Higher Education Expansion and the Rise of the Skill-Intensive Service Sector

(Click here for the latest version)

Khoa Vu*

November, 2024

Abstract

Vietnam established over 100 new universities during 2006 and 2013, drastically improving access to higher education across the country. I examine the implications of this labor supply shock on the rise of the skill-intensive service sector and the general equilibrium implications for the local labor markets. The expansion induces both firms and workers to reallocate away from agriculture and basic services, and into skill-intensive service industries. This reallocation raises the overall employment rate, hourly wages and firm-level total factor productivity. The implied returns to higher education is 41%, but the reallocation effect raises the returns by 4 times.

Keywords: higher education, local labor market, service sector, structural transformation, Vietnam

*I am thankful for the financial support from ERIA to pursue this study. I also benefited from comments from Paul Glewwe, Marc Bellemare, Jason Kerwin, Aaron Sojourner, Nguyen Vuong, Ian Coxhead, Abhijeet Singh, the audience from the Applied Young Economists Webinar, the Arndt-Corden Department of Economics Seminar, the Grand Valley State University Economic Department, the Armenian Economic Association, the TAMU Bush School's International Affairs Department, and many others for their valuable comments. All errors are mine.

1 Introduction

Skill-intensive service industries such as finance and IT are becoming increasingly important in the global economy.[1] The lack of access to higher education thus creates significant barriers for workers seeking to enter these industries, especially in less developed countries or rural areas. This limitation also imposes indirect costs on employers by restricting their access to specialized human capital. Such constraints can prevent firms from entering or expanding in skill-intensive service industries. For example, establishing a bank is more costly without a local supply of college-educated workers trained in banking and finance. High mobility friction, such as the *hukou* systems in China or the caste system in India,[2] can intensify these barriers, as employers cannot attract skilled workers to move with higher wages.

Expanding higher education can alleviate this barrier and shift both workers and firm activities toward skill-intensive service industries. The rise of these industries can subsequently lead to long-term economic development and shape the labor market outcomes for both college-educated and non-college workers.[3] Yet there is limited evidence for the causal impacts of expanding access to higher education on the rise of the skill-intensive service sector, especially on the employer side. We also know little about the general equilibrium implications for employment and college wage premiums in the local labor markets when *both* the supply of skilled labor and the skill-intensive sector expand.

I study these questions in the context of a national expansion of universities in Vietnam between 2006 and 2013, which established over 100 new universities in 63 provinces at different times, resulting in variation in ages of exposure to the expansion at the individual

[1] For example, both China and India went through a rapid structural change over the last two decades. China's export share of GDP fell down substantially between 2005 and 2020, as the country shifted toward business services. India also shifted toward services, but the trend was driven by the growth of consumer services (Fan et al., 2023).

[2] See, e.g., Bosker et al. (2012); Munshi (2019) and Herrendorf and Schoellman (2018).

[3] The reallocation of firm activities in the skill-intensive industries can raise demands for non-college workers since they complement college-educated workers.

level and years of exposure at the market level. My difference-in-differences (DiD) research design leverages these variations along with the policy's abrupt termination in 2013, which occurred due to political pressure and public opinion. The policy's unexpected end left many approved universities unestablished, enabling me to construct an "almost-treated" control group consisting of provinces that would have received universities had the policy continued.

I first use an age-cohort DiD design to document a robust relationship between exposure to the expansion and labor reallocation from agriculture and basic service industries, e.g., food and hospitality, to skill-intensive service sectors, e.g., healthcare, education (besides higher education), and finance, at the worker level. Estimated distributional effects on wages indicate that workers at the bottom of the wage distribution benefit most from the expansion. I apply a falsification test using the almost-treated group as a placebo treatment group to validate the parallel trends assumption.[4] I also test whether the results are robust when relaxing the parallel trends assumption by using the Honest DiD[5] and change-in-changes.[6]

Such results do not capture general equilibrium adjustments, e.g., firms acquiring capital and technology or firms shifting into services.[7] I employ a market-level DiD design comparing provinces over time to examine these general equilibrium adjustments. Using firm-level census data, I find that firms are more likely to supply skill-intensive services. In other words, as the expansion increases the pool of college-educated workers, so does the production of skill-intensive services. Importantly, firm-level total factor productivity rises on average as they move into skill-intensive industries, but firm capital intensity decreases since these industries tend to use less capital.

[4]Since new universities were never actually established in these almost-treated group, the results should be small and statistically insignificant.

[5]This approach extends the conventional pre-trends testing by allowing the post-treatment trends to be violated as much as the pre-treatment trends and estimating a confidence set in which the true effect falls.

[6]Change-in-Changes relies on the assumption that the unobservables can be summarized in one index and the distribution of such index can vary across individuals but not over time. This relaxes the parallel trends assumption and, more importantly, does not depend on the functional form of the dependent variable, which may be a crucial issue for conventional difference-in-differences (Roth and Sant'Anna, 2023).

[7]See, e.g., Duflo (2004); Porzio et al. (2022) and Khanna (2023).

At the local labor market level, I find that the expansion of higher education raises overall employment. The employment rate increases for *both* college and non-college young workers, consistent with the expansion of the skill-intensive service sectors. In other words, as more firms join the skill-intensive service industries, they employ both college and non-college workers, thus raising the overall employment level. This is an important spillover effect of expanding higher education that is often under-discussed in the literature. The expansion has a negative effect on the college premiums, but substantially less for the young cohorts.

These results are consistent with the economic intuition of the skill-biased structural transformation framework (Buera et al., 2022). The expansion generates two opposite forces on the college wage premium. First, as the supply of college-educated workers increases, the college wage premium decreases because non-college workers are imperfect substitutes for college-educated workers. Second, it induces firms to enter the skill-intensive service, which raises labor demands in these sectors, especially for college-educated workers. This increases the college wage premium. The older cohort is unaffected by the rise of the service sector, so their college wage premium is only driven by the negative force. The two forces are both at play for the young cohort, so the net effect on college premiums is smaller.

My findings are analogous to those of the endogenous skill-biased technological change (Acemoglu, 1998, 2007) and skill-biased capital (Lewis, 2013; Khanna, 2023) framework. However, the implications for productivity and capital accumulation are different, allowing me to empirically test these alternative theories. In the endogenous technological change framework, technological change completing skills is induced as the supply of skilled labor increases. In the skill-biased capital framework, firms are induced to accumulate physical capital complementing college-educated workers. My results suggest that firms do not necessarily become more productive, but there are more firms in more productive sectors with lower capital intensity. In other words, the role of higher education in driving the reallocation of labor and firm activities into the skill-intensive sector is empirically more important.

The policy relevance of this study is twofold. First, the expansion leads to more productive firms on average but physical capital declines. Depending on the developmental stage of the economy, this may or may not be the desired policy outcome. Second, the reallocations of firm activities and workers have crucial effects on policy parameters, such as the returns to higher education or the elasticity of substitution between college and non-college workers. In particular, the presence of reallocation tends to inflate the conventional estimates for these parameters.

I devise an empirical approach based on Khanna (2023) to recover the correct returns. Returns to higher education are estimated to be 41% when correcting for the general equilibrium adjustment, which is 4 times lower when not correcting. Similarly, the elasticity of substitution is also inflated; my reduced-form results suggest that correcting for GE adjustment decreases the conventional elasticity of substitution from 1.96 to 1.36. As a benchmark, Buera et al. (2022)'s structural model estimates this elasticity to be 1.53 (after accounting for the labor reallocation) using data for 15 advanced economies for 1970-2005.

