Discussion Paper Series
RIEB
Kobe University

DP2026-17

Human Capital Externalities in Japanese Local Labor Markets

Yudai HIGASHI

May 20, 2026

Research Institute for Economics and Business Administration
Kobe University
2-1 Rokkodai, Nada, Kobe 657-8501 JAPAN

Human capital externalities in Japanese local labor markets[*]

Yudai Higashi[†]
May 20, 2026

Abstract

This study examines the effect of the share of college graduates on average wages in local labor markets, using data covering 1990–2020 in Japan. To address the endogeneity of the local college graduate share arising from labor demand shocks for college graduates, the estimation model uses an instrumental variable approach. According to the results, an increase in the share of college graduates has a positive effect on college graduates' wages during the 2000s and 2010s, while it has no significant effect during the 1990s. However, there is no significant effect on the wages of non-college graduates across all sample periods. These results suggest that human capital externalities were generated among college graduates in local labor markets during the 2000s and 2010s, but no such externalities arose from interactions between college and non-college graduates. Overall, external returns to higher education in terms of local wages appear to be limited in Japan.

Keywords: Human capital externalities, External returns to education, Local labor market

JEL classification: I26, J24, J31, R23

[*] This study utilizes the microdata of the Basic Survey on Wage Structure (BSWS), provided by the Ministry of Health, Labour and Welfare (MHLW) of Japan. The results using the BSWS in this study are calculated by the author; therefore, they are not necessarily consistent with the survey results that the MHLW makes public. This study was supported by JSPS KAKENHI Grant Number JP22K13391 and 25K16677. All remaining errors are my own.

[†] Faculty of Economics, Kyoto Sangyo University, Kyoto, Japan; Research Institute for Economics and Business Administration, Kobe University, Kobe, Japan. Email: yhigashi@cc.kyoto-su.ac.jp

1. Introduction

College enrollment rates in Japan have consistently increased over time. For example, the rate is 57.7% in 2023, while 24.6% in 1990, meaning it has more than doubled over approximately 30 years.[1] How has such a trend affected the Japanese labor market? More educated workers receive higher wages, indicating that private returns to education exit. Simultaneously, education is expected to have positive externalities. To support the policy expanding education, measuring social returns to education, which is the sum of private returns and externalities of education, is significant. This study examines whether human capital externalities, i.e., source of external returns to higher education, in terms of wages have arisen in Japan from 1990 to 2020.[2]

The previous literature examines the external effects of education using regional variations. That is, the effect of the share of educated workers on wages in local labor markets is examined. The literature has led to mixed results. For example, there are significantly positive externalities of college education, the effect of which is larger on wages for non-college graduates than those for college graduates in the US (Moretti 2004a). The positive externalities are found in urban areas in China (Glaeser and Lu 2018). The positive externalities created by college graduates in STEM are stronger than those in non-STEM in the US (Winters 2014). In contrast, no significant externalities are found using the US data after accounting for the endogeneity of regional average years of schooling (Acemoglu and Angrist 2000). Furthermore, the significance of positive externalities varies if the sample period is changed in the US (Sand 2013). A recent meta-

---

[1] The data come from the School Basic Survey, conducted by the Ministry of Education, Culture, Sports, Science and Technology. The value is the number of university-enrolled students, including past-year high school graduates, divided by the number of junior high school students three years before.

[2] While social returns to education can be considered in some aspects, such as crime and voting (Moretti 2004b), these are out of the scope of this study.

analysis reveals that such mixed results reflect that magnitude of externalities slow down with economic development and time, suggesting diminishing external returns to education (Cui and Martins 2021).

Following the literature, this study focuses on variations in local labor markets to identify the effect. Specifically, we regress the regional share of college graduates on the regional average wages. According to the theoretical framework presented by Moretti (2004a, b), the external effects on educated and uneducated workers differ. An increase in the share of educated workers in local labor markets raises the productivity of uneducated workers through imperfect substitution between educated and uneducated workers, as well as through human capital spillovers, thereby increasing their wages. For educated workers, an increase in the share of educated workers in local labor markets has two opposing effects: it decreases wages due to the additional labor supply of similarly educated counterparts, and it increases wages due to human capital spillovers. Consequently, the positive external effect is expected to be larger for uneducated workers than for educated workers. Thus, we separately estimate the effect for non-college and college graduates. The regional average wages are estimated using the worker-level wage data. A concern of endogeneity of the regional share of college graduates is dealt with using the shift-share instrument of regional demographic structure proposed by Moretti (2004a). Confirming a significant positive effect, we can interpret that human capital externalities arise in local labor markets.

Even if a regional variation approach accounting for endogeneity is used, a mechanism labeled as consumption spillovers, distinct from human capital externalities, could still yield a positive effect. Specifically, college graduates who earn high wages tend to demand services related to household production, such as housekeeping and childcare, from the market due to the high opportunity cost of performing these tasks by themselves. Consequently, consumption spillovers arise; that is, an increase in the regional share of college graduates induces labor demand for workers in household-related sectors, who tend to have lower educational attainment (Mazzolari and Ragusa, 2013; Liu and Yang, 2021). Due to this mechanism, a positive effect on non-college graduates can be found.

This study contributes to the literature twofold. First, this study tries to distinguish the mechanism of human capital externalities from labor demand and supply factors. Apart from the mechanism of human capital externalities, the increase in the regional share of college graduates can affect the wages of college graduates through the increase in labor supply of the counterparts, and can affect the wages of non-college graduates through imperfect substitution and consumption spillovers. To isolate these mechanism related to the labor supply and demand, this study estimates regional average wages using the data on the establishment-level starting wages for new graduates (*shoninkyu*) that are equally set among all workers with the same educational backgrounds in the establishments, in addition to the conventional worker-level wages. The starting wages for new graduates do not reflect human capital externalities on the job but labor demand and supply because the workers receiving such wages have no labor market experience.

Second, this study sheds new light on the external returns to education in Japan, which is little known. Exceptionally, Sano (2011), using worker- and regional-level data for 2001, finds positive externalities but does not incorporate the endogeneity of the regional level of education. Sano (2018), using worker- and regional-level data covering 2000–2012, shows that the externalities are significantly positive on wages for college-graduated men if regional college admission capacities are used as the instrument for regional average years of education, suggesting that the externalities are limited. This study contributes to these studies by using data covering 1990–2020 and adopting the instrument typically used by the literature on other countries.

According to the results, the significance of the effect of the regional share of college graduates on average wages varies across periods and across college and non-college graduates. First, an increase in the share of college graduates has a positive effect on the wages of college graduates during the 2000s and 2010s, while it has no significant effect during the 1990s. The positive effect dominates the negative effect due to the increase in labor supply of the counterparts. In contrast, no significant effect on the wages of non-college graduates is confirmed across all periods. These results suggest that the human capital externalities arise across college graduates during the 2000s and 2010s but do not arise between college and non-college graduates. Overall, the human capital externalities are limited in Japan. Additionally, no evidence of consumption spillovers is found.

The remainder of this paper is organized as follows. Section 2 presents the empirical model for estimating the human capital externalities. Section 3 describes the data. Section 4 shows the estimation results. Finally, Section 5 concludes.

2. Empirical model

Following previous literature (e.g., Moretti 2004a; Sand 2013; Liu and Yang 2021), this study adopts a two-step procedure to estimate external returns to education using a variation of local labor markets. In the first step, we obtain average region-year wages adjusted for observable worker characteristics by estimating a worker-level wage function. In the second step, using the variation of local labor markets to estimate human capital externalities, we regress the adjusted wage on the regional college share.

