Revisiting the Returns to Education during the Rapid Structural and Rural Transformation in Thailand: a regression discontinuity approach

Nobuhiko Fuwa* (Graduate School of Public Policy, The University of Tokyo)

Address: 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

E-mail: fuwa@pp.u-tokyo.ac.jp

Upalat Korwatanasakul (Graduate School of Asia-Pacific Studies, Waseda University)

Address: 1-9-5-2111 Nishiwaseda Shinjuku Tokyo 169-0051

E-mail: korwatanasakul.upalat@gmail.com

Abstract:

This paper estimates returns to schooling in Thailand using a regression discontinuity approach applied to the change in compulsory schooling law in 1978. We find that the compulsory schooling law played a role in enhancing human capital investment in the eve of the rapid structural transformation in the 1980s, that the returns to schooling based on our IV estimation was round 8%, while OLS somewhat overestimates (by 20%) such returns, and that returns were higher in urban areas, in services (than in agricultural) sectors and, surprisingly, in underdeveloped Northern regions. Our findings are in sharp contrast with most of the recent studies exploiting similar institutional changes from developed countries, where OLS estimates tend to *under*-estimate returns to schooling with the implication that those school *drop-outs* (whose behavior is altered by compulsory schooling) tend to have *higher* returns than those already in school even before the law change. The conventional notion of 'ability bias' (which we confirm) are more likely to arise in developing (but not so much in developed) countries possibly because parents could be forced to keep only those (among many) of their children with higher ability in school, thereby reinforcing (rather than compensating) inequality among children within the household.

Keywords: returns to education, Mincer equation, ability bias, regression discontinuity, Thailand

JEL classification: I20, I21, I25, I28

1. Introduction

The fundamental importance of human capital formation in the process of economic development is well understood. However, quantitative magnitudes of the causal effects of education on earnings are still intensely debated in both the developed and developing country contexts. Recent studies from developed countries have shown that endogeneity bias in the conventional OLS estimates is quite substantial, and that there is a great deal of heterogeneity in returns to education within population. In developing countries, however, similar studies remain relatively scarce. Obtaining accurate estimates of returns to education is essential for policy making. Without such estimates, governments are poorly informed as to how to allocate scarce resources among different types of capital (e.g., human vs. physical capital) as well as how to allocate their education budgets efficiently (e.g., rural vs. urban areas, different regions in a country, different levels of schooling, primary vs. higher education).

This paper applies a regression discontinuity approach to the incidence of the change in the compulsory schooling law in 1978 (which effectively extended the compulsory years of schooling from 4 to 6 years) in Thailand, a methodology with an increasing number of applications in developed countries but rarely found in developing countries. We find that the compulsory schooling law played a role in enhancing human capital investment in the eve of the rapid structural transformation in the 1980s, that the returns to schooling based on our IV estimation was round 8%, while OLS substantially overestimates (by 20%) such returns, and that returns were higher in urban areas, in services (than in agricultural) sectors and, surprisingly, in underdeveloped Northern regions. Our findings are in contrast with most of the recent studies exploiting similar institutional changes from developed countries, where OLS estimates tend to *under*-estimate returns to schooling with the implication that those school *drop-outs* (whose behavior is altered by compulsory schooling) tend to have *higher* returns than those already in school even before the law change. Traditional ‘ability bias’ (which we confirm) may tend to arise only in developing (but not in developed) countries because parents are forced to keep only those (among many) of their children with higher ability in school, thereby reinforcing (rather than compensating) inequality among children within the household.

The rest of the paper is organized as follows. The next section provides a brief review of the relevant literature. Section 3 discusses the data, methodology, and the identification strategy employed in this study. The empirical results and the interpretations of our estimation of returns to schooling are provided in Section 4. Section 5 concludes.

2. Literature Review

In our attempts to estimate the returns to education in Thailand, we follow the now classic approach developed by Mincer (Mincer, 1958, 1974):

(1) $$\log y_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + e_i$$

where the log of individual earnings, y_i, is a linear function of the years of education an individual *i* has attained (S_i), and a quadratic function of the number of years the individual has worked after completing his/her education (X_i). e_i represents the disturbance term. The returns to education is measured by is the coefficient β_1.

The Mincer specification has been found to fit data reasonably well. Card (1999) shows, for example, that, based on the pooled samples of the 1994-1996 March Current

Population Surveys in the United States, the hourly wage-age profiles for men and women are reasonably well approximated by the Mincer specification and concludes that the Mincerian "human capital earnings function is alive and well." As we will see in the next section, the Mincer equation appears also to fit our Thai data reasonably well.

Despite simplicity of the Mincerian specification and the large number of empirical studies conducted in the past few decades, debates continue on the quantitative magnitudes of the causal effects of schooling on earnings in both developed and developing country contexts. Since the years of schooling is an endogenous variable determined by the choice made by the household (including both parents and children), the association of schooling with earnings does not necessarily represent causal effects but, instead, could also include effects of other factors such as ability of children, heterogeneity in family backgrounds and heterogeneity in school quality (e.g., Beherman 1999). In particular, a major focus in the empirical literature has been on the magnitude of the so called ability bias, the tendency for OLS estimates of the returns to schooling to be (presumably upwardly) biased due to the (presumably positive) correlation between schooling and student ability, since student ability is often unobserved and thus difficult to control for (e. g., Willis, 1986; Schultz, 1988; Card, 1999; Deere and Vesovic, 2006).

Interestingly, a number of recent studies, mostly from developed countries, find that ability bias may not be very serious.[1] In addition to the early skeptics on the ability bias

[1] Apart from the concern about ability bias, there has also been a parallel literature addressing other aspects of heterogeneity, such as family background and school quality. See Behrman (1999) for a review of the literature from developing countries. Behrman (1999) argues that the "standard estimates" that do not address those concerns tend to overestimate the impact of schooling attainment substantially,

concerns (e.g., Becker 1964, Griliches 1977), studies using the incidence of identical twins as instrumental variable tend to show similar results between OLS and IV estimation, suggesting that the magnitude of ability bias is likely to be small, if any (Card 1999). Furthermore, there have been an increasing number of empirical studies utilizing institutional features and changes in law as natural experiments. Such "revisionist" literature finds that OLS estimates of the returns to schooling tend to under-estimate, rather than over-estimate (as the traditional ability bias arguments would predict), the true returns quite substantially (e.g., see Card (1999) and Heckman et al (2006) for literature reviews). Card (1999) concluded that the magnitude of such underestimation could be as large as 20-40%.

