A Study on Regional Return to Education in South Korea: Comparison of Male and Female Wages

JongSoo Lee[a1] and Bit Na Choi[b]

[a]The Department of Economics, The University of Kansa, *Lawrence, Kansas;*

[b]The Department of Mathematics and Statistics, The University of New Hampshire, *Durham, New Hampshire*

abstract

This study examined the return to education in South Korea between metropolitan areas and non-metropolitan areas. In doing so, we used the Korean Labor and Income Panel Study 2018 and 2019 using the Mincer equation with fixed effects. We find that female workers' return to education was higher than male workers. In metropolitan areas, male workers' return to education was 5.4%, and female workers' return to education was 8.1%, while, in non-metropolitan areas, male workers' return to education was 5.4%, and female workers' return to education was 9.0%. Using the Oaxaca decomposition method, we find that private academies increase differential treatment between men and women while public education decreases the gap. RDD method shows that, in the non-metropolitan area, the income difference between respondents who went to universities and those who did not was more significant than in the metropolitan area. It suggests that by region the university premium is significantly different.

Keywords: Human capital, Regional return to education, Oaxaca decomposition, Regression discontinuity design.

Subject classification codes: E24, I26

[1] email: jongsoo@ku.edu, corresponding author.

1. Introduction

According to human capital theory, more years of education are connected to human capital stock accumulation and productivity improvement so paid workers can earn more in the labor market (Becker, 1975; Card, 1999). On average, individuals who invest more in schooling earn a higher income than individuals with less after graduation (Ginther, 2002). In order to earn a higher expected revenue in the future, people invest years of schooling. In the job market, a high level of education signals increased productivity related to higher wages (Spence, 1973). Of course, if people look forward to a high return to education, then the cost of schooling would be higher. However, even though the human capital theory works well, there will be differences in income and quality of education between regions due to inequality in educational opportunities caused by the concentration of higher education institutions in specific areas. In addition, the number of laborers who earn more with their skills instead of education has been increasing recently. This phenomenon could raise a question about human capital theory. A study related to the return to education could answer this question. While studies on the estimation of return to education in South Korea (hereafter Korea) were conducted until the early 2000s, research about this area has not recently been conducted (Kang et al., 2011; Kim & Park, 2016). Therefore, researchers need to study the return to education, and regional analysis is meaningful because previous studies have investigated the regional rate of return on education.

The Mincer equation (Mincer, 1974) has become the traditional workhorse in empirical analyses of estimating the ex-post return of schooling. However, if researchers estimated the rate of return to education using the Mincer-type wage function, they could have underestimated or overestimated problems. These problems are caused by individual non-cognitive abilities (Griliches, 1977) and measurement errors in education data (Schultz, 2003). Many studies tried to solve these endogenous problems. In addition, these studies estimated the different results of the return to education according to demographic characteristics and job status variables. Since the rate of return to education depends on variables, research for the appropriate rate of return to education is continually needed.

The OLS estimation shows that female workers' return to education was higher than male workers. Using the Oaxaca decomposition method, we find that private academies increase differential treatment between men and women while public education decreases the gap. RDD method shows that, in the non-metropolitan area, the income difference between respondents who went to universities and those who did not was greater than in the metropolitan area. It suggests that by region the university premium is significantly different.

The rest of this paper is organized as follows. Section 2 provides previous studies on returns to education. Section 3 discusses the KLIPS data and variables, while Section 4 discusses the methodology for estimating the returns to education, such as income function, Oaxaca decomposition, and RDD model. Section 5 discusses the results, and Section 6 and Section 7 discuss a robust check and conclusion.

2. Literature Review and Background

2.1 Literature Review

The level of education positively influences wages (Mincer, 1974). Researchers think that wage is proportional to the level of education, so researchers are studying the return to education with interest. Prior studies criticized serious bias problems such as underestimation or overestimation from OLS estimation of the traditional wage function (Choi, 2002; Paik & Ihm, 2008). The main issue is to solve endogenous and distorted estimations. To make up for these problems, many studies have used various methods such as proxy variable estimation, instrumental variable estimation, and panel data (e.g., Choi, 2002; Jang & Lee, 2001; Lee, 2006; Han & Cho, 2007; Paik & Ihm, 2008).

First, Angrist and Krueger (1991) studied the effect of compulsory education in the U.S. on the returns to education, and they used birth year as an instrumental variable. The result showed that the estimation of using the instrumental variable increased by 5% more than the one of using the OLS method. When Ashenfender and Krueger (1994) studied the rate of return to education for twins, the result showed that the estimation of the instrumental variable was about twice as high as the estimation of the OLS. Butcher and Case (1994) and Heckman and Lee (2004) studied women's return to education. In Butcher and Case (1994), the number of sisters was used as an instrumental variable because women's education years were affected by whether they had sisters. This research found that the estimation of the instrumental variable was about twice as high as the estimation of the OLS. Heckman and Lee (2004) showed that the instrumental variable method had a higher result than the OLS method. These results are similar to the results of previous studies.

The previous studies shed light on the case of Korea trying to solve the wage estimation function problem by different methods (e.g., Choi, 2002; Han & Cho, 2007; Jang & Lee, 2001; Lee, 2006; Paik & Ihm, 2008). Jang and Lee (2001) and Choi (2002) used various proxy variables. Jang and Lee (2001) estimated the overall return to education investment by adding variables such as job experience, union membership status, regular workers status, marital status, and city residence, regardless of gender. The authors chose the upper and lower ten cities based on human capital. Cities with higher human capital showed higher average education years. Choi (2002) analyzed the return to education between men and women by using the Korean Labor and Income Panel Study (KLIPS) 1998 and 1999. The author used the spouse and father’s education years in this research as variables. As a result, the return to education is 4.78% - 8.2% for men and 6.85%

- 10.69% for women. Therefore, the return has no significant difference between men and women. In addition, there was no sheepskin effect in the domestic labor market. Using KLIPS 2001, Han and Cho (2007) analyzed the return to education by including variables such as the age of workers, regular worker status, and workers' residence in the wage function. The result of this study showed that the average rate of return to education is 5.2% for men and 4.2% for women, so there was no difference between genders. While women's rate of return to education from age 25 to 34 was slightly higher, men's return to education over age 35 was more heightened.

Unlike previous studies that used substitute variables, Paik and Ihm (2008) estimated the average rate of return to education by using KLIPS 2005. They used the spouse and father's level of education as instrumental variables. In this research, the rate of return to education for women was about 2% higher than for men. Lee (2006) criticized that instrumental variable estimation also could not solve the endogeneity problem. The result of this study estimated a much higher rate of return to education than the result of using OLS or instrumental variable estimation.