Relevant Literature

My paper contributes to three main bodies of literature. A growing number of micro-level papers studies the general equilibrium effects of educational expansion (e.g., Duflo, 2004; Bianchi, 2020; Porzio et al., 2022; Khanna, 2023), and of increased supply of skilled labor (e.g., Acemoglu, 1998; Goldin and Katz, 2009; Lewis, 2013). The main emphasis of this literature is technological change and capital deepening. Only a few studies document the effect on the structure of the local economy. My study is closely related to Porzio et al. (2022), whose focus is on basic education and the labor reallocation out of agriculture. My paper focuses on higher education and the expansion of the service sector, including changes on the supply side of the skill-intensive service market. I show that these have important effects on *both* college-educated and non-college workers. More importantly, firm-level TFP increases because of the compositional shift towards skill-intensive sectors.

The paper also contributes to a broader research and policy discussion around the values of universities and access to higher education (Hanushek, 2016; Valero and Van Reenen, 2019). Higher education is traditionally valued based on college graduates' employment and wages. More recent work also focuses on the ability to drive innovation and technological change (Hausman, 2022; Andrews, 2023). I argue in this paper that another fundamental role of higher education is that it unlocks the local supply of skilled labor, lowering the cost for firms to enter skill-intensive industries.[8] Such effect is crucial because it directly affects college graduates' careers and returns to higher education. In a similar setting, China experienced a large national expansion of higher education in the late 1990s, which, however, led to an increase in the unemployment rate and lower relative wages (Knight et al., 2017; Li et al., 2017), likely because the local labor demands for college graduates do not strongly correspond to the supply shock.

Lastly, this study also contributes to the ongoing debate around the rise of the service sector in less developed countries and the implications for long-term economic growth.[9] China and India, for example, both experienced a rapid shift toward services due to a rise of productivity within the service sector (Fan et al., 2023; Chen et al., 2023). Notably, both countries also scaled up access to higher education in the last three decades. This study highlights the role of higher education as a potential driver of both productivity and the rise of the skill-intensive service sector. The evidence from Vietnam is highly relevant, given that Vietnam also underwent a rapid shift in both structural changes and productivity (McCaig and Pavcnik, 2013, 2018) and, more recently, access to higher education.

[8]See also Buera and Kaboski (2012) and Buera et al. (2022).

[9]See, e.g., Rodrik (2016); Fujiwara and Matsuyama (2020); Sen (2023); Baldwin et al. (2024) and Huneeus and Rogerson (2024). See also Nayyar et al. (2021) for an extensive review of this issue.

2 Background

In 2006, the Vietnamese government announced Decree 121/2007, which allowed local governments and private enterprises to apply to establish universities as long as they can show sufficient needs and capacity to operate one. Prior to the expansion, the higher education system of Vietnam relied on three major clusters of universities in Hanoi and Thai Nguyen (in the North), Hue and Da Nang (in the Central), and Ho Chi Minh City and Can Tho (in the South), as shown in Figure 5(a). While this model allowed the government to prioritize resources and regulate education quality in these regions, it created geographic inequality in access to higher education, as students from many provinces had to travel to a different province if they wanted to attend a university. This led to a bottleneck in the supply of college education, as demands for skilled labor rose. The expansion resolved this bottleneck as it led to numerous regions opening a university for the first time as observed in in Figures 5(b) and (c).

A crucial aspect of the expansion that I utilize in my research design is that the policy was intended to be effective between 2006 and 2020, but the central government announced in 2013 that the policy had met the goal early and thus, halted any further establishment of new universities. This unexpected end of the expansion was driven by many reasons: public concerns about the rapid expansion compromising education quality and future career prospects of college graduates. Many universities were announced to be open but eventually were not because of this policy change. Figures 5 (c) shows the provinces in which these universities were supposed to be open.

During the expansion period, both private and public universities were established. Private universities were only established in major cities where there have already been existing universities. These were run by private enterprises or foreign universities with campuses in Vietnam. In contrast, public universities were established by local governments who wanted to increase access to higher education in their areas. Most public universities had relatively low tuition and were funded through the government's budget.

The focus of this paper is provinces with a university for the first time due to the expansion. My control group comprises of provinces without any university during the study period. One potential concern with this design is that these two groups may be very different from each other in terms of observable and unobservable characteristics. Using various sources of data, I compare the province-level characteristics of the treatment group, i.e., provinces with the first university ever, and the never-treated group in Table 1. First, I observe that the two types of provinces do not differ substantially in terms of the share of college and high school graduates. The unemployment and self-employment rates are also similar. However, the treatment group tends to have a lower share of agricultural workers, and higher shares in manufacturing and service. This represents a threat to my identification strategy. it is possible that the labor reallocation already happened prior to the expansion in the treatment provinces. Other economic conditions are relatively similar across the two groups.

To address the concerns about differences in the sectoral compositions of the economy, I also use the "almost-treated" group as the control group. These provinces are a subgroup of the overall control group which never had a university, but they would have had the first university had the policy not been ended in 2013. As expected, the almost-treated provinces are much more similar to the treatment provinces in many aspects. In particular, the economic compositions of the labor force between the two groups are much more comparable. This lends more credibility to this group as an alternative control group for the DiD design.

3 Theoretical Framework

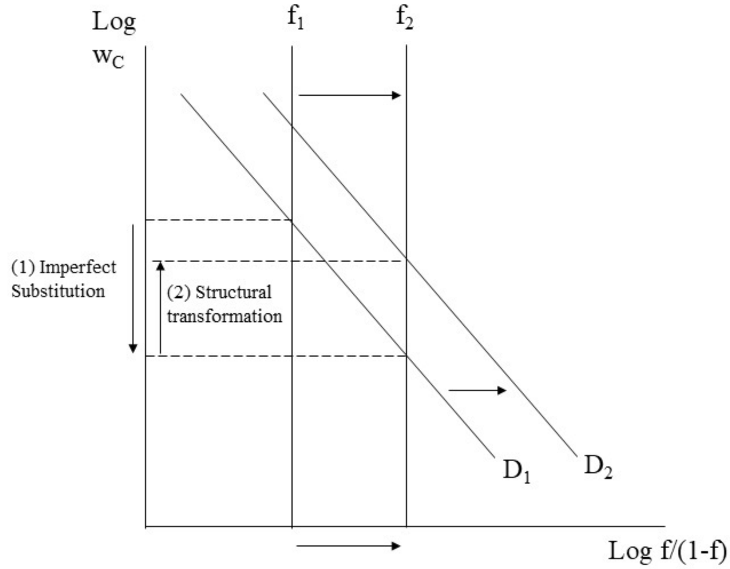
Individuals face the cost of acquiring a university degree when physical access to a university is limited. Expanding access to higher education will substantially reduce this cost, thus increasing the chance that an individual would obtain a college degree. Since skill-intensive sectors tend to pay more for college-educated workers, these workers will naturally shift into these industries.

At the market level, the overall supply of college-educated workers will rise. I use the skill-biased structural change framework proposed by Buera et al. (2022) (see Appendix A) to develop the economic intuition for the general equilibrium. The key implication of this framework is that an increase in the supply of college-educated workers creates two opposite forces on the college wage premium, as summarized in Figure 1.