The LATE (local average treatment effects) interpretations of these recent studies based on the IV estimates exploiting changes in compulsory schooling suggest, rather surprisingly, the possibility of negative, rather than positive, ability bias, where those children whose schooling behavior is affected by the change in compulsory schooling law (i.e., those school drop-outs before the law change) tend to have higher returns to schooling compared to those who were in school even before the law change. Since those empirical findings run counter to the conventional ability bias stories (i. e,. those with higher ability have higher returns to schooling, stay in school longer and earn better), a number of explanations have been proposed to account for such findings. Exogenous constraints, such as credit constraints, could be one possibility, although some question the universal applicability of this explanation across the variety of country contexts where similar

even as much as 40-100%.

findings were obtained (Oreopoulos 2006). Card (1999) develops a theoretical model of schooling choice which suggests that a *negative* correlation between the returns to schooling and the years of schooling can arise if ability differences are not “too important” in the determination of schooling outcomes and if the marginal return to schooling is decreasing. Others have argued that the attenuation bias in OLS estimation due to (classical) measurement errors in the schooling variable could account for parts of the difference between OLS and IV estimates, [2] or that the bias due to discount rate heterogeneity (i.e., those school drop-outs, before the compulsory law change, could include those with higher than average ability and with higher discount rates) could off-set the positive ability bias in OLS estimates (Card 1999, Lang 1993). Alternatively, Heckman et al (2006) interpret the empirical findings of the larger IV estimates than the OLS estimates as suggesting that ability space is multidimensional, rather than single dimensional, as typically assumed in the conventional literature. If there are different types of abilities and different levels of schooling are required for different types of jobs, individuals with different ability and skill mix sort across schooling levels in such a way that the best individuals in one schooling level do not do so well in another levels. With such possibilities, “the idea that individuals with ‘higher ability’ are more likely to enroll in school is no longer obvious. (Heckman et al 2006)” [3]

[2] Card (1999) notes, however, that, based on what is known in the literature on the effects of random measurement errors, the magnitude of the downward bias of OLS estimates found in some studies are too large to be explained by the attenuation bias arising from classical measurement errors.

[3] As an example, if individuals with more schooling become teachers and those with lower schooling become plumbers, “then the latter are better plumbers than the average teachers would be if he become a plumber.” (Heckman et al 2006)

In contrast with the continuing debates in the developed country contexts, similar empirical studies exploiting compulsory law change from developing countries are still relatively scarce. It is possible that some of those explanations for the empirical findings on the OLS and IV estimation results may not apply in developing country contexts; ability differences may be important in explaining schooling outcomes (Card 1999's explanation) or multi-dimensionality of ability or skill space may not be as important in developing counties as in developed countries (Heckman et al 2006' s explanation). It is this lacuna in the literature that this paper intends to address.

3. The Empirical Methodology and the Data

3.1 Methodology and Identification Strategy

This paper follows closely the regression discontinuity (RD) approach by Oreopolous (2006) and others that use the incidence of change in compulsory schooling law as an instrumental variable for estimating the returns to schooling. This paper applies the approach to the incidence of the 1978 Primary Education Act in Thailand. As can be seen from Figure 1 and Figure 2, the timing of the Act came immediately before the rapid growth in per capita GDP starting in the mid-1980s (and lasting until the 1997 Asia Crisis) and the industrialization episode of the 1980s.

The 1978 Primary Education Act was the first education act with which every individual was required to comply. At that time, the whole educational structure was changed from 4-3-3-2 to 6-3-3. Previously, the Thai education system had consisted of four years of lower primary education, three years of upper primary education, and five years of secondary education. The 1978 Primary Education Act reduced the total years of primary education to six years, without division between the lower and upper levels, while secondary education remained the same. In addition, the Government expanded compulsory education from four years to six years of primary education.

With the regular schooling progression (i.e., entering the primary school at age 6), the first birth cohort to be affected by the 1978 compulsory education change was the cohort born in 1968. However, it was relatively commonly observed at the time that some students started enrolling in primary schools as late as age 8. As a result, we consider the cohorts born in 1968 to be the first cohorts to be affected by the 1978 law (Table 1). The law took effect immediately on 5 April 1978, a month before the start of the 1978 academic year. Therefore, all students enrolled in grade four in 1977 were required to move up to grade five in 1978. The possible age range of the fourth year students is between 9 and 11 years old, which correspond to the 1966–1968 cohorts.

One additional complication for our current purposes is the fact that, during the process of compulsory education expansion, the Government provided a five-year adjustment period (1978–1982) to those districts that were not ready for the compulsory education reform. Since the majority of schools had either four years of lower primary education or three years of upper primary education prior to the 1978 reform, these schools found it difficult to comply with the new compulsory education law immediately, and thus needed more time to build more classrooms or combine other schools. The Government nevertheless declared that by 1982 every student and every school must comply with the 1978 Primary Education and Compulsory Education Act.

The 1978 compulsory law change in Thailand affected a large proportion of the population, covering almost half the population of the fourth grade of primary education to stay in school for two more years (until grade six, the final grade of primary education). As a result, similar to the application by Oreopoulos (2006), the estimated LATE in this paper could arguably be closer to the population ATE (average treatment effect) than that of similar studies that affect only relatively small (and arguably peculiar) fractions of the population.

3.2 Data

We use the Thai Labor Force Survey (LFS) conducted by the National Statistical Office (NSO) for the years 1986–2012. The LFS is collected quarterly on about 80,000 randomly selected households for a total of about 200,000 observations per quarter, representing 0.1–0.5 percent of the total Thai population.

The dataset used in this study is constructed by pooling the 27 consecutive annual LFSs. Only the data from the third quarter of the LFS is used in this study to control for the seasonal migration of agricultural labor. In general, agricultural workers move back and forth between the urban manufacturing sector and the rural agricultural sector. Nevertheless, they tend to migrate back to the rural agricultural sector during the rainy season in the third quarter of the year (Sussangkarn and Chalamwong, 1996). Moreover, this study limits the sample to 157,390 wage workers aged 15–60 in the year of interview. The age restriction of 15–60 years is imposed because of the fact that 15 years old is the minimum legal age that individuals can start working and 60 years old is usually the retirement age in Thailand. In addition, we also employ a birth cohort restriction in this study. The analysis is limited to individuals born between 1955 and 1985 since these cohorts are the observations around the cut-off for the regression discontinuity estimation. The set of variables in this study covers age, birth cohort, years of schooling, region of residence, area of residence, industrial sector, and estimated monthly wages.

3.3 Econometric Specification

As discussed in the previous section, we construct a cohort panel using the 1986 to 2012 rounds of the Thai Labor Force Surveys, and compare the schooling and earnings outcomes between those cohorts who were covered by the 1978 compulsory schooling law and those preceding cohorts who were not affected, by introducing into the conventional Mincer equation the dummy variables indicating the coverage by the law as the IVs to control for the endogenous years of schooling.