In previous literature, the estimations using OLS, instrumental variables, and Difference-in-Difference (DID) were the predominant methods for estimating the return to education. These methods are significantly essential to solve the OLS estimation problems, including endogeneity, for a more accurate analysis of the return to education. However, this study used the Regression Discontinuity Design (hereafter RDD) using test scores, a cognitive ability, for two reasons. First, instrumental variables – father's schooling, GPA, and test scores - could control endogeneity but are enough exogenous to proxy the control and treatment status. Second, there has been no attempt to analyze the regional level using RDD test scores. Therefore, a study about the regional return to education in Korea using RDD with test scores will be the first.

2.2 Korean Education System in Cultural Context

For a long time, due to a hierarchical and patriarchal structure in Korea (Oh-Hwang, 1993; Reagan, 1996), the importance of education in Korea led to improving the value of education, which is connected to Confucian ideas and familism. The primary purpose of the Korean education system is Hongik-Ingan (or Humanitarian Ideal), Korea’s founding principle. The meaning of Hongik-Ingan is to benefit all human beings, which is similar to humanitarianism. In order to achieve this idea, people need to develop moral character, develop the ability to succeed in independent living, and play a positive role in a democratic society. Although the Korean education system is based on humanitarianism, the Korean education system is based on rote learning and relies almost exclusively on college entrance exams (Shin & Koh, 2005). Also, Korean parents value their child’s grades more than cognitive development for his or her learning progress. Moreover, they believe those grades more realistically evaluate their child’s academic improvement (Stevenson & Lee, 1991). For this reason, parents try to find an excellent place to help their children make up for their children's lack of public school education (Seth, 2002). Therefore, private education[2], mainly focusing on test preparation, has become a significant industry, driven by the demand to fill gaps in public education.

Because of the unique Korean education culture and expenditure on private education, Korea has the largest educational gap among the OECD countries between parents and children. Most parents consider that going to a high-ranking school is a way to improve their offspring's socioeconomic status in the future. This fact could lead to a higher investment in test prep. If we look at the school-age profiles using KLIPS data, it has similar trends as the graph of the OECD

[2] These private institutions are different from private schools, and are similar to test preps in the U.S.

countries (Figure 1). In other words, the average education level of offspring is much higher than that of parents. Therefore, a large gap between the two groups can be found.

3. Method

3.1. Data

This paper will estimate the return to education using panel data from the Korean Labor and Income Panel Study (KLIPS) from 2018 to 2019. If the data from 2020 is used, the result could be affected by external shocks such as COVID-19. Since the effect of COVID-19 is uncertain, the data from 2019 to 2020 are difficult to compare. Therefore, choosing years before COVID-19 is proper to estimate the return to education. Moreover, a certain number of households dropping out during the survey increased. 5,044 households were added to the 2018 data to compensate for the failure to reflect the character of the entire population. Therefore, this study shed light on the data for 2018 and 2019. The reason for using the panel data is as follows: if only cross-section data are used, it is possible to derive serious overestimation or underestimation from unobservable characteristics. To be specific, when estimating the return to education, only the increase in individual income for one year expected purely due to education investment must be identified. However, analyzing the return using the cross-section data includes not only the effect of education but also that of individual abilities. Since this study uses cross-section data, the family background will be used (Choi, 2002; Paik & Ihm, 2008; Han & Cho, 2007). As independent variables for analysis, schooling, gender, age, marital status, job experience, regular and full-time status, and job types are used. Moreover, in order to reflect the characteristics by region, this study combined the regional proportion of private academies, provided by Statistics Korea.

3.2. Variables

This study will shed light on only paid workers. The proportion of self-employed and employers seems to be more than 20%, but the analysis results, including the self-employed income using KLIPS data, confirmed that there was no significant difference from the analysis results of paid workers targets (Lee, 2017). Incomes corresponding to the dependent variable will use the hourly income function. Using hourly incomes as a dependent variable seems to be a practice in Labor Economics and to solve the endogenous problem of variables that have an important effect on income (Choi, 2002; Hong, 2010).

Since this paper uses the Mincer-type function, independent variables such as schooling, job experience, and the square term of job experience will be used. For schooling, we set up below elementary school=0, elementary school=6, middle school=9, high school=12, university=16, and graduate school=21. If they did not graduate, we subtracted one from the status of the final school using the graduation status because the KLIPS data provided the detailed graduate status.

In order to control regional characteristics for education, the number of private academies per 1,000 people that could affect education will be used. Previous studies used several variables such as labor union and industrial index. This is likely to be analyzed as a wage inequality problem, not the rate of return to education, and if the independent variable increases, a multicollinearity problem can occur (Jang & Lee, 2001).

3.3. Descriptive Statistic

First, the form of workers was analyzed through descriptive statistics (Table 2). The types of workers can be divided into paid workers and non-paid workers. The paid workers include regular, temporary, and daily workers. Moreover, the non-paid workers include employers, self-employed, and unpaid family workers (Households). In the case of male workers, the proportion of regular paid workers is much higher than that of female workers. On the other hand, in the case of female workers, the proportion of temporary workers is much higher than that of male workers. The proportion of male workers self-employed is more than three times higher than that of females self-employed. Therefore, it shows that the ratio of regular and self-employed male workers is high and that the ratio of non-regular workers is high for female workers. It suggests that females have more job opportunities, but the proportion of male workers is still high.

Next, the number of paid workers and average hourly income by region and gender were compared (Table 3). The regions were compared by dividing them into metropolitan areas (Seoul, Gyeonggi, and Incheon) and non-metropolitan areas (Otherwise). Compared to the non-metropolitan areas, the number of paid workers in metropolitan areas such as Gyeonggi and Incheon was significantly higher. However, the hourly wage of workers living in Seoul is lower than in other metropolitan areas because of the expensive housing cost. Since the housing price in Seoul is higher than in other areas, the workers decided to move to Gyeonggi and Incheon.[3] In addition, except for some regions, the total number of paid workers by region decreased in 2019 compared to 2018. While the average hourly income was generally high in the metropolitan areas, the average hourly wages in some non-metropolitan areas such as Chungnam, Ulsan, and Kyungam were high because these areas have industrial complexes. In particular, the male and female workers show high hourly wages in Sejong, where research complexes are gathered along with Ulsan. The highest wage for high skills is one of the important factors in immigration (Borjas, 2001). In order to attract these kinds of workers, higher wages would be needed (Lee, 2017).

[3] According to the Housing Apartment Index in 2018 and 2019, it can be seen that the index of Seoul increased from 2018 (1.11) to 2019 (1.20) and both were significantly higher than that of Gyeonggi (1.01 & 1.01) and Incheon (0.99 & 0.98).

4. Research Model

This paper uses the Mincer-type function (Mincer, 1974) to estimate the returns to education on earnings. After estimation, RDD model will be used to examine discontinuity in returns to education related to the university premium in Seoul. Furthermore, we will exploit the Oaxaca decomposition by region and gender in order to examine productivity and differential treatment gaps.

