-off year of birth 1967 for the pooled year (1991-2000) sample. Therefore, failing to reject the null (p-value greater than the significance level, α) implies that the chosen polynomial order explains the data well.

Figure 5.1: LFS Information on Earnings and Education

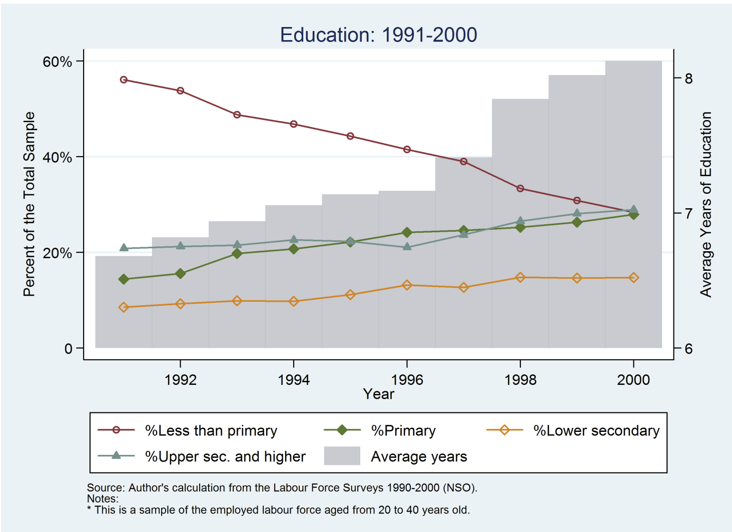
(a) Earnings and Earnings Inequality



(b) Earnings Growth by Earnings Quintile



(c) Educational Attainment



(d) Educational Attainment by Quintile

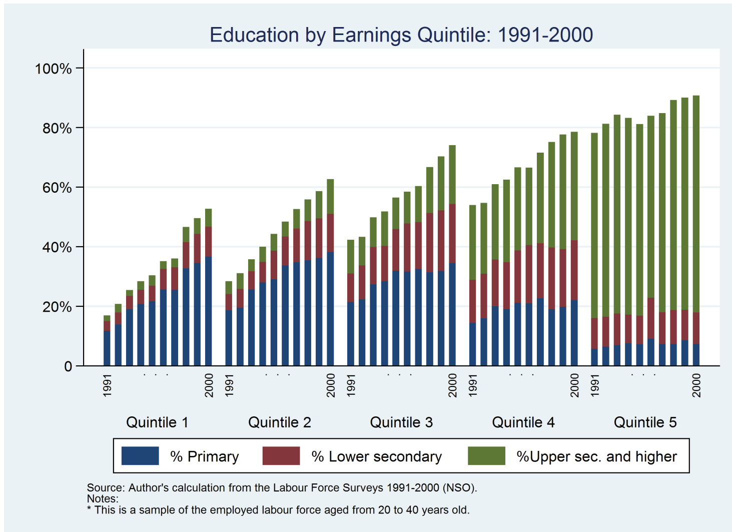
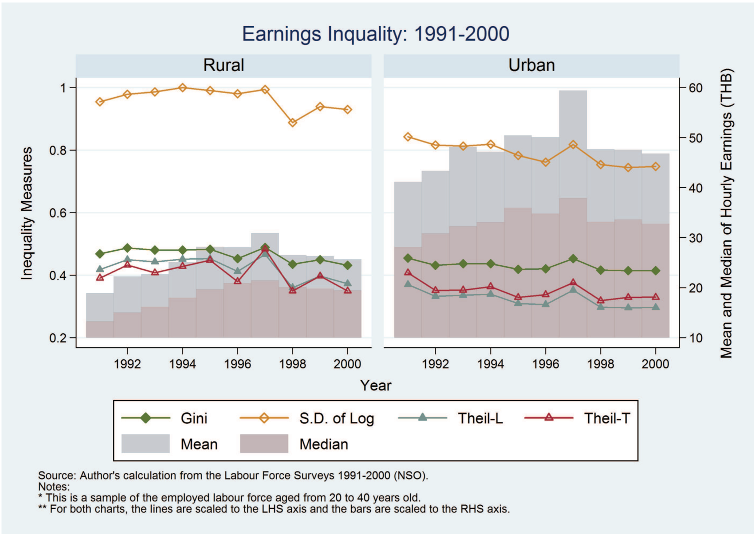
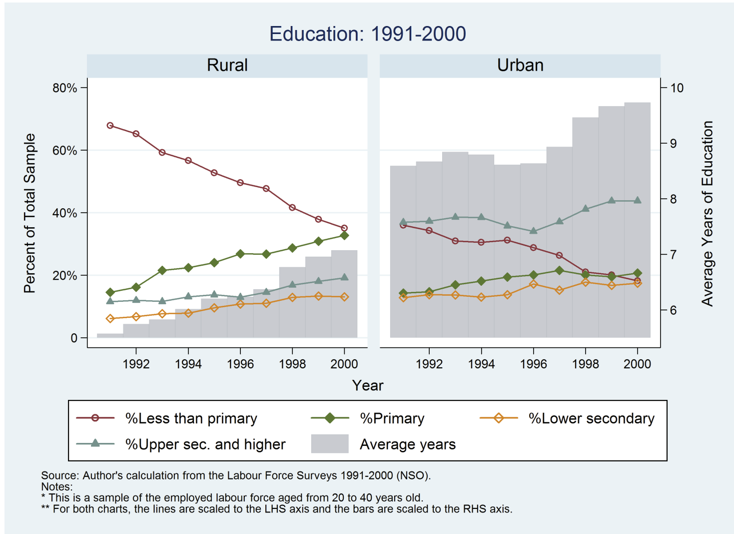


Figure 5.2: LFS Information on Earnings and Education: Rural and Urban

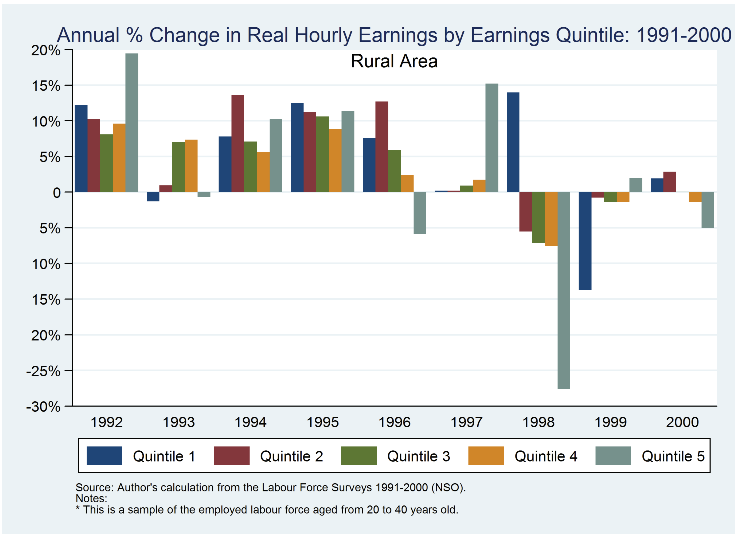
(a) Earnings Inequality: Rural and Urban



(b) Educational Attainment: Rural and Urban



(c) Earnings Growth by Earnings Quintile: Rural and Urban



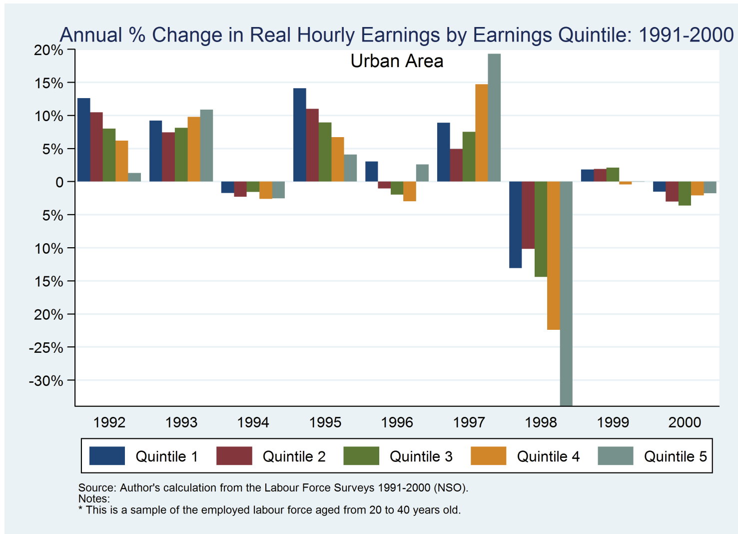
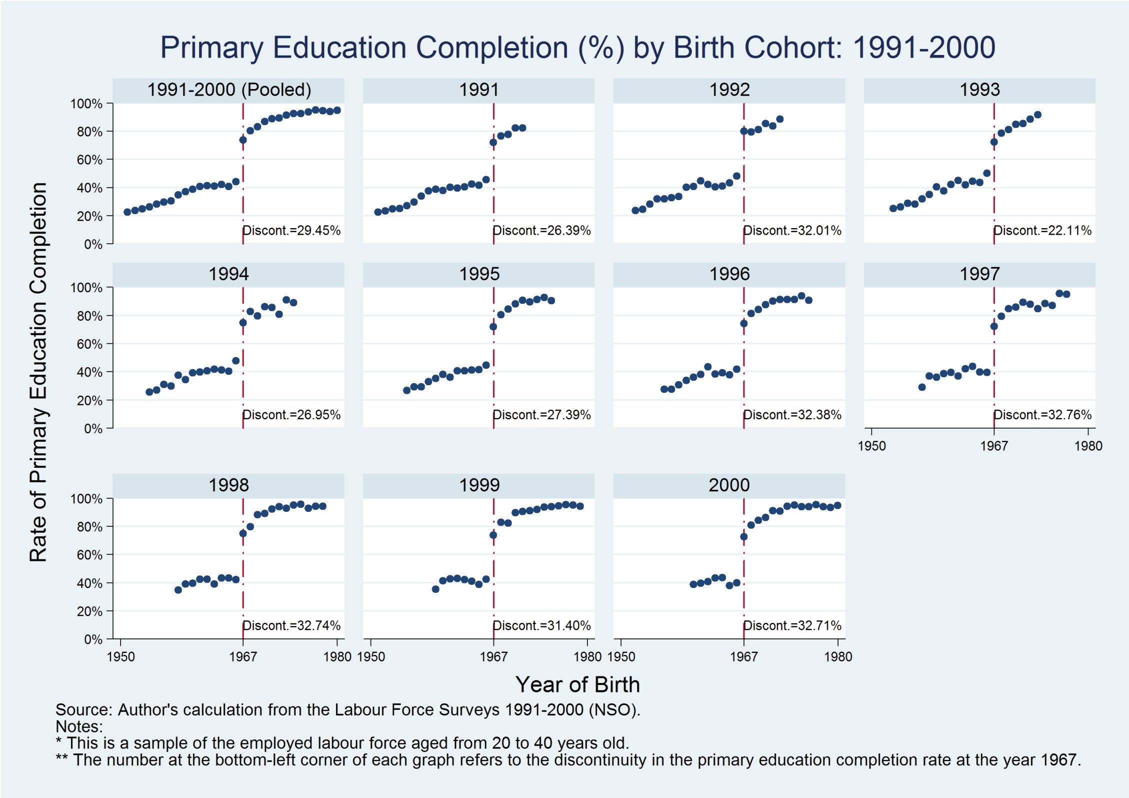


Figure 5.3: Primary Education Completion Rates by Birth Cohort: 1990-2000

(a) Pooled Sample



(b) Rural and Urban

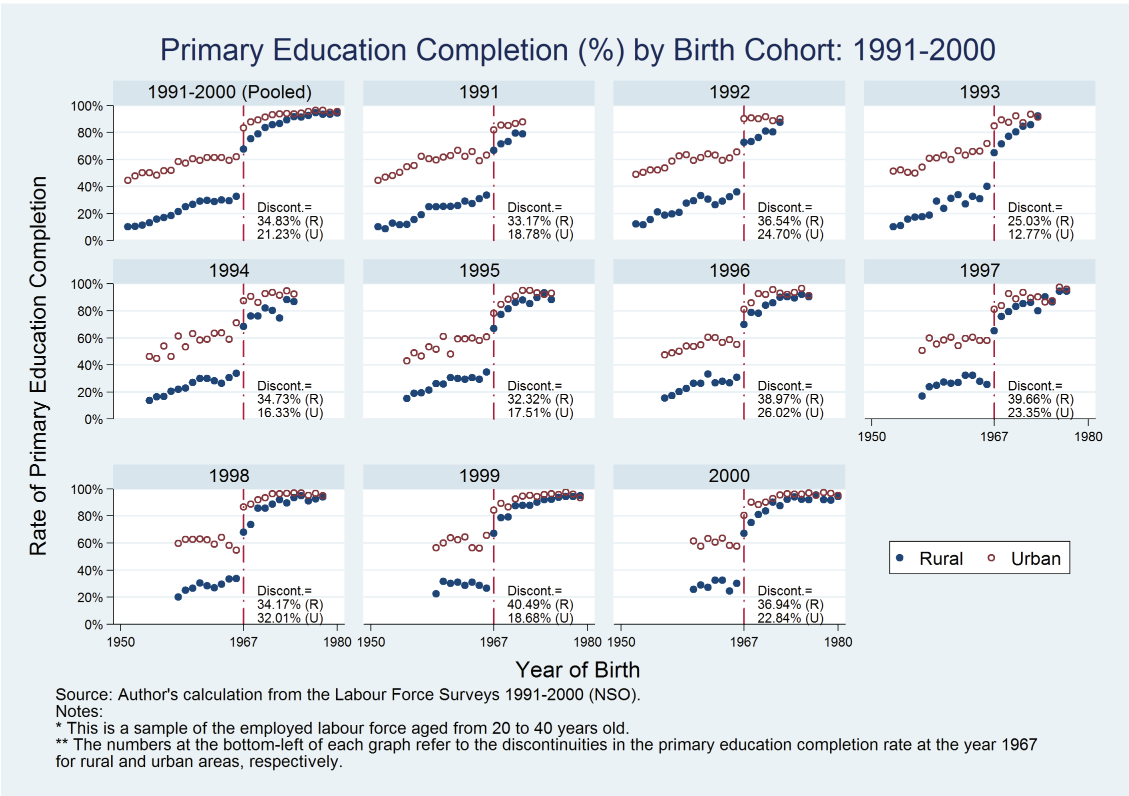
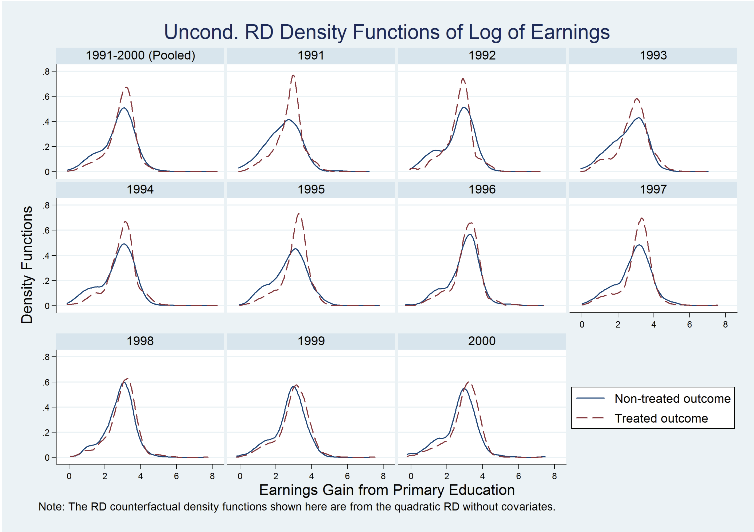
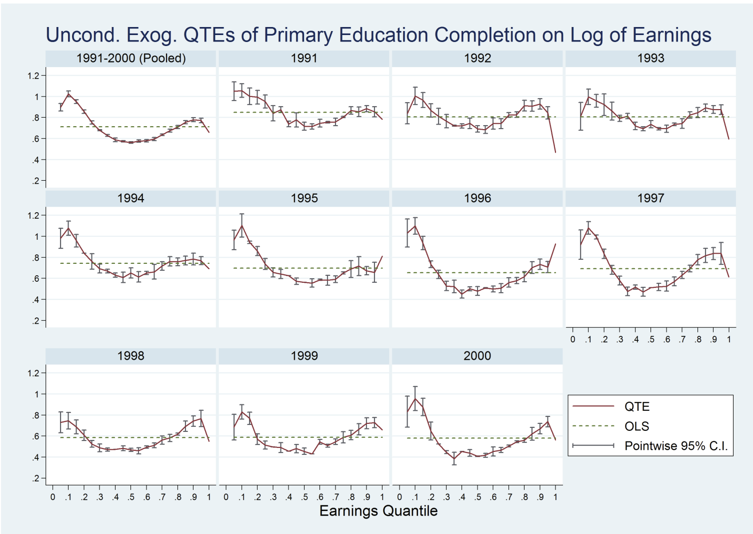