The first-step worker-level wage function using the repeated cross-sectional data is given by

$$\log w_{ict}^{I} = \alpha_{ct}^{I} + \beta_1 X_{ict} + \beta_2 Z_{j(i)ct} + \nu_{ict}^{I}, \tag{1}$$

where $\log w_{ict}^{I}$ is a log of hourly wage for individual i in region c in year t; X_{ict} is a vector of worker characteristics (age, age squared, gender, tenure, and education); $Z_{j(i)ct}$ is a vector of characteristics of establishment j in which worker i is employed (establishment size and its squared); and ν_{ict}^{I} is an error term. Estimating this equation is to yield α_{ct}^{I}, the adjusted region-year average wage. As mentioned in the following Section 3, this equation is estimated using the data from Japan covering almost every 10-year sample period of 1990, 2000, 2010, and 2019.[3] We separately estimate this wage function for subsamples of college and non-college graduates. The sampling weight provided by the data source is used for the estimation.

This study differs from the previous literature by using another type of wage to yield the adjusted region-year average wage, not affected by human capital externalities but by labor demand and supply. In Japan, most firms hire new graduates simultaneously on April 1 of each year. In this hiring process, most students generally find a job before they graduate and start to work immediately after they graduate. Firms set the starting salary for new graduates (*shoninkyu*). Basically, its amount is set equally across new graduate workers with the same educational background in the establishment. Furthermore, the higher amount of the starting wage is generally paid to new graduates with a higher educational degree. For example, in the same establishment, the starting wage for new college graduates is higher than for new high school graduates. This system suggests that the starting wage is not affected by the human capital externalities since the new graduate workers who start immediately after graduation have generally no labor market experience. We can consider that the starting wages are purely determined by the labor demand and supply factors as far as the same education group. Using this type of wage,

[3] See section 3 for why the 2019 data, instead of 2020, are used in detail.

this study tries to isolate the effect of labor demand and supply factors, such as an increase in labor supply of counterparts (college graduates), imperfect substitution between college and non-college graduates, and consumption spillovers. Using this type of wage, another form of the first-step wage function is the establishment level, given by

$$\log w_{jct}^{S} = \alpha_{ct}^{S} + \gamma Z_{jct} + \nu_{jct}^{S}, \quad (2)$$

where $\log w_{jct}^{S}$ is a log of monthly starting wage for new graduates at establishment j in region c in year t; Z_{jct} is a vector of characteristics of the establishment (establishment size and its squared); and ν_{jct}^{S} is an error term. α_{ct}^{S} is the adjusted region-year average starting wage for new graduates. We separately estimate this wage function for subsamples of college and non-college graduates. Accounting for worker characteristics is not necessary for this equation since the starting wages for new graduates are equally set by the establishments, not reflecting differences in skills across individuals as far as their educational backgrounds are the same.

In the second step, we estimate the effect of the regional share of college graduates on the regional adjusted wages estimated by equation (1) or (2) in the first step, using the following regional-level wage function:

$$\Delta\alpha_{ct}^{K} = \delta_0^{K} + \delta_1^{K}\Delta CS_{ct} + \delta_2^{K}\Delta R_{ct} + \varepsilon_{ct}^{K} \ where\ K \in \{I, S\}, \quad (3)$$

where $\Delta\alpha_{ct}^{K}$ is a change in the average adjusted worker-level hourly wage from equation (1) or establishment-level monthly starting wage for new graduates from equation (2) in region c from year $t-10$ to t; ΔCS_{ct} is a change in share of college graduates in city c from year $t-10$ to t; and ΔR_{ct} is a change in regional characteristics in city c from year $t-10$ to t (population share aged 16–24, population share aged 65 and over, female labor force participation rate, and unemployment rate). As mentioned in the following Section 3, the sample periods of the regional data for this equation cover 1990, 2000, 2010, and 2020. However, the last sample period of the microdata for the worker-level and establishment-level wage functions (equations (1) and (2)) is not 2020 but 2019, resulting in a change in the average adjusted wage over nine years from 2010 to 2019. To fill this gap, the 9-year change in the average adjusted wage is converted to a 10-year change as follows: $\Delta\alpha^K_{c,2020} = \left(\alpha^K_{c,2019} - \alpha^K_{c,2010}\right)(10/9)$. The parameter $\delta_1^{K=I}$ when the adjusted wage in the left-hand side is $\Delta\alpha^{K=I}_{ct}$, coming from equation (1), is normally considered to capture the effect of human capital externalities, a source of the external returns to higher education.

It is concerned that equation (3) with $\Delta\alpha^{K=I}_{ct}$ may capture the effect of mechanisms other than human capital externalities. First, a negative effect of the regional college on the wages of college graduates can arise due to the increase in labor supply of their counterparts. Second, a positive effect of the regional college on the wages of non-college graduates can occur due to not only imperfect substitution between college and non-college graduates but also consumption spillovers. That is, an increase in skilled workers with high opportunity costs of household work in a region may lead to demand for unskilled workers because workers in the service sector related to household work are unskilled (Mazzolari and Ragusa 2013; Liu and Yang 2021). We consider that these concerns can be checked based on the parameter $\delta_1^{K=S}$ when the average adjusted wage on the left-hand side is $\Delta\alpha^{K=S}_{ct}$, coming from equation (2), since the starting wages for new graduates are considered to be purely determined by the labor demand and supply factors.

Finally, following Sand (2013), who reveals that the estimation results change depending on the sample period using the US data, this study estimates the equation (3) for subsamples based on 10-year changes for the periods of 1990–2000, 2000–2010, and 2010–2020, in addition to whole sample periods.

Moretti (2004a) claims that the change in the share of college graduates ΔCS_{ct} in equation (3) is considered endogenous because time-varying shocks on local labor markets can simultaneously affect the change in the share of college graduates and the average wages. If labor demand for educated workers increases, ΔCS_{ct} positively correlates with error term ε_{ct}^{K}, resulting in upward bias by the ordinary least squares (OLS) estimations. To deal with this problem, this study uses the following shift-share demographic instrumental variable (IV), proposed by Moretti (2004a):

$$IV_{ct} = \sum_{m} \omega_{mc} \Delta CS_{mt} , \tag{4}$$

where ω_{mc} is a share of gender-age group m in region c in 1980, 10 years before the initial sample period; and ΔCS_{mt} is a change in nationwide share of college graduates for gender-age group m from year $t-10$ to t. The gender-age group is defined by men and women categories and 5-year age classes (15–19, 20–24, …, 80–84, 85 or over). The exclusion restriction is satisfied if the demographic structure in 1980 (ω_{mc}) is independent of the shocks on the local labor market in region c during 1990–2020, such as shocks on demand for educated workers. This demographic IV is adopted by the literature on external returns to education using the regional variation approach (e.g., Moretti 2004a; Sand 2013; Liu and Yang 2021).

3. Data

This study utilizes two main sources of data: worker- and establishment-level microdata on wages to estimate regional average wages and regional-level data on the share of college graduates and other regional characteristics. The overviews of these two data are as follows.

The wage data come from the Basic Survey on Wage Structure (BSWS), conducted by the Ministry of Health, Labour and Welfare (MHLW) of Japan. This survey is conducted annually and covers information as of June 1 of each year. Workers in the BSWS are selected based on a stratified two-step sampling. The first step randomly sampled establishments employing five or more workers from strata of prefecture, industry, and establishment size, and the second step randomly sampled workers at the establishments sampled by the first step. This study uses the waves of BSWS of 1990, 2000, 2010, and 2019 because, as explained below, regional-level data are available every 10 years. The reason why this study uses the 2019 wave rather than 2020 is that the form of the questionnaire was largely changed in 2020. Incidentally, using the 2019 wave allows us to disregard the effects of the COVID-19 pandemic, which started to spread widely in 2020, on the labor market. Each wave this study uses covers approximately 1.4–1.6 million workers at 70–78 thousand establishments.