First, because non-college workers are imperfect substitutes for college-educated workers, a rise in supply of one type would lead to a decline in the wage of the other type. Second, as the economy has more college-educated workers who earn more than non-college workers, the aggregate expenditure towards service consumption also goes up as a result, and, because the service sector is skill-intensive, it also increases demand for college-educated workers. This structural transformation effect, thus, increases the college wage premium.

I derive the formal model in Appendix A. It is useful to note that other theories, such as endogenous technological change or capital-skill complementarity also yield a similar prediction about the ambiguous college wage premium effect. That is, if α_j increases as a result of the increased supply of college-educated workers, sector j becomes more productive (or acquires more capital), thus driving up demand for skilled labor. I test for these alternative theories using firm-level data on total factor productivity and capital intensity.

Figure 1: Skill-Biased Structural Change



4 Data

This study uses several data sources. Data on individual and market-level labor market outcomes are based on the Labor Force Survey (LFS). The General Statistics Office (GSO) of Vietnam has been collected annual data on labor market outcomes annually since 2008, but detailed wage data is only available since 2011. For the individual-level analysis, I focus on college education, hourly wage, and employment as the main outcomes of interest. The LFS asks respondents for their highest educational attainment, which I use to construct a binary variable for whether an individual has obtained a 4-year university degree or more. Employment is measured by whether respondents are employed, as opposed to being self-employed or not working.[10] Hourly wages only include wage compensation. but not other benefits such as bonuses.[11] This information is available for most workers, including those who are self-employed.

[10] I also explore a variation in which employment is measured by whether individuals earn a wage.
[11] This is because questions about bonus are not consistent across years.

For the market-level analysis in the second part of the paper, I aggregate data up to the province-by-year-by age cohort level. Following the theoretical framework discussed in Appendix A, I define relative supply as $\ln{(\frac{C}{N})}$ where C and N denote the numbers of college-educated and non-college workers, respectively. The college wage premium is measured as the log ratio of college to non-college hourly wages, namely, $\ln{(\frac{w_C}{w_N})}$.

I use firm-level census data from the Vietnam Enterprise Census (VEC) for 2006-2019 to study firm adjustment to the labor suppl shock. This data contains detailed accounting information on firms' annual operations, such as short- and long-term assets, total labor, total revenue, and industries. The main outcomes of interest are whether firms are more likely to be in skill-intensive service sectors as the share of college-educated workers increases. Thus, I look at the industry of firms, e.g., agriculture, manufacturing, basic services, and skill-intensive services, as the outcome variables.

To explore the alternative theories such as endogenous technological change and capital-skill complementarity, I look at the effects on firm-level productivity and capital intensity. At the basic level, I use labor productivity measured by value added per labor, where value-added is defined as profit plus wage (Newman et al., 2015). But labor productivity does not necessarily capture technology as a residual of the production function. Thus, I also use total factor productivity at the firm level. To measure the TFP of each firm, I first estimate the relevant production function for all 2-digit industries using Ackerberg et al. (2015)'s approach then use the estimated parameters to obtain the total factor productivity (see Appendix B for a detailed discussion about the estimation process and results). Lastly, capital intensity is measured as the ratio of capital to revenue.

5 Empirical Strategy

I employ two different difference-in-differences research designs that exploit different sources of variation. The first variation is in the year of exposure to a new university at the province level. Within provinces, individuals within the college-going age can benefit from the expansion, while individuals older than the college-going age cannot, so the second variation is in the age of exposure to a new university. Using these sources of variation, I discuss the methods to identify the partial and general equilibrium effects of exposure to the expansion. A conventional approach to measure the labor market effects of the expansion is to use the age of exposure variation, while controlling for location-year and age cohort fixed effects. This approach, as pointed out by Khanna (2023), can only identify the partial equilibrium effect of a school expansion because all age cohorts are affected by the general equilibrium shifts. Thus, I use the variation in the year of exposure to identify the general equilibrium effects by comparing provinces over time.

In both DiD approaches, I compare provinces that had the first university ever to provinces that never had a university during the study period. In each design, I assume that the outcomes would have evolved similarly between the two types of provinces had they not been exposed to the expansion. This assumption is violated if the treatment group and the never-treated group are very different from each other. I use the fact that the expansion ended abruptly and many announced universities did not open because they missed this window to construct a group of provinces that were almost treated. This group comprises the never-treated provinces where universities were announced to be opened but were eventually not because of the abrupt end in 2013. This group would have been similar to the provinces that actually opened a university, had the expansion not happened.

5.1 Age-Cohorts Difference-in-Differences

Formally, let g denote the year that a given province has its first university ever, t the survey year, and c the birth year of a given cohort. Let G_g denote the group of provinces with the same treatment year and $g = \inf$ for the never-treated group. For each group of provinces with the same treatment year G_g in the same survey year, I compare them to the never-treated provinces as the control group across the exposed and unexposed birth cohorts.

To do this, for each G and t, I can form a dataset comprised of provinces treated in year G and inf (i.e., never treated) who appear in survey year t, where $t - G \geq 5$, and estimate the following two-way fixed effects (TWFE) regression:

$$y_{i,p,c}^{G,t} = \delta^{G,t}(T_p^{G,t} \times Exposed_c^{G,t}) + \gamma_p + \eta_c + \epsilon_{i,p,c}$$

where $y_{i,p,c}$ denotes the outcome of individual i in province p of cohort c; T_p indicates whether province p had the first university during the expansion; $Exposed_c$ indicates whether cohort c was 21 years old or younger when the first university was opened in year g. Province and cohort fixed effects are γ_p and η_c. Since we are comparing each treatment group to a never-treated group in a given year, $\delta^{G,t}$ captures the treatment effect of a specific group G in a specific year t. One can then aggregate $\delta^{G,t}$ across all G and t to obtain the average effect for different groups of provinces and for different years.

Alternatively, I can combine all datasets and estimate the following model:

$$y_{i,p,c,s} = \delta(T_{p,s} \times Exposed_{c,p,s}) + \gamma_{p,s} + \eta_{c,s} + \epsilon_{i,p,c,s} \tag{1}$$

where s denotes the subdataset for each G and t, while $\gamma_{p,s}$ and $\eta_{c,s}$ control for province-by-subdataset and cohort-by-dataset fixed effects. Thus, δ captures the weighted average of all $\delta^{G,t}$.

This stacked regression approach, proposed by Cengiz et al. (2019), allows me to compare each treatment group G with a clean control group in each given year, thus avoiding the problem of negative weights in a standard TWFE model (Sun and Abraham, 2021; Goodman-Bacon, 2021a). There are, of course, other different estimators such as Callaway and Sant'Anna (2021), Borusyak et al. (2024), and others. The stacked regression approach is, however, most applicable here, because it allows flexibility in handling two time dimensions, namely, age eligibility and year of establishment.