4.1. Income Function

The Mincer-type equation is the most widely used in empirical work because earnings can be explained simply as a function of schooling and job experience through this model. In addition, hundreds of studies have found that this function provided an accurately reasonable description of age-earning profiles in many countries' labor markets as well as in the United States (Borjas, 2015). Thus, we will conduct an OLS analysis. The wage equations can be derived as follows:

$$ln\, y_{ij} \; = \; y_0 \; + rEDU_{ij} + \beta_1 EXP_{ij} + \beta_2 EXP^2 + \beta_3 AGE_{ij} + \gamma X_i + d_{ij} + e_{ij},$$

where *i* = metropolitan or non-metropolitan areas and *j* = male or female. The term *EDU* represents years of completed education, and e is a statistical residual. If we do not know the number of years of service that an individual has worked since graduating from school, the potential experience can be estimated as *EXP* = age - *EDU* - 6 (Card, 1999).[4] To do so, we assumed that he started school

[4] The potential experience has proved to be very fittable in various empirical analyses (Heckman & Polacheck, 1974; Murphy & Welch, 1990).

at age six and finished *EDU* years of schooling. X_i refers to individual and occupational characteristics. The term d_i refers to unobservable variables.

To reflect the recent data well, the model that adds a quadratic term for schooling or considers the cohort effect will be used because the data do not fit well in an environment where educational achievement is rapidly changing (Card & Lemieux, 2001a; Heckman et al. 2003; Lemieux, 2006). We will add a cohort effect term of age, AGE_{ij}.

Since the 2018 and 2019 results showed similar, we will regress pooled OLS using the year-fixed effect. In order to control the characteristics of individuals and job types, we will use these variables as fixed effects. The problem of the traditional income function could underestimate or overestimate the data because of ability bias problems such as the unobservable individual ability. Solving the problems of measurement errors such as endogeneity is a key point in terms of returns to education. The individual variables, including family background, could be used as a fixed effect.[5] Moreover, we will use the regional apartment index to reflect the regional price and real wage.

4.2. Oaxaca Decomposition Model

This paper will use the Oaxaca decomposition model (Oaxaca, 1973) to estimate the return to education. The differential treatment of return to education (D_e) is as follows. The term m represents male paid worker, and the term f means female paid worker. Two variables, $\underline{I_m}$ and $\underline{I_f}$,

[5] According to Becker (1981), Lam (1988), and Nakosteen and Zeemer (2001), if the income from marriage in the marriage market is a source of public goods for families, it mainly has a positive relationship. Thus, the spouse's education year as a proxy variable should be premised that all wage workers are married. In addition, it is not an appropriate proxy variable that can be used at the current time when the marriage periods of wage workers are delayed. Therefore, this study will use the father and mother's education years as family background variables for unobservable factors.

are the average hourly income of male and female paid workers in metropolitan or non-metropolitan areas due to their return to education. The equation is as follows:

$$D_e = \frac{\underline{I_m} - \underline{I_f}}{\underline{I_f}}$$

After taking a natural log, we can get the difference between labor markets without any differentiation. According to the OLS assumption, the average error term becomes zero and the equation is as below:

$$ln\underline{I_m} = Z'_m\beta_m + \mu_m,$$

$$ln\underline{I_f} = Z'_f\beta_f + \mu_f,$$

$$ln\underline{I_m} = E(Z'_m)\hat{\beta}_m + E(\mu_m),$$

$$ln\underline{I_f} = E(Z'_f)\hat{\beta}_f + E(\mu_f),$$

$$ln\underline{I_m} = \underline{Z'_m}\hat{\beta}_m,$$

and

$$ln\underline{I_f} = \underline{Z'_f}\hat{\beta}_f.$$

Two terms, $ln\underline{I_m}$ and $ln\underline{I_f}$, represent the average hourly income of male and female paid workers in metropolitan and non-metropolitan areas. In addition, $\underline{Z'_m}$ and $\underline{Z'_f}$ are the average variables that can affect the rate of return to education. Therefore, we can get the Oaxaca decomposition assumption based on the wage function of the male paid worker as below:

$$ln\underline{I_m} - ln\underline{I_f} = \underline{Z'_m}\hat{\beta}_m - \underline{Z'_f}\hat{\beta}_f \qquad (1)$$

$$= \underline{Z'_m}\hat{\beta}_m + \underline{Z'_f}\hat{\beta}_m - \underline{Z'_f}\hat{\beta}_m - \underline{Z'_f}\hat{\beta}_f \qquad (2)$$

$$= (\underline{Z'_m} - \underline{Z'_f})\hat{\beta}_m + (\hat{\beta}_m - \hat{\beta}_f)\underline{Z'_f} \qquad (3)$$

In (1), the left side of the Oaxaca decomposition means the average log hourly wage difference between male and female paid workers. In (3), $(\underline{Z'_m} - \underline{Z'_f})\hat{\beta}_m$ represents the difference of the average explanatory variable between male and female workers based on the assumption that both male and female workers earn the same wage. This difference is called the difference in productivity. In the line (3), $(\hat{\beta}_m - \hat{\beta}_f)\underline{Z'_f}$ means the difference in average wage between male and female workers with a fixed explanatory variable in a region. Therefore, these two parts are called the explained part and the unexplained part, respectively. That is, the equation consists of the difference in individual productivity and the difference in reward level. The income gap caused by the difference in reward level is a differential treatment factor used to analyze income inequality. This factor shows what variables cause differential treatment except for factors that determine individuals' productivity, such as educational background and career. The unexplained parts are essential to capture all potential effects of differences in unobservable variables (Jann, 2008). Therefore, if the difference in the rate of return to education by gender within an area is analyzed, then it is possible to analyze what factors cause productivity differences and differential treatment between male and female paid workers living in metropolitan or non-metropolitan areas besides the investment in education.

However, it shows the identification problem of a typically detailed decomposition when the dummy variable is used in the process of choosing a referece,. Even though many studies have used the Oaxaca decomposition, a few considered the detailed issues of the model. If the problem includes the result, the difference in constant variables between groups is significant. Utilizing the ways to solve the identification problem would be needed. Yun (2004, 2005) suggested the average method to solve these assigned problems[6]. We will apply to the average method to normalize the dummy variable.

4.2. Regression Discontinuity Design (RDD)

RDD is a useful statistical tool that allows many researchers to examine the effectiveness of the treatment. RDD began with Thistle and Campbell (1960). The first time RDD appears with Goldberger (1972) in the Economics community is an unpublished econometrics paper. In 1999, notable papers using this method were introduced (Angrist and Lavy, 1999; Black, 1999; Hahn et al., 2001; Van der Klaauw, 2002; Lee and Lemieux, 2010; Hoekstra, 2009). RDD is significantly popular today for two reasons. First, RDD could appeal to many because of its ability to get rid of selection bias convincingly. Moreover, underlying identifying assumptions are viewed by many as easier to evaluate and easier (Cunningham, 2021).