The questionnaire of the BSWS consists of the worker form and establishment form. The worker form covers worker-level information about monthly earnings, monthly working hours, and worker characteristics, while the establishment form covers the starting salary for new graduates and the characteristics of establishments. This study uses both forms. From the worker form, this study obtains variables to estimate worker-level wage function (i.e., hourly wage, age, age squared, gender, tenure, education, and establishment size). Appendix A shows the definition of the variables and of the classification of workers of college and non-college graduates. The sample workers for this study are restricted to a following type of employment. In the BSWS, workers employed for an indefinite period or a definite period of one month or longer are classified as “regular employees,” and those employed for a definite period of less than one month as “temporary employees.” The “regular employees” are further classified into “ordinary workers” and “part-time workers.” The former indicates workers other than the latter, and the latter is those who work for shorter hours in a day or fewer days in a week than general workers in the establishments. The sample for this study is limited to "ordinary workers" because educational background, the prime variable for this study, is surveyed only for this group. Workers lacking other necessary variables are also excluded from the sample for this study. From the establishment form, the variables for estimations of establishment-level wage functions are obtained, namely the starting wages for new senior high school graduates and for college graduates. Establishment size is also obtained.

The regional-level data come from the Population Census, conducted by the Statistics Bureau, Ministry of Internal Affairs and Communications (MIC). This census is conducted every 5 years, whereas education is surveyed only every 10 years. Therefore, this study utilizes the Population Census for 1990, 2000, 2010, and 2020 which survey educational background. The 1980 census is also used to create the shift-share demographic IV, as mentioned in Section 2. From public data from the census, this study uses municipality-level population by educational background to obtain the share of college graduates.[4] Other control variables for estimations of the regional-level wage functions, such as population share by age class, female labor force participation rate, and unemployment rate, are also obtained.

The BSWS, worker- and establishment-level microdata, contains information on establishment location at the municipality level, enabling us to merge the BSWS with the municipality-level aggregate data from the Population Census. Here, the municipalities do not represent units of local labor markets but administrative areas. Thus, to approach a close unit of the local labor market, this study reaggregates municipalities into the Urban Employment Areas (UEA), proposed by Kanemoto and Tokuoka (2002). The UEAs are

---

[4] The public data from the Population Census is available on the website of e-stat (https://www.e-stat.go.jp).

defined based on commuting rates across municipalities. This study uses the 2015 version of the UEA, the newest version available at the time of the study.[5] This version consists of 222 UEAs. Among them, this study drops UEAs that contain less than 10 college graduates or non-college graduates at one period or more for worker-level wage functions and less than 5 establishments reporting the starting wages for new high school or college graduates at one period or more. Finally, this study obtains a sample for the worker-level wage functions containing 1,033,827 college graduates and 2,862,320 non-college graduates in 206 UEAs. For the establishment-level wage functions, the sample contains 34,152 establishments for starting wages for new college graduates and 29,182 establishments for new high school graduates in 125 UEAs. Tables 1, 2, and 3 show the summary statistics for workers, establishments, and regions, respectively.

[Tables 1–3]

4. Results

4.1. Worker-level wages

Figure 1 plots the relationship between a change in regional average wages estimated by the worker-level wage function (equation (1)) and the share of college graduates by decade.[6] Both panels for the average wages of college and non-college graduates show that the relationship varies across periods, including a negative sign. To further explore the effect of college share on the regional average wages, we turn to consider the regression analysis accounting for regional characteristics.

[Figure 1]

---

[5] The list of UEA is available on the website of the Center for Spatial Information Science of the University of Tokyo (https://www.csis.u-tokyo.ac.jp/UEA/).

[6] The estimation results of worker-level wage functions are shown in Table B1 in the Appendix.

First, Table 4 presents the estimation results of the regional-level wage function (equation (3)), pooling all sample years (1990–2020). All models in this table control for year dummies. Column (1) for the regional average wages for college graduates shows the result only controlling for the year dummies. The coefficient of the change in college share is statistically insignificant. This result does not change even after controlling for the changes in regional characteristics, as shown in column (2). For non-college graduates, column (3) shows that the coefficient of the change in college share is statistically negative. After controlling for the changes in regional characteristics, the statistically negative effect is still confirmed. The OLS estimations may suffer from bias due to the endogeneity of the college share. To cope with this problem, columns (5) to (8) show the two-stage least squares (2SLS) estimation results using the demographic IV. For all models, we find no statistically significant results. However, the first-stage F-values for all models are less than 10, indicating the IV may lack relevance to the endogenous variable. Overall, no human capital externalities may exist, but the results are doubtful using all sample periods.

[Table 4]

Next, Table 5 presents the OLS estimation results of the regional-level wage functions by decade. According to columns (1) to (4) for 1990–2000, for both college and non-college graduates, the coefficient of the change in college share is significantly negative in the specification without control variables (columns (1) and (3)), whereas it is not statistically significant after controlling for regional characteristics (columns (2) and (4)). According to columns (5) to (8) for 2000–2010, for college graduates, both specifications in columns (5) and (6) exhibit a significantly positive coefficient of the change in college share. According to column (6), controlling for regional characteristics, a 1% increase in the college share increases the regional average wages for college graduates by 1.09%. For non-college graduates in columns (7) and (8), the change in college share is insignificant regardless of the specifications. For 2010–2020, the change in college share has no significant effect on the regional average wages for college graduates, as shown in columns (9) and (10). For the non-college graduates, both specifications in columns (11) and (12) exhibit that the coefficient of the change in college share is significantly negative. The specification with regional control variables indicates that a 1% increase in the college share decreases the regional average wages for non-college graduates by 1.02%. Overall, after controlling for the regional characteristics, the effect of the change in college share on the regional average wages for both college and non-college graduates is not confirmed during the 1990s. The significantly positive effect on college graduates' wages is shown during the 2000s, whereas the significantly negative effect on non-college graduates' wages is confirmed during the 2010s. If the negative effects for college graduates are shown, they can be explained by the increase in the labor supply of their counterparts. However, those for non-college graduates are unexpected.

[Table 5]

Table 6 shows the 2SLS estimation results accounting for the endogeneity of the change in college share. Except for columns (1) to (4), other models exhibit that the first-stage F-values are larger than 10, indicating that the relevance of the IV is satisfied. Although the results in columns (1) to (4) must be taken care of, the validity of the IV is not a major concern since the coefficient of the change in college share is statistically insignificant, not suggesting any significant effects. Comparing with the OLS results, the coefficient of the change in college share in columns (1) and (3) changes to insignificant, suggesting that no human capital externalities arose during the 1990s. For college graduates during the 2000s, the significantly positive coefficient in columns (5) and (6) remains. Furthermore, such a positive effect is also confirmed in columns (9) and (10).

According to the specification controlling for regional characteristics (columns (6) and (10)), a 1% increase in the college share raises the regional average wages for college graduates by 1.45% during the 2000s and by 2.78% during the 2010s. These results suggest that human capital externalities arose among college graduate workers during the 2000s and the 2010s. The positive effect could dominate the negative effect due to the increase in labor supply of the counterparts. In contrast, such a positive effect on non-college graduates is not confirmed for all periods, suggesting that no human capital externalities occur between college and non-college graduates. During the 2010s, the negative coefficient in column (11) for non-college graduates remains statistically significant, whereas it turns statistically insignificant after controlling for regional characteristics in column (12). Altogether, external returns to higher education in terms of wages exist but are limited to highly educated workers during the 2000s and 2010s.