5.2 Estimating the Distributional Effects of the Expansion

The effects on wages may vary across local labor markets with different economic conditions. I use a change-in-changes estimator to recover the distributional effects of the expansion on wages, by extending Athey and Imbens (2006)'s approach to a staggered timing setting. Let F_{00} and F_{01} denote the distribution functions of the never-treated group for the unexposed and exposed cohorts, and F_{10}, F_{11} the distribution functions of the treatment group for the unexposed and exposed cohorts. The treatment effect is defined as $\delta^{CiC} = F_{Y^1,11} - F_{Y^0,11}$, where $F_{Y^1,11}$ is the distribution of the exposed age cohorts of the treatment provinces and $F_{Y^0,11}$ is the counterfactual distribution of the exposed age cohorts in the treatment group had it not been treated. As shown by Athey and Imbens (2006), one can estimate $F_{Y^0,11}$ as $F^{-1}_{Y,01}(F_{Y,00}(Y_{1,0}))$, where $F_{Y,00}(y)$ is the wage quantile of the unexposed age cohort of the never-treated group and $F^{-1}_{Y,01}(q)$ is the inverse distribution function of the unexposed treatment group.

This approach is also appealing in my context because it relies on an assumption different from the parallel trends assumption of the standard DiD approach. Let U_i denote the distribution of unobserables, the change-in-changes estimator assumes that U_G may vary between the treatment and control groups but not between the exposed and unexposed cohorts. This additional approach is thus not sensitive to whether the outcome variable is measured in log or level, which is not the case for the parallel trends assumption in the standard DiD (Roth and Sant'Anna, 2023).

5.3 Market-Level Difference-in-Differences

I estimate a similar stacked regression at the market level, in which the time dimension is years from being treated. For each treatment year G, I form a sub-dataset s of the treatment group treated in year g and the never-treated group, then stack the sub-datasets together and estimate the following regression:

$$y_{p,t,s} = \delta(T_{p,s} \times Exposed_{t,p,s}) + \gamma_{p,s} + \eta_{t,s} + \epsilon_{p,t,s} \tag{2}$$

where s denotes the subdataset for each G, while $\gamma_{p,s}$ and $\eta_{t,s}$ control for province-by-subdataset and year-by-dataset fixed effects.

Following Porzio et al. (2022) and Khanna (2023), I estimate the general equilibrium effects separately for the young and old age cohorts, with an age cutoff of 30. The young cohort is directly exposed to the expansion and, thus, is more likely to experience the labor reallocation effect than the old cohort. Both cohorts experience the negative effect of the labor supply shock. Thus,

5.4 Robustness Check, Falsification Test, and Pre-Trends Tests

While using the never-treated group allows me to sidestep the issue of treatment effect heterogeneity in a staggered timing DiD design, it raises a different issue about the parallel trends assumption. Provinces that have never had a university before may be very different from those that have had a university for the first time. It is unclear if labor market outcomes would have trended similarly across the two provinces, and thus the parallel trends assumption may be violated.

The almost-treated provinces can be used in two related checks for the parallel trends assumption because they are arguably more similar to the treated provinces than other never-treated provinces. First, I restrict the control group to the almost-treated group and estimate the same model. While this approach is less prone to violating parallel trends assumption, it uses fewer clusters than the main specification, so it is mainly useful as a robustness check. Second, as a falsification test, I use the almost-treated group as the treatment group. Since these provinces only came close to opening a university, the treatment effects of the expansion should be close to zero and statistically insignificant if the parallel trends assumption holds.

I also tap into the more recent development of econometric tools to test for the assumption of parallel trends. First, note that the parameter of interest can be rewritten as

$$\delta = \begin{pmatrix} \tau_{pre} \\ \tau_{post} \end{pmatrix} + \begin{pmatrix} \beta_{pre} \\ \beta_{post} \end{pmatrix}$$

where τ are the treatment effects in pre/post-period and β are the differences in trends between the treatment and never-treated groups in the absence of treatment (Rambachan and Roth, 2023). In other words, β captures the deviation from parallel trends in the respective period. If τ_{pre} is zero under the no anticipation assumption, and β_{post} is zero under the parallel trends assumption, then the DiD estimator captures the treatment effect: $\delta = \tau_{post}$.

The conventional approach to assessing the parallel trends assumption is to assume that $\beta_{pre} = \beta_{post}$ and test for whether $\beta_{pre} = 0$ using event study or falsification tests. Following Goodman-Bacon (2021b), I operationalize this idea by estimating an event study version on detrended outcomes, $Y_{detrended} = Y - \hat{Y}$, where $\hat{Y}$ is the fitted value of estimating Y against the interaction of treatment and a linear time trend for the pre-treatment period.[12]

[12]This regression includes all fixed effects as the main regression. See Goodman-Bacon (2021b).

This assessment can be generalized and formalized using the partial identification approach developed by Rambachan and Roth (2023). Rambachan and Roth (2023) approach estimate a confidence set for δ given some restriction set on trends $\Delta = \{\beta_{pre}, \beta_{post}\}$. Instead of imposing that $\beta_{pre} = 0$, I test whether the results are robust when the post-treatment trends are as large as the pre-treatment trends $\beta_{post} \leq \beta_{pre}$.

6 Results

6.1 Age-Cohort Results

In Table A5, I present the estimated effects on individual labor market outcomes. As described above, the main identification strategy is to compare provinces with a new university for the first time to provinces that never had a university during the study period across age cohorts. I also report two additional sets of results as checks for the parallel trends assumption. First, I replaced the never-treated group (provinces that never had a university) with the almost-treated group (provinces that announced a new university but did not establish one because of the abruption). As argued above, the almost-treated group is more similar to the treatment group than the never-treated group, so this serves as a useful robustness check for my main identification strategy. Second, I replaced the treatment group with the almost-treated group as a falsification test: As the almost-treated provinces never actually had a university, the results should be small and statistically insignificant. The main identification, the robustness check, and the falsification test are presented in Panels A, B, and C, respectively. Each column represents the corresponding outcome variable. I report the estimated treatment effects for age cohorts exposed to a university opening during 19 to 25 years old and age cohorts exposed during 14 to 18 years old.

In column (1), the main result in Panel A suggests that being exposed to a university increases the probability of completing college by 4.1 percentage points for those who were exposed during the age of 14 to 18. For those who were exposed during the age of 19 to 25, the effect is 1.9 percentage points, suggesting that slightly older cohorts can still benefit from a university opening. Panel B shows that the results are robust when using the almost-treated group as the control group: the estimated effects are 4.9 for the main age cohort and 2 for the older cohort. Panel C shows the placebo effect using the almost-treated group. The estimated effects are close to zero and statistically insignificant.

These results strongly suggest that the identifying assumption is not violated for this outcome. Assuming that the almost-treated provinces represent the treatment group had the expansion not happened, the falsification test's null results indicate that the college completion rates would have evolved parallel in the absence of the expansion. The event study in Figure 3 lends further support for this statement. Prior to the exposed age cohorts, the treatment effects are mostly zero. Among the exposed age cohorts, the effects are positive and significant for the main specification and the robustness check. For the exposed age cohorts of the almost-treated group, the effects are still close to zero.

The effects on log hourly wage are also substantial, but not robust. The main result from column (2) is 6.8% and statistically significant, but the robustness result is only 4.5% and insignificant. The falsification test also yields an effect of 4.4%. These results point to a violation of the parallel trends. This is also consistent with the event study in Figure A2, which shows that the wage effect is mostly driven by pretrends.[13] Being exposed to the expansion raises the probability of being employed by 1.16 to 13.4 percentage points, as indicated in column (3). This result is robust to the alternative specification, and is close to zero under the falsification test.