The regression equation of our main interest follows closely Oreoupolos (2006) and takes the form:

$$(2)\ \log y_i = \gamma_0 + \gamma_1 \widehat{S_i} + \gamma_2 C_i^1 + \gamma_3 C_i^2 + \gamma_4 C_i^3 + \gamma_5 C_i^4 + \sum_{\mathrm{k}=16}^{60} \gamma_{6\mathrm{k}} \mathrm{A}_{\mathrm{ki}} + \sum_{\mathrm{l}=1}^{4} \gamma_{7\mathrm{l}} \mathrm{R}_{\mathrm{li}} + \vartheta_i$$

with its first stage equation:

$$(3)\quad S_i = \pi_0 + \pi_1 F_i + \pi_2 C_i^1 + \pi_3 C_i^2 + \pi_4 C_i^3 + \pi_5 C_i^4 + \sum_{\mathrm{k}=16}^{60} \pi_{6\mathrm{k}} \mathrm{A}_{\mathrm{ki}} + \sum_{\mathrm{l}=1}^{4} \pi_{7\mathrm{l}} \mathrm{R}_{\mathrm{li}} + \varepsilon_i$$

where log y_i is the log of the monthly wages of individual i; S_i is to the endogenous years of education of individual i; $\widehat{S_i}$ is the fitted value estimated from the first stage least squares regression; F_i represents a dummy variable, to be used as the instrumental variable, indicating whether an individual must comply with the 1978 compulsory education law (i.e., those individuals who were born in 1966-68 and had not left school at the time of the law change as well as all the subsequent birth-year cohorts, born in 1969 onwards). The control variables include a set of age (as a proxy for working experience) dummies, A_{ki}, ; quartic terms of birth cohort (C_i); and regional dummies (R_{li}). ϑ_i and ε_i (as well as θ_i, and e_i below) represent disturbance terms.

We also estimate the reduced form version:

$$(4) \log y_i = \alpha_0 + \alpha_1 F_i + \alpha_2 C_i + \alpha_3 C_i^2 + \alpha_4 C_i^3 + \alpha_5 C_i^4 + \sum_{k=16}^{60} \alpha_{6k} A_{ki} + \sum_{l=1}^{4} \alpha_{7l} R_{li} + \theta_i$$

The IV estimation results from Equation (2) will be compared with the results based on the OLS regression:

$$(5) \log y_i = \beta_0 + \beta_1 S_i + \beta_2 C_i^1 + \beta_3 C_i^2 + \beta_4 C_i^3 + \beta_5 C_i^4 + \sum_{k=16}^{60} \beta_{6k} A_{ki} + \sum_{l=1}^{4} \beta_{7l} R_{li} + e_i$$

4. Empirical Results

4.1 Some preliminary analysis of the data

Following Card (1999) we first examine the age-wage profiles based on our Thai data, which leads us to a similar conclusion as Card (1999)'s in that the estimates from the Mincer model reasonably well approximate the actual age-wage profiles. Figure 3 shows the age profiles of monthly wages by education level for Thai workers using pooled samples from the 1986–2012 Labor Force Surveys (LFS). The data represents the mean log monthly wage by age for individuals with 4, 6, 9, 12, 16, and 21 years of education. Four years of education refers to the minimum years of education required by the 1962 Compulsory Education Act, whereas six years of education represents the minimum compulsory education level enforced by the 1978 compulsory education law. The other education years refer to the final year of each academic level: the lower secondary level, upper secondary level, undergraduate level, and graduate level. Plotted lines along with the actual means are the fitted values obtained from the Mincer model, which includes only a quadratic term of age. Comparisons of the fitted and actual data suggest that age-earnings profiles for Thai workers are fairly smooth and reasonably well approximated. In contrast to the age-wage profiles from the US, the problem in fitting the precise curvature appears less pronounced in the case of Thailand.

Figure 4 is a graphical representation of the effect of the change in compulsory education law on the fraction of individuals graduating with at most four years of education. The lower line in Figure 4 shows the proportion of adults aged 15–60 who reported their highest attained level of education was at most four years, while the upper line shows the proportion of adults aged 15–60 who report their highest attained level of education was at most six years. Both lines exhibit downward trends. However, by comparing between the two, there appears to be a kink occurring between the 1965 and 1966 cohorts along the lower line (i.e., the fraction graduating at most grade 4) while such kink is not observed along the upper line (i.e., the fraction graduating at most grade 6). The drop in the proportion of individuals with at most four years of education does not quite appear as a vertical shift, however, possibly due to the five-year adjustment period between 1978 and 1983. The dashed line suggests that had there not been the five-year adjustment period, we could have seen a sharp drop in the share of those that graduated with at most four years of education after the 1965 cohort, from 40 percent to less than 10 percent, approximating the magnitude of the discontinuity.

4.2 First Stage Regression Results

Table 2 presents the first stage regression results showing the effects of the change in the compulsory schooling law on educational attainment (equation (3)). Each regression includes as regressors birth cohort quartic polynomial, regional dummies, and a dummy variable indicating whether a cohort faced a new compulsory education law. Columns (3)–(5) also include age controls: a quadratic polynomial or age dummies, where indicated. Each regression includes the sample of those aged 15–60 years old from the 1986–2012 LFSs. Following Oreopoulos (2006), individual-level observations are first aggregated into cell group means by cohort, age, sex, survey year, regions, and industrial sectors of employment, and weighted by cell size. Regressions are clustered by birth cohort, regions, and industrial sectors of employment. The total number of cells is 157,087.

As shown in Table 2, the coefficients on the compulsory education dummy are statistically significant and robust across different specifications. The compulsory education law change, extending the minimum year of schooling from four years to six years, led to roughly 4 additional years of schooling, corresponding to roughly twice the additional schooling required by the law. Some existing studies from developed countries have also found that the impact of compulsory schooling law change went beyond the additional years of schooling imposed by the law change (e. g., Oreoupolos 2003). While the quantitative magnitudes of the impact of compulsory schoolings we see in Thailand appear substantially larger than what was found in the existing literature mostly coming from developed countries (typically ranging between 0.1 to 0.5 year of additional schooling) and a few from China (ranging between 0.8-1.2 year of additional schooling), we should note that direct comparisons among such studies from different countries may not necessarily be warranted, since compulsory education is imposed somewhat differently (e.g., the minimum school leaving age or specified minimum compulsory level of education) in different countries found in the previous studies.

4.3 Reduced Form

Table 3 shows the reduced form equation estimation of the effects of the compulsory education law change on monthly wages. We find that the change in compulsory schooling law led to an approximately 30 percent increase in monthly wages. The relatively large reduced form effects are consistent with the relatively large effects on the years of schooling resulting from the change in compulsory schooling found in the first stage regression results. They appear to be substantially larger compared to the effects of compulsory education law on wages and earnings, especially those in developed countries, found in the existing literature with the major exception of Harmon and Walker (1995) showing similar effects of the change in compulsory education on earnings in the UK.

