Discussion Paper No. 996

MARGINAL RETURNS TO SCHOOLING AND EDUCATION POLICY CHANGE IN JAPAN

Nobuyoshi Kikuchi

March 2017

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Marginal Returns to Schooling and Education Policy Change in Japan*

Nobuyoshi Kikuchi[†]

ISER, Osaka University

March 30, 2017

Abstract

This paper examines the returns to university education in Japan, using tuition, availability of universities, and labor market conditions as instrumental variables. To measure availability of universities, this paper uses total accredited capacity of all universities in the prefecture of residence at the age of 15. This measure captures cross-time and cross-prefecture variations, because birth cohort and prefecture dummies are also controlled. A set of education policy-relevant instruments allows for estimation of the marginal effects for individuals who are induced to enroll in university by policy changes. Using the estimated marginal treatment effect, this paper recovers the average treatment effect parameters. The main empirical result shows that an additional year of university education increases hourly wage by about 9% on the population average. This paper also finds heterogeneous effects by subpopulation groups: the average effect of a year of university education for those enrolled in university is about 17%, but less than 2% for those who did not enroll. Finally, this paper investigates the average returns for those who are induced to enroll in university by a particular policy shift, such as free tuition or an increase in the capacity of local universities. The results suggest that such policy changes bring about positive effects of university education.

JEL Classification: J24, J31, I23, I28
Keywords: returns to education, education policies, marginal treatment effect, policy-relevant treatment effect

*This paper is based on part of my dissertation at the University of Tokyo. I am grateful to my advisor, Hidehiko Ichimura, for his guidance and support. I would also like to thank Hideo Akabayashi, Yuji Genda, Daiji Kawaguchi, Ayako Kondo, Ryo Nakajima, Fumio Ohtake, Hideo Owan, Shinpei Sano, Yasuyuki Sawada, Shintaro Yamaguchi, Kengo Yasui, as well as participants at seminars at the University of Tokyo and Osaka University, and participants at the 2011 HiStat/TCER/Tokyo Labor Economics Workshop Joint Conference on Early Formation of Human Capital at Hitotsubashi University, Kansai Labor Seminar (*Kansai Rodo Kenkyukai*); the 4th Chulalongkorn University and Osaka University Joint Workshop 2016 at Osaka University, ARISH Economics Workshop at Nihon University, Econometrics Workshop; the Third Conference of Education Economics and Policy in Japan at Keio University, and the 2011 Japanese Economic Association Autumn Meeting at the University of Tsukuba for their helpful comments and suggestions on earlier versions of this paper. Any remaining errors are my own. This research was supported by Grant-in-Aid for Research Activity Start-up No. 16H06925 from the Japan Society for the Promotion of Science.

[†]Institute of Social and Economic Research, Osaka University; 6-1, Mihogaoka, Ibaraki, Osaka 567-0047, Japan. E-mail: kikuchi@iser.osaka-u.ac.jp.

1 Introduction

Policy makers and scholars broadly agree about the importance of the government's role in providing equal opportunities for access to higher education. The trend of the university enrollment rate in Japan suggests that this task is progressing satisfactorily: the enrollment rate has risen from around 17% in 1970 to above 50% in 2010, according to the School Basic Survey. In contrast to this absolute increase in enrollment rate, however, providing equal opportunities for access to higher education still requires attention. University enrollment rates are still substantially different between regions, especially between rural and urban regions. Prefectural average enrollment rates range from a low of about 30% in Kagoshima to a high of about 64% in Tokyo in 2015.[1]

Previous studies suggest that the unequal accessibility and affordability of local universities are the main reasons for regional differences in university enrollment rates in Japan. For example, Sasaki (2006), Kobayashi (2009), Nakazawa (2011), and Ueyama (2011, 2012) point out that the number of universities is highly concentrated in urban regions, and show that the capacity of universities in a prefecture is positively correlated with prefectural average enrollment rate. If the number of local universities is insufficient, public supply becomes a key factor in access to higher education. Unfortunately, students also face the burden of high public tuition fees in Japan. For example, OECD (2015) notes that the average annual fee to attend a public tertiary institution in Japan during the 2014–2015 academic year was the highest among OECD countries with available data. Consequently, it is important to question whether education policies address such inequality. It is difficult, however, to answer this question if the returns to schooling are heterogeneous and students simply self-select their schooling based on their idiosyncratic gains. In such a case, supportive policies might not be effective, or may even be inefficient for those who choose not to enroll in university.

The purpose of this paper is quantifying the effects of a counterfactual policy intervention to reduce inequality in the availability of local universities in Japan. For this purpose, I use the program evaluation methods introduced by Heckman and Vytlacil (1999, 2001a, 2005) and Carneiro, Heckman, and Vytlacil (2010, 2011). First, I estimate the probability of university enrollment using the capacity of universities, tuition in public universities, and labor market

[1] University enrollment rates for new upper secondary school graduates. The standard deviation is approximately 8%. All figures are from the author's calculations based on the School Basic Survey.

conditions as instrumental variables (IVs). Second, I estimate the marginal treatment effect (MTE) of Björklund and Moffitt (1987) and Heckman and Vytlacil (1999), using the estimated probability of enrollment. Finally, I estimate the returns to university for individuals induced to enroll in a university by a counterfactual policy intervention as weighted averages of the MTE. Moreover, I examine a set of counterfactual policies in increasing university enrollment: free tuition and an increase in the number of places at local universities. I try to answer policy questions about expanding opportunities for access to affordable universities by examining these counterfactual interventions. To evaluate the effects of university education, I allow a general setting, in which students self-select into university attendance based on their individually heterogeneous returns to schooling.

The contribution of this paper is estimating the impacts of counterfactual policy interventions directly related to actual education policies. The key identification source is variation in capacity of local universities as a supply quantity shifter, changing costs of access to universities. In Japan, the total capacity of universities is regulated by the national government under the School Education Act (*Gakko kyoiku ho*) and the Standards for Establishment of Universities (*Daigaku setchi kijun*). All universities need to pass a one- or two-year screening process by the office of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to change quotas for each department. To capture the exact variation in the availability of local universities, I construct a unique data set that contains information on the opening/closing and increase/decrease of all accredited capacities (quotas) for new students at Japanese universities. I collect this information at department level and total up new student quotas at national, prefectural, municipal, and private co-ed universities by prefecture for each year. This variable is related not only to geographical moving costs or costs of not living at home with parents, but also to the costs of preparation for entrance examinations and the probability of acceptance.

The capacity measure used in this paper reflects both the existence and changes in the size of a university in the area of residence during adolescent years, and therefore more precisely approximates local university availability than the proxies commonly used in the literature. The literature uses the variation in local college accessibility and availability during adolescent years, for example, distance to college (Kane and Rouse, 1993; Carneiro et al., 2011; Doyle and Skinner, 2016; Nybom, 2017), presence of college in the county of residence (Card, 1993; Kling, 2001; Cameron and Taber, 2004; Carneiro, Heckman, and Vytlacil, 2010, 2011; Carneiro, Meghir, and Parey, 2013; Doyle and Skinner, 2016) as a proxy for direct costs of college attendance, and the number of colleges (Currie and Moretti, 2003), the number of campuses (Belskaya, Sabirianova Peter, and Posso, 2014), and the number of admitted students (Kyui, 2016) as more precise measures of local college availability. Because capacity is numeric, I can control birth cohort and prefecture dummy variables when estimating the effects of local capacity on university enrollment decisions. Therefore, the identification of marginal impacts of local capacity is based only on changes within prefecture and year, not on permanent differences between regions over time.

This paper also contributes to the literature on returns to schooling by assessing the causal impacts of university education in Japan. Many previous studies also investigate returns to schooling in Japan. However, Oshio and Seno (2007) and Yasui and Sano (2009) point out that a very limited studies estimate the causal effects of schooling.[2] One exception to this lack is Nakamuro and Inui (2012). Using web-based twin-data, they find significant positive returns to schooling in Japan. Unfortunately, their fixed-effects approach using twins controls for differences in family level only, and thus cannot control for unobserved individual heterogeneity that affects schooling decisions. In this paper, the empirical models are robust to the heterogeneous effects across individuals. To deal with the endogenous schooling decision, I compile data sets on direct and indirect costs of university enrollment for instruments, such as public tuition, capacity of universities, and local labor market conditions.

The main results show that the average effects of a year of university education in Japan are significantly positive, but vary significantly across population subgroups by treatment status. The results also suggest that policies increasing the probability of university enrollment, such as free tuition or an increase in the local capacity of universities, bring about positive effects of university education. However, counterfactual policy simulations of a different level of capacity increase show that larger interventions result in relatively smaller impacts, because individuals with smaller benefits are more affected. An important implication of these results is that the heterogeneity in returns to schooling is key in evaluating an education policy. Impacts of policy interventions depend on who is affected by such policies, thus effects of a general policy that marginally increases the probability of enrollment suggest potential room for policy interventions.

[2]Oshio and Seno (2007) comprehensively survey empirical studies on the economics of education in Japan.

The remainder of this paper is organized as follows. Section 2 explains the empirical framework. Section 3 describes the data. Section 4 presents and discusses the results. Finally, Section 5 concludes.

2 Empirical Framework

The empirical framework of this paper basically follows Heckman and Vytlacil (2005) and Carneiro, Heckman, and Vytlacil (2010, 2011). They consider a standard model of potential outcomes that is firstly applied to schooling by Willis and Rosen (1979).

2.1 Setup

Let consider a linear in the parameters model with two potential outcomes:

$$\begin{aligned} Y_1 &= X\beta_1 + U_1 \\ Y_0 &= X\beta_0 + U_0, \end{aligned} \tag{1}$$

where subscript 0 and 1 correspond to the untreated and treated states. Y_1 is the potential log wage if the individual enrolls in university, and Y_0 is the potential log wage if the individual does not enroll in university. X is a vector of observable characteristics and (U_1, U_0) are unobservable variables. Let $D = 1$ denotes enroll in university; $D = 0$ denotes not enroll in university. The measured outcome variable Y can be written in a potential outcomes framework:

$$Y = DY_1 + (1 - D)Y_0 \tag{2}$$

This equation is related to a latent variable discrete choice model that represents an individual's decision to enroll in a university. I assume the following selection model:

$$
\begin{aligned}
D^* &= \mu_D(Z) - V \\
D &= 1 \;\text{ if }\; D^* \geq 0, \\
&= 0 \;\text{ otherwise,} \\
\text{or }\; D &= \mathbf{1}[\mu_D(Z) - V \geq 0],
\end{aligned}
$$

where V is an unobserved random variable. I assume that V is continuous with a strictly increasing distribution function F_V. Z is a vector of observed random variables that includes some part of X. Z also includes variables that determine the treatment decision but do not directly affect the outcome (the exclusion restriction). I also assume that Z and X are independent of (V, U_1, U_0).