Columns (4) to (7) document the labor reallocation effects of being exposed to the expan-

[13] The downside of using log wage is that those with zero wage are not included in the regression. I follow Mullahy and Norton (2024) and estimate a Poisson regression but the results are qualitatively similar.

sion. Agricultural labor decreases by 9.2 percentage points and is robust, while manufacturing employment increases by 5.3 percentage points but is not robust. The event study further indicates that the manufacturing effect is driven by pretrends. Columns (6) and (7) suggest that the expansion decreases employment in the basic service sector, such as food, hospitality, entertainment, and domestic service, by 3.2 percentage points, while increasing employment in the skill-intensive service sectors, such as health, education, finance, and insurance. Both of these results are robust and pass the falsification test. These results indicate that the expansion reallocates workers from agriculture and less skill-intensive services towards more skill-intensive services.

The surprising result is the lack of effect on monthly wage. As skill-intensive industries tend to have higher pay, it is expected that exposure to the expansion would raise wages via the labor reallocation effect. Part of the explanation is likely the general equilibrium effect: new universities affect wages across different education levels and age cohorts. Thus, the wage gap between college-educated and non-college workers might have already been dampened by the general equilibrium effects. The null result may also hide important heterogeneity across different local labor markets. Local economies with larger skill-intensive sectors such as health care, education, and finance, may have a larger demand for college-educated workers so the wage effect from the reallocation may be larger in such markets.

Figure 7 reports the estimated distributional effects on wages. Exposure to the expansion has very unequal effects across the wage distribution. Effects are much larger towards the bottom quantiles. Part of this is likely due to talent misallocation: For the same level of unobserved ability, individuals at the bottom distribution benefit the most from expanded access to higher education as they do not have the resources to travel to obtain a college degree. Another potential explanation is that non-college workers (who are more representative at the bottom of the distribution) also benefit from the expansion due to the general equilibrium effects. I return to this possibility in the next section.

Another potential explanation for the lack of wage effect may be migration. In Figure 4, I present the event study estimation for the effects on in-migration and out-migration of the expansion. The expansion has a positive effect on in-migration, although the estimates are small and noisy. To check if my results are driven by migration, I estimate the same model to a non-migrant sample and report the results in Table A2. The results are very similar to the main results, suggesting that our findings are not driven by migrant workers.

6.2 Firm-Level Results

I report the results from estimating 2 on the firm-level census data in Table 3. I use the same specifications as reported in the individual-level results. Columns (1) to (4) report the effects on the probability of firms being in the listed sector. First, firms are more likely to produce in the skill-intensive service sector as a result of the expansion; this is consistent with the individual-level findings. There are slight reductions in manufacturing and basic service firms, although these are not statistically significant. Figure 8 indicates that the share of firms operating in the skill-intensive service industries slowly rises during the expansion, while the other sectors do not see any noticeable changes.

The event study results raise some concerns about pretrends, especially for manufacturing and basic services. Thus, I estimate a detrended event study specification, in which I residualize the outcome variables from pre-treatment linear trend interacted with treatment status before estimating the event study. The results are plotted in Figure A4. Controlling for pre-trends shows that firms also move into the manufacturing sector. They also show that firms mostly move out of the basic service sector.

The results in column (5) indicate that the effect of the expansion on TFP is positive. This result is robust to the alternative specification and is null under the falsification test. Figures 9(a) and A5(a) further suggest that these effects are not driven by pretrends. This suggests that the expansion increases firm-level productivity in the long run. One explanation is changes in sectoral composition: firms entering manufacturing and skill-intensive sectors, which tend to be more productive. Or, it can also be driven by technological change that is skill-biased (Acemoglu, 1998). While I cannot rule out the technological change channel, I examine the compositional channel by estimating the effects on TFP in each sector separately. This amounts to controlling for sector-fixed effects.

In Table 4, I report the TFP effects by sector. The agriculture sector experiences little effect on productivity, but partly because there are relatively fewer agricultural firms. In contrast, manufacturing firms experience a large increase in TFP as a result of the expansion, and this result is robust to using the almost-treated group. Figures 9(a) and A5(c) suggest that, however, this is a result of pretrends: Manufacturing firms already become more productive over time even before the expansion.

These sectoral TFP results suggest that the technological change story is unlikely the explanation for the positive effect of the expansion on TFP. Further suggestive evidence from the negative effect on capital-to-output ratio. Generally, technological change is associated with higher capital intensity, as firms use more physical capital relative to other input (Lewis, 2013). The negative effect on capital intensity is more consistent with a composition shift toward the service sector, which tends to use less physical capital.

7 General Equilibrium Effects on Local Labor Markets

The individual-level and firm-level results indicate that the expansion induces both workers and firms into the skill-intensive service sector. The theoretical model suggests that these reallocations will create two opposite forces, so the net effects are theoretically ambiguous. To assess these GE implications, I aggregate the labor survey data at the district level for two age cohorts, those under 30 years old and those above 30 years old.

7.1 Labor Reallocation and College Wage Premiums

I estimate Equation 2 separately for each cohort and report the results in Panels A and B in Table 5. Panel C reports the difference between the two cohorts. I also estimate a similar robustness check as above using the almost-treated group as the control group and report the results in Table A3.[14] The results in columns (1) and (2) capture the main results for the theoretical model. The expansion is found to increase the supply of college-educated workers substantially among the young cohort. The effect on the older cohort is 5 times smaller and not statistically significant.

The expansion decreases the college premiums among both cohorts, but the negative effect is 8.9 percent larger for the old cohort than for the young cohort. The negative effects are driven by the imperfect substitution between college and non-college workers. The significant difference in the wage effects on the two cohorts can be explained by the reallocation effects in Columns (3) to (6). For the older cohort, there is very little labor reallocation except for a positive effect on manufacturing employment. For the younger cohort who experience a large increase in the relative supply of college-educated workers, the effects on manufacturing and skill-intensive service employment are also positive and significant. However, the effects on manufacturing employment are mostly not robust when using the almost-treated group (see Column (4) in Table A3). These results indicate that the young cohort experiences a substantial labor reallocation effect toward the skill-intensive service sector, but the old cohort does not. This reallocation effect offsets the imperfect substitution effect, as predicted by the theoretical model.

[14]See also the event study estimations in Figures 10 and A3.

7.2 Wages and Employment

In Table 6, I report the employment and wage effects by college education status. For the old cohort, non-college workers experience an increase in employment while college-educated workers experience a decrease in employment. The overall employment effect is nonetheless positive. In contrast, the employment effects on the young cohort are positive and significant across college and non-college workers. Therefore, the overall employment effect for the young cohort is also positive. These results indicate that the expansion has a positive effect on the overall employment rate.

The wage effects are less clear. Overall, both cohorts experience an increase in hourly wages because of the expansion. For the old cohort, this is driven by a positive effect on non-college wage offset by a negative effect on college wages. For the younger cohort, non-college wages also increase, but the effect on college wages is negative but small and statistically insignificant. This is consistent with the reallocation effect driving up demands for college-educated workers.

7.3 Elasticity of Substitution

The elasticity of substitution between college-educated and non-college workers is an important parameter in the labor economics literature. For example, it relates to technological change when technology is skill-biased (Goldin and Katz, 2009; Acemoglu, 2011). In the canonical model of skill differentials with one sector, this parameter can be estimated using the effect of the relative supply of college-educated workers on the college wage premium, namely, $\frac{\delta^{supply}}{\delta^{premium}} = -\frac{1}{\rho}$.[15]

However, in a multisector model with the presence of labor reallocation (as derived in

[15]See, e.g., Katz and Murphy (1992); Goldin and Katz (2009) and Acemoglu (2011).