4.4 OLS and IV Estimation

The OLS estimation results on the returns to schooling are shown in Table 4. Each regression equation includes controls for the birth cohort, regional dummies, and age. The rates of return to schooling based on the OLS estimation are approximately 11 percent. Our OLS estimates from Thailand are somewhat higher than those OLS estimates obtained in developed countries in the literature, which range roughly between 8 to 10%. Our OLS estimation results of 11 percent are the same as those obtained by Warunsiri and McNown (2010) for Thailand.

The IV estimates of returns to schooling are shown in Table 5. We find that one additional year of schooling is associated with an approximately 8 percent increase in monthly wages, which is somewhat lower than the OLS estimates. The IV estimates from this study are less than, but somewhat consistent with, those of Canada and the UK, which are approximately 10 percent. While the OLS estimates of the returns to schooling are the same between Warunsiri and McNown (2010) and our studies, our IV estimates are lower than the estimates obtained by the cohort panel analysis (using cohort fixed effects) of 14 percent by Warunsiri and McNown (2010), which is similar to estimates obtained from the US.[4]

It is quite intriguing to find that our IV estimates are lower than OLS estimates by the order of 20 percent. There are a few other studies from developing countries that also suggest OLS may be overestimating the returns to schooling (e.g., China, Turkey), consistent with Behrman's (1999) view. In contrast, the relatively recent studies based on compulsory schooling law (mostly from developed countries) find that their IV estimates of the returns to schooling are substantially higher than OLS estimates. A number of explanations have been raised in the literature explaining the latter set of results. Among them, however, Card (1999) and Heckman et al (2006) stand out in that they develop

[4] Warunsiri and McNown (2010) conclude that the estimates from IV and panel fixed effects are similar in magnitude; therefore, the problem of endogeneity bias is fixed in their estimations. Different from other studies, they use the existence of a university or a teacher training college within a province as the instrument in the IV estimation. Meghir and Rivkin (2011) argues that such IV adopted by Warunsiri and McNown (2010) may be correlated with individual ability due to the non-random nature of individuals and school allocation.

formal (though relatively simple) models to explore how such outcome can arise. Card (1999) argues that a negative correlation between schooling and the returns to schooling (and thus lower OLS estimates than IV estimates) could arise if ability differences are "not too important" in the determination of the years of schooling[5]. Such a story appears to be plausible in explaining why the positive ability bias is absent in developed counties while it could be relatively more important in developing countries. Unless resource (financial) constraints are severe, parents could make every effort to educate their children regardless of their ability in developed societies. Since fertility tends to be low in those countries, parents are unlikely to be under pressure to select only better ability children to be sent school. In contrast, in developing countries where many households are resource constrained and the number of children tends to be large, parents may not be able to keep all of their (numerous) children in school up to their perceived optimum. Hard pressed to keep a subset of their children in school while putting the rest to work, parents may try to keep only those children who have relatively better ability in school[6]. As a country grows richer, however, severe resource constraints are likely to be gradually lifted, and fertility also likely to declines at the same time. As a result, parents are less likely to have to make such choices and more likely to keep all of their (smaller number of) children in school

[5] Card (1999) also argues that, while OLS attenuation bias due to classical measurement errors could partially account for the pattern, the quantitative magnitudes of the (negative) OLS bias is not likely to be measurement errors alone.

[6] Theoretically, it is not obvious that parents in poor households invest more in human capital of better endowed (higher ability) children, thereby enhancing, rather than compensating, inequality among children in endowments (Becker 1991). Beherman, Pollak and Taubman (1982) show that whether parents compensate or enhance inter-sibling inequality depends on parental preferences (utility function) over their relative priority on ensuring equity among their children. Our empirical findings appear to be consistent with the possibility that parental preferences toward equity among children are not strong.

regardless of ability heterogeneity. The relative importance of ability heterogeneity in determining schooling outcomes could decline, consistent with Card (1999).

Heckman et al (2006) propose an alternative explanation for the absence of (positive) ability bias. They argue that ability is multidimensional, where different types of ability or skills and different levels of schooling are required by different types of jobs or in different industries. According to this view, "individuals sort themselves across schooling levels in such a way that the best individuals in one schooling level are the worst in the other, and vice versa" (Heckman et al 2006; 374). In relatively industrialized and diversified economies, such a story would be quite plausible. In less diversified and predominantly low-skilled economies, however, such possibilities may be arguably less plausible. Based on the single dimensional skill/ability space view, on the other hand, the conventional positive ability bias in the determination of schooling could become quite important.

Thus based on both Card (1999) and Heckman et al (2006)'s views of why positive ability bias may not be important in developed countries, the role of conventional positive ability bias in OLS estimates of the returns to schooling can become relatively more important in developing country contexts, which is consistent with our empirical results.

4.5 Disaggregated Analysis of Returns to Schooling

In addition to the overall estimates of returns to schooling, another important issue is potential heterogeneity in educational returns across individuals. This section presents initial results from our exploratory analysis of the returns to schooling disaggregated by demographic and geographic aspects, including gender, birth cohort, area of residence (urban and rural areas), region and industrial sector of employment.[7]

We find that returns to schooling (based on the IV estimation) are similar between women and men, both at roughly 8 percent (8.3% for female and 7.9% for male). This appears in line with the conventional view that gender disparity is much less pronounces in Southeast Asia, compared to, say, South or East Asia (e.g., Atkinson and Errington 1990). We have also disaggregated the sample between (relatively) earlier (1955-70) versus more recent (1961-1985) birth cohort; the returns are slightly higher for the earlier, rather than later, birth cohort (8.6 percent vs. 8.2 percent).[8] This is rather surprising since during the rapid structural transformation and economic development, returns to schooling could be increasing over time, and thus higher among later cohorts than among earlier ones. On the other hand, not surprisingly, we find that the returns to schooling in urban areas are substantially higher (8.3 percent) than in rural areas (6.8 percent). While the difference between the urban and rural returns appears to be sizable, it is not immediately clear, however, to what extent this magnitude (or, even the sign?) could be causal, due to selective rural-to-urban migration. In addition, consistent with the urban-rural gap in the returns to schooling is the finding that returns are higher in non-agricultural sectors than in

[7] Simply splitting the sample into geographical or/and demographic subsamples could lead to the classic problems of self-selection. Schultz (1988) argues, for example, that stratifications of samples based on heterogeneous demographic characteristics lead to selection bias in estimating returns to education, especially in case of developing countries, due to the prevalence of self-employment and non-wage labor, imbalanced economic development among different areas and regions, and gender-based segregation in occupational choices. Dahl (2002) is an example of directly addressing the self-selection problem due to migration, albeit in the context of the US, and finding the magnitudes of bias being quite modest. We intend to explore those aspects in our future work.