In order to solve the endogeneity problems, we can use the test score as a running variable with RDD model (Hoekstra, 2009). Fortunately, the KLIPS data set has categorical test scores data for college entrance exams. Under the continuity assumption, any discontinuous jump in log wages at the normalized score cut-off would be appropriately interpreted as the causal effect of higher test scores on wages. Since the distribution of observations shows the density is smooth across the

[6] Kim (2013) insisted that the average method may not answer the identification problem. As an alternative, he suggested a grand-mean centering method.

test score cutoff, the test score is a continuous variable (Figure 2). To formally test for RDD in the returns to education, we estimate the following equation:

$$ln\, wage_i = \beta_0 + \beta_1 D(X_i \geq c_0) + \beta_2 X_i + \beta_3 \gamma + \varepsilon_i$$

where outcome of paid workers is an independent variable and =1 if $D = 1$ if $X_i \geq 0$ (Normalized test score)[7] or D=0 if $X_i < 0$. Estimating the wage equation allowed us to estimate potential outcomes from the above equation that could be used to figure out the graph.

In order to solve the self-selection problem, we will compare pre- and post-entrance scores as a cut-off point (c_0). To normalize the cut-off point, we will subtract the minimum admission scores of universities in Seoul. Because degrees from high-ranked universities could boost the future expected earnings, most Korean students' academic goals are to get a higher score on the university entrance exam and to enter high-ranked universities. Most high-ranked universities in Korea are in metropolitan areas and private universities. Discontinuity would have happened in wage differences when comparing the minimum scores for universities to the ones that are not.

It is hard to get the data unless the four-year university provides official admission scores. Instead, from the 1980s, private academies posted the estimated admission scores of universities in the newspaper every year based on their students' scores. From 1982 to 1992, the maximum score for the exam was 340. However, this exam was criticized because this exam is composed of

[7] In order to make the adjusted score following the Hoekstra (2009), we subtract the actual test score from the universities' admission minimum score. The recentered minimum score would be zero. In addition, an individual score who entered a university shows positive value.

excessive memorization questions, which not only put a disproportionate burden on the students but could not select the candidates qualified for college. After that, to replace this test, the research of a college entrance test that modeled the SAT in the U.S. was carried out. The College Scholastic Ability Test (CSAT) was implemented in 1993 through seven experimental evaluations from 1990 to 1992. The maximum CSAT score was 200 from 1993 to 1996 and was 400 after 1997. Since the KLIPS data in 2002 includes the categorical university entrance examination data, we analyzed it based on the range corresponding to the admission score of universities in Seoul.

5. The Results

5.1. The Return to Education between Metropolitan and Non-metropolitan Areas

Since we found no difference in the return on education between male and female workers between 2018 and 2019, this study could use the pooled OLS. We find that female workers' return to education was higher than male workers. The female workers' return to education was 8.1% in metropolitan areas and 9.0% in non-metropolitan areas, while the male workers' return to education was 5.4% in metropolitan areas and in non-metropolitan areas. If we look at the characteristics of occupations, it can be seen that male workers have a higher return in non-metropolitan areas as they have regular jobs, which can be seen as the effect of the decentralization policy. Due to its decentralization policy in 2012, Korea has transferred many companies and research institutes to non-metropolitan areas. Therefore, companies and research institutions in this region pay higher wages to attract workers from metropolitan areas.

We will analyze the age cohort effect. Unlike males, females are more active in the labor market at a young age. Women enter society faster than men in Korea because most males have to go to military service for 18 months to 27 months during university or after graduation from high school. However, female workers leave the labor market as they grow older and rejoin it a few years later (U-shaped). This is because the primary factors of women's career disruption are associated with marriage and childbearing (Kim & Oh, 2022). As a result, female workers are affected by various effects at this time. They have relatively low human capital due to the lack of educational opportunities and job experience during their break and find lower-level jobs compared to their last jobs. Therefore, females who re-entered the labor market show a relatively low income, and the income gap between female workers with different education levels becomes smaller. For these reasons, the return to education is significantly lower for female workers in high-age groups.

5.2. Oaxaca Decomposition Analysis

The results of the Oaxaca decomposition analysis by grouping males and females living in the metropolitan and non-metropolitan areas are shown in Table 5. The term $(\underline{Z'_m} - \underline{Z'_f})\hat{\beta}_m$ is the factor that causes the difference in productivity between males and females under the assumption that they all receive the same level of wage, and the term $(\hat{\beta}_m - \hat{\beta}_f)\underline{Z'_f}$ is the differential treatment coefficient that refers to the factor that causes differential treatment between men and women under the assumption that their productivity is the same.

In metropolitan and non-metropolitan areas, the proportions of the explained parts were about 29.9% (0.096) and 25.5% (0.077), which were much less than the unexplained parts, which are about 70.1% (0.225) and 74.5% (0.226). It implies that the factors that cause productivity in the rate of return to education between male and female workers are much smaller than the factors that lead to differential treatment.

Surprisingly, this study showed that while differences in levels of education can cause a productivity gap between males and females, they can be a factor that can decrease the differential treatment. Another interesting fact is that participating in private academic institutions allows male workers to increase differential treatment in metropolitan areas (0.146). In Korea, it is suggested that private education has a significant influence as a complementary good to public education. According to data released by the Korean National Statistics and the Ministry of Education, the average monthly cost of private education in South Korea is 32.1% of the total average salary as of 2018, and the participation rate in private education is 74.8%. Therefore, in Korea, while public education narrows the differential treatment gap between men and women, private academies play a role in widening the gap. Heo (2007) pointed out that it can act as discriminatory factors according to academic years. Even if women are well educated and have the same labor competitiveness, promotions and salaries are often limited compared to men. Therefore, the glass ceiling issue in this society should be continuously discussed, and research to examine the social structure and problems by gender should be continued.

In sum, in this study, the factors that cause productivity and differential treatment between men and women by region were analyzed through the Oaxaca decomposition method. As the non-explanatory variables were higher than the explanatory ones, the factors that cause the significant differential treatment between men and women were found. Participating in private academic institutions causes gender treatment differential.

5.3. The Differences in Outcome Using the Test Scores

Figure 3 provides a graphical representation of the effect of minimum CSAT scores for university admission in Seoul on outcomes. The scores are centered at zero because we normalized the scores. The results indicated that individuals who got the minimum university admission scores in Seoul earned higher wages than those who did not. It implies that there is a university premium. According to the areas, the difference in outcomes with cut-off is significantly different.

In the non-metropolitan area, the outcome difference between those who went to universities and those who did not was greater than in the metropolitan area. This is primarily due to differences in factors between regions. Since most good quality jobs are concentrated in the metropolitan area, university premiums fall that much as well-educated workers flock to the metropolitan area. At the same time, the higher the score, the greater the likelihood of entering a good school, so you can see the wage increase after the cut-off. On the other hand, job shortages and low wages are the biggest problems in the non-metropolitan area. Therefore, it is not easy to find a good workforce. As a result, there is a large difference in wages between those who did not graduate from university and those who graduated from college. However, compared to the metropolitan area, the higher the SAT score, the less the wage increases.