Figure 5.4: Unconditional RD Density Functions and Unconditional Quantile Treatment Effects

(a) Unconditional RD Density Functions of Log Hourly Earnings



(b) Unconditional Exogenous Quantile Regressions



(c) Unconditional Endogenous (RD) QTEs

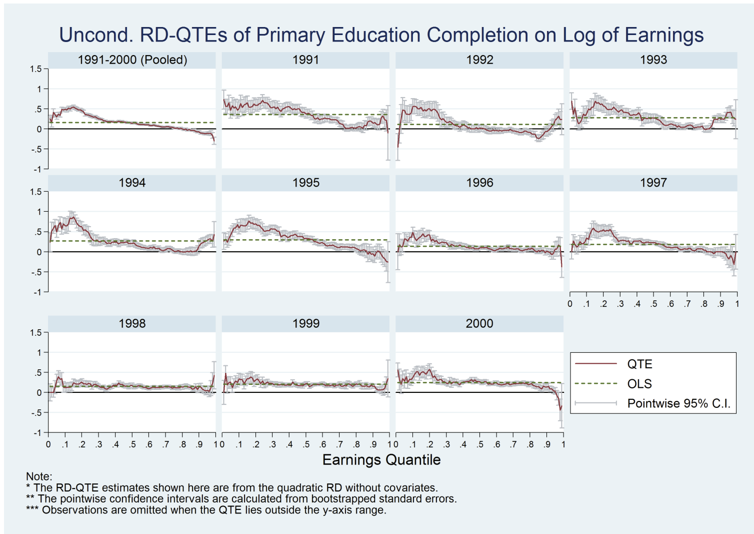
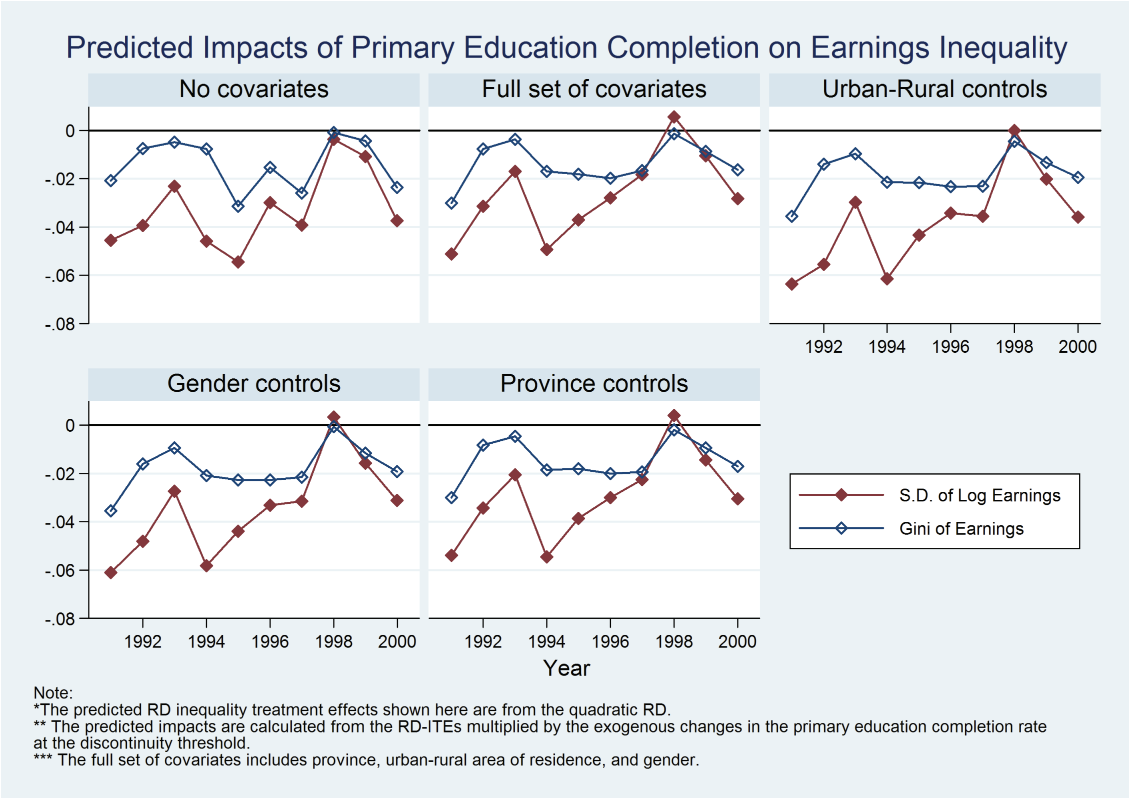


Figure 5.5: Unconditional Endogenous (RD) Predicted Inequality Treatment Effects

(a) Predicted RD ITEs: SD of Log and Gini



(b) Predicted RD ITEs: Theil-L and Theil-T

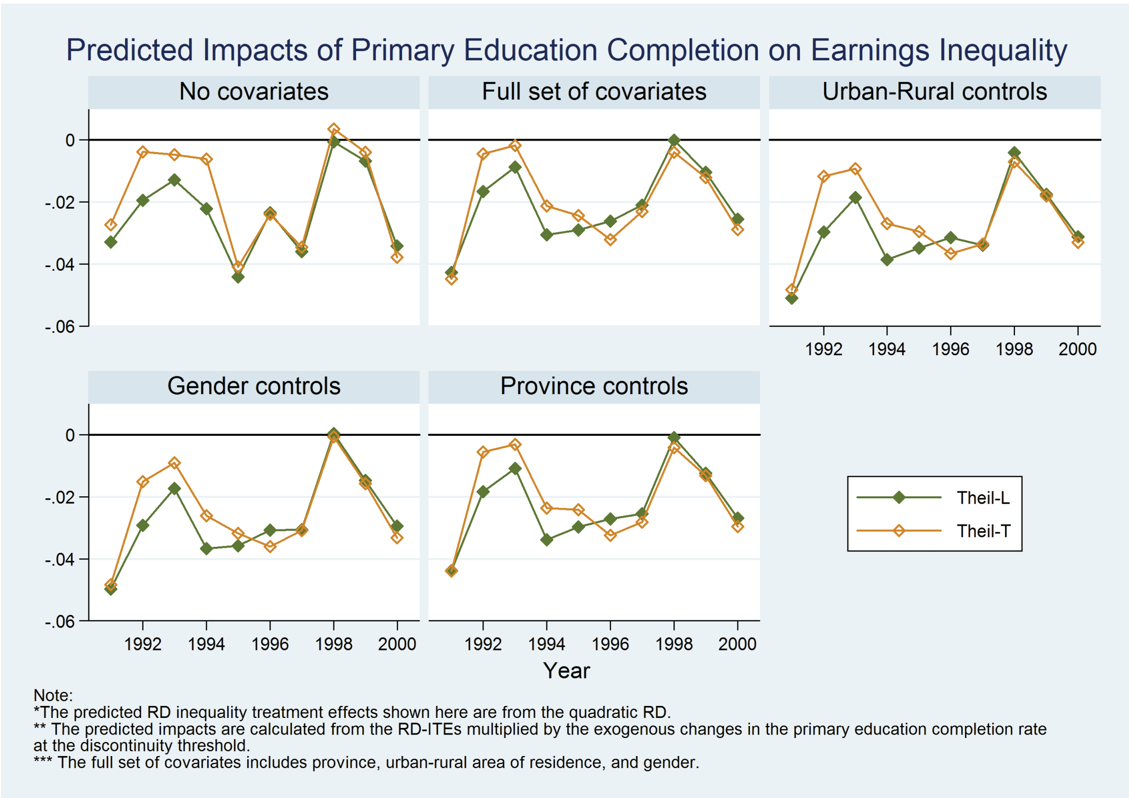
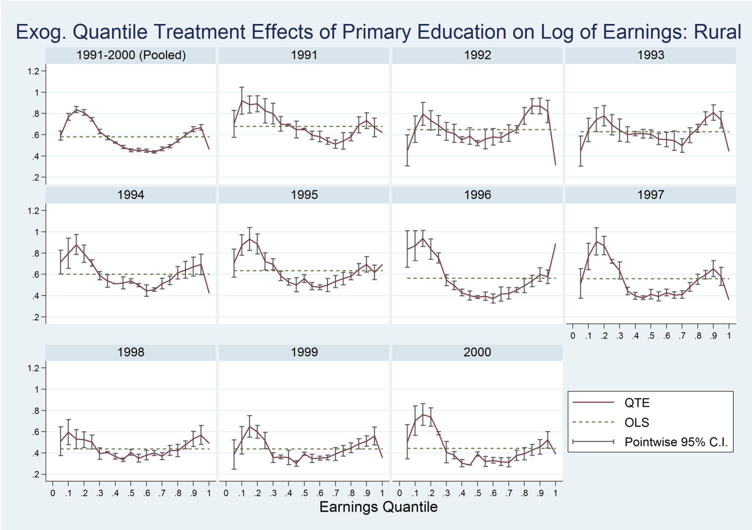
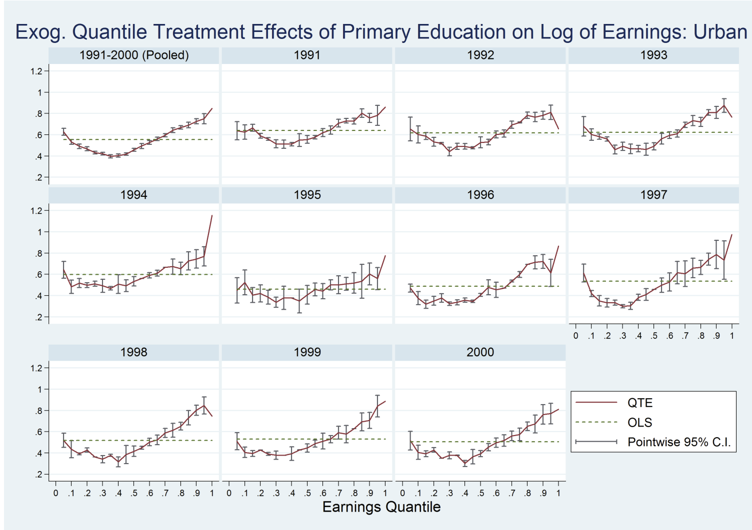


Figure 5.6: Conditional Quantile Treatment Effects: Rural and Urban

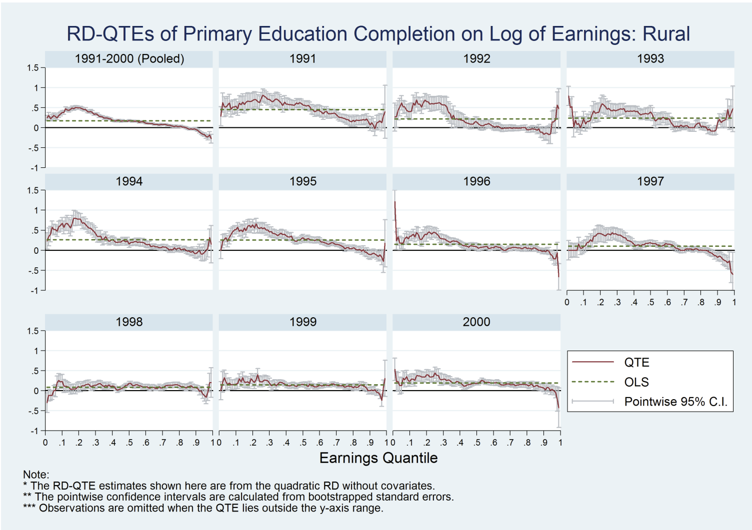
(a) Conditional Exogenous Quantile Regressions: Rural