Let $Pr(D = 1|Z)$ denotes the probability of university enrollment conditional on Z. It is innocuous to rewrite the the selection equation for convenience as follows:

$$
\begin{aligned}
D &= \mathbf{1}[F_v(\mu_D(Z)) \geq F_v(V)] \\
&= \mathbf{1}[P(Z) \geq U_D],
\end{aligned}
$$

with

$$
\begin{aligned}
U_D &\stackrel{\text{def}}{=} F_v(V) \sim Unif[0,1] \\
P(Z) &\stackrel{\text{def}}{=} F_v(\mu_D(Z)) = Pr[D = 1 \mid Z].
\end{aligned}
$$

Different values of U_D correspond to different quantiles of V, and thus U_D is the normalized latent variable of the unobserved resistance to enroll in university.

Using equations 1 and 2, the observed outcome can be written as:

$$
\begin{aligned}
Y &= X\beta_0 + DX(\beta_1 - \beta_0) + D(U_1 - U_0) + U_0 \\
&= X\beta_0 + DX\beta + \epsilon.
\end{aligned}
\tag{3}
$$

This equation indicates that the effect of university enrollment varies across individuals for differences in their X and U_1, U_0. If the enrollment decision depends on the unobservable gain $U_1 - U_0$, a dummy variable D is not independent of the disturbance ϵ. In this case, neither ordinary least squares (OLS) nor simple linear IV estimates recover the standard average effect parameters, such as the average treatment effect (ATE): $E(Y_1 - Y_0)$, the ATE on the treated (ATT): $E(Y_1 - Y_0 \mid D = 1)$, and the ATE on the untreated (ATUT): $E(Y_1 - Y_0 \mid D = 0)$. Heckman and Vytlacil (1999, 2001a, 2005) establish that these treatment effect parameters of interest can be identified as weighted averages of the MTE of Björklund and Moffitt (1987) and Heckman and Vytlacil (1999).[3] The MTE is defined as:

$$MTE(x, u_D) \stackrel{\text{def}}{=} E(Y_1 - Y_0 \mid X = x, U_D = u_D).$$

The MTE indicates the effects of university enrollment for individuals with $X = x$ who would be indifferent between enrollment or not if they were exogenously assigned a value of Z such as $U_D = u_D$.

2.2 Estimating Marginal Treatment Effect

Heckman and Vytlacil (1999, 2001a, 2005) show that the MTE can be identified by the local instrumental variables. Using the equation 3, the conditional expectation of Y given $X = x$, and $P(Z) = p$ is

$$\begin{aligned} E[Y|X = x, P(Z) = p] &= x\beta_0 + x(\beta_1 - \beta_0)p + K(p) \\ &= x\beta_0 + \int_0^p MTE(x, u_D) du_D, \end{aligned} \tag{4}$$

where,

$$\begin{aligned} K(p) &= E(U_1 - U_0 \mid D = 1, P(Z) = p) \\ &= \int_{-\infty}^{\infty} \int_0^p (u_1 - u_0) f(u_1 - u_0 \mid X = x, U_D = u_D) du_D d(u_1 - u_0), \end{aligned}$$

[3]Definitions of weights for the treatment effect parameters are in Appendix B.

where $f(u_1 - u_0 \mid X = x, U_D = u_D)$ is the conditional density of $U_1 - U_0$. Therefore, the MTE is identified by differentiating the equation 4 with respect to p,

$$\frac{\partial}{\partial p} E[Y \mid X = x, P(Z) = p] = E[Y_1 - Y_0 \mid X = x, U_D = u_D]. \tag{5}$$

The equation 4 can be estimated using the model of a semiparametric approach, such as the partially linear model of Robinson (1988), and the equation 5 can be estimated in nonparametrically. One of the disadvantages of a semiparametric approach is that the MTE can only be estimated over the empirical support of $P(Z)$, because the MTE is identified over the support of $P(Z)$. If the empirical support does not cover full unit interval of $P(Z)$, it is impossible to recover conventional treatment parameters. An alternative way of estimating the MTE is a parametric approach that assumes joint normal distribution of the unobservables (U_0, U_1, V). With this additional assumption, Heckman, Tobias, and Vytlacil (2001) show that the MTE can be written as follows:

$$MTE(x, u_D) = x(\beta_1 - \beta_0) - (\sigma_{1V} - \sigma_{0V})\Phi^{-1}(u_D), \tag{6}$$

where, $\sigma_{jV} = Cov(U_j, V)$, $j = 0, 1$, and $\Phi(\cdot)$ is CDF of standard normal. This parametric approach is a conventional way of estimating the equations and is related to switching regression models (Björklund and Moffitt, 1987; Willis and Rosen, 1979).[4] Following the literature, I estimate the outcome and selection equations together using the method of maximum likelihood.[5] An advantage of specifying the normality assumption is that it helps to estimate the MTE over the full unit interval of $P(Z)$ and to recover the treatment effect parameters of interest.

2.3 Policy-Relevant Treatment Effects

Once the MTE is estimated, the parameters that are directly relevant to the policy questions can also be estimated as weighted averages of it. I compute the policy-relevant treatment effects (PRTE) introduced by Heckman and Vytlacil (2001b) and marginal version of PRTE (MPRTE)

[4]See also Heckman, Tobias, and Vytlacil (2001) for further description.

[5]In the estimation, I normalize the variance of V to 1.

proposed by Carneiro, Heckman, and Vytlacil (2010). Let D^*, Y^*, and P^* denote the treatment state, outcome, and probability of university enrollment after a policy change. Heckman and Vytlacil (2005, 2007) define the PRTE when $E(D^*) \neq E(D)$ as,

$$\frac{E(Y^*) - E(Y)}{E(D^*) - E(D)} \stackrel{\text{def}}{=} \int_0^1 MTE(u_D)\omega_{PRTE}(u_D)du_D,$$

where,

$$\omega_{PRTE}(u_D) = \frac{F_p(u_D) - F_{p^*}(u_D)}{E_{F_{p^*}}(P) - E_{F_p}(P)},$$

where F_p^* and F_p are the distribution of P^* and P, respectively.[6]

The MPRTE is defined as the limit of PRTE with a sequence of alternative policies indexed by a scalar variable α such that $lim_{\alpha \to 0} P_\alpha^*(Z) = P(Z)$. I consider three policy sequences as defined in Carneiro, Heckman, and Vytlacil (2010, 2011): (1) a policy intervention that has an effect similar to a shift in one of the components of Z, say Z^k, such that $Z_\alpha^k = Z^k + \alpha$ and $Z_\alpha^j = Z^j$ for $j \neq k$; (2) a policy that increases the probability of university enrollment by α so that $P_\alpha^* = P + \alpha$; and (3) a policy that changes each individual's probability of university enrollment by the proportion $(1 + \alpha)$, so that $P_\alpha^* = (1 + \alpha)P$.

3 Data

3.1 The Japanese General Social Survey

The main analysis data are the Japanese General Social Survey (JGSS).[7] The JGSS is repeated cross-section data for men and women aged 20–89, as of September 1 of each survey year. This

[6] To simplify the notation, I suppress control variables.

[7] The Japanese General Social Surveys (JGSS) are designed and carried out by the JGSS Research Center at Osaka University of Commerce (Joint Usage / Research Center for Japanese General Social Surveys accredited by Minister of Education, Culture, Sports, Science and Technology), in collaboration with the Institute of Social Science at the University of Tokyo. The data for this secondary analysis, the JGSS, the JGSS Research Center, was provided by the Social Science Japan Data Archive, Center for Social Research and Data Archives, Institute of Social Science, The University of Tokyo.

paper uses data from surveys conducted in 2000, 2001, 2002, 2005, 2006, 2008, and 2010, and pools male respondents from all waves.

From the pooled original data, I exclude observations in the following four steps. First, I limit the sample by age. I exclude individuals who were younger than 28 years old in the survey year, because they might not have completed their academic schooling. Second, I drop individuals who answered that either of their parents was absent at the age of 15, because their single-parent structure might substantially differ from a family with two parents, and thus, unobservable effects cannot be controlled. Third, I only use individuals who had reached their first university enrollment decision after the 1972 academic year due to the availability of tuition data. Finally, I use individuals whose observational characteristics match the comparison information on the instruments and covariates for the estimation explained below. The remaining sample after restrictions contains male workers were born between 1953 and 1979 who were 28 to 54 years old at the date of the survey.

The JGSS has the advantage of including information about workers' annual earned income and working hours per week. For the outcome variable, I calculate the worker's individual hourly wage as the worker's annual income from his main job divided by annual working hours.[8] Unfortunately, the JGSS measures income in terms of 19 categories. Following Oshio and Kobayashi (2009) and Sano and Yasui (2009), I assign the median value of each category and evaluate it in 2005 consumer prices, transforming it into a logarithm.[9]

One of the disadvantages of the JGSS is that the data set contains limited information on residence at the time of the university enrollment decision. I use information on the prefecture where individuals resided at the age of 15 and assume that in the year of university enrollment, their residence (at least their parental residence) was still in that prefecture. The control variables from the JGSS are parents' education level (and their squares), number of siblings (and its square), age at the survey date (and its square), urban residence dummy and rural (farm or fishing village) residence dummy at the age of 15, a set of dummies for prefecture resided in at the age of 15, cohort dummies, and survey year dummies. I do not control for experience, which is commonly included in the literature, because experience is endogenous and thus captures a

[8] Annual working hours are defined as reported working hours per week $\times$ 52.

[9] For the lowest category (less than 700,000 Japanese Yen [JPY]), I assign 700,000. For the highest category (over 23,000,000 JPY), I assign 23,000,000. When I exclude the people in the lowest and the highest categories, the estimated results are basically same.

part of the returns to schooling. For the educational attainment of respondents and their parents, I use the last school attended and assign the standard years of schooling in Japan.