How will regional outcomes of workers differ if respondents get a score to get into top-school? In principle, the more they get into a better university, the better productivity they get, which can affect your expected future earnings. The cut-off was changed to the minimum admission score at Seoul National University, Korea's top school, to examine the presence of regional discontinuity (Figure 4). As with the previous results, no top school premium was found in the metropolitan area. First, many workers flock to the metropolitan area to get good jobs, and their profits will be reduced by that much. Second, even if you enter a top university, the profits may vary depending on the major. We suppose that there are respondents who entered the humanities department of

Seoul National University and those who entered the STEM related department of a slightly lower-ranked university. As there is no top-school premium, there is a high possibility that there will be a difference in wages depending on the major. As a result, the latter person is more likely to get better jobs. This study did not look at the differences in majors, but it will be left as a future study.

6. Robust Check

In Korea, the age of employment is increasing due to high job competition. In general, male worker's entry into society is delayed than female workers' because they are obligated to go to the military service. Therefore, this analysis was conducted except for those under the age of 29 to consider if the volatility affects the results of return to education (Table 6). If this volatility affects a lot, the OLS estimate obtained earlier will be biased. Focusing only on workers from the age of 30 showed that the return to education was 5.6 percent for men in the metropolitan area, 8.6 percent for women, 5.8 percent for men in non-metropolitan areas and 9 percent for women. Therefore, no significant difference was found compared to the previous results. As a result, the OLS estimate obtained earlier is a robust result.

7. Conclusion

This paper analyzed the return to education for male and female paid workers by region using KLIPS data in 2018 and 2019. With fixed effects of individual, job-types, and region, we found that the return to education of female workers was higher than that of male workers. Using a reubst check with being except for the workers under 29, the pooled OLS results are robust.

Using the Oaxaca decomposition method by gender and region, we found that the proportions of the unexplained parts were much less than the explained parts. It means there is less weight to productivity gaps than differential treatment factors. The differential treatment between male and female workers arises from participating in test preps. On the other hand, the public education between male and female workers decreases in differential treatment. Through this study, surprisingly, we confirmed that private and public schooling have different effects regarding differential treatment and productivity gap by gender and region.

Through RDD model, we found a difference in wages based on the minimum score for going to a four-year university in Seoul. This was more pronounced in the non-metropolitan area than in the metropolitan area. This suggests that it is caused by the difference in the density of workers and the opportunity to get a job by region. On the other hand, the top-school premium was not found, but rather the opposite effect appeared in the non-metropolitan areas.

This study has several limitations. First, it was analyzed using cross-section data due to the changed sample data from 2018. Therefore, it could analyze the return to education more accurately using panel data over a long period before 2018. The panel data has the advantage of solving endogenous and measurement error problems (Park & Yoo, 2013; Shin, 2011). Moreover, this study analyzed the return on education with a few variables. However, there is a possibility that other factors will affect the return on male and female education. Therefore, an accurate analysis will be needed by selecting variables without endogenous problems in the future.

This study focused on the return to education. Still, it was found that there are differential treatment factors between male and female paid workers. In other words, it can be seen that there is a glass ceiling that is not yet visible in our society. Through this study, we confirmed that types of education could be critical factors that can increase or reduce the invisible differential treatment between men and women. Therefore, research on the return to education to resolve differences in male and female productivity and differential treatment gap problems should continue at the regional level.

Acknowledgement

We would like to thank Donna Ginther, David Slusky, and ChangHwan Kim for helpful feedback on earlier versions of this paper. We would also like to extend our special thanks to Zongwu Cai, Shahnaz Parsaeian, Youjin Hahn, Jinwook Jeong, Dongwoo Kang, Jack Osman, Llamosas Rosas Irving, Antonella Mancino, participants at the 62nd Western Regional Science Association Conference, 87th Midwest Economics Association, 60th Missouri Valley Economic Association, and seminar participants at the University of New Hampshire for their helpful comments and suggestions. We are also grateful to Ronald Oaxaca for his precious comments.

Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Angrist J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics* 106(4): 979-1014.

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics* 114(2): 533–575.

Becker, G. S. (1975). *Human Capital*. New York: NBER, Columbia University.

Becker, G. S. (1981). *A Treatise On The Family, Cambridge, Mass*. Harvard University Press.

Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. The *Quarterly Journal of Economics* 114(2): 577–599.

Borjas, G. J. (2001). Does immigration grease the wheels of the labor market? *Brookings Paper on Economics Activity* 1: 69-119.

Borjas, G. J. (2015). *Labor Economics*, 7E, McGraw Hill Education.

Butcher, K. F., and Case, A. (1994). The effect of sibling sex composition on Women's education and earnings. *The Quarterly Journal of Economics* 109(3): 531-563.

Card, D.(1999). The causal effect of education on earnings. *Handbook of Labor Economics (North Holland, Amstredam)* 3: 1801-1863.

Choi, K. (2002). An analysis of returns to schooling in Korea. *Korea Review of Applied Economics* 4(2): 229-257.

Cunningham, S. (2021). *Causal Inference: The mixtape*. Yale University Press Book.

Department of Statistics. (2022). Results of the 2022 Elementary, Middle, and High School Private Education Expenditure Survey. *Research Report*.

Fairlie, R. W. (2005). An extension of the blinder-Oaxaca decomposition technique to logit and probit Models. *Journal of Economic and Social Measurement* 30(4): 305-316.

Ginther, D. K. (2002). Alternative estimates of the effect of schooling on earnings. *Journal of Labor Economics* 34(2): 509-543.

Griliched, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica* 45(1): 1-22.

Han, S., and Cho, I. (2007). The return to education and Sheepskin effect in Korea: Comparison of male and female workers. *Journal of Labour Economics* 30(1): 1-30.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica* 69(1): 201–209.

Heckman, J. J., and Li, X. (2003). Selection bias, comparative advantage and heterogeneous returns to education: Evidence from China in 2000. *Pacific Economic Review* 9(3): 155-171.

Heo, S. (2007). Regional wage differentials in Korea: The case of metropolitan and non-metropolitan areas. *Journal of Korean Industrial Economic Association* 20(1): 1-16.

Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *Review of Economics and Statistics* 91(4): 717-724.

Hong, M. (2010). A discussion on estimating wage differentials between regular-nonregular works. *Korean Journal of Labor Studies* 16(2): 141-161.

Jang, S., and Lee, B. (2001). Distribution of human capital across Korean cities and industries, and external economies of human capital. *Journal of Labour Economics* 24(1): 1-33.

Jann, B. (2008). A Stata Implementation of the Blinder-Oaxaca Decomposition, *The Stata Journal* 8(4): 1-33.

Kang, S., Yoon, S., and Park, S. (2011). *Performance analysis of human capital investment in Korea*. Korea Labor Institution.

Kim, C. (2013). Detailed wage decompositions: revisiting the identification problem. *Sociological Methodology* 43(1): 346-363.

Kim, C., and Oh, B. (2022). Tested-based Gender Discrimination in South Korea, *Social Science Research* 104: 1-16.