(b) Conditional Exogenous Quantile Regressions: Urban



(c) Conditional Endogenous (RD) QTEs: Rural



(d) Conditional Endogenous (RD) QTEs: Urban

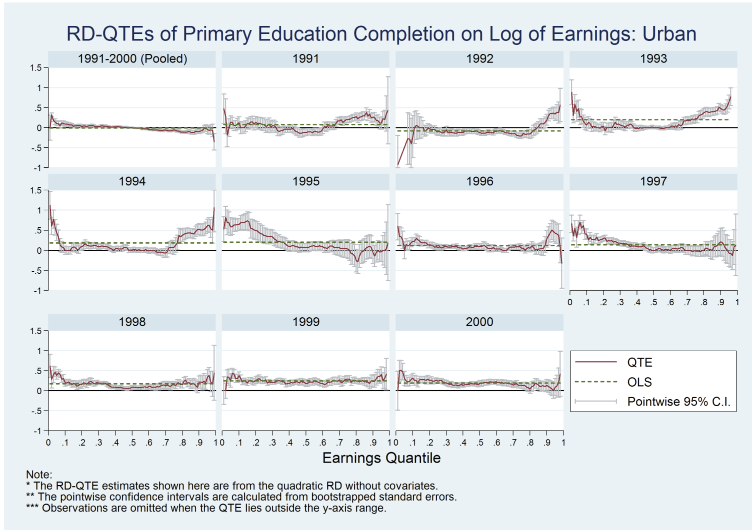
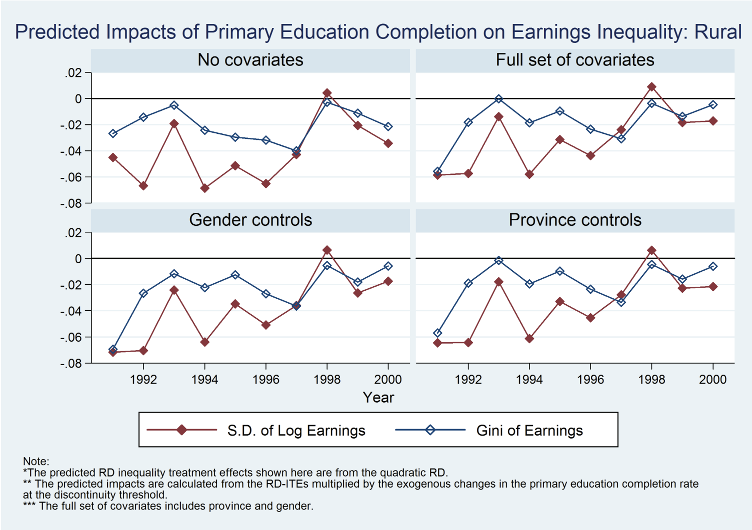
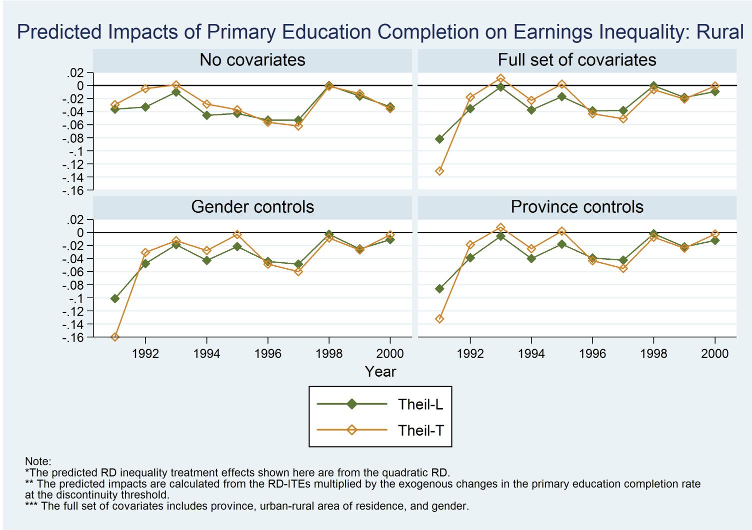


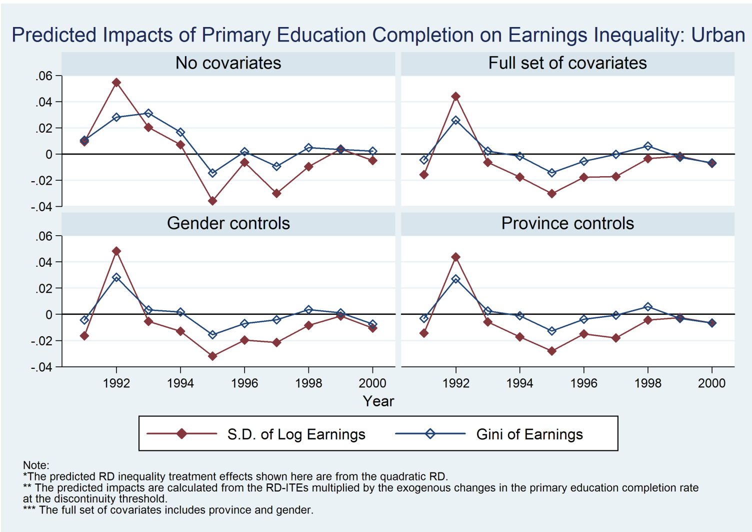
Figure 5.7: Conditional Endogenous (RD) Predicted Inequality Treatment Effects: Rural and Urban

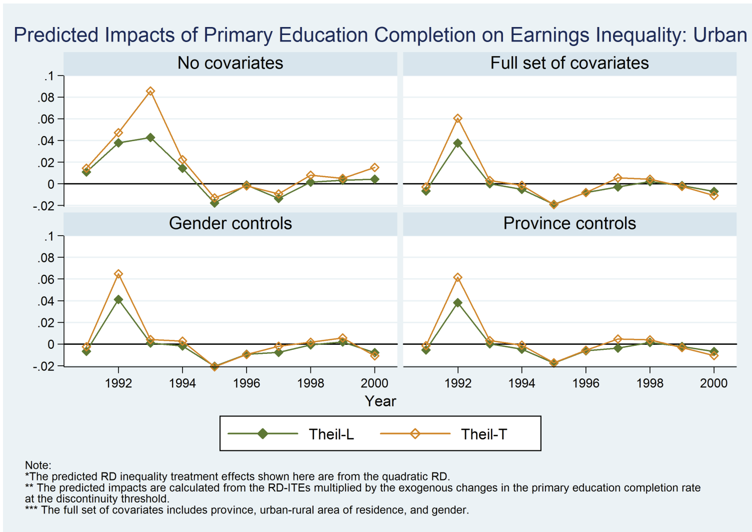
(a) Predicted RD ITEs: Rural





(b) Predicted RD ITEs: Urban





5.A Chapter Appendix

5.A.1 The Proof for the Identification of the RD Counterfactual Distributions, Taken from Frandsen et al. (2012)

Equations (5.5) and (5.6)

From Equation (5.5),

$$F_{Y^1|C}(y) = \frac{\lim_{r\to r_0^+} E\left[\mathbb{1}\{Y\le y\}\cdot D \mid R=r\right] - \lim_{r\to r_0^-} E\left[\mathbb{1}\{Y\le y\}\cdot D \mid R=r\right]}{\lim_{r\to r_0^+} E\left[D \mid R=r\right] - \lim_{r\to r_0^-} E\left[D \mid R=r\right]}$$

First, the probability of a local complier is identified as the denominator of $F_{Y^1|C}(y)$ as follows.

$$\begin{aligned}
\lim_{r\to r_0^+} E\left[D \mid R=r\right] - \lim_{r\to r_0^-} E\left[D \mid R=r\right] &= E\left[\lim_{r\to r_0^+} D(r) \mid R=r\right] - E\left[\lim_{r\to r_0^-} D(r) \mid R=r\right] \\
&= E\left[D^1 \mid R=r_0\right] - E\left[D^0 \mid R=r_0\right] \\
&= E\left[D^1 - D^0 \mid R=r_0\right] \\
&= \Pr\left[D^1 \ge D^0 \mid R=r_0\right] \\
&= \Pr\left[\text{Complier} \mid R=r_0\right]
\end{aligned}$$

The first equality follows from the local smoothness assumption (**I2**) and the monotonicity assumption (**I3**). The fourth equality follows from the monotonicity assumption (**I3**), $\Pr\left[\text{Indefinite} \mid R=r_0\right]=0$, and the fact that D^1-D^0 equals to one when $D^1>D^0$ and equals to zero when $D^1=D^0$.

Second, with similar derivations for the numerator of $F_{Y^1|C}(y)$,

$$\begin{aligned}
&\lim_{r\to r_0^+} E\left[\mathbb{1}\{Y\le y\}\cdot D \mid R=r\right] - \lim_{r\to r_0^-} E\left[\mathbb{1}\{Y\le y\}\cdot D \mid R=r\right] \\
&= \lim_{r\to r_0^+} E\left[\mathbb{1}\{Y^1\le y\}\cdot D \mid R=r\right] - \lim_{r\to r_0^-} E\left[\mathbb{1}\{Y^1\le y\}\cdot D \mid R=r\right] \\
&= E\left[\mathbb{1}\{Y^1\le y\}\cdot \lim_{r\to r_0^+} D(r) \mid R=r_0\right] - E\left[\mathbb{1}\{Y^1\le y\}\cdot \lim_{r\to r_0^-} D(r) \mid R=r_0\right] \\
&= E\left[\mathbb{1}\{Y^1\le y\}\cdot D^1 \mid R=r_0\right] - E\left[\mathbb{1}\{Y^1\le y\}\cdot D^0 \mid R=r_0\right] \\
&= E\left[\mathbb{1}\{Y^1\le y\}\cdot \left(D^1-D^0\right) \mid R=r_0\right] \\
&= E\left[\mathbb{1}\{Y^1\le y\} \mid D^1>D^0,\ R=r_0\right]\cdot \Pr\left[D^1\ge D^0 \mid R=r_0\right] \\
&= E\left[\mathbb{1}\{Y^1\le y\} \mid D^1>D^0,\ R=r_0\right]\cdot \Pr\left[\text{Complier} \mid R=r_0\right] \\
&= F_{Y^1|C}(y)\cdot \Pr\left[\text{Complier} \mid R=r_0\right]
\end{aligned}$$

The second equality follows from the definition of the potential treatment status, the local smoothness assumption (**I2**) and the monotonicity assumption (**I3**).

The identification of $F_{Y^0|C}(y)$ (Equation (5.6)) is analogous.

5.A.2 The Proof for the RD Counterfactual Distributions with Covariates, Taken from Frölich and Melly (2010b)

Equations (5.11) and (5.12)

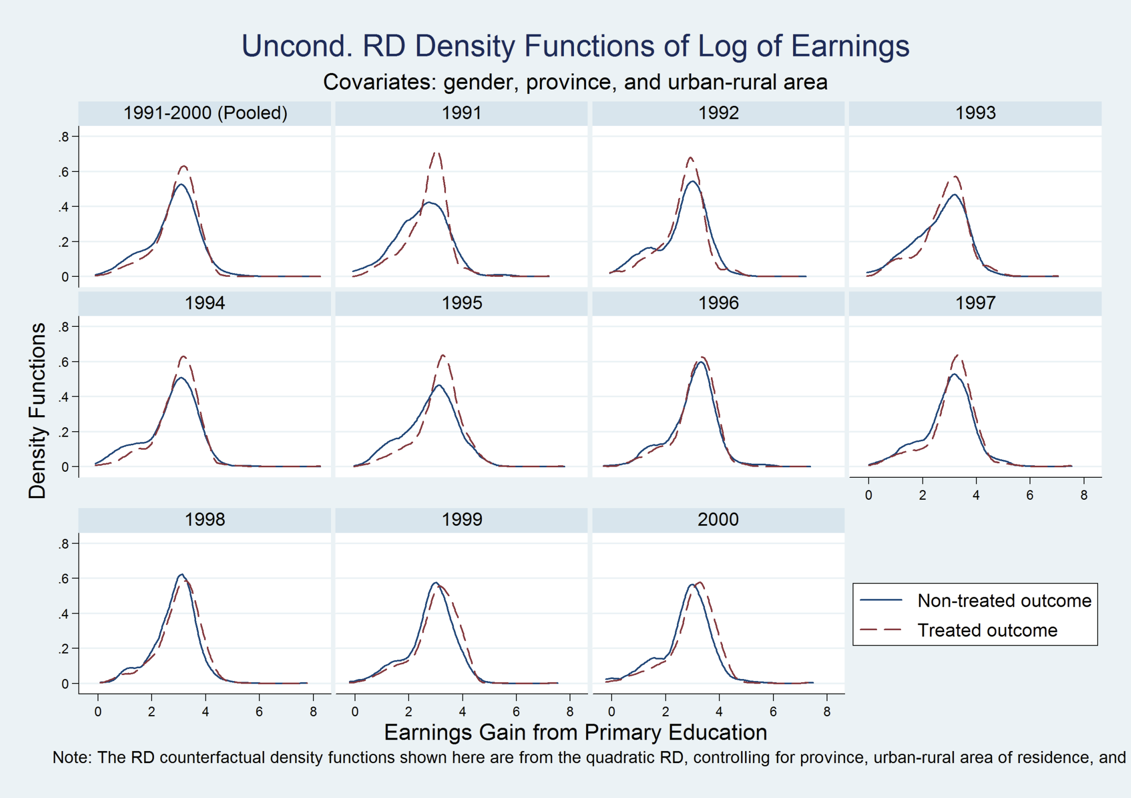
Let $N_\varepsilon = (r_0 - \varepsilon, r_0 + \varepsilon)$ and $p_\varepsilon(x) = \Pr(R \geq r_0 \mid X = x,\, R \in N_\varepsilon)$ for $\varepsilon > 0$. Then, for any real measurable and absolutely integrable function $g(y)$,

$$
\begin{aligned}
&\int \left(\lim_{r \to r_0^+} E\left[g(y) \cdot D \mid X,\, R = r\right] - \lim_{r \to r_0^-} E\left[g(y) \cdot D \mid X,\, R = r\right] \right) dF_{X \mid R \in N_\varepsilon} \\
&= \int E\left[\frac{g(y) \cdot D \cdot \mathbb{1}\{R \geq r_0\}}{p_\varepsilon(x)} - \frac{g(y) \cdot D \cdot (1 - \mathbb{1}\{R \geq r_0\})}{1 - p_\varepsilon(x)} \mid X,\, R \in N_\varepsilon \right] dF_{X \mid R \in N_\varepsilon} \\
&= E\left[g(y) \cdot D \cdot \frac{\mathbb{1}\{R \geq r_0\} - p_\varepsilon(x)}{p_\varepsilon(x) \cdot (1 - p_\varepsilon(x))} \mid R \in N_\varepsilon \right] \\
&= \Pr(D = 1 \mid R \in N_\varepsilon) \cdot E\left[g(y) \cdot \frac{\mathbb{1}\{R \geq r_0\} - p_\varepsilon(x)}{p_\varepsilon(x) \cdot (1 - p_\varepsilon(x))} (2D - 1) \mid R \in N_\varepsilon,\, D = 1 \right]
\end{aligned}
$$