I also control for long-term trends in the active job openings to applications ratio (*yuko kyujin bairitsu*) from the reports of the Employment Service Agency (*Shokugyou Antei Gyoumu Toukei*) and the annual average monthly total cash earnings of the Monthly Labor Survey (*Maigetsu Kinro Tokei Chosa*), and estimated population size at ages 15–19 from the Population Census (and its square) for the prefecture where individuals resided when they were 15 years old.

In the analysis, I consider the binary treatment decision for university enrollment at the completion of upper secondary education. Therefore, I separate individuals into two groups: (1) individuals who graduated from high school or completed upper secondary education, and (2) individuals who had at least some university education.[10]

3.2 Instrumental Variables

This paper uses four IVs that capture differential changes in direct and opportunity costs of university attendance across prefectures and cohorts, while controlling for both permanent differences and aggregate trends. The IVs are local university availability measures covering all universities, average tuition at a public university in the fresh-man year, and local labor market conditions in the high school years, in the prefecture where the individual resided when he was 15 years old.[11]

3.2.1 Capacity of Universities

Local college availability measures are first used by Card (1993) and Kane and Rouse (1993) as a proxy for direct costs of college attendance, and are subsequently widely used in the literature. Kane and Rouse (1993) use a distance to college measure as an instrument for schooling, and followed by Carneiro et al. (2011), Nybom (2017), and Doyle and Skinner (2016) in recent

[10]For respondents, completion of their last level of schooling is available. When I construct the variable of years of schooling, I reduce the number by one year from the standard years for those who dropped out before finishing their last school. However, I am unable to know whether technical college dropouts completed their upper secondary education. In the analysis, I assume that they have completed their upper secondary education and include them in the analysis sample. If I exclude them from the analysis sample, the results are similar.

[11]The construction of these data is amply described in Appendix A.

years. An indicator of presence of college in the county of residence is used by Card (1993) as a substitute for distance to college, and is commonly used in the literature, for example, by Kling (2001), Cameron and Taber (2004), Carneiro, Heckman, and Vytlacil (2010, 2011), Carneiro, Meghir, and Parey (2013), and Doyle and Skinner (2016). Unfortunately, the JGSS only has information on the prefecture of residence at the age of 15, and because during the analyzed years, all prefectures in Japan contain at least one university, I am unable to use this local presence measure.

Currie and Moretti (2003) use the number of two- and four-year colleges and Belskaya, Sabirianova Peter, and Posso (2014) use the number of campuses as instruments for college attendance. These measures are superior to the indicator definition because their continuous variations across residential areas and years allow the researcher to control for permanent differences across residential areas. Although the number of colleges or campuses measure takes into account quantitative differences in college availability across residential areas to a degree, it is too rough to precisely capture differences in local opportunities for college education. Because colleges differ in size, each college has different effects on local availability of colleges. The number of admitted students (e.g. Kyui, 2016) and the number of enrollments are also widely used as a substitute for capacity of colleges.[12] However, as Currie and Moretti (2003) point out, not only the supply of college places but also the demand for these places determines these variables, and thus number of admitted students or enrollment might not be a valid instrument in this case.

In this paper, I construct a unique data set that contains information on all accredited capacity (quotas) for new university students by the national government in Japan. I collect the information at the department level and total up new student quotas of national, prefectural, municipal, and private co-ed universities by prefecture in each year. This measure is merged with the individual data based on the high school graduate's standard college examination year and the residence at the age of 15.[13] I assume that this measure of local availability of universities is a proxy of easy access to a local university; i.e., costs of geographical moving or costs of not living at home with parents. This measure also reflects the potential costs of preparation

[12]For example, Sasaki (2006), Kobayashi (2009) use the number of enrollment as a proxy of the potential capacity of a prefecture. Doyle and Skinner (2016) also use distance-weighted enrollment for an instrument to analyze postsecondary education in United States.

[13]When I merge the year at age of 18, the acquired results are basically same.

for taking examinations and the probability of acceptance.[14] I also control cohort dummies and prefecture dummies when I estimate the effect of capacity on university enrollment. Therefore, the identification relies on unexpected changes of capacity within local prefecture and year of the university enrollment decision, but not on permanent differences.

One potential problem with this capacity measure is that changes in cohort size are likely to have an impact on the availability of universities given any fixed amount of capacity (Card and Lemieux, 2001; Currie and Moretti, 2003). To avoid this problem, I control for local cohort size at ages 15–19 when the individual was aged 18, in both selection and outcome equations.

Another concern regarding the use of this measure is that it is affected by new institution openings or by increase in size of a university reflecting an expected increase in local demand for university education.[15] Although I am unable to completely rule out these possibilities, the Japanese centralized educational system to some degree mitigates this concern: The School Education Act (*Gakko kyoiku ho*), prior to the revision in 2003, prescribed that all openings and closings of university departments are required to be approved in advance by the national government. Private universities also need approval in advance for changes in capacity at the department level.[16] To increase capacity, a university needs to pass a one or two years screening and investigative process including a preliminary interview by the office of the MEXT based on the Standards for Establishment of Universities (*Daigaku setchi kijun*). Therefore, no university can freely control its capacity in response to the expected local demand for university education.

3.2.2 Average Tuition at Public Universities

The tuition measure is created as accredited capacity weighted averages over all public co-ed universities in a prefecture, or at the regional level if there is no public university in the prefecture. Kane and Rouse (1993), Cameron and Heckman (1998, 2001), Carneiro, Heckman, and Vytlacil (2010, 2011), Carneiro, Meghir, and Parey (2013), and Doyle and Skinner (2016) use tuition as

[14] Although I integrate the capacity of all universities into one measure, changes in capacities might have heterogeneous effects on college attendance by student's major field of study. In this paper, because the available data set has no information on the major field of study, I cannot analyze the impacts of college major choice in detail, which I leave to future research.

[15] Currie and Moretti (2003) point out this concern in the case of number of colleges.

[16] Public universities need to notice their changes in capacity in advance. Before 1974, private universities are allowed to increase their capacity with a notification to the government in advance, therefore capacity of private universities might not be a valid instrument in these periods. For a robustness check, I exclude individuals who were born before 1955 and re-estimate the analysis. The results are almost similar but more imprecise for the smaller sample size.

an instrument to predict college attendance.

One concern regarding this approach is that the tuition variable is highly correlated with the quality of the university (Cameron and Heckman, 2001). If the measure captures both cost of college attendance and college quality, it directly affects wage differences. To mitigate this concern, I only include entrance and course fees at prefectural and municipal universities in the tuition measure.[17] These are specified by local governments based on the fees set by the national government for national universities, and individual universities are not permitted to change the amount charged.[18] Therefore, the public tuition measure relies on fees determined at local government level, and basically reflects a variation at the prefecture level. I do not include other fees in public tuition, because each university is allowed to determine the amount of such fees at the department level. The same thing applies to tuition at private universities. It varies from university to university, and thus might reflect the quality of the university. I exclude tuition of national universities from the public tuition measure because it was unified across the country until 2003. The differences in tuition at national universities are captured by the cohort dummies.

Using the local tuition at prefecture level, I presume that the variable influences schooling choice of the individual. One could argue that individuals might move to a different prefecture for their university education to avoid high tuition costs at local public universities (Carneiro, Meghir, and Parey, 2013). However, it appears reasonable to consider that prefectural variation matters in Japan. Because prefectural and municipal universities usually set a lower tuition for intra-regional students, movers are not only prevented from the option of living at home, but also disadvantaged by paying higher tuition fees as extra-regional students.

3.2.3 Local Labor Market Conditions

Labor market conditions in the high school years are used as instruments to assess university enrollment decisions. In Japan, previous studies find that temporary shocks in the labor market significantly affect schooling decisions, by using aggregate data.[19] I use the active job openings

[17] In Japan, new university enrollments require the payment f entrance fees, course fees, and other fees for school expenses. Other fees inclued, for example, training fees, fees for facilities and equipment.

[18] Before 2003, entrance and course fees at national universities were specified by the national government. In 2004, national universities were incorporated by the the National University Corporation Act (*Kokuritsu Daigaku Hojin Ho*).

[19] For example, see Yano and Hamanaka (2006), Ueyama (2011), and Ogawa (2015).

to applications ratio and the annual average monthly total cash earnings in the prefecture of residence at age 15. Local labor market conditions are used as instruments by Cameron and Heckman (1998), and followed by Cameron and Taber (2004), Arkes (2010), Carneiro, Heckman, and Vytlacil (2010, 2011), and Carneiro, Meghir, and Parey (2013) among others.

I construct these measures as a three-year average over the high school period and merge them with the individual data for the year the individual was 18. I presume that local earnings capture temporary shocks to family income. Local job openings reflect the speed of job transition or of finding a job if unemployed, and thus they also are related to temporary variation in family resources. In addition, local earnings might capture foregone earnings as opportunity costs of additional schooling. A potential problem of using local labor market conditions is that long-run trends of labor market conditions might affect both these measures and residential choice at age 15. If local active job openings to applications ratio and local earnings in high school are correlated with the unobservables in the outcome equations, these would not be valid instruments. To avoid this concern, I include trends in local labor market conditions at ages 13–18 averaged over 6 years in both selection and outcome equations to control for residential choice at age 15 and long-run differences in labor market conditions in the prefecture of residence during adolescent years. Additionally, I control for a set of dummy variables of the prefecture of residence at age 15, allowing for permanent or aggregate differences in prefecture characteristics.

Table 1 shows summary statistics for the instruments, with outcomes and covariates. It shows that individuals with some university education have, on average, higher wages than those without university education. The difference between the two groups is about 3.81 years of schooling. Using this figure, all estimates of treatment effects reported below are annualized. Individuals with some university education have more-educated parents, fewer siblings, and have lived in prefectures with better labor market conditions in both adolescence and the survey year. Their residence at the age of 15 was more likely to be in an urban area and in a prefecture with larger number of adolescents. They are less likely to come from rural areas than those without university education. Regarding the instruments, individuals with some university lived in prefectures where with greater university capacity and better labor market conditions when they were in high school or upper secondary education. However, these individuals also lived in prefectures where the average tuition of public universities was higher than among those without university education.