Kim, H. and Park, S. (2016). Estimation for rate of return for education in Korea: OLS VS Quantile regression. *Korea Journal of Public Finance* 9(1): 1-25.

Lee, J. (2017). A study on Regional Wage Differentials in Korea. *Quarterly Journal of Labor Policy* 17(2): 143-171.

Lee, S. (2006). Estimating the Income Effect of Educational Background Using Panel Data. *Labor Review* 33-42.

Lemieux, T. (2006) Postsecondary Education and Increasing Wage Inequality, *American Economic Review* 96(2): 195-199.

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

Murphy, K. M., and Welch, F. (1990). Empirical age-earnings profile. *Journal of Labor Economics* 8: 202-229.

Nakosteen, R. A., and Zimmer, M. (2001). Spouse selection and earnings: Evidence of marital sorting. *Economic Inquiry* 39(2): 201-213.

Oaxaca, R. L. (1973). Male-female wage differentials in urban labor markets. *International Economics Review* 14: 693-709.

Oh-Hwang, Y. (1993). Linkage between home environments, child psychosocial maturity, and child academic achievement. *Gifted International* 8(1): 32-37.

Paik, I. and Ihm, J. (2008). Rates of return to higher education for women OLS and instrumental variables approach. *Journal of Educational Administration* 26(3): 75-94.

Park, S. and Yu, J. (2013). Return to schooling: Panel data analysis. *International Economic Journal* 19(1): 135-160.

Reagan, T. (1996). *Non-Western educational traditions: alternative approaches to educational thought a proactive*. Mahwah, NJ: Lawerence Erlbaum Associates Publications.

Schultz, T. P. (2003). Human capital, schooling and health. *Economics and Human Biology* 1: 207-221.

Seth, M. J. (2002). *Education fever: society, politics, and the pursuit of schooling in South Korea*. Honolulu: the University of Hawaii Press.

Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics* 87: 355-374.

Stevenson, H. W. and Lee, S. (1991). Context of achievement: A study of American Chinese and Japanese children. *Monographs of the Society for Research in Child Development* 55 (1-2 Serial No. 221).

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology* 51: 309–317

Van Der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review* 43(4): 1249–1287.

Yun, M. (2004). Decomposing differences in the first moment. *Economic Letters* 82(2): 275-280.

Yun, M. (2005). A simple solution to the identification problem in detailed wage decompositions. *Economics Inquiry* 43(4): 766-772.

Tables

Table 1. Variables

Variable	Unit	Explanations
Log Hourly Wage	won	Log hourly real wage of paid workers
Education	years	Final academic standards
Job Experience	years	Potential job experience=Age-Education-6
The Square of Job Experience/100	years	The square of the year of job experience for paid workers
Father's Year of Education	years	Schooling according to father's personal education
Mother's Year of Education	years	Schooling according to mother's personal education
Gender	dummy	Male=1 & Female=0
Age	dummy	15~29=1, 30~39=2, 40~49=3, 50~60=4, over 60=0
Marital Status	dummy	Married=1 & Non-married=0
Regular Status	dummy	Regular=1 & Non-regular=0
Full-time Status	dummy	Full-time=1 & Non-full-time=0
Job Types	dummy	Agriculture=1, Manufacture=2, Management=3, Construction=4, Finance=5, Service=6, Profession=7, Simple labor=0
The Number of Private Academies	units	The number of private academies per 1,000 people

Table 2. The number of Paid Workers and Non-Paid Workers

			2018		2019	
			Population	**Ratio**	**Population**	**Ratio**
Male	**Paid Worker**	**Regular**	4,493	34.7%	4,374	35.1%
		Temporary	610	4.7%	619	5.0%
		Daily	529	4.1%	464	3.7%
	Non-Paid Worker	**Employers /Self-employed**	1,989	15.4%	1,861	14.9%
		Households	95	0.7%	73	0.6%
	Subtotal		7,716	59.6%	7,390	59.3%
Female	**Paid Worker**	**Regular**	2,757	21.3%	2,700	21.7%
		Temporary	999	7.7%	998	8.0%
		Daily	245	1.9%	238	1.9%
	Non-Paid Worker	**Employers /Self-employed**	741	5.7%	692	5.6%
		Households	486	3.8%	451	3.6%
	Subtotal		5,228	40.4%	5,080	40.7%
Total			12,944	100.0%	12,470	100.0%

Note: We consider the survey weight provided by the KLIPS data.

Table 3. The number of Paid Workers and Average Hourly Income by Region and Gender

Region		2018					2019				
		Population		Average Hourly Income/1000		Gender Relative Income (F/M)	Population		Average Hourly Income/1000		Gender Relative Income (F/M)
		Total	Paid Worker	Female	Male		Total	Paid Worker	Female	Male	
Metro-politan areas	Seoul	2,260	1,597	10.97	13.32	0.82	4,174	1,576	9.76	12.81	0.76
	Gyeonggi	3,251	2,050	10.29	16.37	0.63	5,603	2,102	11.02	16.73	0.66
	Incheon	741	465	10.19	15.78	0.65	1,303	514	11.11	16.35	0.68
Non-Metro-politan areas	Chungbuk	426	276	10.46	14.68	0.71	761	272	14.89	16.52	0.90
	Chungnam	586	325	9.84	18.55	0.53	925	327	11.22	19.73	0.57
	Daejeon	375	234	9.25	13.32	0.69	678	253	9.25	14.42	0.64
	Daegu	615	402	11.81	13.85	0.85	1,153	398	12.81	13.96	0.92
	Pusan	892	554	11.45	16.25	0.70	1,626	535	12.15	17.47	0.70
	Ulsan	343	229	12.56	20.86	0.60	579	225	13.12	21.68	0.61
	Kyungbuk	716	378	10.45	16.08	0.65	1,273	371	10.97	16.57	0.66
	Kyungnam	914	488	11.27	16.60	0.68	1,585	522	12.77	18.74	0.68
	Gangwon	355	152	10.10	14.66	0.69	624	161	11.44	14.78	0.77
	Gwangju	311	188	9.50	14.11	0.67	618	213	10.18	13.54	0.75
	Jeonbuk	463	245	12.96	15.05	0.86	826	229	11.23	16.52	0.68
	Jeonnam	467	214	9.86	16.41	0.60	804	221	10.48	16.84	0.62
	Jeju	153	62	9.64	11.99	0.80	235	60	10.12	13.60	0.74
	Sejong	76	52	14.70	16.79	0.88	120	54	13.45	17.62	0.76