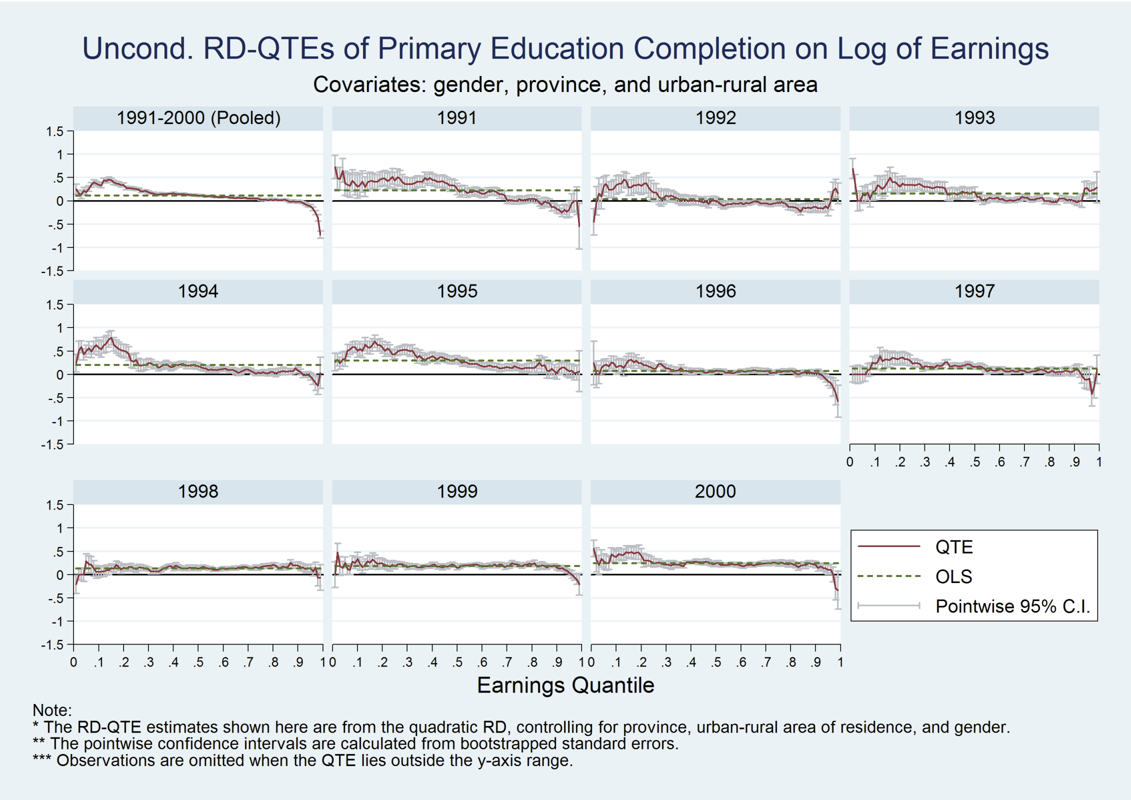
Equations (5.11) and (5.12) can be obtained by applying this result to the numerators and denominators of Equations (5.5) and (5.6) respectively.

Figure 5.A.1: Unconditional RD Density Functions and Unconditional Quantile Treatment Effects – With Controls

(a) Unconditional RD Density Functions of Log Hourly Earnings – With Controls



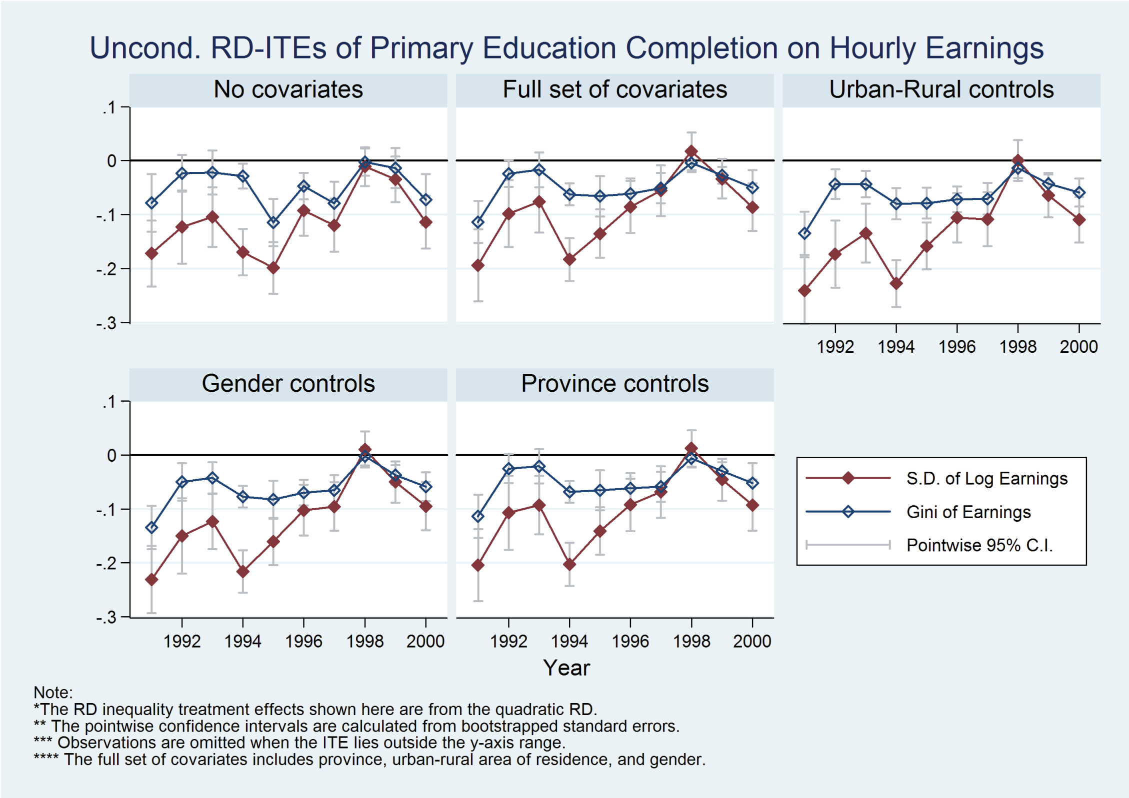
(b) Unconditional Endogenous (RD) Quantile Treatment Effects – With Controls



Notes: * Figure 5.A.1a (Figure 5.A.1b) is a robustness check for Figure 5.4a (Figure 5.4c), when the full set of controls is included. ** The controls include gender, province, and rural-urban area.

Figure 5.A.2: Unconditional Endogenous (RD) Inequality Treatment Effects

(a) RD ITEs: SD of Log and Gini Coefficient



(b) RD ITEs: Theil-L and Theil-T

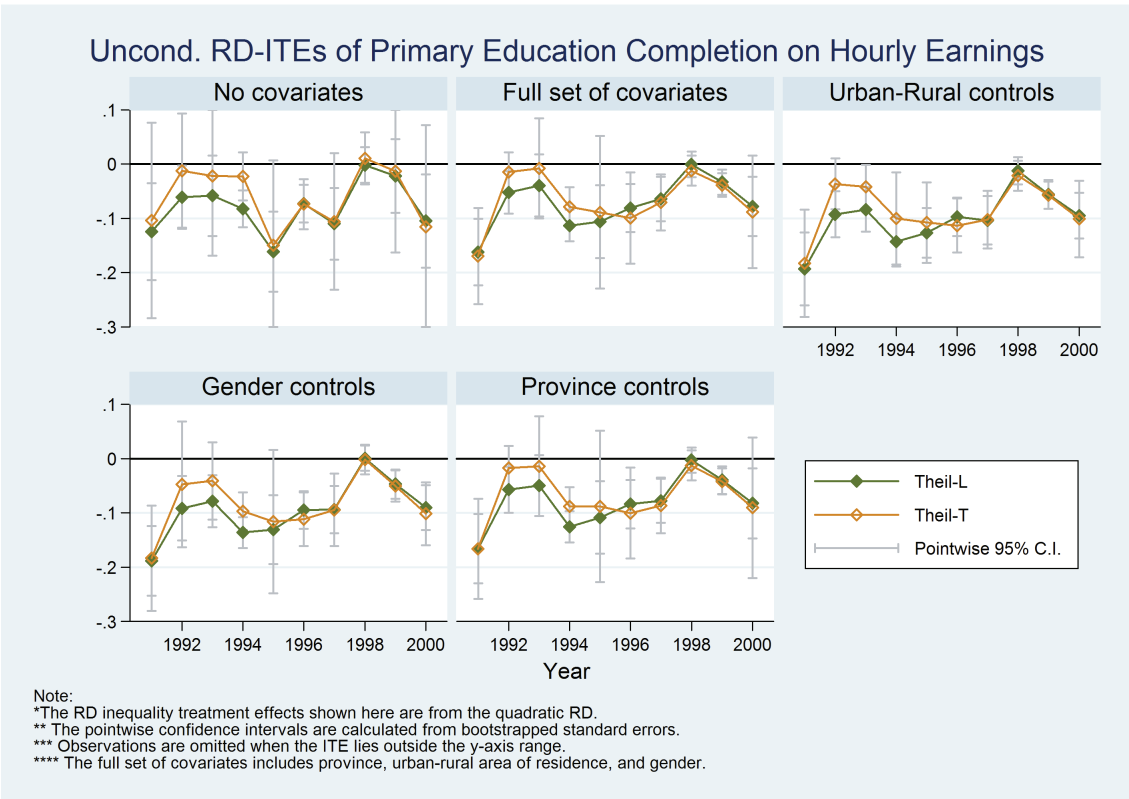
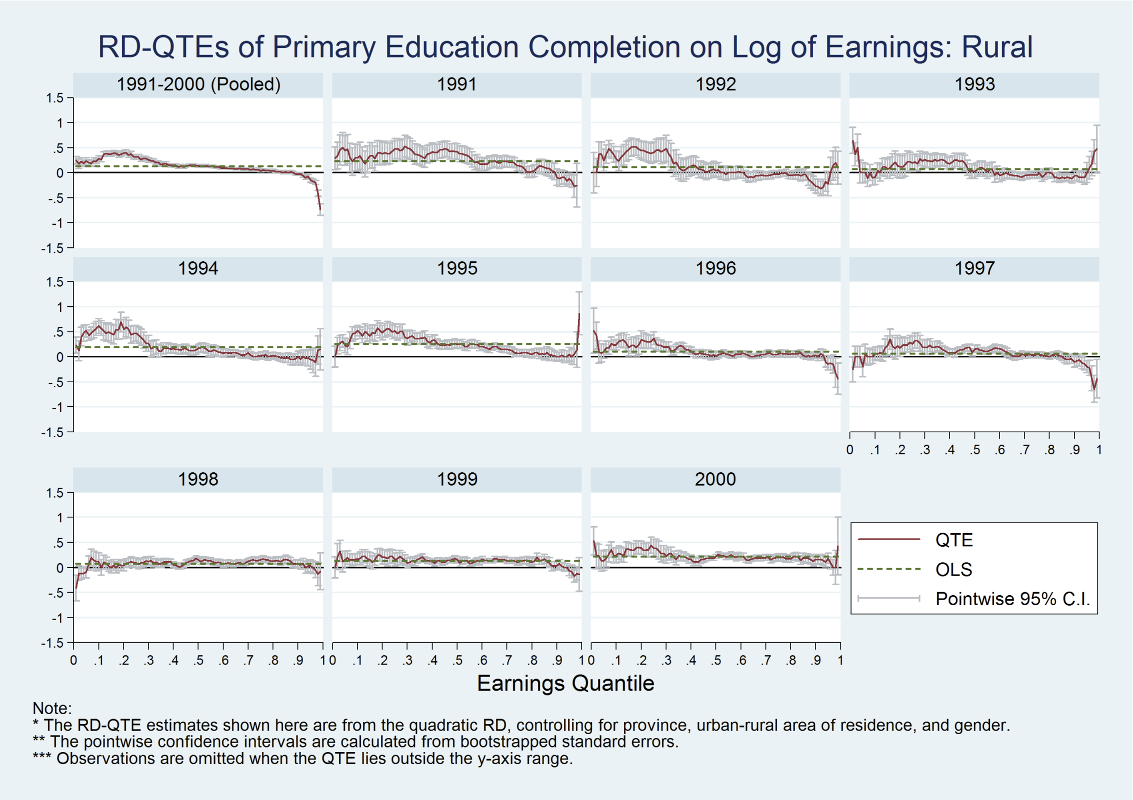
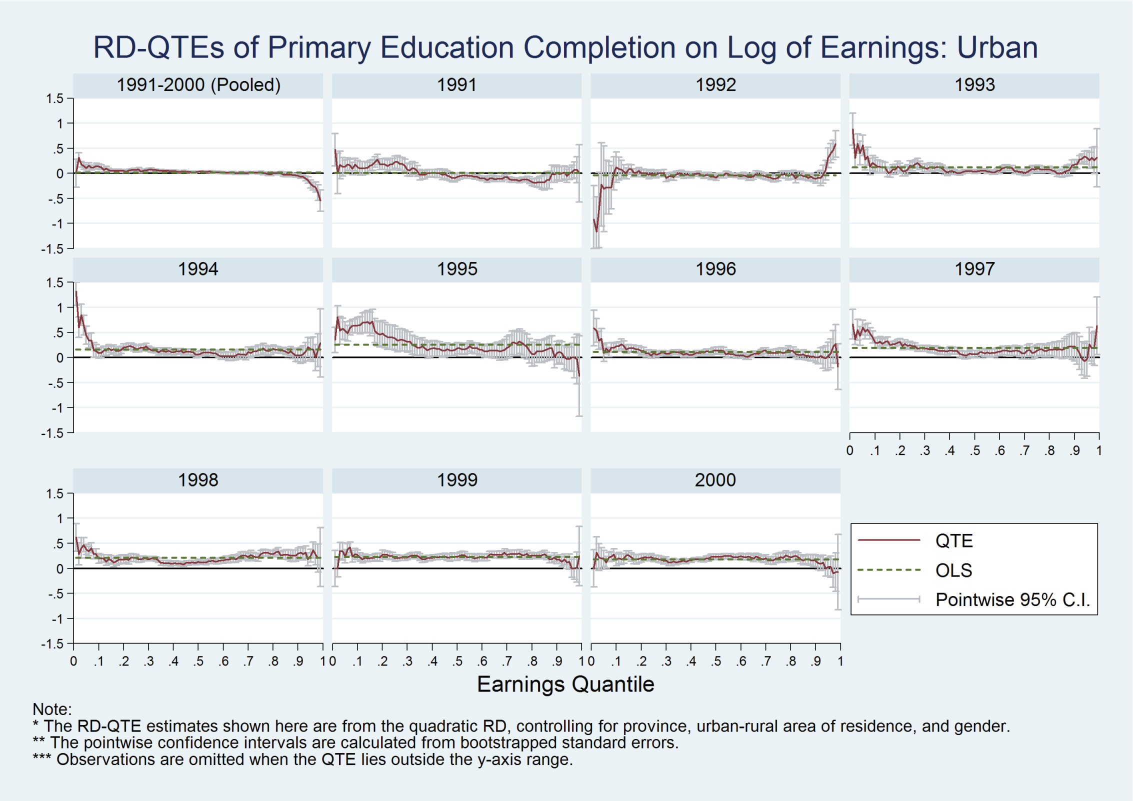


Figure 5.A.3: Conditional Endogenous (RD) Quantile Treatment Effects – With Controls: Rural and Urban