4 Results

4.1 University Enrollment Decision

Table 2 presents the estimates of the selection model of university enrollment. I estimate the probit model where the dependent variable is a dummy variable that equals one if the individual has ever attended university, and I report the marginal effects at the mean value for each variable. All controls reported in the table perform well, and estimates show the expected signs reported in previous studies except for the father's years of schooling. For example, individuals who have a mother with more years of schooling are more likely to enroll in university than those whose mother has lower levels of education. The instruments are jointly strong predictors of university enrollment, although local active job openings to applications ratio (local job openings) is not individually significant. Local capacity of universities is an important determinant of university enrollment. If local capacities increase by 1,000 places, the probability of university enrollment increases by about 1%. Local tuition in public universities also has a statistically significant effect on university enrollment. If local tuition rises by 10,000 JPY, it decreases university enrollment by about 0.7%. Local earnings play the role of an opportunity cost variable for university enrollment. If local earnings averaged over high school years increase by 1%, university enrollment decreases by about 2.5%. Finally, a better local active job openings to applications ratio at university enrollment increases probability of university enrollment. Specifically, a 0.1 point improvement in local active job openings to applications ratio in high school period increases probability of university enrollment by about 1.3%, but this in not a statistically significant effect.

Using the estimates of the selection model, Figure 1 shows the distributions of predicted probability of university enrollment, $P(Z)$, for those with university and for those without university. This figure indicates the empirical common support of $P(Z)$ over which the MTE can be identified in the semiparametric model.[20] Although the overlapping support of predicted

[20]Following the literature, I define common support as the intersection of empirical support of these distributions.

$P(Z)$ for these two distributions covers a wide range of unit interval, it lacks top and bottom tails.[21] Therefore, it is necessary to estimate the model with joint normal assumptions to recover the treatment effects that need full support.

4.2 OLS and IV

In Table 3, I present standard OLS and IV estimates, and compare them with the estimates in previous studies using these methods. The OLS estimate shows that annualized returns to university education are about 5.49%. The magnitude of the OLS estimate in this paper is much smaller than the estimates derived in previous studies using Japanese micro-data. These studies report around 7–11% of OLS estimates of returns to schooling (e.g., Tachibanaki, 1988; Trostel, Walker, and Woolley, 2002; Ono, 2004; Sano and Yasui, 2009; Yasui and Sano, 2009; Nakamuro and Inui, 2012). The differences in the magnitudes of the literature and this paper's estimates might be explained by the differences in sample restriction and in model specification. There are three main differences. First, I only use male observations. Trostel, Walker, and Woolley (2002), Sano and Yasui (2009), and Yasui and Sano (2009) suggest a larger magnitude of the returns for females. I exclude females from the analysis to avoid selection bias from labor market participation decisions, as including female observations may provide larger estimates of returns to schooling. Second, I consider binary treatment of university enrollment, thus excluding individuals with less than upper secondary education. It is possible to consider a model with multiple levels of treatment or with continuous treatment of schooling, but more restrictions and additional instruments would be needed to identify the treatment parameters of interest.[22] Finally, I specify the model with observational heterogeneity by interacting the treatment and control variables. Such interaction terms are not included in previous studies. Controlling for this observable heterogeneities might weaken the magnitude of the coefficient of university enrollment.

Table 3 shows that the IV estimate of returns to university education is 11.63%. In line with the literature, the IV is larger than the OLS estimate. Card (2001) suggests that such a finding indicates returns to schooling are heterogeneous and higher for individuals who are induced

[21] The empirical common support is $0.06951 \leq P(Z) \leq 0.94474$. I restrict the empirical estimates to the common support when estimating the model with semiparametric assumptions.

[22] See, for example, Heckman, Urzua, and Vytlacil (2006), Heckman and Vytlacil (2007), and Florens et al. (2008).

to enroll in university by changes in the instruments than for those who have average returns. This interpretation is related to the local average treatment effect (LATE) parameter of Imbens and Angrist (1994). However, the IV estimates do not necessarily reflect the original LATE parameter if instruments are multiple and the model includes a set of controls, as is the case in this paper.[23] Interpreting an IV estimate is not always straightforward. Even fort LATE, this indicates the policy effect of interest only if the variation of the instrument corresponds exactly to the policy variation (Heckman, Urzua, and Vytlacil, 2006; Carneiro, Heckman, and Vytlacil, 2011).

4.3 Marginal Treatment Effect

Here, I estimate the MTE with both parametric and semiparametric approaches. In Figure 2, I estimate the MTE assuming joint normality of (U_1, U_0, V) and plot it with 90% confidence interval bands. To depict the MTE depending only on U_D, I evaluate it at the mean values of control variables. When the U_D is particularly low, for individuals who are more likely to enroll in university, the marginal effects are high at around 40%, but for those who have low values of U_D, unwilling to enroll in university, the effects are substantially lower around −20%. These individuals incur negative gains from attending university. The MTE monotonically declines as U_D increases, suggesting substantial heterogeneity of marginal effects of university education. A simple test of selection on unobserved heterogeneity is a test of whether the slope of the MTE is zero. For the normal selection model this is equivalent to a test of whether $\sigma_{1v} - \sigma_{0v} = 0$ in the equation 6. I estimate that $\sigma_{1v} - \sigma_{0v} = -0.3064$ with a standard error of 0.1192. This supports that the effects are heterogeneous and students decide to enroll or not with at least partial knowledge of their idiosyncratic gains. Carneiro, Heckman, and Vytlacil (2011) note that these results imply individuals self-select university enrollment based on their comparative advantage with respect to their gains.

Figure 3 shows the semiparametrically estimated MTE with 90% confidence interval bands computed from the bootstrap.[24] The semiparametric estimates produce the same pattern as the

[23]See Angrist and Imbens (1995) and Heckman and Urzua (2010) for their discussion on the identification and interpretation of IV estimates.

[24]Following Carneiro, Heckman, and Vytlacil (2011), I use Robinson (1988)'s method for estimating the MTE. I run the Nadayara-Watson kernel regressions of Y and X on P using a Gaussian kernel and a bandwidth of 0.04, and estimate the linear term using a linear regression with 1.5% trimming. Then, I run a local quadratic regression to estimate $K(P)$ using a Gaussian kernel and a bandwidth of 0.35. The results are robust to choices of both larger

parametric estimates, implying that estimates of the MTE are free from parametric assumption. Figure 3 suggests that unobservable heterogeneity has an important role in the effects of university education. From the value of more than 40% for those with low U_D, the MTE monotonically decreases to less than 0% for those with high U_D. Unfortunately, the confidence intervals on the semiparametrically estimated MTE are too wide, and thus include negative effects for low U_D and positive effects for high U_D.

4.4 Average and Policy-Relevant Treatment Effects

The IV and MTE estimates show that university education has substantial and heterogeneous effects on future pecuniary outcome. However, it is not clear how large the average impacts are for different subpopulations, or how these effects are related to educational policies. To examine these issues, I show the treatment effect parameters in Table 3, which are constructed from the MTE using the weights presented in Heckman and Vytlacil (2005) and Carneiro, Heckman, and Vytlacil (2010, 2011).

Average effects are substantially different by subpopulation group. The ATE shows that an additional year of university education increases hourly wages by 8.89%. The ATT is larger than the ATE, and suggests that the return to one year of university education is 17.43% for those who enrolled in university. The ATUT is much smaller than the ATE and the ATT, and shows that the effect is 1.67% for those who did not enroll in university, that is, the gain if they had enrolled.

Conventional average treatment effects parameters are important by themselves, but these parameters only answer policy questions in extreme cases. For example, the ATUT indicates the effects of a policy forcing an entire population to receive university education. In contrast, the MPRTE parameter of Carneiro, Heckman, and Vytlacil (2010, 2011) answers questions about marginal gains from specific policies in a more general case. Table 3 presents estimates of different definitions of the MPRTE, where the policy is considered as a marginal change in the probability of university enrollment. The MPRTE of a policy intervention that has an effect similar to a shift in one of the components of Z, is 9.46%. The MPRTE with a policy that increases the probability of university enrollment by an amount α, is 9.39%. A policy

and smaller bandwidths.

that changes the probability of university enrollment by the proportion $(1 + \alpha)$ provides slightly smaller effects. The MPRTE of such a policy is 6.64%, but standard errors are large.

I also calculate the average returns for those induced to enroll in university by a particular policy shift as a counterfactual policy simulation. Table 3 reports the PRTE of two counterfactual policies: (1) a policy of free tuition in public universities and, (2) a policy that increases university capacity by 500 places if the number of places in the prefecture is less than the sample median value of places. The PRTE of free public tuition has an effect of 8.17%, which is similar to the magnitude of the ATE. The PRTE of increasing capacity suggests a larger impact, 10.73%. A free public tuition policy is a large intervention, but it provides relatively smaller effects than that of the capacity intervention. This is because it heavily weights on those with smaller MTE. Figure 4 shows the weights on MTE for the PRTE parameters, all evaluated at the mean of X. The capacity policy weights mainly the middle section of the MTE. The free public tuition policy over-weights individuals with high values of U_D because its effect on enrollment is larger for those with already high levels of P. These results suggest that the magnitude of the PRTE is related to how a population is targeted. To further investigate this point, I estimate the distribution of the PRTE by different capacity-increasing policies. In this policy simulation, I increase capacity by Δ places if the number of places in the prefecture is less than the π-percentile value, where Δ takes 200 to 1000 places, and π takes 10 to 100 percentile. Figure 5 shows that the magnitude of PRTE as decreasing if the intervention level is increasing. Specifically, the PRTE reduces when the population affected increases.

Finally, I semiparametrically estimate the treatment effect parameters to examine the robustness of the estimates with the parametric assumptions. These parameters cannot be identified semiparametrically because the empirical support of $P(Z)$ lacks the full unit interval. However, these parameters can be estimated over the empirical support when I restrict the weights to integrate to one in the support. In Table 3, the semiparamteric estimates suggest substantial heterogeneity in the treatment effects of university education. The ATE is 9.18%, which is similar to the magnitude of the parametric estimate. The ATT is 28.32%, which is much larger than the parametric estimate. The ATUT now shows a large negative effect, -8.15%. The MPRTE that has an effect similar to a shift in one of the components of Z, is 9.35%. The MPRTE with a policy that increases the probability of university enrollment by an amount α, is 9.47%. These estimates are very similar to the parametric estimates. These results validate the advantage of

MPRTE: the identification does not require a large support condition of $P(Z)$. The MPRTE with a policy that changes the probability of university enrollment by the proportion $(1 + \alpha)$ has a much smaller effect than the parametric estimate. The effect of such a policy is negative, -3.10%. The PRTE of two counterfacutual policies are also examined. The semiparametric estimates of PRTE are slightly larger than the parametric estimates. The PRTE of free public tuition is 10.07%. The PRTE of increasing capacity is 11.37%. In general, the semiparametric estimates produce the same pattern as the parametric estimates. This suggests that the parametric estimates are basically free from the assumptions, although the semiparametrically estimated standard errors are too large to draw a confident conclusion.