Table 4. OLS Estimation Results by Region

	Metropolitan		Non-Metropolitan	
	(1)	**(2)**	**(3)**	**(4)**
VARIABLES	**Male**	**Female**	**Male**	**Female**
Education	0.054***	0.081***	0.054***	0.090***
	(0.006)	(0.006)	(0.006)	(0.008)
Experience	0.019***	-0.002	0.024***	-0.005
	(0.006)	(0.006)	(0.007)	(0.007)
Experience/100	-0.043***	0.007	-0.038***	0.031***
	(0.010)	(0.009)	(0.012)	(0.011)
Marry	0.240***	0.077***	0.194***	0.027
	(0.027)	(0.027)	(0.028)	(0.033)
Age				
15-29	-0.091	0.082	0.021	0.202
	(0.137)	(0.138)	(0.146)	(0.179)
30-39	-0.083	0.127	0.071	0.303**
	(0.110)	(0.115)	(0.121)	(0.149)
40-49	0.039	0.175*	0.167*	0.317***
	(0.089)	(0.093)	(0.101)	(0.123)
50-59	0.151**	0.119*	0.178**	0.247***
	(0.069)	(0.069)	(0.075)	(0.095)
Regular	0.194***	0.136***	0.234***	0.170***
	(0.029)	(0.026)	(0.030)	(0.032)
Full-time	0.005	-0.110***	-0.106*	-0.123***
	(0.052)	(0.033)	(0.063)	(0.041)
Constant	7.543***	7.590***	8.208***	7.613***
	(0.259)	(0.277)	(0.240)	(0.325)
Observations	1,963	1,598	1,755	1,306
R-squared	0.366	0.287	0.329	0.253
Year FE	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES
Job Type FE	YES	YES	YES	YES
Region FE	YES	YES	YES	YES

Standard errors in parentheses, *** p<0.01, ** p<0

Table 5. Oaxaca Decomposition Results by Region (2018)

	Metropolitan Areas		**Non-Metropolitan Areas**	
	Endowments $(\underline{Z'_m} - \underline{Z'_f})\hat{\beta}_m$	**Coefficients** $(\hat{\beta}_m - \hat{\beta}_f)\underline{Z'_f}$	**Endowments** $(\underline{Z'_m} - \underline{Z'_f})\hat{\beta}_m$	**Coefficients** $(\hat{\beta}_m - \hat{\beta}_f)\underline{Z'_f}$
Education	0.063	-0.228	0.073	-0.254
Academic	0.002	0.146	0.000	0.096
Total	0.096	0.225	0.077	0.226
Proportion (%)	29.9	70.1	25.5	74.5
$\boldsymbol{D_e}$	0.371		0.388	

Note: Endowment is explained part and cofficients is unexplain part.

Table 6. Robust check

VARIABLES	**(1)** Metro male	**(2)** Metro female	**(3)** NonMetro male	**(4)** NonMetro female
Education	0.056***	0.086***	0.058***	0.090***
	(0.007)	(0.008)	(0.007)	(0.010)
Observations	1,688	1,217	1,485	1,016
R-squared	0.334	0.308	0.261	0.262
Year FE	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Region FE	YES	YES	YES	YES

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Excludes wokers under 29 to verify the OLS result.

Figures

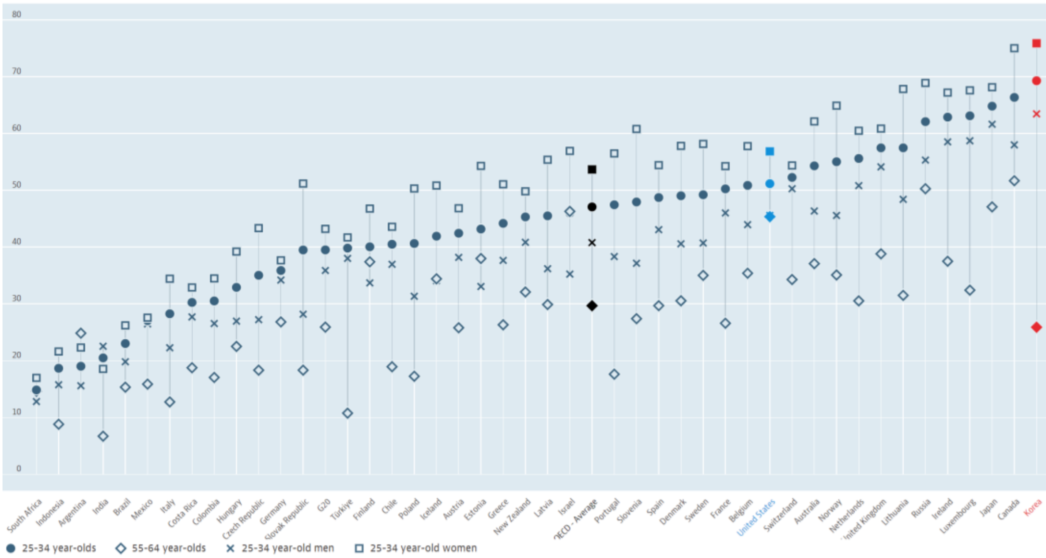

Figure 1. Population with Tertiary Education (Korea)
https://data.oecd.org/eduatt/population-with-tertiary-education.htm#indicator-chart

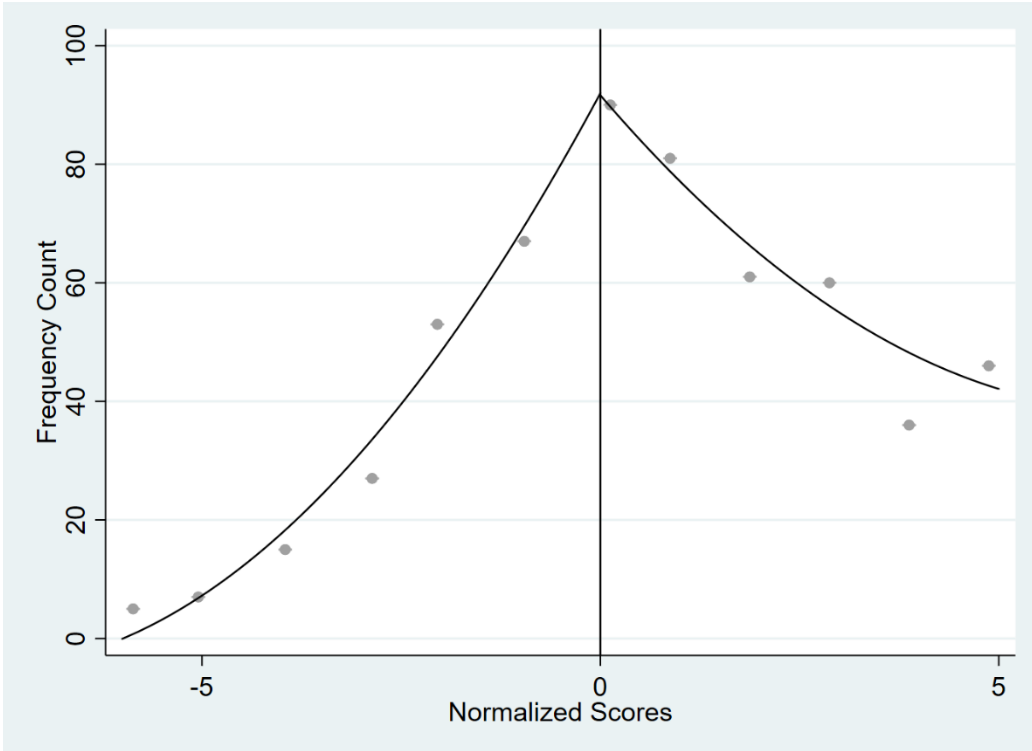

Figure 2. Distribution of University Entrance Exam Scores Near Minimum Admission Score