Appendix A), this effect reflects a different parameter:

$$\tilde{\rho} = \rho E(h_S, e_S, \epsilon, \frac{dA_G}{A_G} - \frac{dA_S}{A_S})$$

where $E(h_S, e_S, \epsilon, \frac{dA_G}{A_G} - \frac{dA_S}{A_S})$ is the general equilibrium (GE) adjustment from the labor reallocation. This GE adjustment is a function of the skill intensity of the service sector, the service share of aggregate expenditure, utility elasticity of substitution, and sectoral productivity change.

I do not attempt to estimate E, but the results imply that the true elasticity is smaller than the conventional measure. In particular, $\tilde{\rho}$ can be estimated using the supply and premium effects on the young cohort in a Wald estimator: $\tilde{\rho} = \frac{\delta_y^{supply}}{\delta_y^{premium}}$, which is roughly 1.96. My estimate is higher than the usual benchmark of 1.4 in Katz and Murphy (1992) and closer to 1.6 to 1.7 in more recent studies (Acemoglu, 2011; Autor et al., 2020). To see why this measure is likely inflated, I assume that the college wage premium effect on the old cohort is driven entirely by imperfect substitution, so the implied elasticity of substitution is 1.36. Thus, the GE effect of the expansion inflates the true elasticity of substitution by 44%.

7.4 Returns to Higher Education

A conventional approach is to take the ratio of the wage effect to the college completion effect to obtain the returns to higher education among those who completed college because of the expansion: $r = \frac{\hat{W}_{DiD}}{\hat{C}_{DiD}}$, where $\hat{W}_{DiD}$ and $\hat{C}_{DiD}$ reflect the DiD estimates for the treatment effects of being exposed to the expansion on wages and college completion, respectively. This is equivalent to using the exposure to the expansion as an instrument for completing college using a Wald estimator, assuming that exposure only affects wages through college completion.

But this assumption is likely incorrect, as pointed out by Khanna (2023). The expansion also affects the wages of the older cohorts and non-college workers, as I discussed above. To see how these forces drive the wage effect observed in the age-cohort DiD estimator, rewrite the wage effect as:

$$W_{DiD} = (\log w_{y,T=1} - \log w_{y,T=0}) - (\log w_{o,T=1} - \log w_{o,T=0})$$

$$= f_{y,T=1}\delta^w_{y,C} + (1 - f_{y,T=1}).\delta^w_{y,N} + \delta^C_y.r - (f_{o,T=1}\delta^w_{o,C} + (1 - f_{o,T=1}).\delta^w_{o,N})$$

where $f_{y,T=1}$ and $f_{o,T=1}$ are the shares of college-educated workers among the young and old cohorts, respectively; $\delta^w_{y,C}$, $\delta^w_{y,N}$, $\delta^w_{o,C}$, $\delta^w_{o,N}$ are the wage effects on college and non-college workers of the young and old cohorts. The δ^C_y term measures the share of compliers due to the expansion, namely, those who complete college because of the expansion, and thus is captured by $\hat{C}_{DiD}$.

The correct returns to higher education can be recovered by subtracting the GE effect from the DiD wage estimate: $r = \frac{W_{DiD} - GE}{\hat{C}_{DiD}}$ where

$$GE = [f_{y,T=0}\delta^w_{y,C} + (1 - f_{y,T=0}).\delta^w_{y,N}] - [f_{o,T=1}\delta^w_{o,C} + (1 - f_{o,T=1}).\delta^w_{o,N}]$$

is the GE effect on wages. The first term in the bracket is the weighted average wage effect on the young cohort, and the second term is that on the old cohort. The pre-treatment shares of college-educated workers for the young and old cohorts are 8.62% and 7.85%, respectively. The wage effects can be obtained from Table 6. The long-term GE effect is roughly 5.1%.

The conventional age-cohort DiD estimate for W_{DiD} and C_{DiD} are roughly 6.8% and 4.1% (see Table A5). The conventional Wald estimator yields an inflated return of 165%. Correcting for the GE adjustment thus yields a rate of return of 41%. In other words, the general equilibrium effect drives up the estimated returns by 4 times.

8 Conclusion

The national expansion of higher education in Vietnam represents a substantial increase in access to higher education. On the labor supply side, it lowers the cost of obtaining a college degree, allowing workers to move into skill-intensive service industries. On the labor demand side, it allows firms to expand or scale up in the skill-intensive service industries.

The reallocation of labor and firm activities into the skill-intensive service sector has three important implications. First, it raises the average total factor productivity at the firm level. Second, it raises the employment of both college and non-college workers as the skill-intensive service sector expands. Third, it has important implications for the returns to higher education and the elasticity of substitution between the two types of workers.

The main focus of this paper is the increased supply of college-educated workers. It does not address other important dimensions of higher education. For example, as the scale of higher education supply increases, the quality of these universities is also of concern. While the education workforce is also scaled up as more staff and professors are trained, it is not clear whether the overall quality has changed. Furthermore, as higher education becomes more accessible, the underlying skill distribution also matters more. This paper also does not account for differences in college majors, which is an important driver of labor market outcomes. Future studies should examine how these factors interact with the expansion.

References

Acemoglu, D. (1998). Why do new technologies complement skills? directed technical change and wage inequality. The quarterly journal of economics 113(4), 1055–1089.

Acemoglu, D. (2007). Equilibrium bias of technology. Econometrica 75(5), 1371–1409.

Acemoglu, D. (2011). Skills, tasks and technologies: Implications for employment and earnings.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. Econometrica 83(6), 2411–2451.

Andrews, M. J. (2023). How do institutions of higher education affect local invention? evidence from the establishment of us colleges. American Economic Journal: Economic Policy 15(2), 1–41.

Athey, S. and G. W. Imbens (2006). Identification and inference in nonlinear difference-in-differences models. Econometrica 74(2), 431–497.

Autor, D., C. Goldin, and L. F. Katz (2020). Extending the race between education and technology. In AEA papers and proceedings, Volume 110, pp. 347–351. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Baldwin, R., R. Freeman, and A. Theodorakopoulos (2024). Deconstructing deglobalization: the future of trade is in intermediate services. Asian Economic Policy Review 19(1), 18–37.

Bianchi, N. (2020). The indirect effects of educational expansions: Evidence from a large enrollment increase in university majors. Journal of Labor Economics 38(3), 767–804.

Borusyak, K., X. Jaravel, and J. Spiess (2024). Revisiting event-study designs: robust and efficient estimation. Review of Economic Studies, rdae007.

Bosker, M., S. Brakman, H. Garretsen, and M. Schramm (2012). Relaxing hukou: Increased labor mobility and china's economic geography. Journal of Urban Economics 72(2-3), 252–266.

Buera, F. J. and J. P. Kaboski (2012). The rise of the service economy. American Economic Review 102(6), 2540–2569.

Buera, F. J., J. P. Kaboski, R. Rogerson, and J. I. Vizcaino (2022). Skill-biased structural change. The Review of Economic Studies 89(2), 592–625.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. Journal of econometrics 225(2), 200–230.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). The effect of minimum wages on low-wage jobs. The Quarterly Journal of Economics 134(3), 1405–1454.

Chen, X., G. Pei, Z. Song, and F. Zilibotti (2023). Tertiarization like china. Annual Review of Economics 15.