[Table 6]

4.2. Establishment-level starting wages for new graduates

This subsection investigates the effects of regional college share on average wages through the labor demand and supply mechanism rather than human capital externalities, using the starting wages for new graduates. If the effect on the average starting wages for new high school graduates is significantly positive, we interpret that the imperfect substitution between college and high school graduates or the consumption spillovers arise in the local labor markets. Figure 2 plots the relationship between a change in regional average starting wages for new graduates estimated by the establishment-level wage function (equation (2)) and the share of college graduates by decade.[7] Like the

[7] The estimation results of establishment-level wage functions are shown in Table B2 in the Appendix.

cases for conventional wages in the previous subsection, the relationship varies across periods for both new college and new high school graduates.

[Figure 2]

Table 7 presents the estimation results of regional-level wage functions using all sample periods (1990–2020). According to the 2SLS estimations, which are more preferable results, column (5) shows that the coefficient of college share is significantly positive at a 10% level for new college graduates. This result turns insignificant after controlling for the regional characteristics (column (6)). For new high school graduates, regardless of specifications, the coefficient is statistically insignificant (columns (7) and (8)).

[Table 7]

Table 8 shows the OLS estimation results of regional-level wage functions separately estimated by decade. For 1990–2000, the coefficient of the change in college share on starting wages for new college graduates is significantly negative in the specification without regional control variables (column (1)), whereas it turns positive and insignificant after controlling for regional characteristics (column (2)). Similar results are confirmed for new high school graduates (columns (3) and (4)). For 2000–2010 for new college graduates, the specifications without regional control variables exhibit a significantly positive coefficient of college share (columns (5)), but it with regional control variables show no significant effect (columns (6)). With regard to new high school graduates, the coefficient is statistically insignificant, regardless of specifications (column (7) and (8)). For 2010–2020, column (9) for new college graduates shows that the coefficient of college share is significantly negative. After controlling for regional characteristics, the coefficient is still negative but statistically insignificant, as shown in column (10). No significant effect of college share is confirmed on the average starting wages for new high school graduates, regardless of the specifications (columns (11) and (12)).

[Table 8]

Next, we examine whether the above OLS estimation results are robust after incorporating the endogeneity of the college share. Table 9 presents the 2SLS estimation results of the regional-level wage functions using the demographic IV. The first-stage F-values of all models for 1990–2000 (columns (1) to (4)) are less than 10, indicating that the IV is not valid due to lack of relevance to the endogenous variable. While care should be taken in discussing these estimation results, the concern is limited since the coefficients of college share are statistically insignificant in columns (1) to (4). Columns (5) to (12), the first-stage F-values of which specifications are larger than 10, the estimation results in this table are similar to those by OLS in Table 8, except for column (10). Focusing on ideal models with regional control variables, only column (10) for new college graduates for 2010–2020 shows the statistically negative effect of college share on the average starting wages. According to its magnitude, a 1% increase in regional college share decreases the average starting wages for new college graduates by 2.63%. This result suggests that college graduates are substitutes for their counterparts in local labor markets during the 2010s. Nevertheless, our results suggest that this negative effect is dominated by the positive effect of human capital externalities among college graduates, as the corresponding result for college graduates' average wage exhibits a significantly positive effect of college share, containing both human capital externalities and labor demand and supply factors (column (10) of Table 6). Finally, if a significantly positive effect is confirmed for high school graduates, we interpret that the imperfect substitution between college and high school students or the consumption spillovers arise. However, such evidence is not confirmed for all periods.

[Table 9]

5. Conclusion

This study examined whether external returns to higher education with regard to wages existed in Japan from 1990 to 2020, using variations of the share of college graduates across local labor markets. Specifically, we regress the regional average wages adjusted for individual and establishment characteristics on the regional share of college graduates, a measure of the average regional level of skilled workers. To deal with the endogeneity of the share of college graduates, demographic shift-share IV is used. The mechanism of the positive effect of college share on regional average wages is considered human capital externalities, a source of external returns to higher education. However, the regional share of college graduates can also affect wages through labor demand and supply factors. To distinguish the mechanism of human capital externalities with labor demand and supply factors, regional average starting wages for new graduates, purely reflecting labor demand and supply, are also utilized.

According to the results, the effect of college share on wages varies across the sample periods and across workers with different educational backgrounds. An increase in the share of college graduates significantly raised the wages of college graduates during the 2000s and 2010s, suggesting that human capital externalities arise across college graduates. This positive effect exceeds the negative effect stemming from the increase in supply of their counterparts. However, no significant effect was found during the 1990s. There is no significant effect on the wages of the non-college graduates across all periods, suggesting that human capital externalities did not occur between college and non-college graduates. Overall, the external returns to higher education were limited to the wages of college graduates during the 2000s and 2010s in Japan.

Appendix A. Definitions of variables

Worker-level hourly wage: it is calculated based on information about monthly earnings and working hours in the worker form of the BSWS as follows:

$$hourly\ wage = \frac{Contractual\ Cash\ Earnings + Annual\ Special\ Cash\ Earnings/12}{Actual\ Number\ of\ Scheduled\ Hours\ Worked + Actual\ Number\ of\ Overtime\ Worked}.$$

It is adjusted for inflation using the Consumer Price Index (CPI) for all items with 2020 as the base year, issued by the Statistics Bureau of the MIC.

Education: The worker form of the BSWS contains workers' highest degree in four categories: "junior high schools," "senior high schools," "higher professional schools or junior colleges," and "universities or graduate schools." For the estimations of worker-level wage functions, these four categories are treated as dummy variables.

Classification of college graduates and non-college graduates: Using the above four categories of educational background, workers are classified as follows: college graduates are defined as workers holding a university or graduate school degree, while non-college graduates as workers holding a degree from junior high schools, senior high schools, or higher professional schools or junior colleges.

Establishment size: The number of regular employees in the BSWS.

Establishment-level monthly starting wages for new senior high school and college graduates: The establishment form of the BSWS asks establishments for monthly starting wages for new graduates by a school category from which the corresponding workers graduated. Specifically, all four waves of the BSWS for this study (1990, 2000, 2010, and 2019) inquired about the starting wage for new graduates and the number of new graduate workers from senior high schools, higher professional schools or junior colleges, and universities. The waves of 1990 and 2000 additionally inquired about those from junior high schools, whereas the waves of 2010 and 2019 additionally inquired about those from graduate schools. In other words, the waves of 1990 and 2000 lack those from graduate schools, whereas the waves of 2010 and 2019 lack those from junior high schools. This gap prevents us from calculating consistent average values of starting wages for college and non-college graduates, the classification of which is defined above. Thus, for the estimations of wage functions using starting wages for new graduates, college graduates and senior high school are adopted for the subsamples. The starting wage for new graduates and the number of new graduate workers are separately inquired by gender. Additionally, those from universities are separately inquired for clerical and technical workers. Those from senior high school in the waves 1990 and 2000 are separately inquired for clerical and production workers. This study pools these worker types in each school category. Average starting wages for new senior high school and college graduates are calculated weighted by the number of each type of worker. These wages are adjusted for inflation by the 2020 base CPI for all items.