5 Conclusion

This paper investigates the efficacy of a counterfactual policy intervention that reduces inequality in availability for local universities in Japan. For this objective, I estimate the MTE using instruments that reflect the direct costs of college attendance: total accredited capacity of all universities and public tuition in the prefecture of residence at age 15. These measures capture cross-time and cross-prefecture variation, because I also control for a set of birth cohorts and prefecture dummies. I use an empirical framework that students' self-select based on their heterogeneous returns, and I recover the average and marginal returns to university education as weighted average of MTE. Specifically, I analyze the impacts of a set of counterfactual simulations as the PRTE, which show the effects of university education for those induced to enroll in university due to a counterfactual policy change.

The main results of this paper are as follows. First, MTE varies across individuals. This implies substantial heterogeneity of the marginal effects of university education. The test of whether the slope of the MTE is zero for the normal selection model shows that $\sigma_{1v} - \sigma_{0v} = -0.3064$. This figure suggests that heterogeneity in Japan is larger than in the US (Carneiro, Heckman, and Vytlacil, 2011) or Sweden (Nybom, 2017), but smaller than in Russia (Belskaya, Sabirianova Peter, and Posso, 2014).

Second, the estimates of a conventional ATE parameters show that additional university schooling increases hourly wages. The ATE is 8.89%, which indicates the effects for the population. Additionally, this paper finds heterogeneous average effects by subpopulation group.

The average effect for those enrolled in university, ATT, is 17.43%, but for the ATUT, the effect is 1.67% for those who did not enroll.

Finally, I find that counterfactual policy interventions increasing the probability of university enrollment provide significant positive effects of university education. The MPRTE regarding a general policy that marginally increases enrollment probability also show the positive effects of university education, but the effects are substantially different by affected population. However, the results of the PRTE of a different level of capacity increase show that larger interventions result in relatively smaller impacts, because individuals with smaller the MTE are more affected. These results imply that there is a potential room for policy interventions in university enrollment opportunities, as individual benefits of university education induced by such policy interventions are unstable due to the heterogeneity in returns to schooling.

I acknowledge that there are several limitations of the present paper. First, the program evaluation framework rules out general equilibrium effects or peer effects. For example, if there is a large increase in capacity of university by a policy change, then there is a large increase in the supply of skilled labor. This lowers the relative price of skilled labor, and thus violates the assumption of identification. As such, the distribution of potential outcomes is changed by a policy intervention. Therefore, the results of the PRTE might overstate the impacts of policy interventions. The estimates of the MPRTE are relatively robust to this violation, because the MPRTE evaluates marginal gains corresponding to a marginal change in the probability of university enrollment by a policy shift. Second, I paid less attention to the quality or heterogeneity of university due to data limitations. The estimation model implicitly assumes that the marginal impact of local capacity is homogeneous across the individuals, conditional on the control variables. This is an invalid assumption if marginal impacts of local university increases are heterogeneous across students' abilities. This paper is unable to control for the respondent's ability measures or skill measures relating to cognitive and noncognitive skills, and thus the estimates might include the so-called ability bias. However, in Japan, before the national achievement test held in 2007, there was no uniform criteria-based national academic indicator for students under the secondary educational level. In addition, students had no obligation to take an IQ test or tests for personal traits. Therefore, it is difficult to analyze the role of ability in detail. Nybom (2017) shows the returns to schooling substantially vary with respect to abilities in Sweden. The heterogeneous effects of policy intervention in local college availability by degree of students' ability may be a fruitful topic for future research.

References

Angrist, Joshua D. and Guido W. Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association* 90 (430):431–442.

Arkes, Jeremy. 2010. "Using Unemployment Rates as Instruments to Estimate Returns to Schooling." *Southern Economic Journal* 76 (3):711–722.

Belskaya, Olga, Klara Sabirianova Peter, and Christian Posso. 2014. "College Expansion and the Marginal Returns to Education: Evidence from Russia." IZA Discussion Paper 8735, Institute for the Study of Labor (IZA).

Björklund, Anders and Robert Moffitt. 1987. "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models." *Review of Economics and Statistics* 69 (1):42–49.

Cameron, Stephen V. and James J. Heckman. 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *Journal of Political Economy* 106 (2):262–333.

———. 2001. "The Dynamics of Educational Attainment for Black, Hispanic, and White Males." *Journal of Political Economy* 109 (3):455–499.

Cameron, Stephen V. and Christopher Taber. 2004. "Estimation of Educational Borrowing Constraints Using Returns to Schooling." *Journal of Political Economy* 112 (1):132–182.

Card, David. 1993. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." Working Paper 4483, National Bureau of Economic Research, Cambridge, MA.

———. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69 (5):1127–1160.

Card, David and Thomas Lemieux. 2001. "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis." *Quarterly Journal of Economics* 116 (2):705–746.

Carneiro, Pedro, James J. Heckman, and Edward Vytlacil. 2010. "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin." *Econometrica* 78 (1):377–394.

———. 2011. "Estimating Marginal Returns to Education." *American Economic Review* 101 (6):2754–2781.

Carneiro, Pedro, Michael Lokshin, Cristobal Ridao-Cano, and Nithin Umapathi. 2011. "Average and Marginal Returns to Upper Secondary Schooling in Indonesia." IZA Discussion Papers 6162, Institute for the Study of Labor (IZA).

Carneiro, Pedro, Costas Meghir, and Matthias Parey. 2013. "Maternal Wducation, Home Environments, and the Development of Children and Adolescents." *Journal of the European Economic Association* 11 (s1):123–160.

Currie, Janet and Enrico Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings." *Quarterly Journal of Economics* 118 (4):1495–1532.

Doyle, William R. and Benjamin T. Skinner. 2016. "Estimating the Education-Earnings Equation Using Geographic Variation." *Economics of Education Review* 53:254–267.

Florens, Jean-Pierre, James J. Heckman, Costas Meghir, and Edward Vytlacil. 2008. "Identification of Treatment Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects." *Econometrica* 76 (5):1191–1206.

Heckman, James J., Justin L. Tobias, and Edward Vytlacil. 2001. "Four Parameters of Interest in the Evaluation of Social Programs." *Southern Economic Journal* 68 (2):210–223.

Heckman, James J. and Sergio Urzua. 2010. "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify." *Journal of Econometrics* 156 (1):27–37.

Heckman, James J., Sergio Urzua, and Edward Vytlacil. 2006. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *Review of Economics and Statistics* 88 (3):389–432.

Heckman, James J. and Edward Vytlacil. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences* 96 (8):4730–4734.

———. 2001a. "Local Instrumental Variables." In *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, edited by Cheng Hsiao, Kimio Morimune, and James L Powell. New York: Cambridge University Press, 1–46.

———. 2001b. "Policy-Relevant Treatment Effects." *American Economic Review* 91 (2):107–111.

———. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica* 73 (3):669–738.

———. 2007. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments." In *Handbook of Econometrics*, vol. 6, Part B, chap. 71. Amsterdam: Elsevier, 4875–5143.

Imbens, Guido W. and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62:467–475.

Kane, Thomas J. and Cecilia E. Rouse. 1993. "Labor Market Returns to Two-and Four-Year Colleges: Is a Credit a Credit and Do Degrees Matter?" Working Paper 4268, National Bureau of Economic Research, Cambridge, MA.

Kling, Jeffrey R. 2001. "Interpreting Instrumental Variables Estimates of the Returns to Schooling." *Journal of Business & Economic Statistics* 19 (3):358–364.

Kobayashi, Masayuki. 2009. *Daigaku Shingaku no Kikai (An Opportunity of University Enrollment)*. Tokyo, Japan: the University of Tokyo Press. (in Japanese).

Kyui, Natalia. 2016. "Expansion of Higher Education, Employment and Wages: Evidence from the Russian Transition." *Labour Economics* 39:68–87.

Nakamuro, Makiko and Tomohiko Inui. 2012. "Estimating the Returns to Education Using the Sample of Twins: The Case of Japan." Discussion Paper 12-E-076, RIETI Discussion Paper Series.

Nakazawa, Wataru. 2011. "Shussin Chiiki ni yoru Kosotsugo Shingaku Kikai no Hubyodo (Inequality of Opportunity for the Progression to Higher Education on the Basis of Residential Areas)." Discussion Paper 43, Panel Survey Project Discussion Paper Series, Institute of Social Science, the University of Tokyo. (in Japanese).

Nybom, Martin. 2017. "The Distribution of Lifetime Earnings Returns to College." *Journal of Labor Economics* 35 (4). Forthcoming.

OECD. 2015. *Education at a Glance 2015: OECD Indicators 2015 – Country note for Japan*. Paris: OECD Publishing.

Ogawa, Kazuo. 2015. "Shitsugyo to Gakko Kyoiku niokeru Jinteki Shihon Keisei: Todofuken-betsu Panel Data niyoru Keiryo Bunseki (Unemployment and Human Capital Formation in School Education: An Econometric Analysis of Prefectural Panel Data)." *The Japanese Journal of Labour Studies* 656:37–53. (in Japanese).

Ono, Hiroshi. 2004. "College Quality and Earnings in the Japanese Labor Market." *Industrial Relations: A Journal of Economy and Society* 43 (3):595–617.

Oshio, Takashi and Miki Kobayashi. 2009. "Happiness, Self-Rated Health, and Income Inequality: Evidence from Nationwide Surveys in Japan." PIE/CIS Discussion Paper 451, Center for Intergenerational Studies, Institute of Economic Research, Hitotsubashi University.

Oshio, Takashi and Wataru Seno. 2007. "The Economics of Education in Japan: A Survey of Empirical Studies and Unresolved Issues." *Japanese Economy* 34 (1):46–81.

Robinson, Peter M. 1988. "Root-N-Consistent Semiparametric Regression." *Econometrica* 56 (4):931–54.

Sano, Shinpei and Kengo Yasui. 2009. "Nihon ni okeru Kyoiku no Return no Suikei (Estimating the Returns to Education in Japan)." *Kokumin Keizai Zassi* 200 (5):71–86. (in Japanese).

Sasaki, Yosei. 2006. “Kyoiku Kikaino Chiikikan Kakusa (Regional Gaps in Educational Opportunities).” *The Journal of Educational Sociology* 78:303–320. (in Japanese).