[8] Two cohorts compared here are overlapping because we need to include the birth year cohort in the neighborhood of the compulsory law change (i.e., 1966 cohort) in order to apply our IV estimation.

the agricultural sector. The estimated return to schooling are 5.8 percent, 6.8 percent and 8.1 percent in agricultural, industrial and service sectors, respectively. The relatively lower returns in the agricultural sector seem consistent with the conventional views on the process of economic development, while the higher returns to services sector than in the industrial sector is somewhat puzzling in light of the rapid industrialization in the 1980s (Figure 2). Like the case of the urban-rural gap, those estimates based on sectoral disaggregation are potentially subject to self-selection bias, however, and thus causal inferences may not be warranted. We also attempted disaggregation across geographical regions (i.e., Bangkok, North, Northeast, South and Center). Rather surprisingly, again, the returns to schooling appear to be relatively low in the Bangkok area at 7.4 percent (second lowest only next to South) and relatively high in the relatively underdeveloped regions of North (9.3 percent) and Northeast (9.7 percent). This pattern is quite puzzling. There would be no question that relatively higher return opportunities were located in the Bangkok area than in the northern area. Furthermore, while the self-selection problem is a potential issue, the conventional wisdom would predict the opposite direction in bias, where those with higher returns would migrate to the Bangkok area. While there may be a possibility that a rapid increase in the labor supply in Bangkok due to rural-to-urban migration could have depressed the wages (and thus the returns to schooling) during the period under analysis, the extent of the magnitude would still appear to be quite large.

According to Table 6, the overall results are similar to the main result reported earlier in the sense that OLS estimates of the returns to schooling are higher than the IV estimates. There appear to be some variations in terms of the magnitude of the difference between OLS and IV estimates, which could suggest the severity of the endogentity biases. The differences appear relatively more pronounced for the older age cohort (12.5 percent, based on OLS, and 8.6 percent based on IV), rather than younger cohort, for rural areas (10.4 percent, based on OLS, and 6.8 percent, based on IV), for North region (14.0 percent, based on OLS, and 9.3 percent, based on IV) and for the agricultural sector (10.0 percent, based on OLS, and 5.8 percent, based on IV). Consistent with our interpretation of the main results, there appears to be a tendency that the extent of the ability bias in OLS estimates are larger in geographical areas or the economic sector where agents (households) are relatively more resource constrained.

Conclusion

This paper estimates returns to schooling in Thailand using a regression discontinuity approach applied to the change in compulsory schooling law in 1978. We find that the compulsory schooling law played a role in enhancing human capital investment in the eve of the rapid structural transformation in the 1980s, that the returns to schooling based on our IV estimation was round 8%, while OLS somewhat overestimates (by 20%) such returns, and that returns were higher in urban areas, in services (than in agricultural) sectors and, surprisingly, in the relatively underdeveloped Northern regions. Our findings are in sharp contrast with most of the recent studies exploiting similar institutional changes from developed countries, where OLS estimates tend to *under*-estimate returns to schooling with the implication that those school *drop-outs* (whose behavior is altered by compulsory schooling) tend to have *higher* returns than those already in school even before the law change. The conventional notion of ‘ability bias’ (which we confirm) are more likely to arise in developing (but not so much in developed) countries because parents could be forced to keep only those (among many) of their children with higher ability in school, thereby reinforcing (rather than compensating) inequality among children within the household.

References

Atkinson, Jane Monnig, and Errington, Shelly.(eds.) (1990). *Power and Difference: Gender in Island Southeast Asia*. Stanford, CA: Stanford University Press.

Aydemir, A. & Murat, K. (2015). Low wage returns to schooling in developing country: evidence from a major policy reform in Turkey. (IZA Discussion Paper No. 9274). Bonn: Institute for the Study of Labour.

Angrist, J.D. & Krueger, A.B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics, 106*(5), 979-1014.

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

Becker, G. (1991). *A Treatise on the Family.* Cambridge: Harvard University Press.

Behrman, J.R. (1990). The action of human resources and poverty on one another: What we have yet to learn. (Living Standards Measurement Study Working Paper No. 74). Washington D.C.: World Bank.

Behrman, J.R. (1999). Labor markets in developing countries. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics: vol. 3B* (pp. 2859–2939). Amsterdam: Elsevier.

Behrman, J.R. & Birdsall, N. (1983). The quality of schooling: Quantity alone is misleading. *American Economic Review,* 73(5), 928-946.

Behrman, J.R., Pollak, R., & Taubman, P. (1982). Parental Preferences and Provision for Progeny. *Journal of Political Economy*, 90(1), 52-73.

Behrman, J.R. & Srinivasan, T.N. (1995). Part 9: Policy reform, stabilization, structural adjustment and growth. In J.R. Behrman & T.N. Srinivasan (eds.), *Handbook of development economics: vol.* 3B (pp. 2467–2496). North-Holland: Elsevier.

Bound, J. & Jaeger, D. (1996). On the validity of season of birth as an instrument in wage equations: A comment on Angrist and Krueger's 'Does compulsory school attendance affect schooling and earnings? (NBER Working Paper No. 5835). Cambridge, MA: National Bureau of Economic Research.

Card, D. (1994). Earnings, schooling, and ability revisited. (NBER Working Paper No. 4832). Cambridge, MA: National Bureau of Economic Research.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics: vol. 3A* (pp. 1801-1863). Amsterdam: Elsevier.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica,* 69(5), 1127– 1160.

Dahl, G. (2002) Mobility and the Returns to Education: Testing a Roy Model with Multiple Markets. *Econometrica*, 70 (6), 2367-2420.

Deere, D.R. & Vesovic, J. (2006). Chapter 6 educational wage premiums and the U.S. income distribution: A survey. In E.A. Hanushek & F. Welch (Eds.), *Handbook of the economics of education: vol. 1* (pp. 255–306). Amsterdam: Elsevier.

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.

Fang, H., Eggleston, K.N., Rizzo, J.A., Rozelle, S., & Zeckhauser R.J. (2012). The returns to education in China: Evidence from the 1986 compulsory education law. (NBER Working Paper No. 18189). Cambridge, MA: National Bureau of Economic Research.

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, 1-22.

Harmon, C. & Walker, I. (1995). Estimates of the economic return to schooling for the United Kingdom. *American Economic Review*. 85(5), 1278–86.

Heckman, J.J., Lochner, L.J., & Todd, P.E. (2006). chapter 7 earnings functions, rates of return and treatment effects: The Mincer equation and beyond. In E.A. Hanushek & F. Welch (Eds.), *Handbook of the economics of education: vol. 1* (pp. 307-458). Amsterdam: Elsevier.