College share: it is calculated using data from the Population Census as follows:

$$College\ share = \frac{college}{graduates - schools\ unknow} \times \frac{graduates}{population - status\ unknown}$$

where $college$ is a the number of university or graduate school graduates; $graduates$ is the number of all graduates; $schools\ unknow$ is the number of graduates whose schools graduated are unknown; $population$ it the population aged 15 or over; $status\ unknown$ is the population whose schooling or graduation status is unknown. This calculation method follows Statistics Bureau, Ministry of Internal Affairs and Communications (2024).

Population share aged 16–24: It is a faction of a population aged 16–24 to a population for all ages minus a population whose ages are unknown, using data from the

Population Census.

Population share aged 65 and over: It is the same as the above.

Appendix B. Supplemental tables

[Tables B1 and B2]

References

Acemoglu, D., & Angrist, J. (2000). How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws. *NBER Macroeconomics Annual*, *15*, 9–59. https://doi.org/10.1086/654403

Cui, Y., & Martins, P. S. (2021). What drives social returns to education? A meta-analysis. *World Development*, *148*, 105651. https://doi.org/10.1016/j.worlddev.2021.105651

Glaeser, E., & Lu, M. (2018). Human-Capital Externalities in China. NBER Working Paper No. 24925. https://doi.org/10.3386/w24925.

Kanemoto, Y., & Tokuoka, K. (2002). Proposal for the standards of metropolitan areas of Japan. *Journal of Applied Regional Science,* 7, 1–15 (in Japanese).

Liu, S., & Yang, X. (2021). Human capital externalities or consumption spillovers? The effect of high-skill human capital across low-skill labor markets. *Regional Science and Urban Economics*, *87*, 103620. https://doi.org/10.1016/j.regsciurbeco.2020.103620

Mazzolari, F., & Ragusa, G. (2013). Spillovers from High-Skill Consumption to Low-Skill Labor Markets. *Review of Economics and Statistics*, *95*(1), 74–86. https://doi.org/10.1162/REST_a_00234

Moretti, E. (2004a). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, *121*(1–2), 175–212. https://doi.org/10.1016/j.jeconom.2003.10.015

Moretti, E. (2004b). Chapter 51—Human Capital Externalities in Cities. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2243–2291). Elsevier. https://doi.org/10.1016/S1574-0080(04)80008-7

Sand, B. M. (2013). A re-examination of the social returns to education: Evidence from U.S. cities. *Labour Economics*, *24*, 97–106. https://doi.org/10.1016/j.labeco.2013.07.001

Sano, S. (2011). The Social Return to Education in Japan. *Kokumin-Keizai Zasshi*, 204(4), 63–74 (in Japanese).

Sano, S. (2018). "Estimating the social returns to education in Japan." Economics Association for Chiba University WORKING PAPER SERIES, #18E067 (in Japanese).

Statistics Bureau, Ministry of Internal Affairs and Communications. (2024). Final report of the 2020 Population Census: Population and Households of Japan. (in Japanese). https://www.stat.go.jp/data/kokusei/2020/kekka/pdf/all.pdf

Winters, J. V. (2014). STEM graduates, human capital externalities, and wages in the U.S. *Regional Science and Urban Economics*, *48*, 190–198. https://doi.org/10.1016/j.regsciurbeco.2014.07.003

Table 1: Summary statistics of variables for the worker-level wage function

	College graduates				Non-college graduates			
	1990	2000	2010	2019	1990	2000	2010	2019
Hourly wage (JPY)	3193.82	3263.84	3179.58	3145.56	1997.87	2211.72	2078.13	2114.43
	(2399.43)	(2223.81)	(2258.87)	(2102.87)	(1308.59)	(1574.69)	(1178.24)	(1188.13)
Age	36.08	37.37	39.06	40.48	38.35	40.11	41.87	43.95
	(9.94)	(10.42)	(11.21)	(11.88)	(12.75)	(12.80)	(12.61)	(12.88)
Female	0.09	0.14	0.21	0.27	0.36	0.34	0.37	0.40
	(0.28)	(0.35)	(0.41)	(0.44)	(0.48)	(0.48)	(0.48)	(0.49)
Years of tenure	9.98	10.77	10.74	11.36	10.65	11.88	11.78	12.19
	(8.62)	(9.21)	(9.78)	(10.40)	(9.77)	(10.66)	(11.01)	(11.17)
Junior high school					0.23	0.13	0.07	0.04
					(0.42)	(0.34)	(0.26)	(0.21)
High school					0.65	0.67	0.67	0.66
					(0.48)	(0.47)	(0.47)	(0.47)
Higher professional school/Junior college					0.12	0.20	0.26	0.29
					(0.32)	(0.40)	(0.44)	(0.45)
College	1.00	1.00	1.00	1.00				
	(0.00)	(0.00)	(0.00)	(0.00)				
Establishment size	621.05	526.48	600.90	554.89	424.76	324.24	374.36	346.62
	(1539.34)	(1333.30)	(1402.75)	(1275.58)	(1437.51)	(1025.62)	(1158.52)	(1153.47)
Observations	175854	259705	268411	329857	911730	786445	583936	580209

Notes: Standard deviations are in parentheses. Sampling weights in the BSWS are used.

Table 2: Summary statistics of variables for the establishment-level wage function

	College graduates				High school graduates			
	1990	2000	2010	2019	1990	2000	2010	2019
Monthly starting wage for new graduates (JPY in hundred)	1818.56	1950.79	2050.80	2044.73	1417.36	1570.33	1667.40	1682.22
	(185.92)	(216.30)	(241.83)	(218.92)	(176.53)	(185.83)	(198.51)	(178.71)
Establishment size	136.13	106.65	105.60	110.61	93.14	94.49	116.47	122.66
	(353.11)	(273.05)	(296.95)	(286.82)	(265.16)	(272.17)	(327.98)	(281.99)
Observations	8096	8700	8092	9264	13078	5905	4417	5782

Notes: Standard deviations are in parentheses. Sampling weights in the BSWS are used.

Table 3: Summary statistics of variables for the regional-level wage function

	Sample for estimation using worker-level hourly wage				Sample for estimation using establishment-level monthly starting wage for new graduate			
	1990	2000	2010	2020	1990	2000	2010	2020
College share	0.067	0.089	0.119	0.156	0.076	0.102	0.135	0.176
	(0.021)	(0.028)	(0.039)	(0.047)	(0.021)	(0.028)	(0.037)	(0.044)
Population share aged 16–24	0.128	0.114	0.090	0.083	0.136	0.119	0.094	0.088
	(0.022)	(0.017)	(0.014)	(0.013)	(0.018)	(0.013)	(0.010)	(0.009)
Population share aged 65 and over	0.146	0.207	0.261	0.330	0.137	0.193	0.246	0.311
	(0.028)	(0.037)	(0.040)	(0.046)	(0.024)	(0.031)	(0.030)	(0.032)
Female labor force participation rate	0.503	0.495	0.489	0.521	0.503	0.499	0.496	0.530
	(0.043)	(0.034)	(0.031)	(0.032)	(0.045)	(0.034)	(0.029)	(0.028)
Unemployment rate	0.029	0.044	0.066	0.039	0.029	0.044	0.065	0.038
	(0.012)	(0.013)	(0.016)	(0.007)	(0.012)	(0.012)	(0.014)	(0.006)
Observations	206	206	206	206	125	125	125	125

Table 4: Estimation results of regional-level wage function based on worker-level wages for all sample periods