Tachibanaki, Toshiaki. 1988. “Education, Occupation, Hierarchy and Earnings.” *Economics of Education Review* 7 (2):221–229.

Trostel, Philip, Ian Walker, and Paul Woolley. 2002. “Estimates of the Economic Return to Schooling for 28 Countries.” *Labour Economics* 9 (1):1–16.

Ueyama, Kojiro. 2011. “Daigaku Shingakuritsu no Todofukenkan Kakusa no Youin to Sono Henyo: Taboshudan Path Kaiseki ni yoru 4jiten Hikaku (Factor Structure of the Prefectural Gap in College Entrance Rate and its Transformation: A Four Time Point Comparison Using Multi-population Path Analysis).” *The Journal of Educational Sociology* 88:207–227. (in Japanese).

———. 2012. “Koto Kyoiku Shingakuritsu ni okeru Chikikan Kakusa no Saikensho (Re-verification of Regional Disparities in Higher Education Enrollment Rate).” *Gendai Shakaigaku Kenkyu* 25:21–36. (in Japanese).

Willis, Robert J. and Sherwin Rosen. 1979. “Education and Self-Selection.” *Journal of Political Economy* 87 (5, Part 2):S7–S36.

Yano, Masakazu and Junko Hamanaka. 2006. “Naze? Daigaku ni Shingaku Shinainoka: Kenzaiteki Juyo to Senzaiteki Juyo no Kettei Yoin (Why Don’t High School Students Go to University?: Determinants of the Demand for Higher Education).” *The Journal of Educational Sociology* 79:85–104. (in Japanese).

Yasui, Kengo and Shinpei Sano. 2009. “Kyoiku ga Chingin ni Motarasu Ingateki na Koka ni tsuite (Causal Effects of Education on Wages).” *The Japanese Journal of Labour Studies* 588:16–33. (in Japanese).

A Data Construction

This appendix describes the construction of the instrumental variables. The data source for capacity of universities is *Zenkoku Daigaku Ichiran*. This book is published with a list of all accredited national, prefectural, municipal and private universities in each academic year by the *Bunkyo Kyokai*. It contains the detailed information on the accredited capacity, the location, and the date of opening and closing by the department of the university. I collected the total quota for new students offered in each prefecture of an academic year in the department level. If the department is located in more than one prefecture, I take the prefecture where students of the department stay in longer. If general education courses are collectively offered in the other prefecture, I assign each department to the prefecture where they offer an upper-level or a specialized course. I exclude the universities that offer only in correspondence, Internet learning and a graduate school. I also exclude following categories of departments: art, music, religious, home economics, and physical education. An important problem is the conversion of female to co-ed universities. *Zenkoku Daigaku Ichiran* provides the information on conversion of single-sex to co-ed in university or college level, but not in department level. I search the history of the university in the official web cite or in an official report published by the university, and I identify the department started to offer courses to male. If I am unable to identify which department was changed to co-ed, I assume that all departments offer co-ed courses based on the information of *Zenkoku Daigaku Ichiran*.

Tuition data are based on *Keisetsu Jidai* for 1972-2000, published by Obunsha. I define tuition is sum of entrance fees and course fees. Some prefectural and municipal universities have price discrimination by residential area of students. I assign minimum price of tuition for intra-regional students if prefectural or municipal universities are available and assign maximum price of tuition for extra-regional students if there is no prefectural or municipal universities in the prefecture of residence at the age of 15. I construct the measure as accredited capacities weighted averages over prefectural and municipal universities in a prefecture, or at the regional level if prefectural and municipal universities are not available. The region is based on the definition of region code of the Labor Force Survey (*Rodo-Ryoku Chosa*) of the Statistics Bureau of the Ministry of Internal Affairs and Communications.

The active job openings to applications ratio (*yuko kyujin bairitsu*) is based on the report on employment service (*shokugyou antei gyoumu toukei*) of the Bureau of Employment Security of the Ministry of Health, Labour and Welfare. It is defined as number of active job openings per number of active applications. I use the ratio that excludes new graduates and part-timers. I construct average local earnings in high school years from the annual average monthly total cash earnings (establishments with 30 employees or more) of the Monthly Labor Survey (*Maigetsu Kinro Tokei Chosa*). Local earning is evaluated at the 2005 consumer prices and transform it in logarithm.

B Definitions of Weights for the Treatment Effect Parameters

The treatment effect parameter (j) is weighted averages of the MTE with a weight (ω_j) conditional on $X = x$ that can be estimated as,

$$\int_0^1 MTE(x, U_D)\omega_j(x, u_D)du_D,$$

where j = {ATE, ATT, ATUT, PRTE, MPRTE}.

Heckman and Vytlacil (2005) and Carneiro, Heckman, and Vytlacil (2010, 2011) provide the weights:

$$\omega_{ATE}(u_D) = 1$$

$$\omega_{ATT}(u_D) = \frac{\int_{u_D}^1 f(p)dp}{E(P)}$$

$$\omega_{ATUT}(u_D) = \frac{\int_0^{u_D} f(p)dp}{E(1-P)}$$

$$\omega_{PRTE}(u_D) = \frac{F_P(u_D) - F_{P^*}(u_D)}{E_{F_{P^*}}(P) - E_{F_P}(P)}$$

$$\omega_{MPRTE}(u_D) = \frac{f_P(u_D)f_V\left(F_V^{-1}(u_D)\right)}{E\left(f_V\left(\mu_D(Z)\right)\right)} \quad \text{for } Z_\alpha^k = Z^k + \alpha$$

$$\omega_{MPRTE}(u_D) = f_P(u_D) \quad \text{for } P_\alpha = P + \alpha$$

$$\omega_{MPRTE}(u_D) = \frac{u_D f_P(u_D)}{E(P)} \quad \text{for } P_\alpha = (1+\alpha)P,$$

where f is the desity of $P(Z)$ and conditional on $X = x$ is implicit.

Table 1: Summary Statistics

Variables	High school (D = 0)	University (D = 1)
Years of schooling	12.3299	16.1439
	(0.7384)	(0.6304)
Log hourly wage	7.4813	7.7428
	(0.5436)	(0.5369)
Control Variables		
Mother's years of schooling	10.0539	11.4722
	(2.2403)	(2.2391)
Father's years of schooling	10.1556	12.2285
	(2.5825)	(3.0751)
Number of siblings	1.6836	1.3527
	(1.0304)	(0.8185)
Age at the survey date	39.7728	39.5835
	(7.2197)	(7.0449)
Urban residence at age 15	0.0871	0.1937
	(0.2822)	(0.3955)
Rural residence at age 15	0.4346	0.2575
	(0.4960)	(0.4375)
Local population of ages 15–19	28.5065	34.4926
(in 10,000)	(22.3170)	(25.8994)
Local log earnings at ages 13–18	12.5964	12.6493
	(0.2311)	(0.2265)
Local job openings at ages 13–18	1.0370	1.0819
	(0.7304)	(0.7093)
Instrumental Variables		
Capacity of universities	11.0961	17.3315
(in 1,000)	(18.4653)	(26.8027)
Tuition in public universities	25.0574	25.9429
(in 10,000)	(18.3435)	(17.7490)
Local log earnings in high school	12.6435	12.6926
	(0.1996)	(0.1971)
Local job openings in high school	1.0296	1.0402
	(0.7328)	(0.7067)
Number of observations	964	862

Notes: This table reports summary statistics of the analysis data. Local job openings indicate the active job openings to applications ratio. Standard deviations are in parentheses.

Table 2: University Enrollment Decision

Dependent variable:	University Enrollment
Control Variables	
Father's years of schooling	-0.0643
	(0.0414)
Mother's years of schooling	0.0568
	(0.0551)
Number of siblings	-0.1174
	(0.0340)
Age at the survey date	0.0325
	(0.0403)
Urban residence at age 15	0.1159
	(0.0482)
Rural residence at age 15	-0.0978
	(0.0310)
Local population of ages 15–19	-0.0250
(in 10,000)	(0.0108)
Local log earnings at ages 13–18	2.0878
	(1.0922)
Local job openings at ages 13–18	-0.1479
	(0.0941)
Instrumental Variables	
Capacity of universities	0.0114
(in 1,000)	(0.0047)
Tuition in public universities	-0.0069
(in 10,000)	(0.0023)
Local log earnings in high school	-2.4672
	(0.9765)
Local job openings in high school	0.1299
	(0.0908)
Test for joint significance of IVs	
Chi-square	19.461
p-value	0.0006

Notes: This table reports the marginal effects evaluated at the mean value of each variable from the probit model of university enrollment (a dummy variable that is equal to one if an individual has ever attended university, and equal to zero if he has never attended university but has completed upper secondary education). Local job openings indicate the active job openings to applications ratio. Survey year dummies, cohort dummies, and a set of dummies for prefecture of residence at the age of 15 are also controlled in the model but not reported. Robust standard errors are in parentheses. Chi-square and p-values indicate the results of the test of joint significance of coefficients on the instrumental variables.

Table 3: Estimates of Returns to a Year of University Education

Parameters		
OLS	0.0549	
	(0.0057)	
IV	0.1163	
	(0.0507)	
	Parametric	Semiparametric
ATE	0.0889	0.0918
	(0.0412)	(0.1905)
ATT	0.1743	0.2832
	(0.0818)	(0.2381)
ATUT	0.0167	-0.0815
	(0.0474)	(0.2426)
MPRTE		
$Z_{\alpha}^{k} = Z^{k} + \alpha$	0.0946	0.0935
	(0.0416)	(0.1885)
$P_{\alpha} = P + \alpha$	0.0939	0.0947
	(0.0421)	(0.1885)
$P_{\alpha} = (1 + \alpha)P$	0.0664	-0.0310
	(0.0394)	(0.2042)
PRTE		
Free tuition	0.0817	0.1007
	(0.0381)	(0.1915)
Increase in	0.1073	0.1137
capacities of universities	(0.0451)	(0.1828)

Notes: This table reports estimates of returns to university education: average treatment effect (ATE), average treatment on the treated (ATT), average treatment on the untreated (ATUT), marginal policy-relevant treatment effect (MPRTE), and policy-relevant treatment effect (PRTE). The PRTE corresponds to the two counterfactual policies: (1) free tuition: a policy of free tuition in public universities, (2) increase in capacities of universities: a policy that increases capacities of universities by 500 places if the prefecture has less than median value of places. The IV estimate uses $P(Z)$ as the instrument (probit model for first stage with all instruments). Parametric column indicates the estimates for the normal selection model. Semiparametric column indicates the estimates for the semiparametric model. Standard errors (in parentheses) are obtained by the bootstrap method (250 replications). All estimates are annualized (divided by 3.81 years).