Krueger, A. & Lindahl, M. (2001). Education and growth: Why and for whom? *Journal of Economic Literature*. 39, 1101-1136.

La, V. 2014. Does schooling pay? Evidence from China. (MPRA Working Paper No. 54578). Munich: Munich Personal RePEc Archive.

Lang, K. (1993). Ability bias, discount rate bias, and the return to education. *Unpublished Discussion Paper*. Boston: Boston University Department of Economics.

Lochner, L. & Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review*. 94(1), 155-189.

Meghir, C. & Rivkin, S. (2011). Chapter 1 – econometric methods for research in education. In E.A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education: vol. 3* (pp. 1–87). Amsterdam: Elsevier.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*. 66(4), 281-302.

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

Orazem, P.F. & King, E.M. (2008). Chapter 55 Schooling in developing countries: The roles of supply, demand and government policy. In T.P. Schultz & J. Strauss (Eds.), *Handbook of development economics, volume 4*. (pp. 3475-3559). North-Holland: Elsevier.

Oreopoulos, P. (2003). Do dropouts drop out too soon? International evidence from changes in school-leaving laws. (NBER Working Paper No. 10155). Cambridge, MA: National Bureau of Economic Research.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*. 96(1), 152–75.

Oreopoulos, P. (2008). Estimating average and local average treatment effects of education when compulsory schooling laws really matter: Corrigendum. https://www.aeaweb.org/aer/contents/corrigenda/corr_aer.96.1.152.pdf

Patrinos, H. & Psacharopoulos, G. (2011). Education: Past, present and future global challenges. (The World Bank Policy Research Working Paper No. 5616), Washington D.C.: World Bank.

Pons, E., & Gonzalo, M.T. (2002). Returns to schooling in Spain: How reliable are instrumental variable estimates? *Labour*. 16(4), 747-770.

Psacharopoulos, G. & Patrinos, H. (2002). Returns to investment in education: A further update. (The World Bank Policy Research Working Paper No. 2881). Washington D.C.: World Bank.

Schultz, T.P. (1988). Education investment and returns. In H. Chenery & T.N. Srinivasan (Eds.), *Handbook in development economics: vol. 1* (pp. 543-630). North-Holland: Elsevier.

Strauss, J. & Thomas, D. (1995). Human resources: Empirical modeling of household and family decisions. In J.R. Behrman & T.N. Srinivasan (eds.), *Handbook of development economics: vol. 3A* (pp. 1883–2023). North-Holland: Elsevier.

Stephens Jr., M. & Yang, D.Y. (2014). Compulsory education and the benefits of schooling.

American Economic Review. 104(6), 1777–1792.

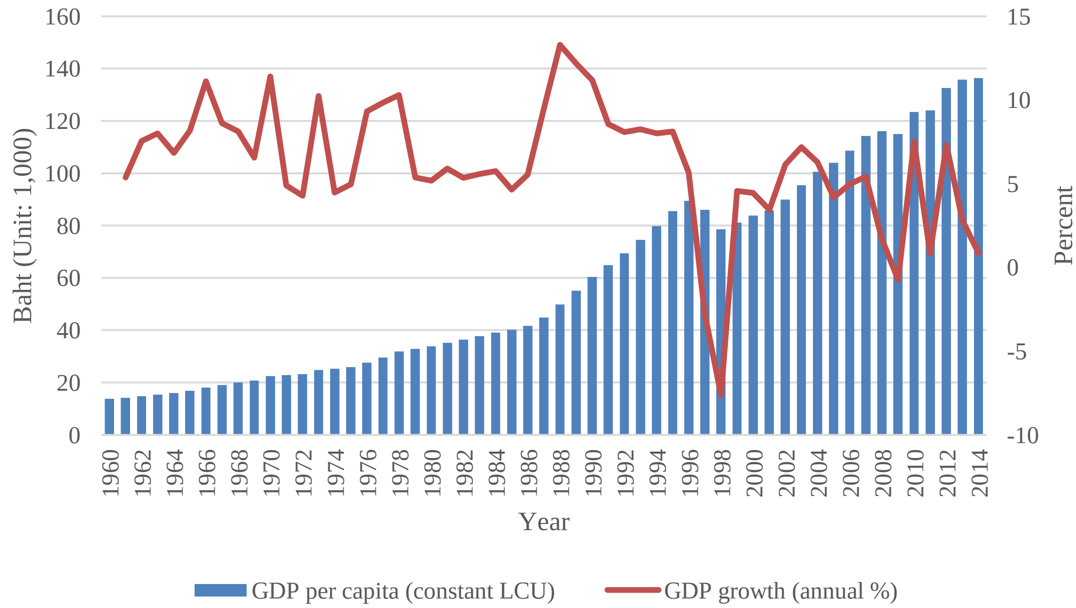
Sussangkarn, C. & Chalamwong, Y. (1996). Thailand development strategies and their impacts on labour markets and migration. In D. O'Connor & L. Farsakh (Eds.), *Development strategy, employment, and migration*, Paris: OECD.

Warunsiri, S. & McNown, R. (2010). The returns to education in Thailand: A pseudo-panel approach. *World Development*. 38(11), 1616–1625.

Willis, R.J. (1986). Human capital and the rise and fall of families: Comment. *Journal of Labor Economics*. 4(3), S40-47.

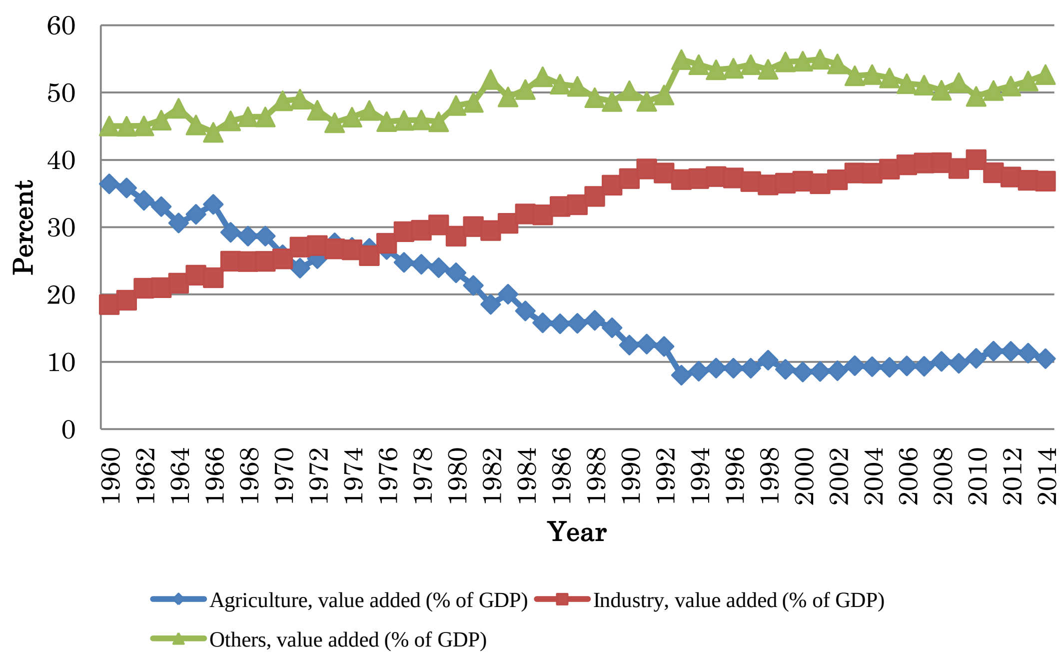
World Bank. (1989). *Educational development in Thailand: The role of World Bank lending*. Washington D.C.: World Bank.