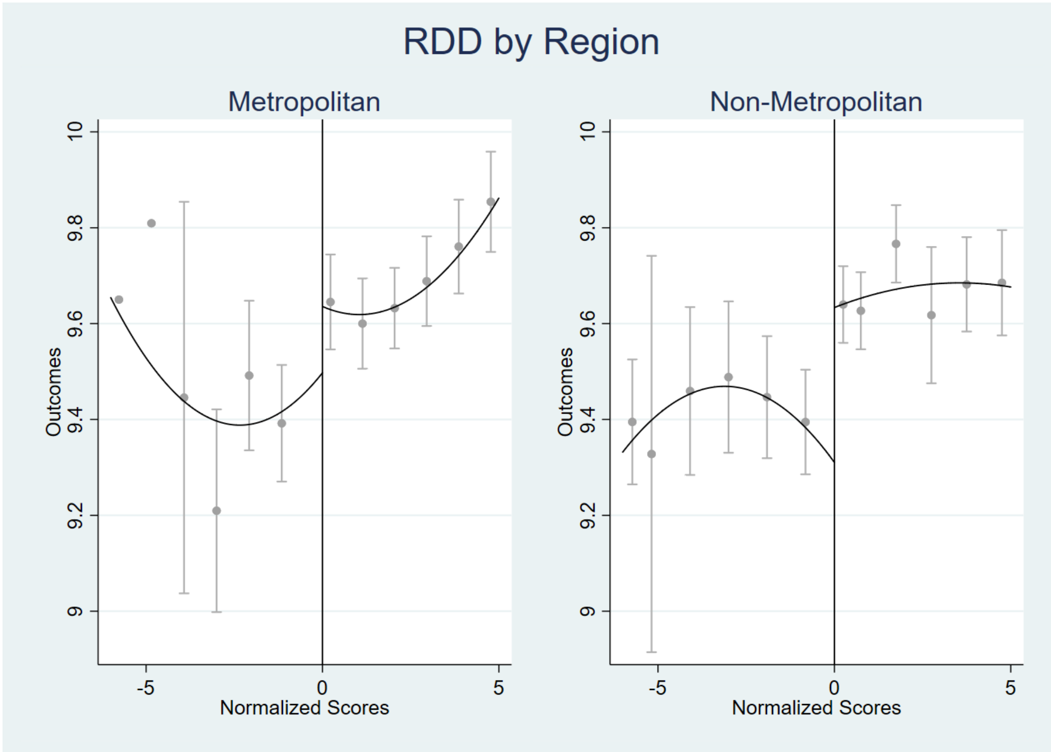

Figure 3. Regression Discontinuity Esimates of the Regional Effect of University on Outcomes

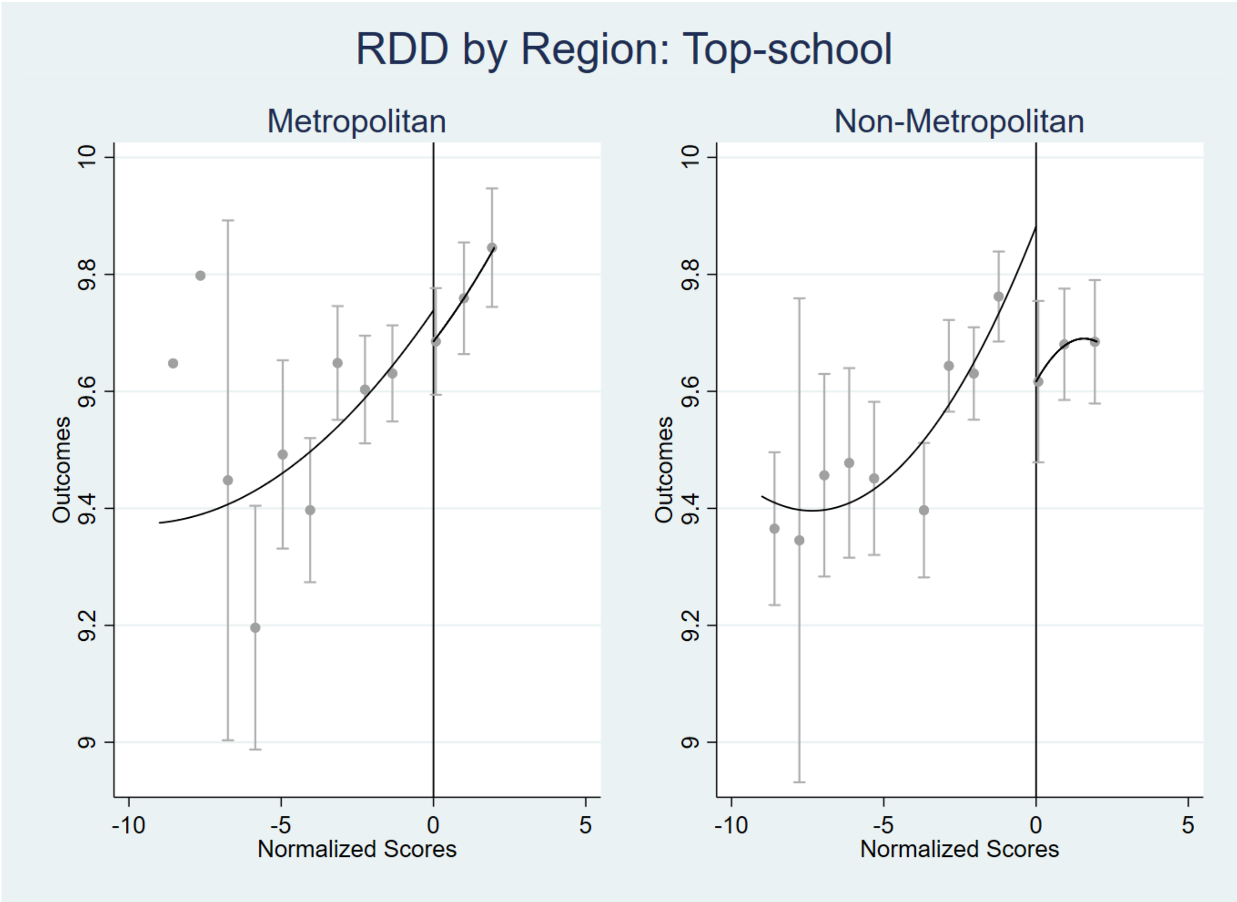

Figure 4. Regression Discontinuity Estimates of the Regional Effect of Top-school on Outcomes

The List of Figures