Duflo, E. (2004). The medium run effects of educational expansion: Evidence from a large school construction program in indonesia. Journal of Development Economics 74(1), 163–197.

Fan, T., M. Peters, and F. Zilibotti (2023). Growing like india—the unequal effects of service-led growth. Econometrica 91(4), 1457–1494.

Fujiwara, I. and K. Matsuyama (2020). A technology-gap model of premature deindustrialization.

Goldin, C. and L. F. Katz (2009). The race between education and technology. harvard university press.

Goodman-Bacon, A. (2021a). Difference-in-differences with variation in treatment timing. Journal of econometrics 225(2), 254–277.

Goodman-Bacon, A. (2021b). The long-run effects of childhood insurance coverage: Medicaid implementation, adult health, and labor market outcomes. American Economic Review 111(8), 2550–2593.

Hanushek, E. A. (2016). Will more higher education improve economic growth? Oxford Review of Economic Policy 32(4), 538–552.

Hausman, N. (2022). University innovation and local economic growth. Review of Economics and Statistics 104(4), 718–735.

Herrendorf, B. and T. Schoellman (2018). Wages, human capital, and barriers to structural transformation. American Economic Journal: Macroeconomics 10(2), 1–23.

Huneeus, F. and R. Rogerson (2024). Heterogeneous paths of industrialization. Review of Economic Studies 91(3), 1746–1774.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963–1987: supply and demand factors. The quarterly journal of economics 107(1), 35–78.

Khanna, G. (2023). Large-scale education reform in general equilibrium: Regression discontinuity evidence from india. Journal of Political Economy 131(2), 549–591.

Knight, J., Q. Deng, and S. Li (2017). China's expansion of higher education: The labour market consequences of a supply shock. China Economic Review 43, 127–141.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. The review of economic studies 70(2), 317–341.

Lewis, E. (2013). Immigration and production technology. Annu. Rev. Econ. 5(1), 165–191.

Li, H., Y. Ma, L. Meng, X. Qiao, and X. Shi (2017). Skill complementarities and returns to higher education: Evidence from college enrollment expansion in china. China Economic Review 46, 10–26.

McCaig, B. and N. Pavcnik (2013). Moving out of agriculture: structural change in vietnam. Technical report, National Bureau of Economic Research.

McCaig, B. and N. Pavcnik (2018). Export markets and labor allocation in a low-income country. American Economic Review 108(7), 1899–1941.

Mullahy, J. and E. C. Norton (2024). Why transform y? the pitfalls of transformed regressions with a mass at zero. Oxford Bulletin of Economics and Statistics 86(2), 417–447.

Munshi, K. (2019). Caste and the indian economy. Journal of Economic Literature 57(4), 781–834.

Nayyar, G., M. Hallward-Driemeier, and E. Davies (2021). At your service?: The promise of services-led development. World Bank Publications.

Newman, C., J. Rand, T. Talbot, and F. Tarp (2015). Technology transfers, foreign investment and productivity spillovers. European Economic Review 76, 168–187.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. Econometrica 64(6), 1263–1297.

Porzio, T., F. Rossi, and G. Santangelo (2022). The human side of structural transformation. American Economic Review 112(8), 2774–2814.

Rambachan, A. and J. Roth (2023). A more credible approach to parallel trends. Review of Economic Studies 90(5), 2555–2591.

Rodrik, D. (2016). Premature deindustrialization. Journal of economic growth 21, 1–33.

Roth, J. and P. H. Sant'Anna (2023). When is parallel trends sensitive to functional form? Econometrica 91(2), 737–747.

Sen, K. (2023). Varieties of structural transformation: Patterns, determinants, and consequences. Cambridge University Press.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of econometrics 225(2), 175–199.

Valero, A. and J. Van Reenen (2019). The economic impact of universities: Evidence from across the globe. Economics of Education Review 68, 53–67.

Tables

Table 1: Differences in Provincial Characteristics between Treatment and Control Groups

	Treatment (N = 22)	Never Treated (N = 14)		Almost Treated (N = 6)	
	Mean	Mean	Difference	Mean	Difference
% college graduates	0.039	0.034	0.005	0.024	0.015
% high school graduates	0.249	0.247	0.002	0.227	0.022
% college enrolment	0.082	0.051	0.032	0.062	0.020
% self employed	0.860	0.888	-0.028	0.879	-0.019
% employed	0.140	0.112	0.028	0.121	0.019
% agricultural worker	0.594	0.749	-0.155	0.647	-0.054
% manufacturing worker	0.165	0.076	0.089	0.159	0.006
% service worker	0.216	0.145	0.072	0.173	0.043
Log income per capita	8.763	8.550	0.214	8.693	0.071
Urban	0.217	0.150	0.067	0.140	0.077
% in poverty	0.134	0.178	-0.044	0.128	0.006
% age 0-5	0.214	0.236	-0.022	0.214	0.000
% age 6-18	0.179	0.175	0.004	0.172	0.007
Log TFP	1.020	0.830	0.190	0.930	0.090
Log Value Added per Worker	16.963	16.913	0.050	16.905	0.058
Log Capital-Labor Ratio	19.171	19.065	0.105	18.863	0.308
Total Provinces	36				

This table shows the means of pre-treatment, province-level characteristics by treatment status and the results from their balance tests. The *Mean* columns display the mean values of these covariates for each treatment group. The *Difference* columns show the results from estimating a regression with the characteristics as the dependent variable and the treatment status dummy variables as the independent variables. Data on the pre-treatment characteristics are aggregated from the 2004-2006 VHLSS data.

Table 2: Effects of Higher Education Expansion on Individual Labor Market Outcomes

	Complete College	Hourly Wage	Employed	Sectoral Employment: Agri-culture	Manu-facturing	Basic Service	Skill-Intensive Service
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Main Specification (Never-Treated as Control)							
Age exposed:							
19 to 25	0.019***	0.034*	0.051***	-0.019	0.037**	-0.023***	-0.005
	(0.007)	(0.019)	(0.016)	(0.016)	(0.015)	(0.006)	(0.007)
14 to 18	0.041***	0.068**	0.134***	-0.092***	0.053**	-0.032***	0.025***
	(0.010)	(0.029)	(0.034)	(0.030)	(0.026)	(0.011)	(0.006)
N	4,200,772	2,646,254	3,919,516	3,921,836	3,921,836	4,200,772	4,197,347
Panel B: Alternative Specification (Almost-Treated as Control)							
Age exposed:							
19 to 25	0.020**	0.049**	0.042*	-0.022	0.029	-0.018**	0.001
	(0.008)	(0.019)	(0.022)	(0.020)	(0.019)	(0.008)	(0.010)
14 to 18	0.049***	0.045	0.116**	-0.086*	0.041	-0.026*	0.029***
	(0.013)	(0.038)	(0.051)	(0.046)	(0.034)	(0.013)	(0.009)
N	1,854,802	1,223,391	1,713,156	1,714,029	1,714,029	1,854,802	1,852,397
Panel C: Falsification Tests (Almost-Treated as Treatment)							
Age exposed:							
19 to 25	-0.000	-0.013	0.018	-0.002	0.015	-0.009	-0.006
	(0.009)	(0.021)	(0.027)	(0.022)	(0.022)	(0.008)	(0.010)
14 to 18	-0.011	0.044	0.037	-0.020	0.024	-0.009	-0.002
	(0.016)	(0.045)	(0.059)	(0.052)	(0.038)	(0.015)	(0.011)
N	3,516,136	2,187,297	3,305,228	3,307,170	3,307,170	3,516,136	3,513,612

The table reports DiD estimates for impacts on individual outcomes using stacked regression, where treated provinces are compared to never-treated provinces across birth cohorts. I control for age, age squared, gender, province, birth, and cohort fixed effects. All standard errors are clustered at the province level. Data is drawn from LFS 2010-2018, but the sample is limited to 5 years after treatment or more.