	OLS				2SLS			
	College graduates		Non-college graduates		College graduates		Non-college graduates	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔCollege share	-0.158	0.474	-1.048***	-0.532***	-1.071	8.952	2.092	-0.170
	(0.156)	(0.329)	(0.099)	(0.137)	(3.363)	(9.017)	(2.105)	(2.643)
ΔPopulation share aged 16–24		1.577***		1.671***		3.566*		1.756**
		(0.325)		(0.203)		(2.067)		(0.703)
ΔPopulation share aged 65 and over		0.665**		0.428**		3.008		0.528
		(0.288)		(0.187)		(2.756)		(0.661)
ΔFemale labor force participation rate		-0.107		-0.009		-3.979		-0.174
		(0.302)		(0.153)		(4.020)		(1.218)
ΔUnemployment rate		-1.161**		-1.194***		0.508		-1.122
		(0.577)		(0.405)		(2.013)		(0.712)
Constant	0.048***	0.059**	0.109***	0.139***	0.079	-0.304	0.005	0.123
	(0.013)	(0.025)	(0.008)	(0.014)	(0.114)	(0.403)	(0.065)	(0.110)
First stage								
Demographic IV					-0.890***	0.601	-0.890***	0.601
					(0.305)	(0.465)	(0.305)	(0.465)
Observations	618	618	618	618	618	618	618	618
Adj. R2	0.222	0.277	0.686	0.745				
F-value (First stage)					8.513	1.668	8.513	1.668

Notes: Standard errors clustered by UEA are reported in parentheses. ***, **, and * represent statistical significance at 1%, 5%, and 10% level, respectively. Regressions are weighted by labor force of the corresponding UEA. All models contain year dummies.

Table 5: OLS estimation results of regional-level wage function based on worker-level wages by changes over a decade

	1990–2000				2000–2010				2010–2020			
	College graduates		Non-college graduates		College graduates		Non-college graduates		College graduates		Non-college graduates	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ΔCollege share	-3.581***	-1.330	-3.660***	0.309	1.048***	1.088**	-0.238	0.182	0.072	-0.581	-0.698***	-1.024**
	(0.486)	(0.970)	(0.310)	(0.535)	(0.274)	(0.543)	(0.239)	(0.443)	(0.350)	(0.810)	(0.253)	(0.438)
ΔPopulation share aged 16–24		0.483		1.589***		1.228		3.268***		-0.168		-0.461
		(0.413)		(0.242)		(1.193)		(0.762)		(1.618)		(0.844)
ΔPopulation share aged 65 and over		0.141		1.557***		0.744		0.200		-0.827		-0.132
		(0.788)		(0.433)		(0.503)		(0.259)		(0.852)		(0.505)
ΔFemale labor force participation rate		-1.097**		-0.530**		0.672		0.613**		0.262		0.796
		(0.469)		(0.231)		(0.412)		(0.264)		(0.788)		(0.484)
ΔUnemployment rate		-3.235***		-1.912***		1.483		1.063*		-0.932		-0.983**
		(0.946)		(0.511)		(0.938)		(0.618)		(0.759)		(0.452)
Constant	0.162***	0.153**	0.195***	0.063*	-0.093***	-0.135***	-0.062***	-0.025	-0.016	0.031	0.064***	0.031
	(0.018)	(0.062)	(0.010)	(0.037)	(0.014)	(0.043)	(0.010)	(0.025)	(0.020)	(0.084)	(0.013)	(0.047)
Observations	206	206	206	206	206	206	206	206	206	206	206	206
Adj. R2	0.187	0.242	0.420	0.588	0.053	0.058	0.003	0.085	-0.005	-0.007	0.026	0.078

Notes: Standard errors clustered by UEA are reported in parentheses. ***, **, and * represent statistical significance at 1%, 5%, and 10% level, respectively. Regressions are weighted by labor force of the corresponding UEA. All models contain year dummies.

Table 6: IV estimation results of regional-level wage function based on worker-level wages by changes over a decade

	1990–2000				2000–2010				2010–2020			
	College graduates		Non-college graduates		College graduates		Non-college graduates		College graduates		Non-college graduates	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ΔCollege share	-7.845	50.769	-4.100	10.497	1.359***	1.484**	0.098	0.620	1.293*	2.779**	-0.920**	-1.038
	(5.271)	(181.548)	(3.234)	(38.878)	(0.301)	(0.719)	(0.360)	(0.637)	(0.736)	(1.400)	(0.438)	(0.697)
ΔPopulation share aged 16–24		18.253		5.064		1.543		3.616***		-0.761		-0.459
		(62.599)		(13.401)		(1.108)		(0.778)		(1.486)		(0.833)
ΔPopulation share aged 65 and over		22.307		5.891		0.721		0.175		1.072		-0.141
		(78.144)		(16.643)		(0.492)		(0.263)		(1.099)		(0.580)
ΔFemale labor force participation rate		-4.712		-1.237		0.588		0.520*		-0.884		0.801
		(12.512)		(2.741)		(0.443)		(0.288)		(0.944)		(0.508)
ΔUnemployment rate		1.223		-1.040		1.873*		1.494**		-1.358*		-0.981**
		(19.411)		(3.972)		(0.992)		(0.702)		(0.797)		(0.446)
Constant	0.302*	-2.285	0.209*	-0.414	-0.107***	-0.148***	-0.077***	-0.040	-0.077**	-0.220*	0.076***	0.032
	(0.174)	(8.583)	(0.109)	(1.831)	(0.016)	(0.049)	(0.014)	(0.031)	(0.036)	(0.123)	(0.022)	(0.061)
First stage												
Demographic IV	-1.802*	0.269	-1.802*	0.269	-10.539***	-6.674***	-10.539***	-6.674***	5.245***	3.147***	5.245***	3.147***
	(0.946)	(0.949)	(0.946)	(0.949)	(1.115)	(0.776)	(1.115)	(0.776)	(1.082)	(0.252)	(1.082)	(0.252)
Observations	206	206	206	206	206	206	206	206	206	206	206	206
F-value (First stage)	3.628	0.080	3.628	0.080	89.375	73.963	89.375	73.963	23.488	156.408	23.488	156.408

Notes: Standard errors clustered by UEA are reported in parentheses. ***, **, and * represent statistical significance at 1%, 5%, and 10% level, respectively. Regressions are weighted by labor force of the corresponding UEA. All models contain year dummies.

Table 7: Estimation results of regional-level wage function based on establishment-level starting wages for new graduates for all sample periods

	OLS				2SLS			
	College graduates		High school graduates		College graduates		High school graduates	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔCollege share	-0.082	0.277	0.029	0.392*	6.297*	-160.245	3.209	-39.524
	(0.093)	(0.200)	(0.128)	(0.212)	(3.377)	(2305.853)	(2.802)	(581.754)
ΔPopulation share aged 16–24		-0.084		0.913***		-39.970		-9.006
		(0.264)		(0.305)		(573.905)		(144.709)
ΔPopulation share aged 65 and over		0.865**		0.383		-45.197		-11.071
		(0.335)		(0.448)		(668.674)		(168.452)
ΔFemale labor force participation rate		-0.342*		-0.131		73.049		18.119
		(0.197)		(0.211)		(1053.459)		(265.539)
ΔUnemployment rate		-0.625		-0.911*		-33.674		-9.130
		(0.461)		(0.516)		(480.872)		(121.202)
Constant	0.078***	0.031	0.111***	0.124***	-0.141	6.945	0.002	1.843
	(0.006)	(0.024)	(0.006)	(0.032)	(0.119)	(99.773)	(0.089)	(25.159)
First stage								
Demographic IV					-1.216***	0.039	-1.216***	0.039
					(0.363)	(0.559)	(0.363)	(0.559)
Observations	375	375	375	375	375	375	375	375
Adj. R2	0.393	0.422	0.513	0.550				
F-value (First stage)					11.199	0.005	11.199	0.005

Notes: Standard errors clustered by UEA are reported in parentheses. ***, **, and * represent statistical significance at 1%, 5%, and 10% level, respectively. Regressions are weighted by labor force of the corresponding UEA. All models contain year dummies.