(a) Conditional Endogenous (RD) Quantile Treatment Effects – With Controls: Rural



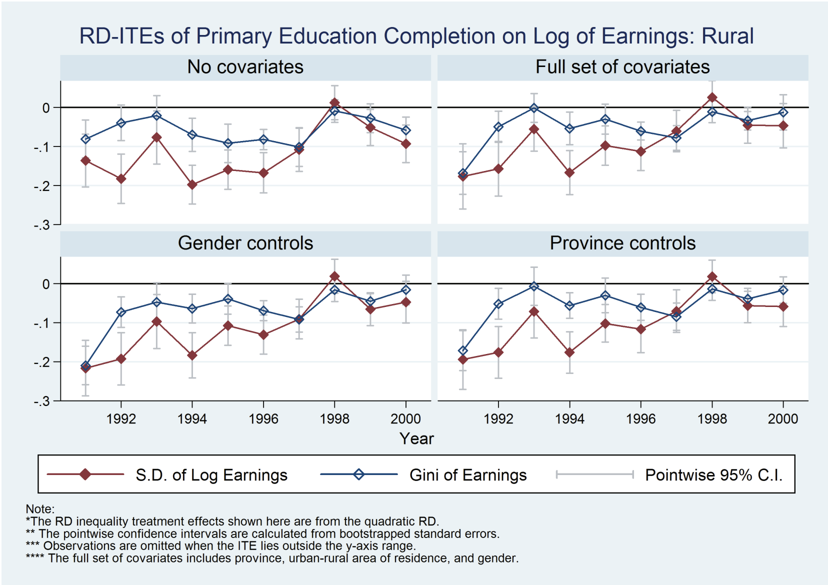
(b) Conditional Endogenous (RD) Quantile Treatment Effects – With Controls: Urban

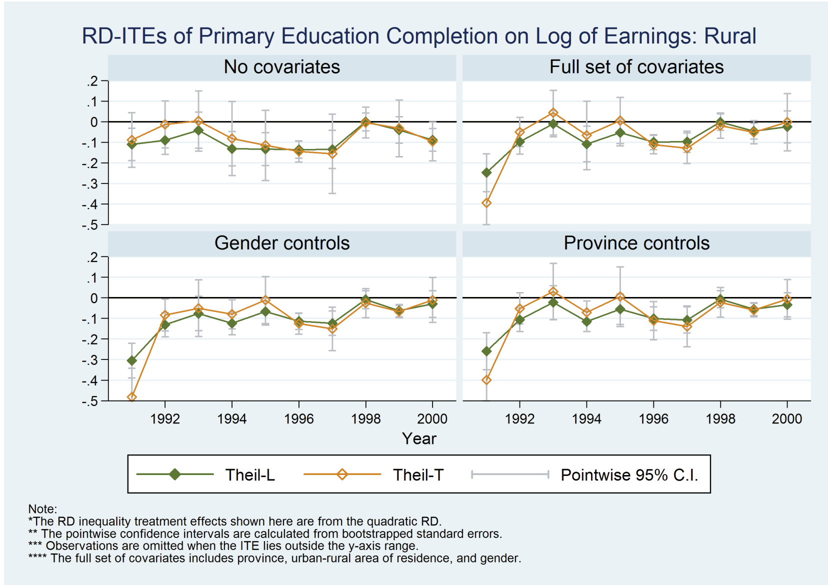


Notes: * Figure 5.A.3a (Figure 5.A.3b) is a robustness check for Figure 5.6c (Figure 5.6d), when the full set of controls is included. ** The controls include gender and province.

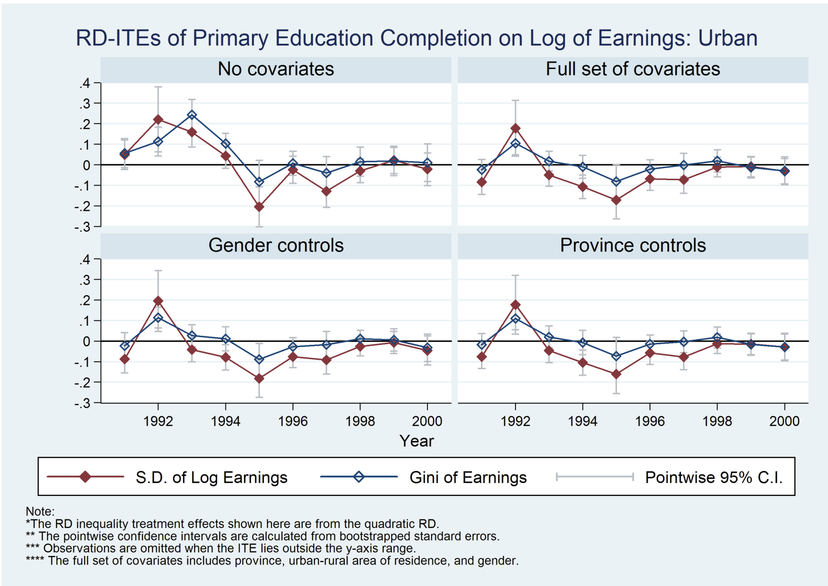
Figure 5.A.4: Conditional Endogenous (RD) Inequality Treatment Effects: Rural and Urban

(a) RD ITEs: Rural





(b) RD ITEs: Urban



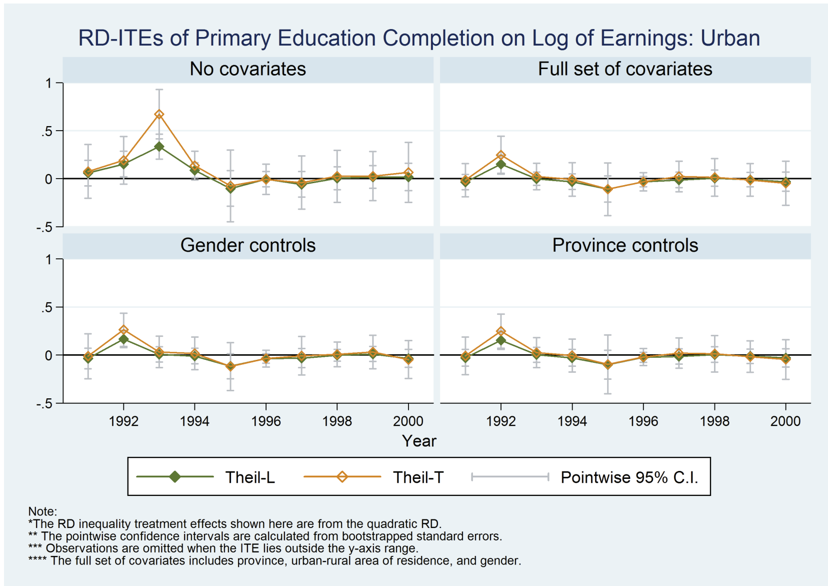
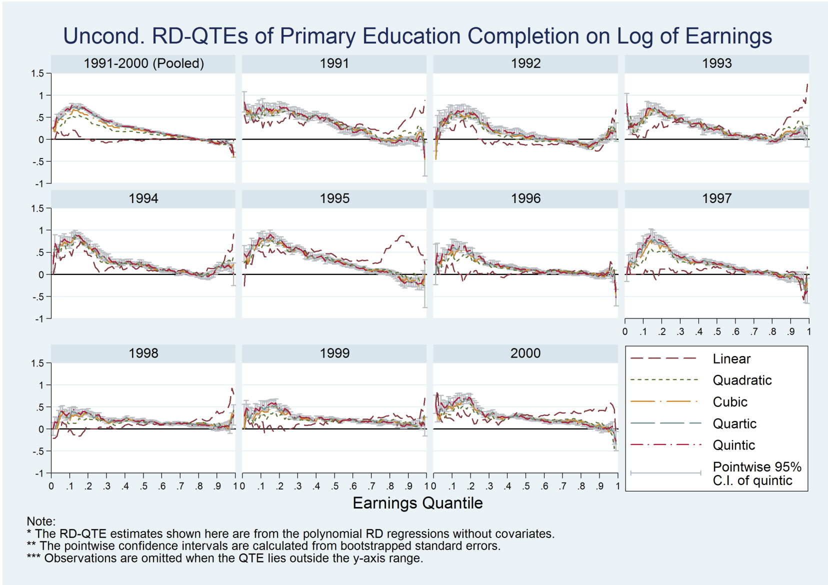
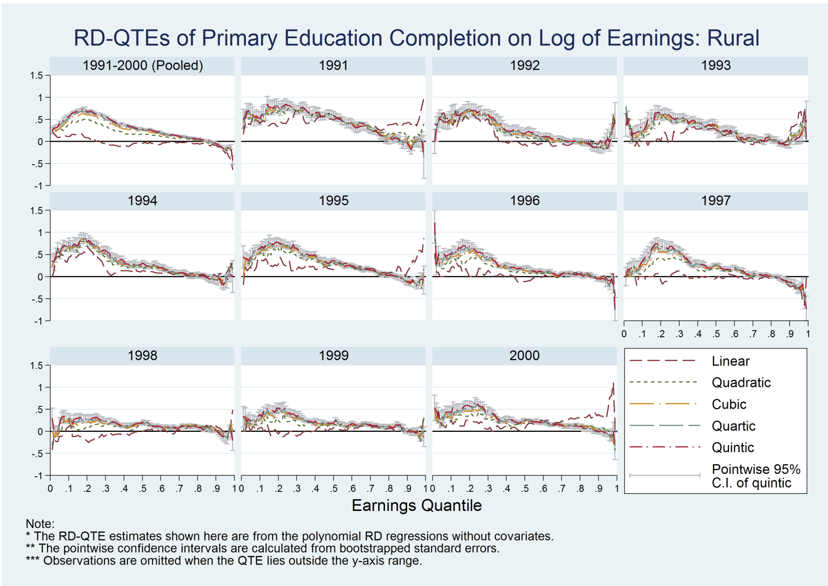


Figure 5.A.5: Robustness Checks of the RD Quantile Treatment Effects

(a) Pooled Sample



(b) Rural-Urban



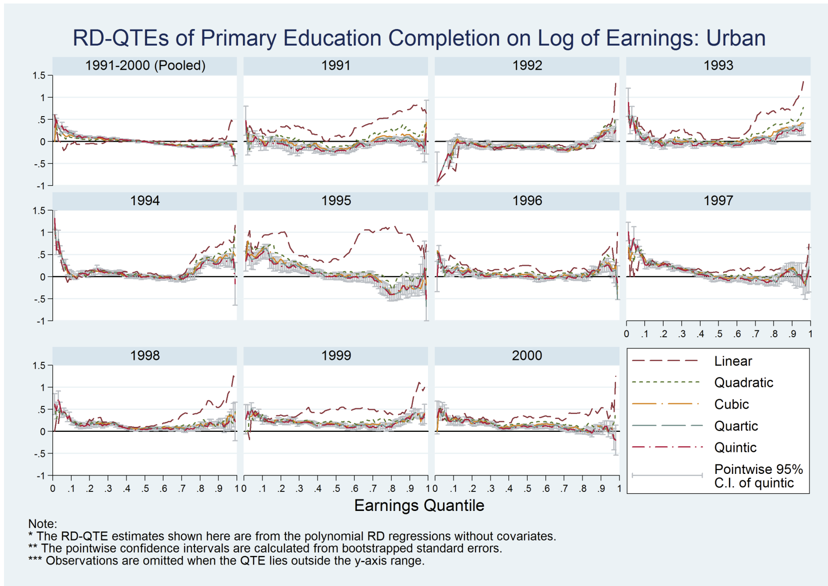
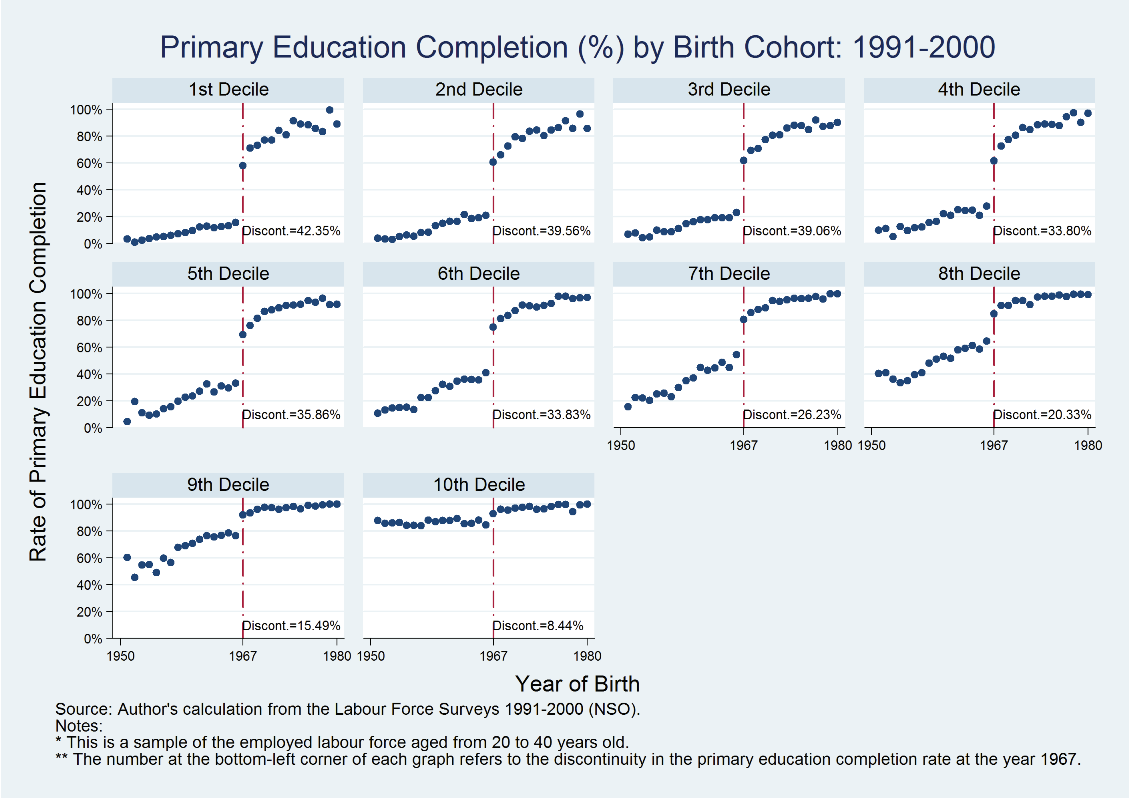
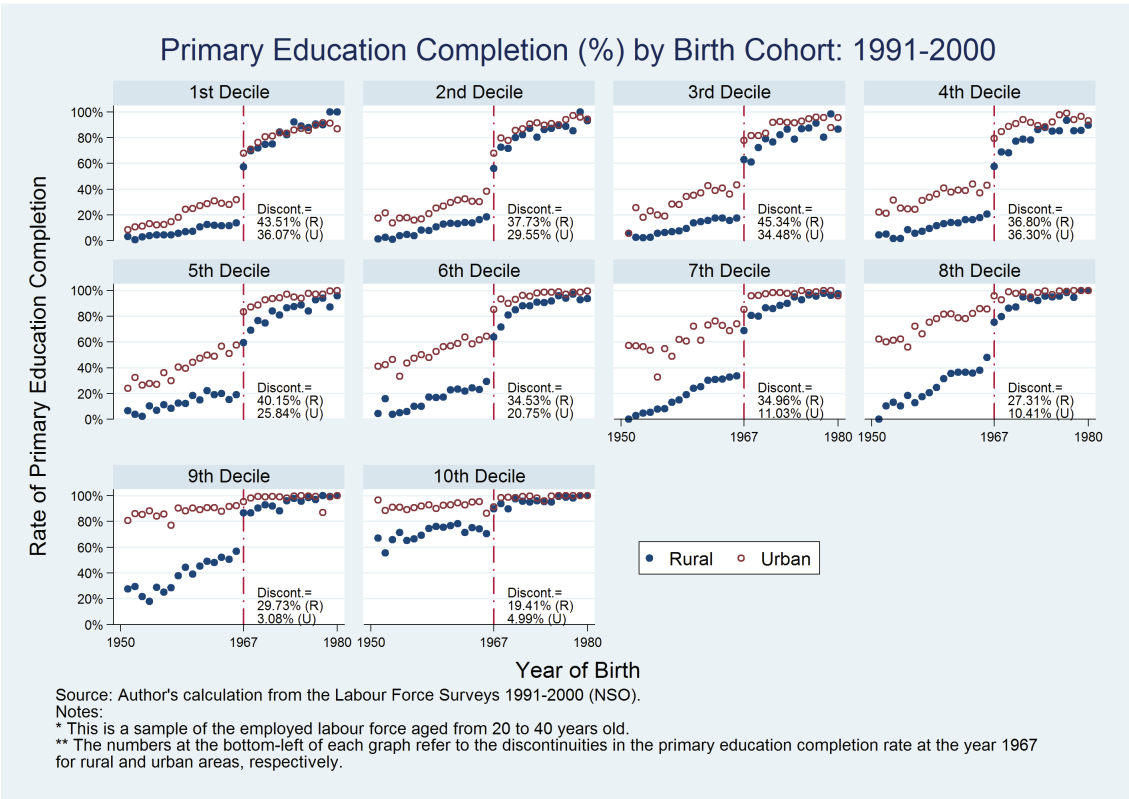