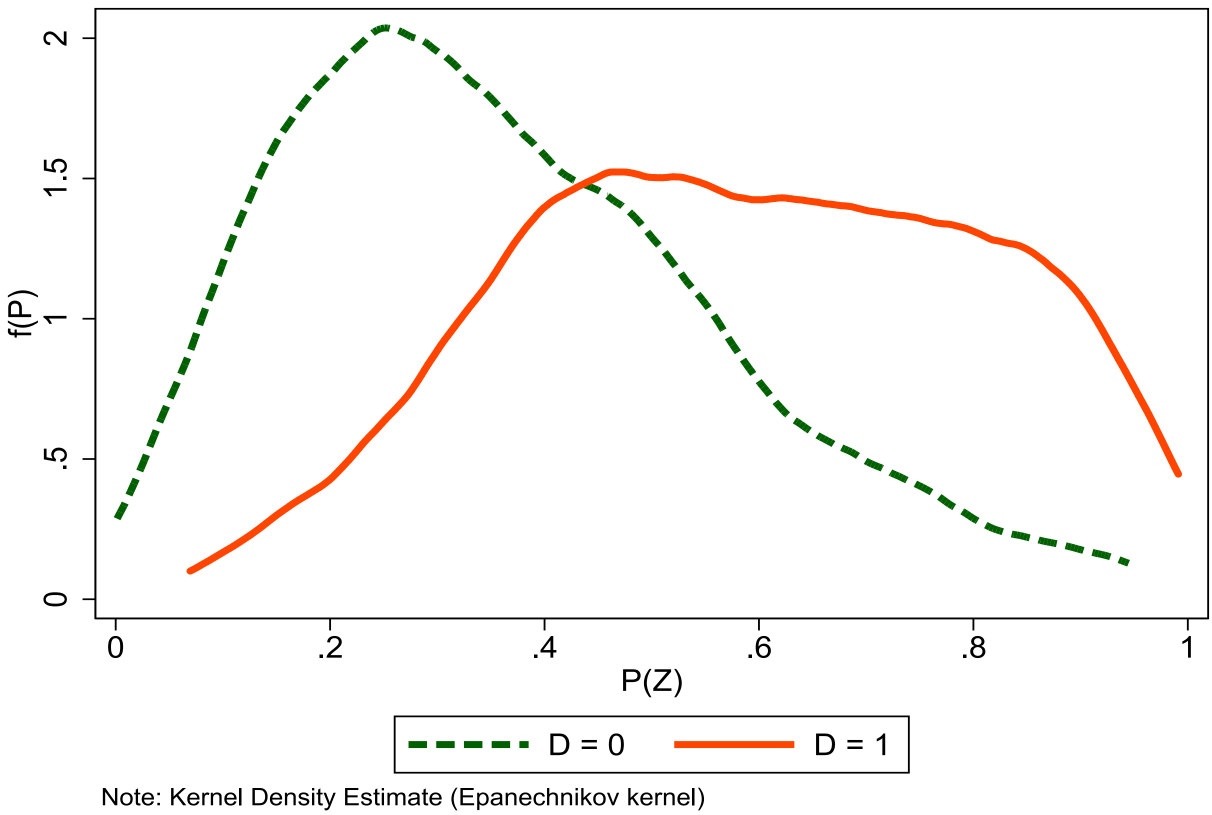

Figure 1: Support of $P(Z)$ for $D = 1$ and $D = 0$

Notes: This figure shows the support of $P(Z)$ for those who enroll in university ($D = 1$) and for those who do not ($D = 0$). $P(Z)$ is predicted using the estimates from a probit model of university enrollment (a dummy variable that is equal to one if an individual has ever attended university, and equal to zero if he has never attended university but has completed upper secondary education) on mother's and father's years of schooling, number of siblings, age at the survey date, dummy variables of survey years, dummy variables of urban and of rural residence at the age of 15, local population of age 15–19, local log earnings at age 13–18, local job openings at age 13–18, capacity of local universities, tuition in local public universities, local log earnings in high school, the active job openings to applications ratio in high school, cohort dummies, and a set of dummies for prefecture of residence at the age of 15 (see Table 2).

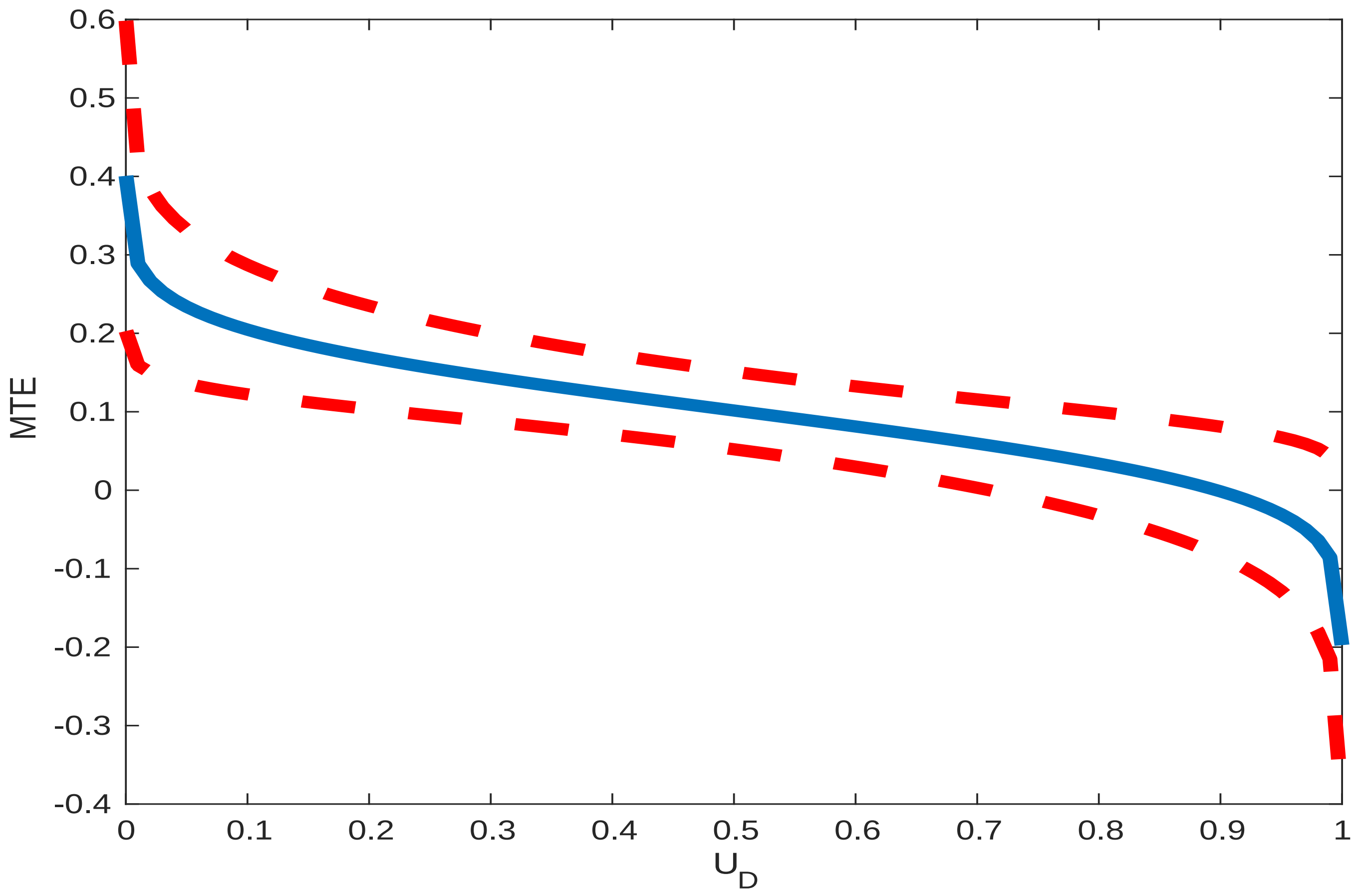

Figure 2: Marginal Treatment Effect

Notes: This figure depicts the marginal treatment effect (MTE) with assumptions on normally distributed unobservables. The solid line indicates the estimated effect. Dashed lines indicate 90% confidence intervals.