Figure 1 Real GDP Growth and Per Capita GDP in Thailand, 1960-2014



Note: 1 USD = 35 Baht (as of 2016).
Source: Author's compilation based on data from World Development Indicators, 2016.

Figure 2 Structural Transformation: Net Output as % of GDP (%), 1960 - 2014

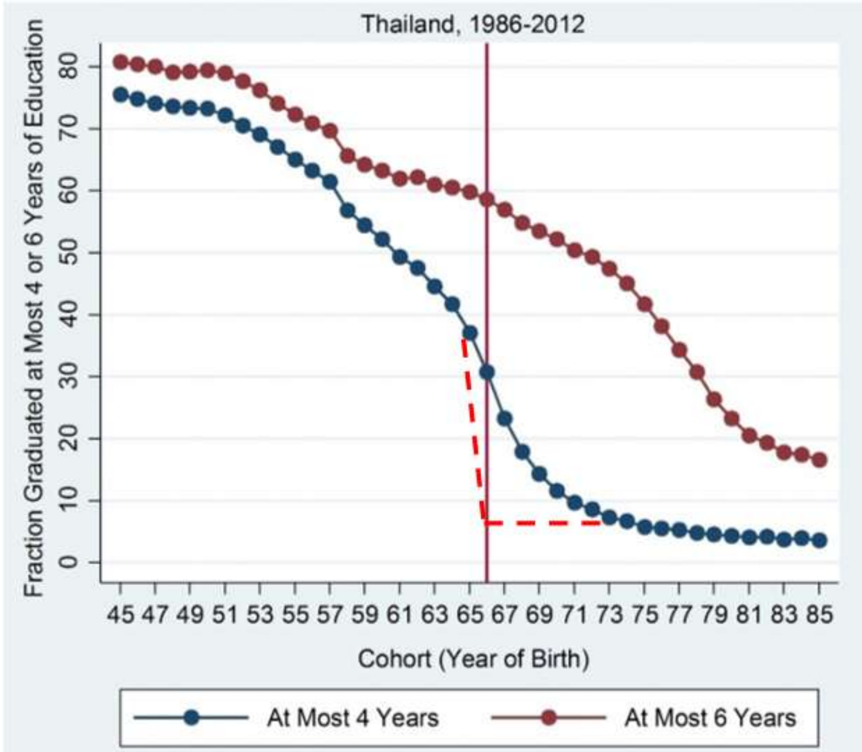


Source: Author's compilation based on data from World Development Indicators, 2016.

Figure 3 Age Profiles of Monthly Wages, Thailand

Note: The regression models include a linear education term and a quadratic in age.
Source: Author's compilation based on LFS data, 1986-2012

Figure 4 Fraction Graduating at Most Four and Six Years of Education, 1986-2012



Note: The lower line shows the proportion of adults aged 15 to 60 from 1986 to 2012 LFSs who report the highest attained level of education is at most four years. The upper line shows the proportion of adults aged 15 to 60 who report the highest attained level of education is at most six years. The vertical line indicates the first cohort affected by the 1978 compulsory education law. The dash line indicates that, in case there was no five-year-adjustment period, it is expected to see a sharp drop of the fraction graduated at most four years of education from 40 percent to less than 10 percent. Hence, the vertical dash line represents the discontinuity.
Source: Author's compilation based on LFS data 1986-2012.

Table 1 The Identification of the First Cohorts Affected by the 1978 Compulsory Education

Cohort	Year													
	1965	**1966**	**1967**	**1968**	**1969**	**1970**	**1971**	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**	**1978**
School Grade										1	2	3	4	5
1969					0	1	2	3	4	5	6	7	8	9
1968				0	1	2	3	4	5	6	7	8	9	10
1967			0	1	2	3	4	5	6	7	8	9	10	11
1966		0	1	2	3	4	5	6	7	8	9	10	11	12
1965	0	1	2	3	4	5	6	7	8	9	10	11	12	13

Source: Author's compilation.

Table 2 First Stage Regression Results: Estimated Effect of Compulsory Education Law on Education Attainment

Chap	(1)	(2)	(3)	(4)	(5)
			First Stage		
		Dependent Variable: Number of Years of Schooling			
Compulsory Education	4.356***	4.294***	4.259***	4.270***	4.046***
	(0.392)	(0.391)	(0.365)	(0.364)	(0.313)
Fixed Effects:					
Regional Controls	No	Yes	Yes	Yes	Yes
Birth Cohort	Quartic	Quartic	Quartic	Quartic	Quartic
Additional Controls:	None	None	Age Dummy	Age Dummy	Age Dummy
				Gender	Gender
					Urban
Initial sample size	1,308,519	1,308,519	1,308,519	1,308,519	1,308,519
R-squared	0.091	0.104	0.128	0.129	0.184

Note: The dependent variables are number of years of schooling. Each regression includes controls for a birth cohort quartic polynomial, regional dummies (except for the models with explicit region variables), and an indicator whether a cohort faced a new compulsory education law (six years of compulsory education). Column (3) to (5) also include age dummy variables. Each regression includes the sample of 15 to 60 years old from the 1986 through 2012 LFSs. Data are first aggregated into cell means and weighted by cell size. Regressions are clustered by birth cohort, regions, and industrial sectors of employment. Robust standard errors in parentheses. ***, **, and * indicate $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively.
Source: Author's compilation based on LFS 1986-2012.