Figure 5.A.6: Primary Education Completion Rates by Earnings Decile

(a) Pooled Sample



(b) Rural and Urban



Chapter 6

Conclusion

In this thesis, I have examined the importance of human capital for economic development in Thailand between 1985 and 2000. This period is particularly interesting as the Thai economy experienced high economic growth and underwent a process of structural transformation. More specifically, I have estimated the effects of education on three labour market outcomes – namely earnings, sector of employment, and earnings inequality – using the Labour Force Survey (LFS) data. I addressed the endogeneity of education using a change in the compulsory schooling law as a source of exogenous variation in education. In doing so, I investigated how the presence of endogeneity, sample selection, and heterogeneity in returns to education across individuals, levels of education, sectors of work, and the earnings distribution affects the estimation of the returns to education.

The second chapter reviewed the existing literature on the returns to education on earnings and sector of work as well as the effects of education on inequality in Thailand. I noted that most of the earlier studies focused only on wage-employed workers, which accounted for less than half of the working population, while ignoring the endogeneity and sample selection problems. In addition, the effects of education are often constrained to be linear and homogeneous across sectors of work and individuals. The overall descriptive statistics from the LFS suggest that these assumptions are unlikely to be met, particularly during the structural transformation, as human capital endowments and earnings differ between agriculture and non-agriculture. The descriptive statistics also highlight a substantial, but declining, number of unpaid family workers, which potentially has implications for labour market studies. Further research is therefore called for to clarify the roles of human capital in economic development during the structural transformation period.

I started the investigation, in the third chapter, by estimating the overall effects of education on earnings and the probability of being in the non-agricultural sector. In this chapter, the earnings and sectoral sorting models allow for the endogenous choice of education and heterogeneous education effects between gender. As the change in the compulsory schooling law influences individuals who were born in different cohorts in a discontinuous way, a regression discontinuity (RD) framework is adopted to identify the average return to education and the average impact of education on the sector of work.

First, I showed that the change in the compulsory schooling law has a significant positive impact on education at all levels, and especially around the bottom and middle of the education distribution. Second, I found significant effects of education on both earnings and the sectoral sorting process. Using the RD to estimate the average education effects, I found that the returns to education are around 8 per cent, lower than the ordinary least squares (OLS) estimates and much lower than the estimates for the wage-employed workers from the existing literature. Meanwhile, the RD impacts of education on the probability of working in agriculture are around 3 per cent, slightly lower than their OLS estimates. If the compliers are representative of the population, the lower-than-OLS education effects would suggest a positive correlation between education and unobserved individual heterogeneity, such as ability and preferences, among paid workers included in the sample. Otherwise, the lower RD estimates could imply relatively lower returns among those who were affected by the change in the compulsory schooling law. The latter explanation is more likely as the education reform did not affect the entire population in the same way. Third, by estimating the education effects for men and women separately, I confirmed the importance of heterogeneity across gender which is masked by the pooled sample. While the returns to education are higher for women, the effects on working in non-agriculture are stronger for men. These findings, together with the substantial proportion of female unpaid workers in agriculture call for further analysis which takes into account the sample selection issue as well as the heterogeneity by gender. It is also important to emphasise that the RD technique restricts the estimated effects of education to be linear.

Comparing returns to education across sectors of employment and levels of education could be informative about the role of human capital development in facilitating structural transformation. In the fourth chapter, I estimated the effects of education on sectoral sorting and earnings processes in a more flexible way than previously done under the RD framework. First, I allowed the sectoral sorting process to be correlated with the selection into paid employment as suggested by the LFS descriptive statistics. Second, the returns to education were allowed to be heterogeneous across sectors, levels of education, and individuals. The heterogeneous returns across sectors can be studied by estimating the earnings functions for agriculture and non-agriculture separately. Given potential non-random selection into both the sector of work and type of employment, a double selection technique is employed to correct for the possible bias. Third, I attempted to estimate these education effects for the entire population using a control function (CF) approach, which allows for the endogeneity of education and heterogeneous returns across individuals.

My results in the fourth chapter emphasise the crucial role of human capital in enhancing economic growth during the structural transformation period. I found that education has significant positive effects on reallocating workers towards the non-agricultural sector for all workers, and on moving workers towards paid employment particularly for female workers. The correlation between the selection into paid employment and the agricultural sector is significant and negative as anticipated. In addition, the effects of education on earnings are found to be non-linear and differ greatly across sectors. The education-earnings profiles are estimated to be concave in the agricultural sector and convex in the non-agricultural sector. More specifically, at a low level of education, the agricultural returns to education are relatively higher but they are decreasing in education and eventually become negative after 6 years of primary education. On the other hand, the non-agricultural returns to education are very low at the beginning but increase with education and become higher than the agricultural returns at around 6 years of primary school. The returns to education are also found to be heterogeneous across gender, as women's education-earnings profile is relatively more concave in the agricultural sector and less convex in the non-agricultural sector.

In the fifth chapter, I used the RD framework to investigate the returns to primary education when they are allowed to be heterogeneous across the earnings distribution. I explored whether this type of heterogeneity, if it exists, results in higher or lower earnings inequality. In addition to the overall distributional effects, I also analysed the effects of primary education completion on earnings for the rural and urban population separately.

I found that the increased primary education completion rate, induced by the change in compulsory schooling, reduces earnings inequality as the returns to primary education are larger for the poor, compared to the rich. The equalising effects of primary education are also found to be higher in the first half of the 1990s and among workers in rural areas. The insignificant distributional effects of primary education in urban areas could be attributed to the smaller number of individuals who are affected by the change in the compulsory schooling, particularly at the top of the earnings distribution.

To conclude, this thesis has used micro-level labour data to study the labour market characteristics of the growth process. I have confirmed the importance of human capital for economic development and structural transformation in Thailand. Equally importantly, I have shed light on the importance of accounting for various types of heterogeneity when analysing micro-level labour data on the earnings and sectoral sorting processes.

Bibliography

Aakvik, A., K. G. Salvanes, and K. Vaage: 2003, 'Measuring Heterogeneity in the Returns to Education in Norway Using Educational Reforms'. IZA Discussion Papers 815, Institute for the Study of Labor (IZA).

Abadie, A.: 1997, 'Changes in Spanish Labor Income Structure during the 1980's: A Quantile Regression Approach'. *Investigaciones Economicas* **21**(2), 253–272.

Abadie, A., J. Angrist, and G. Imbens: 2002, 'Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings'. *Econometrica* **70**(1), 91–117.

Acemoglu, D.: 1998, 'Why Do New Technologies Complement Skills? Directed Technical Change And Wage Inequality'. *The Quarterly Journal of Economics* **113**(4), 1055–1089.

Acemoglu, D. and J. Angrist: 2001, 'How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws'. In: *NBER Macroeconomics Annual 2000, Volume 15*, NBER Chapters. National Bureau of Economic Research, Inc, pp. 9–74.

ADB: 2007, *Asian Development Outlook 2007: Growth Amid Change.* Asian Development Bank.

Ahluwalia, M. S.: 1976, 'Income Distribution and Development: Some Stylized Facts'. *American Economic Review* **66**(2), 128–35.

Anand, S.: 1983, *Inequality and Poverty in Malaysia: Measurement and Decomposition*, World Bank research publication. Published for the World Bank by Oxford University Press.

Anand, S. and S. M. R. Kanbur: 1993, 'Inequality and Development: A Critique'. *Journal of Development Economics* **41**(1), 19–43.

Angrist, J. D. and G. W. Imbens: 1995, 'Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity'. *Journal of the American Statistical Association* **90**(430), pp. 431–442.

Angrist, J. D., G. W. Imbens, and D. B. Rubin: 1996, 'Identification of Causal Effects Using Instrumental Variables'. *Journal of the American Statistical Association* **91**(434), pp. 444–455.

Angrist, J. D. and A. B. Krueger: 1991, 'Does Compulsory School Attendance Affect Schooling and Earnings?'. *The Quarterly Journal of Economics* **106**(4), 979–1014.

Appleton, S., L. Song, and Q. Xia: 2012, 'Understanding Urban Wage Inequality in China 1988-2008: Evidence from Quantile Analysis'. IZA Discussion Papers 7101, Institute for the Study of Labor (IZA).

Arabsheibani, G. R. and A. Mussurov: 2007, 'Returns to Schooling in Kazakhstan'. *Economics of Transition* **15**(2), 341–364.

Arrazola, M. and J. de Hevia: 2008, 'Three Measures of Returns to Education: An Illustration for the Case of Spain'. *Economics of Education Review* **27**(3), 266–275.

Arrazola, M., J. D. Hevia, M. Risueno, and J. F. Sanz: 2003, 'Returns to Education in Spain: Some Evidence on the Endogeneity of Schooling'. *Education Economics* **11**(3), 293–304.

Ashenfelter, O., C. Harmon, and H. Oosterbeek: 1999, 'A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias'. Working Papers 804, Princeton University, Department of Economics, Industrial Relations Section.

Ashenfelter, O. and A. Krueger: 1994, 'Estimates of the Economic Return to Schooling from a New Sample of Twins'. *The American Economic Review* **84**(5), pp. 1157–1173.

Ashenfelter, O. and D. J. Zimmerman: 1997, 'Estimates Of The Returns To Schooling From Sibling Data: Fathers, Sons, And Brothers'. *The Review of Economics and Statistics* **79**(1), 1–9.

Autor, D. H., L. F. Katz, and A. B. Krueger: 1998, 'Computing Inequality: Have Computers Changed The Labor Market?'. *The Quarterly Journal of Economics* **113**(4), 1169–1213.

Aydemir, A. B. and M. G. Kirdar: 2013, 'Estimates of the Return to Schooling in a Developing Country: Evidence from a Major Policy Reform in Turkey'. Mpra paper, University Library of Munich, Germany.

Bank of Thailand: 2013, *Exchange Rates 1985 to Present*. (Accessed - November 2013).

Barro, R. J.: 1999, 'Inequality, Growth, and Investment'. NBER Working Papers 7038, National Bureau of Economic Research, Inc.

Barro, R. J.: 2000, 'Inequality and Growth in a Panel of Countries'. *Journal of Economic Growth* **5**(1), 5–32.

Barro, R. J. and J.-W. Lee: 1993, 'International Comparisons of Educational Attainment'. *Journal of Monetary Economics* **32**(3), 363–394.

Barro, R. J. and J. W. Lee: 1996, 'International Measures of Schooling Years and Schooling Quality'. *American Economic Review* **86**(2), 218–23.

Becker, G. S.: 1964, 1994, *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition)*, No. beck94-1 in NBER Books. National Bureau of Economic Research, Inc.

Becker, G. S. and B. R. Chiswick: 1966, 'Education and the Distribution of Earnings'. *The American Economic Review* **56**(1/2), pp. 358–369.

Behrman, J. R. and P. Taubman: 1976, 'Intergenerational Transmission of Income and Wealth'. *American Economic Review* **66**(2), 436–40.

Berman, E., J. Bound, and Z. Griliches: 1994, 'Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufacturers'. *The Quarterly Journal of Economics* **109**(2), pp. 367–397.

Blackburn, M. L. and D. Neumark: 1995, 'Are OLS Estimates of the Return to Schooling Biased Downward? Another Look'. *The Review of Economics and Statistics* **77**(2), pp. 217–230.

Blanchflower, D. and P. Elias: 1993, 'Ability, Schooling and Earnings: Are Twins Different?'. Working papers, Dartmouth College.

Blom, A., L. Holm-Nielsen, and D. Verner: 2001, 'Education, Earnings, and Inequality in Brazil, 1982-98 : Implications for Education Policy'. Policy Research Working Paper Series 2686, The World Bank.

Blundell, R., L. Dearden, and B. Sianesi: 2004, 'Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS'. CEE Discussion Papers 0047, Centre for the Economics of Education, LSE.

Bound, J. and G. Johnson: 1992, 'Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations'. *American Economic Review* **82**(3), 371–92.

Bound, J. and G. Solon: 1999, 'Double Trouble: On the Value of Twins-based Estimation of the Return to Schooling'. *Economics of Education Review* **18**(2), 169–182.

Brunello, G., M. Fort, and G. Weber: 2009, 'Changes in Compulsory Schooling, Education and the Distribution of Wages in Europe'. *Economic Journal* **119**(536), 516–539.