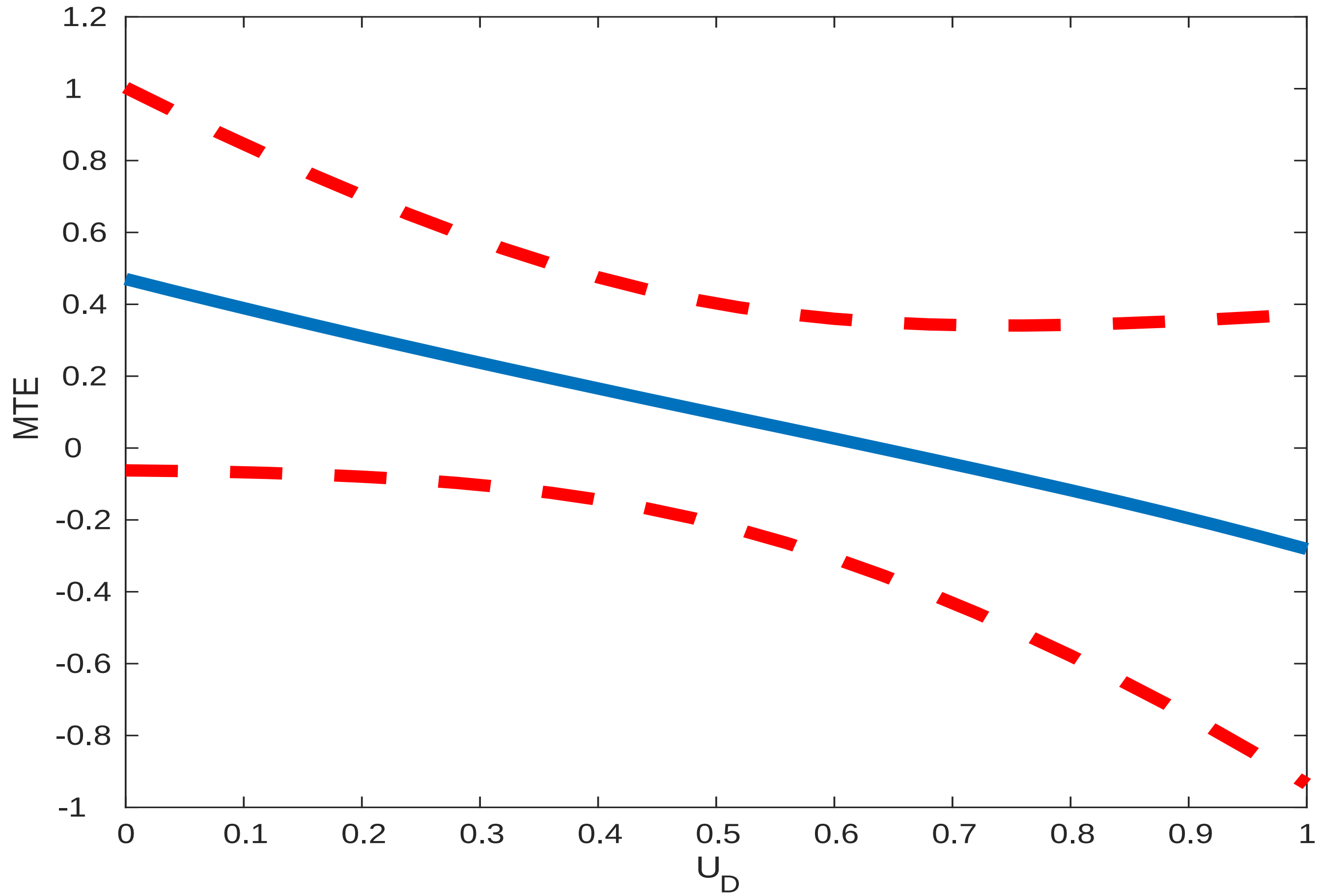

Figure 3: Marginal Treatment Effect

Notes: This figure depicts the marginal treatment effect (MTE) with a semiparametric model. The solid line indicates the estimated effect. Dashed lines indicate 90% confidence intervals obtained by the bootstrap method (250 replications).

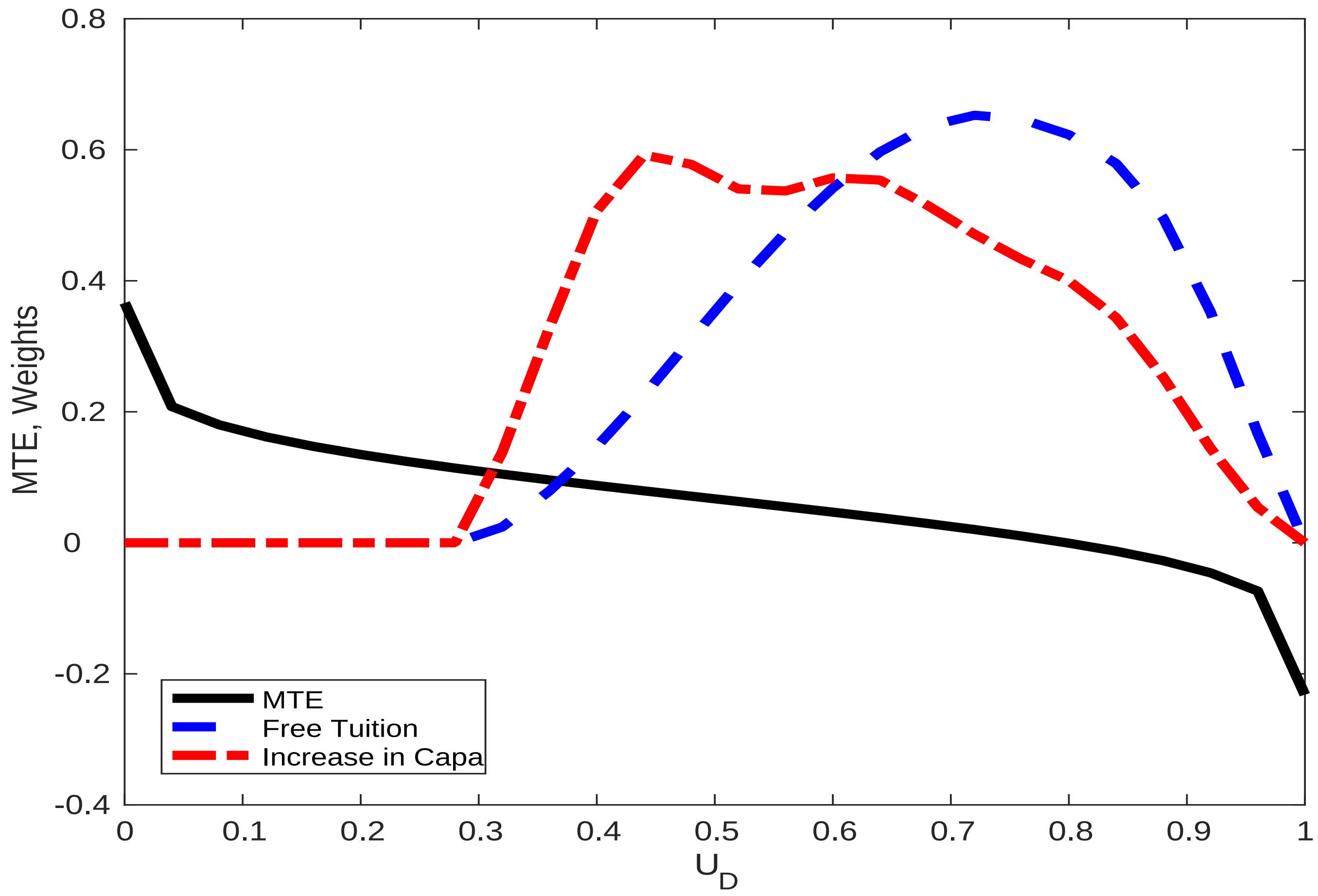

Figure 4: Weights for the PRTE

Notes: This figure depicts weights for the policy-relevant treatment effect (PRTE) of two conterfactual policies: (1) a policy of free tuition in public universities and, (2) a policy that increases university capacity by 500 places if the prefecture has less than the median value of places. The scale of the vertical axis is the scale of the marginal treatment effect (MTE), not the scale of the weights, which are scaled to fit the picture.

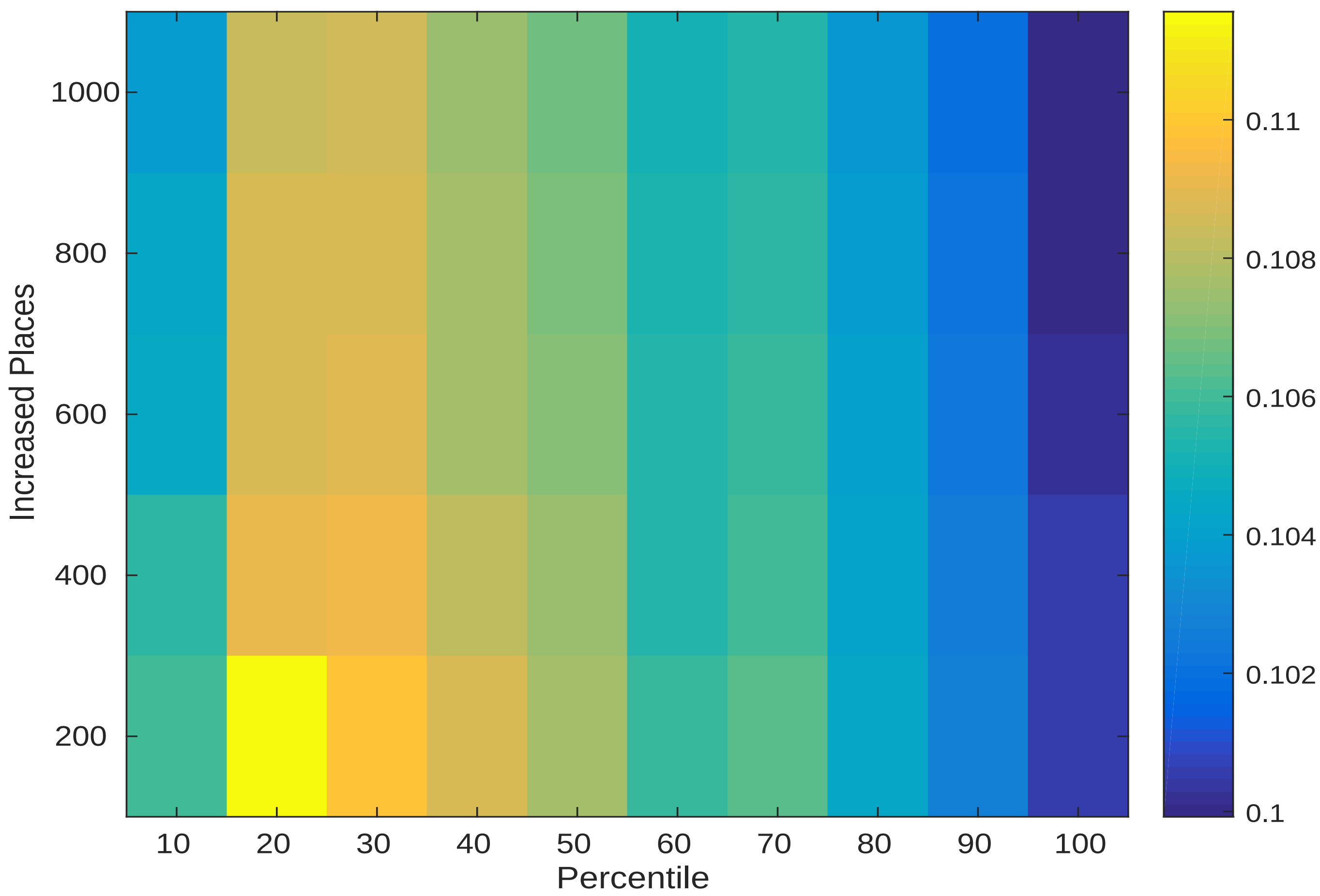

Figure 5: Distribution of PRTE by Different Levels of Capacity Increase

Notes: This figure depicts the policy-relevant treatment effect (PRTE) of a policy that increases university capacity by Δ places if the prefecture has less than the π-percentile value of places. The horizontal axis indicates the π-percentile affected. The vertical axis indicates the number of increased places Δ.