Table 3 Reduced Form Equation Results: Estimated Effect of Compulsory Education Law on Log Monthly Wages

	(1)	(2)	(3)	(4)	(5)
	Reduced Form Dependent Variable: Log Monthly Wages				
Compulsory Education	0.354***	0.343***	0.355***	0.348***	0.310***
	(0.0590)	(0.0559)	(0.0585)	(0.0592)	(0.0497)
Fixed Effects:					
Regional Controls	No	Yes	Yes	Yes	Yes
Birth Cohort	Quartic	Quartic	Quartic	Quartic	Quartic
Additional Controls:	None	None	Age Dummy	Age Dummy Gender	Age Dummy Gender Urban
Initial sample size	1,308,519	1,308,519	1,308,519	1,308,519	1,308,519
R-squared	0.017	0.082	0.126	0.134	0.200

Note: The dependent variables are log monthly wages. Each regression includes controls for a birth cohort quartic polynomial, regional dummies (except for the models with explicit region variables), and an indicator whether a cohort faced a new compulsory education law (six years of compulsory education). Column (3) to (5) also include age dummy variables. Each regression includes the sample of 15 to 60 years old from the 1986 through 2012 LFSs. Data are first aggregated into cell means and weighted by cell size. Regressions are clustered by birth cohort, regions, and industrial sectors of employment. Robust standard errors in parentheses. ***, **, and * indicate $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively.
Source: Author's compilation based on LFS 1986-2012.

Table 4 Returns to Schooling Estimates for Log Monthly Wages (OLS)

	(1)	(2)	(3)	(4)	(5)
			OLS		
			Dependent Variable: Log Monthly Wages		
Year of Schoolings	0.113***	0.111***	0.112***	0.112***	0.109***
	(0.00184)	(0.00172)	(0.00186)	(0.00182)	(0.00165)
Fixed Effects:					
Regional Controls	No	Yes	Yes	Yes	Yes
Birth Cohort	Quartic	Quartic	Quartic	Quartic	Quartic
Additional Controls:	None	None	Age Dummy	Age Dummy Gender	Age Dummy Gender Urban
Initial sample size	1,308,519	1,308,519	1,308,519	1,308,519	1,308,519
R-squared	0.527	0.567	0.603	0.614	0.621

Note: The dependent variables are log monthly wages. Each regression includes controls for a birth cohort quartic polynomial, regional dummies (except for the models with explicit region variables), and an indicator whether a cohort faced a new compulsory education law (six years of compulsory education). Column (3) to (5) also include age dummy variables. Each regression includes the sample of 15 to 60 years old from the 1986 through 2012 LFSs. Data are first aggregated into cell means and weighted by cell size. Regressions are clustered by birth cohort, regions, and industrial sectors of employment. Robust standard errors in parentheses. ***, **, and * indicate $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively.

Source: Author's compilation based on LFS 1986-2012.

Table 5 IV Returns to Schooling Estimates for Log Monthly Wages (IV Estimation)

	(1)	(2)	(3)	(4)	(5)
			IV		
	Dependent Variable: Log Monthly Wages				
Years of Schooling	0.0818***	0.0799***	0.0832***	0.0807***	0.0767***
	(0.00772)	(0.00680)	(0.00767)	(0.00790)	(0.00751)
Fixed Effects:					
Regional Controls	No	Yes	Yes	Yes	Yes
Birth Cohort	Quartic	Quartic	Quartic	Quartic	Quartic
Additional Controls:	None	None	Age Dummy	Age Dummy Gender	Age Dummy Gender Urban
Initial sample size	1,308,519	1,308,519	1,308,519	1,308,519	1,308,519
R-squared	0.487	0.528	0.571	0.575	0.584

Note: The dependent variables are log monthly wages. Each regression includes controls for a birth cohort quartic polynomial, regional dummies (except for the models with explicit region variables), and an indicator whether a cohort faced a new compulsory education law (six years of compulsory education). Column (3) to (5) also include age dummy variables. Each regression includes the sample of 15 to 60 years old from the 1986 through 2012 LFSs. Data are first aggregated into cell means and weighted by cell size. Regressions are clustered by birth cohort, regions, and industrial sectors of employment. Robust standard errors in parentheses. ***, **, and * indicate $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively.

Source: Author's compilation based on LFS 1986-2012.

Table 6 Disaggregated Analysis of OLS and IV Returns to Schooling

Dependent Variables	OLS	IV	Bias Gap	Sample Size	Comparison: Returns to Schooling	Comparison: Bias Gap
Log monthly wages, all workers	0.112***	0.0832***	0.0288	1,308,519		
	(0.00186)	(0.00767)				
Log monthly wages, male	0.108***	0.0790***	0.029	663,501	Female > Male	Female > Male
	(0.00185)	(0.00932)				
Log monthly wages, female	0.116***	0.0831***	0.0329	645,018		
	(0.00190)	(0.00718)				
Log monthly wages, cohort 1955-1970	0.125***	0.0860***	0.039	813,981	Old > Young	Old > Young
	(0.00205)	(0.00549)				
Log monthly wages, cohort 1961-1985	0.101***	0.0816***	0.0194	1,017,586		
	(0.00183)	(0.00623)				
Log monthly wages, urban	0.108***	0.0834***	0.0246	857,828	Urban > Rural	Rural > Urban
	(0.00159)	(0.00546)				
Log monthly wages, rural	0.104***	0.0680***	0.036	450,691		
	(0.00225)	(0.00907)				
Log monthly wages, BKK	0.0953***	0.0737***	0.0216	162,399	Northeast, North > Others	Northeast, North > Others
	(0.00145)	(0.00391)				
Log monthly wages, North	0.124***	0.0965***	0.0275	256,447		
	(0.00309)	(0.0114)				
Log monthly wages, Northeast	0.140***	0.0925***	0.0475	298,457		
	(0.00379)	(0.0293)				
Log monthly wages, South	0.0926***	0.0709***	0.0217	222,181		
	(0.00309)	(0.00959)				
Log monthly wages, Centre	0.0972***	0.0748***	0.0224	369,035		
	(0.00295)	(0.00708)				
Log monthly wages, Agricultural sector	0.100***	0.0583***	0.0417	428,987	Service > Manufacture > Agriculture	Agriculture > Manufacture > Service
	(0.00299)	(0.00536)				
Log monthly wages, Manufacturing sector	0.0936***	0.0682***	0.0254	238,514		
	(0.00252)	(0.00307)				
Log monthly wages, Service sector	0.102***	0.0812***	0.0208	638,080		
	(0.00184)	(0.00225)				

Notes: The dependent variables are log monthly wages. Each regression includes controls for a birth cohort dummies (except for the models with explicit cohort variables), regional dummies (except for the models with explicit region variables), and an indicator whether a cohort faced a new compulsory education law (6 years of compulsory education). Moreover, each model also includes age dummy variables. Each regression includes the sample of 15 to 60 years old from the 1986 through 2012 Labour Force Surveys. Data are first aggregated into cell means and weighted by cell size. Regressions are clustered by birth cohort, regions, and industrial sectors of employment. ***, **, and * indicate $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively. Bias gap refers to the difference between OLS estimate and the IV estimate.

Source: Author's compilation based on LFS 1986-2012.