Buchinsky, M.: 1994, 'Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression'. *Econometrica* **62**(2), pp. 405–458.

Buchinsky, M.: 2001, 'Quantile Regression with Sample Selection: Estimating Women's Return to Education in the U.S'. *Empirical Economics* **26**(1), 87–113.

Burger, R.: 2010, 'Estimating the Shape of the South African Schooling-Earnings Profile'. Ph.D. thesis, Department of Economics, University of Oxford.

Buscha, F. and M. Dickson: 2012, 'The raising of the school leaving age: Returns in later life'. *Economics Letters* **117**(2), 389–393.

Card, D.: 1993, 'Using Geographic Variation in College Proximity to Estimate the Return to Schooling'. NBER Working Papers 4483, National Bureau of Economic Research, Inc.

Card, D.: 1994, 'Earnings, Schooling, and Ability Revisited'. Working Papers 710, Princeton University, Department of Economics, Industrial Relations Section.

Card, D.: 1999, 'The Causal Effect of Education on Earnings'. In: O. Ashenfelter and D. Card (eds.): *Handbook of Labor Economics*, Vol. 3, Part A. Elsevier, 1 edition, Chapt. 30, pp. 1801–1863.

Card, D.: 2001, 'Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems'. *Econometrica* **69**(5), 1127–60.

Card, D. and J. E. DiNardo: 2002, 'Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles'. *Journal of Labor Economics* **20**(4), pp. 733–783.

Carneiro, P. and J. J. Heckman: 2003, 'Empirical Estimates of the Returns to Schooling'. Working papers, University of Chicago.

Chalamwong, Y. and S. Amornthum: 2001, 'Rate of Return to Education'. Working papers, TDRI.

Chamberlain, G.: 1991, 'Quantile Regression, Censoring, and the Structure of Wages'. Harvard institute of economic research working papers, Harvard - Institute of Economic Research.

Chamberlain, G. and Z. Griliches: 1975, 'Unobservables with a Variance-Components Structure: Ability, Schooling, and the Economic Success of Brothers'. *International Economic Review* **16**(2), 422–49.

Checchi, D.: 2000, 'Does Educational Achievement Help to Explain Income Inequality?'. Research Paper 208, World Institute for Development Economics Research.

Chen, G. and S. Hamori: 2009, 'Economic Returns to Schooling in Urban China: OLS and the Instrumental Variables Approach'. *China Economic Review* **20**(2), 143–152.

Chernozhukov, V. and C. Hansen: 2005, 'An IV Model of Quantile Treatment Effects'. *Econometrica* **73**(1), 245–261.

Chiswick, B. R.: 1969, 'Minimum Schooling Legislation and the Cross-Sectional Distribution of Income'. *Economic Journal* **79**(315), 495–507.

Chiswick, C. U.: 1977, 'On Estimating Earnings Functions for LDCs'. *Journal of Development Economics* **4**(1), 67 – 78.

Clay, K., J. Lingwall, and J. Melvin Stephens: 2012, 'Do Schooling Laws Matter? Evidence from the Introduction of Compulsory Attendance Laws in the United States'. NBER Working Papers 18477, National Bureau of Economic Research, Inc.

Co, C. Y., I. N. Gang, and M.-S. Yun: 1999, 'Switching Models with Self-Selection: Self-Employment in Hungary'. Departmental Working Papers 199912, Rutgers University, Department of Economics.

Cowell, F.: 2011, *Measuring Inequality*, London School of Economics Perspectives in Economic Analysis. OUP Oxford.

Das, M., W. K. Newey, and F. Vella: 2003, 'Nonparametric Estimation of Sample Selection Models'. *Review of Economic Studies* **70**(1), 33–58.

De Gregorio, J. and J.-W. Lee: 2002, 'Education and Income Inequality: New Evidence from Cross-Country Data'. *Review of Income and Wealth* **48**(3), 395–416.

Deininger, K.: 2003, 'Does Cost of Schooling Affect Enrollment by the Poor? Universal Primary Education in Uganda'. *Economics of Education Review* **22**(3), 291–305.

Deininger, K. and L. Squire: 1996, 'A New Data Set Measuring Income Inequality'. *World Bank Economic Review* **10**(3), 565–91.

Deininger, K. and L. Squire: 1998, 'New Ways of Looking at Old Issues: Inequality and Growth'. *Journal of Development Economics* **57**(2), 259–287.

Del Bono, E. and F. Galindo-Rueda: 2004, 'Do a Few Months of Compulsory Schooling Matter? The Education and Labour Market Impact of School Leaving Rules'. IZA Discussion Papers 1233, Institute for the Study of Labor (IZA).

Devereux, P. J. and R. A. Hart: 2010, 'Forced to be Rich? Returns to Compulsory Schooling in Britain'. *Economic Journal* **120**(549), 1345–1364.

Dobkin, C. and F. Ferreira: 2010, 'Do School Entry Laws Affect Educational Attainment and Labor Market Outcomes?'. *Economics of Education Review* **29**(1), 40–54.

Duflo, E.: 2001, 'Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment'. *American Economic Review* **91**(4), 795–813.

Eckstein, Z. and I. Zilcha: 1994, 'The Effects of Compulsory Schooling on Growth, Income Distribution and Welfare'. *Journal of Public Economics* **54**(3), 339–359.

Falco, P., A. Kerr, P. Paci, and B. M. J. Rijkers: 2014, *Working toward Better Pay: Earning Dynamics in Ghana and Tanzania*. The World Bank.

Fallon, P. R. and R. E. Lucas: 2002, 'The Impact of Financial Crises on Labor Markets, Household Incomes, and Poverty: A Review of Evidence'. *World Bank Research Observer* **17**(1), 21–45.

Fang, H., K. N. Eggleston, J. A. Rizzo, S. Rozelle, and R. J. Zeckhauser: 2012, 'The Returns to Education in China: Evidence from the 1986 Compulsory Education Law'. Working Paper 18189, National Bureau of Economic Research.

Fields, G. S.: 1980, *Education and Income Distribution in Developing Countries: A Review of the Literature*. Washington, DC: The World Bank.

Firpo, S., N. Fortin, and T. Lemieux: 2009, 'Unconditional Quantile Regressions'. *Econometrica* **77**(3), 953–973.

Firpo, S., N. Fortin, and T. Lemieux: 2011, *Decomposition Methods in Economics*, Vol. 4 of *Handbook of Labor Economics*, Chapt. 1, pp. 1–102. Elsevier.

Firpo, S. P. and C. Pinto: 2011, 'Identification and Estimation of Interventions Using Changes in Inequality Measures'. Textos para discussão 214, Escola de Economia de São Paulo, Getulio Vargas Foundation (Brazil).

Fishe, R. P., R. Trost, and P. M. Lurie: 1981, 'Labor Force Earnings and College Choice of Young Women: An Examination of Selectivity Bias and Comparative Advantage'. *Economics of Education Review* **1**(2), 169 – 191.

Fitzenberger, B., R. Hujer, T. E. MaCurdy, and R. Schnabel: 2001, 'Testing for Uniform Wage Trends in West-Germany: A Cohort Analysis Using Quantile Regressions for Censored Data'. *Empirical Economics* **26**(1), 41–86.

Fofack, H. and A. Zeufack: 1999, 'Dynamics of Income Inequality in Thailand: Evidence from Household Pseudo-Panel Data'. Working papers, World Bank.

Fort, M.: 2012, 'Unconditional and Conditional Quantile Treatment Effect: Identification Strategies and Interpretations'. Working Papers wp857, Dipartimento Scienze Economiche, Universita' di Bologna.

Frandsen, B. R., M. Frölich, and B. Melly: 2012, 'Quantile Treatment Effects in the Regression Discontinuity Design'. *Journal of Econometrics* **168**(2), 382–395.

Frölich, M.: 2007, 'Regression Discontinuity Design with Covariates'. IZA Discussion Papers 3024, Institute for the Study of Labor (IZA).

Frölich, M. and B. Melly: 2008, 'Unconditional Quantile Treatment Effects under Endogeneity'. IZA Discussion Papers 3288, Institute for the Study of Labor (IZA).

Frölich, M. and B. Melly: 2010a, 'Estimation of Quantile Treatment Effects with Stata'. *Stata Journal* **10**(3), 423–457.

Frölich, M. and B. Melly: 2010b, 'Quantile Treatment Effects in the Regression Discontinuity Design: Process Results and Gini Coefficient'. IZA Discussion Papers 4993, Institute for the Study of Labor (IZA).

Frölich, M. and B. Melly: 2012, 'Unconditional Quantile Treatment Effects under Endogeneity'. IZA Discussion Papers 3288, Institute for the Study of Labor (IZA).

García, J., P. J. Hernández, and A. López-Nicolás: 2001, 'How Wide is the Gap? An Investigation of Gender Wage Differences using Quantile Regression'. *Empirical Economics* **26**(1), 149–167.

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

Gelman, A. and G. Imbens: 2014, 'Why High-order Polynomials Should not be Used in Regression Discontinuity Designs'. NBER Working Papers 20405, National Bureau of Economic Research, Inc.

Ghinetti, P.: 2014, 'A New Look at the Public Wage Premium in Italy: The Role of Schooling Endogeneity'. *Labour* **28**(1), 87–111.

Girma, S. and A. Kedir: 2005, 'Heterogeneity in returns to schooling: Econometric evidence from Ethiopia'. *The Journal of Development Studies* **41**(8), 1405–1416.

Glomm, G. and B. Ravikumar: 1992, 'Public versus Private Investment in Human Capital Endogenous Growth and Income Inequality'. *Journal of Political Economy* **100**(4), 818–34.

Grenet, J.: 2013, 'Is Extending Compulsory Schooling Alone Enough to Raise Earnings? Evidence from French and British Compulsory Schooling Laws'. *Scandinavian Journal of Economics* **115**(1), 176–210.

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

Griliches, Z. and W. M. Mason: 1972, 'Education, Income, and Ability'. *Journal of Political Economy* **80**(3), S74–S103.

Gronau, R.: 1974, 'Wage Comparisons: A Selectivity Bias'. *Journal of Political Economy* **82**(6), pp. 1119–1143.

Guiteras, R.: 2008, 'Estimating Quantile Treatment Effects in a Regression Discontinuity Design'. Working paper, Department of Economics, MIT.

Hahn, J., P. Todd, and W. Van der Klaauw: 2001, 'Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design'. *Econometrica* **69**(1), 201–09.

Hall, R. E. and C. I. Jones: 1999, 'Why Do Some Countries Produce So Much More Output Per Worker Than Others?'. *The Quarterly Journal of Economics* **114**(1), 83–116.

Ham, J. C.: 1982, 'Estimation of a Labour Supply Model with Censoring Due to Unemployment and Underemployment'. *Review of Economic Studies* **49**(3), 335–54.

Harmon, C., H. Oosterbeek, and I. Walker: 2003, 'The Returns to Education: Microeconomics'. *Journal of Economic Surveys* **17**(2), 115–156.

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

Harris, J. R. and M. P. Todaro: 1970, 'Migration, Unemployment & Development: A Two-Sector Analysis'. *American Economic Review* **60**(1), 126–42.

Hawley, J. D.: 2004, 'Changing Returns to Education in Times of Prosperity and Crisis, Thailand 1985-1998'. *Economics of Education Review* **23**(3), 273–286.

Heckman, J., J. L. Tobias, and E. Vytlacil: 2001, 'Four Parameters of Interest in the Evaluation of Social Programs'. *Southern Economic Journal* **68**(2), 210–223.

Heckman, J. J.: 1974, 'Shadow Prices, Market Wages, and Labor Supply'. *Econometrica* **42**(4), 679–94.

Heckman, J. J.: 1979, 'Sample Selection Bias as a Specification Error'. *Econometrica* **47**(1), pp. 153–161.

Heckman, J. J. and J. Smith: 1997, 'Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts'. *Review of Economic Studies* **64**(4), 487–535.

Hernandez, Z.: 2004, 'Industrial Policy in East Asia: In Search for Lessons'. Technical report, World Bank. Background paper prepared for the WDR 2005: A Better Investment Climate for Everyone.

Heshmati, A.: 2004, 'Inequalities and Their Measurement'. IZA Discussion Papers 1219, Institute for the Study of Labor (IZA).

Holland, P. W.: 1986, 'Statistics and Causal Inference'. *Journal of the American Statistical Association* **81**(396), pp. 945–960.

Imbens, G. and J. M. Wooldridge: 2007a, 'Control Functions and Related Methods'. Working papers, National Bureau of Economic Research, Inc.

Imbens, G. and J. M. Wooldridge: 2007b, 'What's New in Econometrics'. Working papers, NBER Summer Lectures.

Imbens, G. W. and J. D. Angrist: 1994, 'Identification and Estimation of Local Average Treatment Effects'. *Econometrica* **62**(2), pp. 467–475.

Imbens, G. W. and T. Lemieux: 2008, 'Regression Discontinuity Designs: A Guide to Practice'. *Journal of Econometrics* **142**(2), 615–635.

Imbens, G. W. and D. B. Rubin: 1997, 'Estimating Outcome Distributions for Compliers in Instrumental Variables Models'. *The Review of Economic Studies* **64**(4), pp. 555–574.

Inada, K.-I.: 1963, 'On a Two-Sector Model of Economic Growth: Comments and a Generalization'. *The Review of Economic Studies* **30**(2), pp. 119–127.

Jimenez, E.: 1986, 'The Public Subsidization of Education and Health in Developing Countries: A Review of Equity and Efficiency'. *World Bank Research Observer* **1**(1), 111–29.

Katz, L. F. and K. M. Murphy: 1992, 'Changes in Relative Wages, 1963-1987: Supply and Demand Factors'. *The Quarterly Journal of Economics* **107**(1), 35–78.

Kerr, A. and S. Quinn: 2010, 'Returns to Education in Tanzania: Exploiting a Natural Experiment'. Working papers, CSAE, Oxford.

Khitarishvili, T.: 2010, 'Assessing the Returns to Education in Georgia'. Economics Working Paper Archive WP608, The Levy Economics Institute.

Knight, J. and L. Song: 2003, 'Increasing urban wage inequality in China'. *Economics of Transition* **11**(4), 597–619.

Knight, J. B. and R. H. Sabot: 1983, 'Educational Expansion and the Kuznets Effect'. *American Economic Review* **73**(5), 1132–36.

Koop, G. and D. J. Poirier: 1997, 'Learning about the Across-Regime Correlation in Switching Regression Models'. *Journal of Econometrics* **78**(2), 217–227.

Krongkaew, M., S. Chamnivickorn, and I. Nitithanprapas: 2006, *Economic Growth, Employment, and Poverty Reduction Linkages: The Case of Thailand*, Issues in employment and poverty, discussion paper. ILO.

Krueger, A. B.: 1993, 'How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989'. *The Quarterly Journal of Economics* **108**(1), 33–60.