Table 8: OLS estimation results of regional-level wage function based on establishment-level starting wages for new graduates by changes over a decade

	1990–2000				2000–2010				2010–2020			
	College graduates		High school graduates		College graduates		High school graduates		College graduates		High school graduates	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ΔCollege share	-0.870*	0.128	-2.041***	0.726	0.709**	0.526	0.573	-0.264	-0.862***	-1.065	0.391	0.469
	(0.510)	(0.910)	(0.365)	(0.714)	(0.283)	(0.504)	(0.435)	(0.674)	(0.295)	(0.679)	(0.384)	(0.579)
ΔPopulation share aged 16–24		-0.252		0.998***		0.191		0.489		-3.699***		-3.585***
		(0.454)		(0.350)		(1.015)		(1.362)		(1.183)		(0.990)
ΔPopulation share aged 65 and over		0.650		1.137*		1.261**		0.553		-1.116*		-1.036
		(0.641)		(0.660)		(0.587)		(0.496)		(0.576)		(0.647)
ΔFemale labor force participation rate		-0.795**		-0.776**		0.053		0.586		-0.408		-0.635
		(0.340)		(0.342)		(0.407)		(0.418)		(0.619)		(0.687)
ΔUnemployment rate		-2.118**		-0.852		0.434		-0.574		-0.052		-0.846
		(1.061)		(0.740)		(1.137)		(0.817)		(0.655)		(0.618)
Constant	0.105***	0.070	0.182***	0.078*	0.013	-0.054	0.028	0.051	0.044***	0.115*	-0.004	0.037
	(0.017)	(0.051)	(0.013)	(0.046)	(0.013)	(0.037)	(0.018)	(0.044)	(0.016)	(0.059)	(0.019)	(0.059)
Observations	125	125	125	125	125	125	125	125	125	125	125	125
Adj. R2	0.025	0.099	0.172	0.283	0.046	0.081	0.025	0.046	0.043	0.116	0.004	0.105

Notes: Standard errors clustered by UEA are reported in parentheses. ***, **, and * represent statistical significance at 1%, 5%, and 10% level, respectively. Regressions are weighted by labor force of the corresponding UEA. All models contain year dummies.

Table 9: IV estimation results of regional-level wage function based on establishment-level starting wages for new graduates by changes over a decade

	1990–2000				2000–2010				2010–2020			
	College graduates		High school graduates		College graduates		High school graduates		College graduates		High school graduates	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ΔCollege share	-7.636	1.736	-3.632	-2.844	1.017**	0.832	0.750	-0.160	-2.052***	-2.633**	-0.699	-0.824
	(9.948)	(13.840)	(5.391)	(17.413)	(0.445)	(0.688)	(0.542)	(0.855)	(0.757)	(1.140)	(0.883)	(1.021)
ΔPopulation share aged 16–24		0.293		-0.211		0.429		0.570		-3.532***		-3.448***
		(4.666)		(5.846)		(1.030)		(1.271)		(1.136)		(0.956)
ΔPopulation share aged 65 and over		1.347		-0.411		1.240**		0.546		-1.983***		-1.751**
		(5.993)		(7.707)		(0.586)		(0.491)		(0.769)		(0.737)
ΔFemale labor force participation rate		-0.938		-0.459		-0.005		0.566		0.210		-0.126
		(1.243)		(1.535)		(0.412)		(0.434)		(0.754)		(0.756)
ΔUnemployment rate		-1.947		-1.233		0.737		-0.471		0.113		-0.710
		(1.679)		(1.671)		(1.179)		(0.978)		(0.650)		(0.609)
Constant	0.336	-0.007	0.236	0.248	-0.001	-0.064	0.019	0.048	0.105***	0.227***	0.052	0.129
	(0.341)	(0.660)	(0.184)	(0.834)	(0.019)	(0.041)	(0.022)	(0.050)	(0.037)	(0.086)	(0.041)	(0.080)
First stage												
Demographic IV	-0.983	0.454	-0.983	0.454	-10.168***	-6.989***	-10.168***	-6.989***	4.967***	2.782***	4.967***	2.782***
	(1.008)	(1.123)	(1.008)	(1.123)	(1.129)	(0.830)	(1.129)	(0.830)	(1.227)	(0.278)	(1.227)	(0.278)
Observations	125	125	125	125	125	125	125	125	125	125	125	125
F-value (First stage)	0.951	0.163	0.951	0.163	81.079	70.946	81.079	70.946	16.397	100.265	16.397	100.265

Notes: Standard errors clustered by UEA are reported in parentheses. ***, **, and * represent statistical significance at 1%, 5%, and 10% level, respectively. Regressions are weighted by labor force of the corresponding UEA. All models contain year dummies.

Table B1: Estimation results of worker-level wage functions

	College graduates	Non-college graduates
	(1)	(2)
Age	0.068***	0.048***
	(0.006)	(0.002)
Age squared/100	-0.069***	-0.057***
	(0.006)	(0.002)
Female	-0.113***	-0.229***
	(0.005)	(0.013)
Tenure	0.034***	0.032***
	(0.002)	(0.000)
Tenure squared/100	-0.047***	-0.027***
	(0.004)	(0.001)
Establishment size/100	0.011***	0.009***
	(0.001)	(0.001)
Establishment size/100 squared/100	-0.006***	-0.004***
	(0.000)	(0.001)
Education (Reference: Junior high school)		
High school		0.149***
		(0.005)
Higher professional school/Junior college		0.301***
		(0.005)
Observations	1033827	2862320
Adj. R2	0.494	0.541