Krueger, J. J.: 2008, 'Productivity and Structural Change: A Revieew of the Literature'. *Journal of Economic Surveys* **22**(2), 330–363.

Kuchiki, A.: 2007, 'Industrial Policy in Asia'. Technical Report 128, Institute of Developing Economies.

Kuznets, S.: 1955, 'Economic Growth and Income Inequality'. *The American Economic Review* **45**(1), pp. 1–28.

Lang, K.: 1993, 'Ability Bias, Discount Rate Bias and the Return to Education'. MPRA Paper 24651, University Library of Munich, Germany.

Lathapipat, D.: 2008, 'The Changing Educational Distribution and its Impact on the Evolution of Wages in Thailand, 1987-2006'. EABER Working Papers 21951, East Asian Bureau of Economic Research.

Leckcivilize, A.: 2013, 'Essays on Labour Economics'. Ph.D. thesis, The London School of Economics and Political Science.

Lee, D. S. and D. Card: 2008, 'Regression Discontinuity Inference with Specification Error'. *Journal of Econometrics* **142**(2), 655–674.

Lee, D. S. and T. Lemieux: 2010, 'Regression Discontinuity Designs in Economics'. *Journal of Economic Literature* **48**(2), 281–355.

Lemieux, T.: 2006, 'Post-secondary Education and Increasing Wage Inequality'. *American Economic Review* **96**(2), 195–199.

Lewis, W. A.: 1954, 'Economic Development with Unlimited Supplies of Labour'. *The Manchester School* **22**(2), 139–191.

Lewis, W. A.: 1979, 'The Dual Economy Revisited'. *The Manchester School of Economic & Social Studies* **47**(3), 211–29.

Lucas, R. J.: 1988, 'On the Mechanics of Economic Development'. *Journal of Monetary Economics* **22**(1), 3–42.

Mankiw, N. G., D. Romer, and D. N. Weil: 1992, 'A Contribution to the Empirics of Economic Growth'. *The Quarterly Journal of Economics* **107**(2), 407–37.

Marin, A. and G. Psacharopoulos: 1976, 'Schooling and Income Distribution'. *The Review of Economics and Statistics* **58**(3), pp. 332–338.

Martins, P. S. and P. T. Pereira: 2004, 'Does education reduce wage inequality? Quantile regression evidence from 16 countries'. *Labour Economics* **11**(3), 355–371.

Mata, J. and J. A. F. Machado: 2005, 'Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression'. *Journal of Applied Econometrics* **20**(4), 445–465.

McMillan, M. S. and D. Rodrik: 2011, 'Globalization, Structural Change and Productivity Growth'. NBER Working Papers 17143, National Bureau of Economic Research, Inc.

Meghir, C. and M. Palme: 1999, 'Assessing the effect of schooling on earnings using a social experiment'. IFS Working Papers W99/10, Institute for Fiscal Studies.

Meghir, C. and M. Palme: 2005, 'Educational Reform, Ability, and Family Background'. *American Economic Review* **95**(1), 414–424.

Mephokee, C.: 2003, 'Thai Labour Market in Transition toward a Knowledge-Based Economy'. Working Papers Chapter 1, pp.1-40, Institute of Developing Economies, Japan External Trade Organization.

Messinis, G.: 2013, 'Returns to education and urban-migrant wage differentials in China: IV quantile treatment effects'. *China Economic Review* **26**(0), 39 – 55.

Miguel, E., S. Satyanath, and E. Sergenti: 2004, 'Economic Shocks and Civil Conflict: An Instrumental Variables Approach'. *Journal of Political Economy* **112**(4), 725–753.

Mincer, J. A.: 1974, *Schooling, Experience, and Earnings*, No. 74-1 in NBER Books. National Bureau of Economic Research, Inc.

Ministry of Commerce of Thailand, *Regional Price Indices 1985 to Present*. (Accessed - November 2013).

Ministry of Education of Thailand: 1978, *Education Reform Act, 1978*. (Accessed - November 2013).

Ministry of Education of Thailand: 2010, *Education Reforms in Thailand*. (Accessed - November 2013).

Moenjak, T. and C. Worswick: 2003, 'Vocational education in Thailand: A Study of Choice and Returns'. *Economics of Education Review* **22**(1), 99–107.

Motonishi, T.: 2006, 'Why Has Income Inequality in Thailand Increased?: An Analysis using Surveys from 1975 to 1998'. *Japan and the World Economy* **18**(4), 464–487.

Mwabu, G. and T. P. Schultz: 1996, 'Education Returns Across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa'. *The American Economic Review* **86**(2), pp. 335–339.

National Economic and Social Development Board: 2010, *National Economic and Social Development Plans (from 1961 to Present)*. (Accessed - April 2010).

National Statistical Office of Thailand, *Labour Force Survey 1985-2000*. (Accessed - November 2013).

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

Numnak, G.: 2006, 'Development and Structural Diversification of the Thai Economy: The Cause of Labour Immigration'. Working papers, University of New England.

Oosterbeek, H. and D. Webbink: 2007, 'Wage Effects of an Extra Year of Basic Vocational Education'. *Economics of Education Review* **26**(4), 408–419.

Oreopoulos, P.: 2006a, 'The Compelling Effects of Compulsory Schooling: Evidence from Canada'. *Canadian Journal of Economics* **39**(1), 22–52.

Oreopoulos, P.: 2006b, 'Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter'. *American Economic Review* **96**(1), 152–175.

Oreopoulos, P.: 2008, 'Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter: Corrigendum'. Working papers, Department of Economics, University of British Columbia.

Park, K. H.: 1996, 'Educational Expansion and Educational Inequality on Income Distribution'. *Economics of Education Review* **15**(1), 51 – 58.

Patrinos, H. A.: 1995, 'Education and Earnings Differentials'. mimeo, World Bank.

Patrinos, H. A., C. Ridao-Cano, and C. Sakellariou: 2006, 'Estimating the Returns to Education: Accounting for Heterogeneity in Ability'. Policy Research Working Paper Series 4040, The World Bank.

Patrinos, H. A., C. Ridao-Cano, and C. Sakellariou: 2009, 'A Note on Schooling and Wage Inequality in the Public and Private Sector'. *Empirical Economics* **37**(2), 383–392.

Patrinos, H. A. and C. Sakellariou: 2011, 'Quality of Schooling, Returns to Schooling and the 1981 Vouchers Reform in Chile'. *World Development* **39**(12), 2245–2256.

Pekkarinen, T. and M. Pellicer: 2013, 'Education and Allocation of Skills in Tunisia: Evidence from an Education Reform'. *IZA Journal of Labor & Development* **2**(1), 14.

Pischke, J.-S. and T. von Wachter: 2008, 'Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation'. *The Review of Economics and Statistics* **90**(3), 592–598.

Prasad, E. S.: 2004, 'The Unbearable Stability of the German Wage Structure: Evidence and Interpretation'. *IMF Staff Papers* **51**(2), 354–385.

Psacharopoulos, G.: 1977, 'Unequal Access to Education and Income Distribution'. *De Economist* **125**(3), 383–392.

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

Puhani, P.: 2000, 'The Heckman Correction for Sample Selection and Its Critique'. *Journal of Economic Surveys* **14**(1), 53–68.

Ram, R.: 1981, 'Inequalities in Income and Schooling: A Different Point of View'. *De Economist* **129**(2), 253–261.

Ram, R.: 1989, 'Can educational expansion reduce income inequality in less-developed countries?'. *Economics of Education Review* **8**(2), 185 – 195.

Ram, R.: 1990, 'Educational Expansion and Schooling Inequality: International Evidence and Some Implications'. *The Review of Economics and Statistics* **72**(2), 266–74.

Ranis, G. and J. C. H. Fei: 1961, 'A Theory of Economic Development'. *The American Economic Review* **51**(4), pp. 533–565.

Ranis, G. and J. C. H. Fei: 1964, *Development of the Labor Surplus Economy: Theory and Policy*, A Publication of the Economic Growth Center, Yale University Series. R. D. Irwin.

Ranis, G. and J. C. H. Fei: 1997, *Growth and Development from an Evolutionary Perspective*. Malden, Mass. [u.a.]: Blackwell.

Rijkers, B., C. R. Laderchi, and F. Teal: 2010, 'Who Benefits from Promoting Small Enterprises? Some Empirical Evidence from Ethiopia'. *World Development* **38**(4), 523 – 540.

Rock, M. T.: 1995, 'Thai Industrial Policy: How Irrelevant Was It to Export Success?'. *Journal of International Development* **7**(5), 745–757.

Rodriguez-Pose, A. and V. Tselios: 2009, 'Education and Income in Equality in the Regions of the European Union'. *Journal of Regional Science* **49**(3), 411–437.

Romer, P. M.: 1989, 'Human Capital And Growth: Theory and Evidence'. NBER Working Papers 3173, National Bureau of Economic Research, Inc.

Roy, A. D.: 1951, 'Some Thoughts on the Distribution of Earnings'. *Oxford Economic Papers* **3**(2), pp. 135–146.

Saint-Paul, G. and T. Verdier: 1993, 'Education, democracy and growth'. *Journal of Development Economics* **42**(2), 399–407.

Schultz, T. P.: 1994, 'Human Capital Investment in Women and Men: Micro and Macro Evidence of Economic Returns'. Working papers, U.S. Agency for International Development.

Schultz, T. W.: 1963, *The Economic Value of Education*. New York and London: Columbia University Press.

Söderbom, M., F. Teal, A. Wambugu, and G. Kahyarara: 2006, 'The Dynamics of Returns to Education in Kenyan and Tanzanian Manufacturing'. *Oxford Bulletin of Economics and Statistics* **68**(3), 261–288.

Sergenti, E.: 2006, 'A Probit Model and Other Extensions to Miguel, Satyanth, and Sergenti (2004) - Economic Shocks and Civil Conflict: An Instrumental Variables Approach'. Working papers, New York University.

Shapiro, C. and J. E. Stiglitz: 1984, 'Equilibrium Unemployment as a Worker Discipline Device'. *American Economic Review* **74**(3), 433–44.

Shorrocks, A. F.: 1984, 'Inequality Decomposition by Population Subgroups'. *Econometrica* **52**(6), 1369–85.

Simione, F.: 2012, 'Labor Market Returns to Education in Mozambique: Learning From a Quasi-Experiment'. Working papers, Government of Mozambique.

Siriprachai, S.: 1996, *Inconsistencies and Inequalities in Thai Industrialisation*. Longman.

Solow, R. M.: 1956, 'A Contribution to the Theory of Economic Growth'. *The Quarterly Journal of Economics* **70**(1), 65–94.

Spence, A. M.: 1973, 'Job Market Signaling'. *The Quarterly Journal of Economics* **87**(3), 355–74.

Stiglitz, J. E.: 1973, 'Education and Inequality'. *Annals of the American Academy of Political and Social Science* **409**, pp. 135–145.

Sussangkarn, C. and Y. Chalamwong: 1994, 'Development Strategies and Their Impacts on Labour Market and Migration: Thai Case Study'. Working paper, TDRI.

Sussangkarn, C. and Y. Chalamwong: 1996, *Development Strategies and Their Impacts on Labour Market and Migration: Thai Case Study*. Paris: OECD.

Sussangkarn, C., F. Flatters, and S. Kittiprapas: 1999, 'Comparative Social Impacts of the Asian Economic Crisis in Thailand, Indonesia, Malaysia and the Philippines: A Preliminary Report'. *TDRI Quarterly Review* **14**(1), 3–9.

Sylwester, K.: 2002a, 'Can Education Expenditures Reduce Income Inequality?'. *Economics of Education Review* **21**(1), 43 – 52.

Sylwester, K.: 2002b, 'A Model of Public Education and Income Inequality with a Subsistence Constraint'. *Southern Economic Journal* **69**(1), pp. 144–158.

Tangtipongkul, K.: 2013, 'Rates of Return to Schooling in Thailand'. Working papers, Thammasat University.

Theil, H.: 1967, *Economics and I nformation Theory*, Studies in mathematical and managerial economics. North-Holland Pub. Co.

Tunali, I.: 1986, 'A General Structure for Models of Double-Selection and an Application to a Joint Migration/Earnings Process with Remigration'. *Research in Labor Economics* **8**, 235–282.

Van Der Klaauw, W.: 2008, 'Regressioon Discontinuity Analysis: A Survey of Recent Developments in Economics'. *LABOUR* **22**(2), 219–245.

Vijverberg, W. P. M.: 1993a, 'Educational Investments and Returns for Women and Men in Côte d'Ivoire'. *Journal of Human Resources* **28**(4), 933–974.

Vijverberg, W. P. M.: 1993b, 'Measuring the Unidentified Parameter of the Extended Roy Model of Selectivity'. *Journal of Econometrics* **57**(1-3), 69–89.

Walker, I. and Y. Zhu: 2001, *The Returns to Education: Evidence from the Labour Force Surveys*, Research report (Great Britain Department for Education and Skills). Great Britain Department for Education and Skills.

Wang, L.: 2011, 'How Does Education Affect the Earnings Distribution in Urban China?'. IZA Discussion Papers 6173, Institute for the Study of Labor (IZA).

Warunsiri, S. and R. McNown: 2010, 'The Returns to Education in Thailand: A Pseudo-Panel Approach'. *World Development* **38**(11), 1616 – 1625.

Weber, A. M. and A. Ammermüller: 2003, 'Education and Wage Inequality in Germany: A Review of the Empirical Literature'. ZEW Discussion Papers 03-29, ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research.

Wetzels, C. and A. Zorlu: 2003, 'Wage Effects of Motherhood: A Double Selection Approach'. NIMA Working Papers 22, Núcleo de Investigação em Microeconomia Aplicada (NIMA), Universidade do Minho.

Willis, R. J.: 1986, 'Human Capital and the Rise and Fall of Families: Comment'. *Journal of Labor Economics* **4**(3), S40–47.

Winegarden, C. R.: 1979, 'Schooling and Income Distribution: Evidence from International Data'. *Economica* **46**(181), pp. 83–87.

Wooldridge, J. M.: 1997, 'On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model'. *Economics Letters* **56**(2), 129–133.

Wooldridge, J. M.: 2002, *Econometric Analysis of Cross Section and Panel Data*, Vol. 1 of *MIT Press Books*. The MIT Press, 1 edition.

Wooldridge, J. M.: 2010, *Econometric Analysis of Cross Section and Panel Data*, Vol. 1 of *MIT Press Books*. The MIT Press, 2 edition.

World Bank: 1993, *The East Asian Miracle: Economic Growth and Public Policy: Summary*, A World Bank policy research report. World Bank.

World Bank's WDI: 2013, *World Development Indicators (WDI).* (Accessed - November 2013).

Zhang, J.: 1996, 'Optimal Public Investments in Education and Endogenous Growth'. *The Scandinavian Journal of Economics* **98**(3), pp. 387–404.