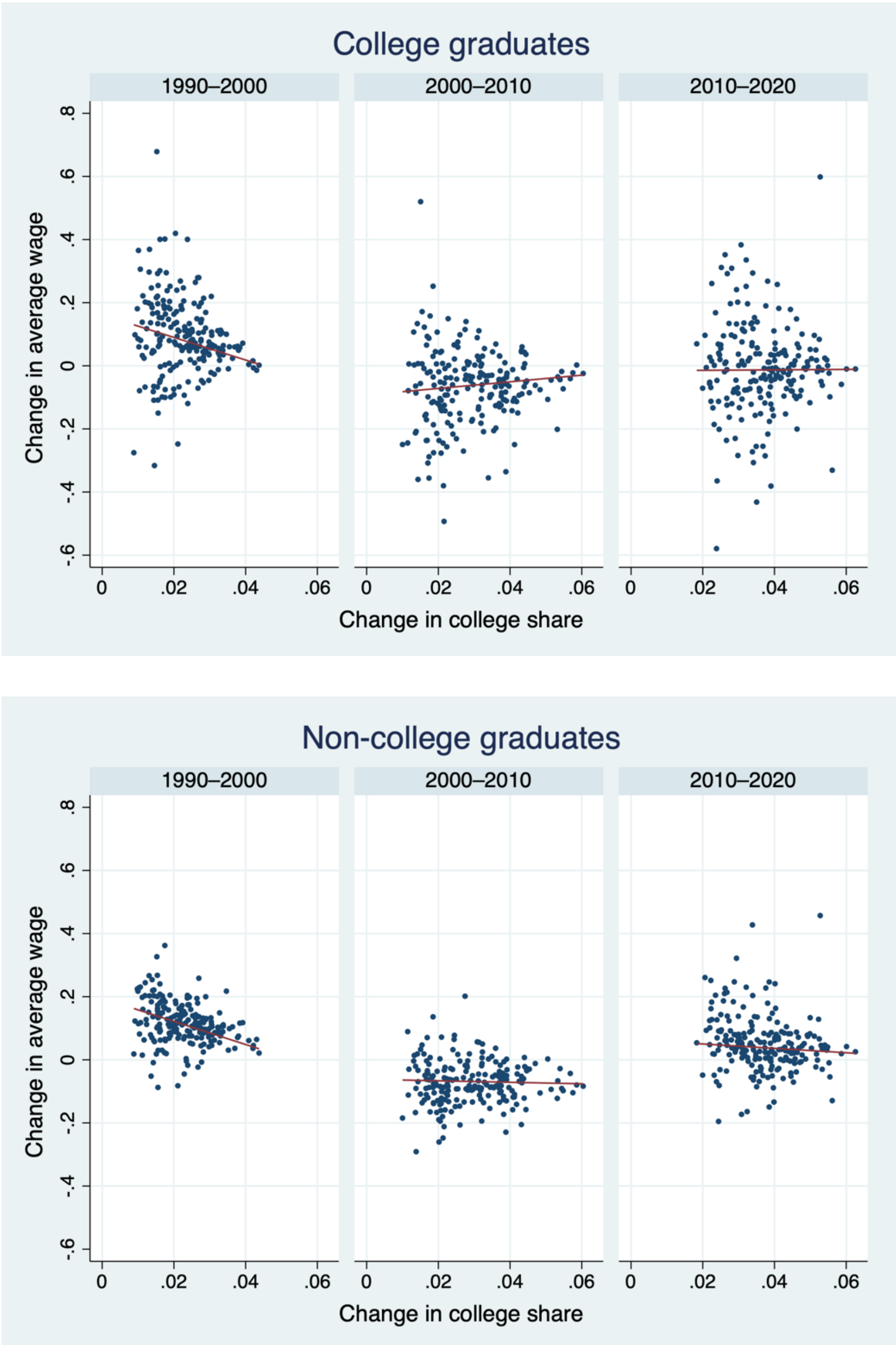
Notes: Standard errors clustered by UEA are reported in parentheses. ***, **, and * represent statistical significance at 1%, 5%, and 10% level, respectively. Sampling weights in the BSWS are used. All models contain a constant term and interaction terms of UEA dummy with year dummy.

Table B2: Estimation results of establishment-level wage functions using starting wage for new graduates

	College graduates	High school graduates
	(1)	(2)
Establishment size/100	0.003***	0.001*
	(0.001)	(0.001)
Establishment size/100 squared/100	-0.002***	-0.001
	(0.000)	(0.001)
Observations	34152	29182
Adj. R2	0.291	0.462

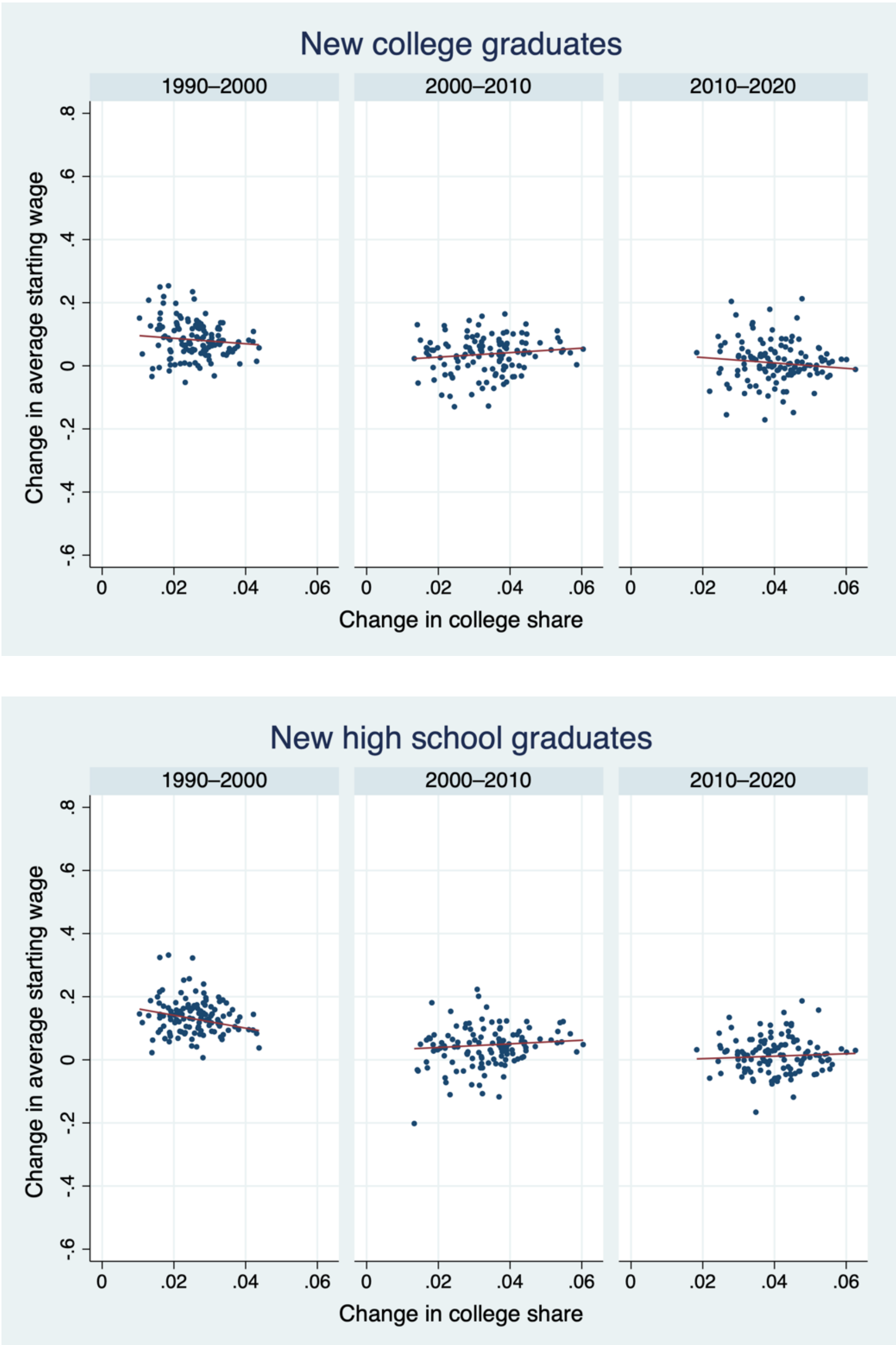
Notes: Standard errors clustered by UEA are reported in parentheses. ***, **, and * represent statistical significance at 1%, 5%, and 10% level, respectively. Sampling weights in the BSWS are used. All models contain a constant term and interaction terms of UEA dummy with year dummy.

Figure 1: Relationship between a change in regional average wages and college share based on worker-level wages



Notes: Each dot represents UEA. The changes in regional wages are estimated by the worker-level wage function (equation (1)).

Figure 2: Relationship between a change in regional average wages and college share based on establishment-level starting wages for new graduates



Notes: Each dot represents UEA. The changes in regional wages are estimated by the establishment-level wage function using the starting wages for new graduates (equation (2)).