Table 3: Effects of Higher Education Expansion on Firm-Level Outcomes

	Sector						
	Ag	Manu-facturing	Basic Service	Skill-Intensive Service	TFP	Labor Productiv-ity	Capital Intensity
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Main Specification (Never-Treated as Control)							
Years since university opening							
5 to 9	0.006	-0.011	-0.008	0.013***	0.266	0.037	-0.089**
	(0.008)	(0.021)	(0.016)	(0.004)	(0.179)	(0.038)	(0.043)
10 to 13	0.008	-0.022	-0.019	0.033***	0.692**	0.093	-0.254***
	(0.012)	(0.029)	(0.022)	(0.009)	(0.311)	(0.091)	(0.072)
N	586,076	586,076	586,076	586,076	586,076	586,076	586,076
Panel B: Alternative Specification (Almost-Treated as Control)							
Years since university opening							
5 to 9	0.011*	0.001	-0.022	0.010***	0.280	0.039	-0.127*
	(0.005)	(0.026)	(0.022)	(0.003)	(0.186)	(0.074)	(0.072)
10 to 13	0.009*	0.007	-0.043	0.027**	0.586***	0.086	-0.338***
	(0.005)	(0.028)	(0.028)	(0.011)	(0.154)	(0.193)	(0.103)
N	467,636	467,636	467,636	467,636	467,636	467,636	467,636
Panel C: Falsification Tests (Almost-Treated as Treatment)							
Years since university opening							
5 to 9	-0.008	-0.002	0.008	0.002	-0.083	-0.013	0.083
	(0.012)	(0.032)	(0.026)	(0.006)	(0.311)	(0.081)	(0.072)
10 to 13	-0.003	-0.022	0.020	0.005	0.061	0.036	0.144
	(0.021)	(0.046)	(0.040)	(0.014)	(0.529)	(0.205)	(0.104)
N	213,395	213,395	213,395	213,395	213,395	213,395	213,395

The table reports DiD estimates for impacts on firm-level outcomes using stacked regression. In the main specification in Panel A, the DiD compares provinces with the first university ever to provinces that never had a university during the study period. In the alternative specification in Panel B, the control group is provinces with universities being announced but not open because the expansion ended early, namely the almost-treated group. Panel C provides the falsification test results in which the almost-treated group is defined as a placebo treatment group while the rest of the never-treated group is the control. Standard error is clustered at the province level. Columns (1)-(4) report the effects on the industry of firms. Column (5) reports the effect on total factor productivity. Column (6) reports the result for labor productivity, measured as log value added per worker. Column (7) reports the results for capital-to-output ratio.

Table 4: Effects of Higher Education Expansion on Firm-Level TFP - By Sector

	TFP By Sector				
	Ag	Manu- facturing	Basic Service	Skill- Intensive Service	Overall TFP
	(1)	(2)	(3)	(4)	(5)
Panel A: Main Specification (Never-Treated as Control)					
Years since university opening					
5 to 9	0.022	0.186**	-0.093	-0.018	0.266
	(0.177)	(0.070)	(0.157)	(0.288)	(0.179)
10 to 13	0.065	0.349***	0.104	-0.224	0.692**
	(0.246)	(0.078)	(0.307)	(0.457)	(0.311)
N	20,860	253,377	278,033	33,806	586,076
Panel B: Alternative Specification (Almost-Treated as Control)					
Years since university opening					
5 to 9	0.042	0.270***	-0.204	0.300	0.280
	(0.326)	(0.053)	(0.277)	(0.299)	(0.186)
10 to 13	0.076	0.423***	-0.051	0.417	0.586***
	(0.428)	(0.060)	(0.457)	(0.550)	(0.154)
N	15,100	208,237	213,933	30,362	467,636
Panel C: Falsification Tests (Almost-Treated as Treatment)					
Years since university opening					
5 to 9	-0.003	-0.157**	0.117	-0.503	-0.083
	(0.342)	(0.066)	(0.331)	(0.376)	(0.311)
10 to 13	0.045	-0.121	0.249	-1.174	0.061
	(0.421)	(0.078)	(0.612)	(0.775)	(0.529)
N	11,435	85,953	109,665	6,342	213,395

The table reports DiD estimates for impacts on firm-level outcomes using stacked regression. In the main specification in Panel A, the DiD compares provinces with the first university ever to provinces that never had a university during the study period. In the alternative specification in Panel B, the control group is provinces with universities being announced but not open because the expansion ended early, namely the almost-treated group. Panel C provides the falsification test results in which the almost-treated group is defined as a placebo treatment group while the rest of the never-treated group is the control. Standard error is clustered at the province level.

Table 5: General Equilibrium Effects of the Expansion on Local Labor Markets

	Relative Supply	College Premium	Sectoral Employment: Agriculture	Sectoral Employment: Manu-facturing	Sectoral Employment: Basic Service	Sectoral Employment: Skill-Intensive Service
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Old Cohort						
Year since university opening						
5 to 9	0.153	-0.204**	-0.032	0.010	0.016	-0.001
	(0.106)	(0.096)	(0.020)	(0.006)	(0.013)	(0.005)
10 to 13	0.181	-0.373***	-0.049*	0.033**	0.016	-0.002
	(0.109)	(0.125)	(0.025)	(0.013)	(0.018)	(0.007)
N	3,996	3,996	3,996	3,996	3,996	3,996
Panel B: Young Cohort						
Year since university opening						
5 to 9	0.317***	-0.159**	-0.037*	0.031*	-0.007	0.009*
	(0.103)	(0.063)	(0.019)	(0.017)	(0.014)	(0.005)
10 to 13	0.506***	-0.258**	-0.046	0.056**	-0.031	0.016**
	(0.147)	(0.097)	(0.027)	(0.025)	(0.026)	(0.007)
N	2,997	2,988	2,997	2,997	2,997	2,997
Panel C: Young - Old						
Year since university opening						
5 to 9	0.153*	0.029	-0.005	0.028*	-0.033***	0.012***
	(0.083)	(0.041)	(0.020)	(0.016)	(0.007)	(0.004)
10 to 13	0.389***	0.089*	-0.003	0.043	-0.063***	0.016**
	(0.123)	(0.046)	(0.027)	(0.028)	(0.013)	(0.007)
N	6,993	6,984	6,993	6,993	6,993	6,993

Note: This table presents the results of the market-level DiD analysis, in which each observation is at the province-by-year-by age cohort level. Relative supply is the ratio of log college-educated workers to log non-college workers. Treatment group is provinces with the first university ever. Control group is provinces that never had a university. College premium is measured as the ratio of log college wage to log non-college wage. Sectoral employment outcomes are the share of workers in a given sector. Basic service includes food, hospitality, entertainment, domestic, and other services that do not typically require a college degree. Skill-intensive service includes sectors that typically require a college degree, including health, education, finance, insurance, and science. Panels A and B are for Old and Young Cohorts, respectively. Panel C test the difference between the two age cohorts.

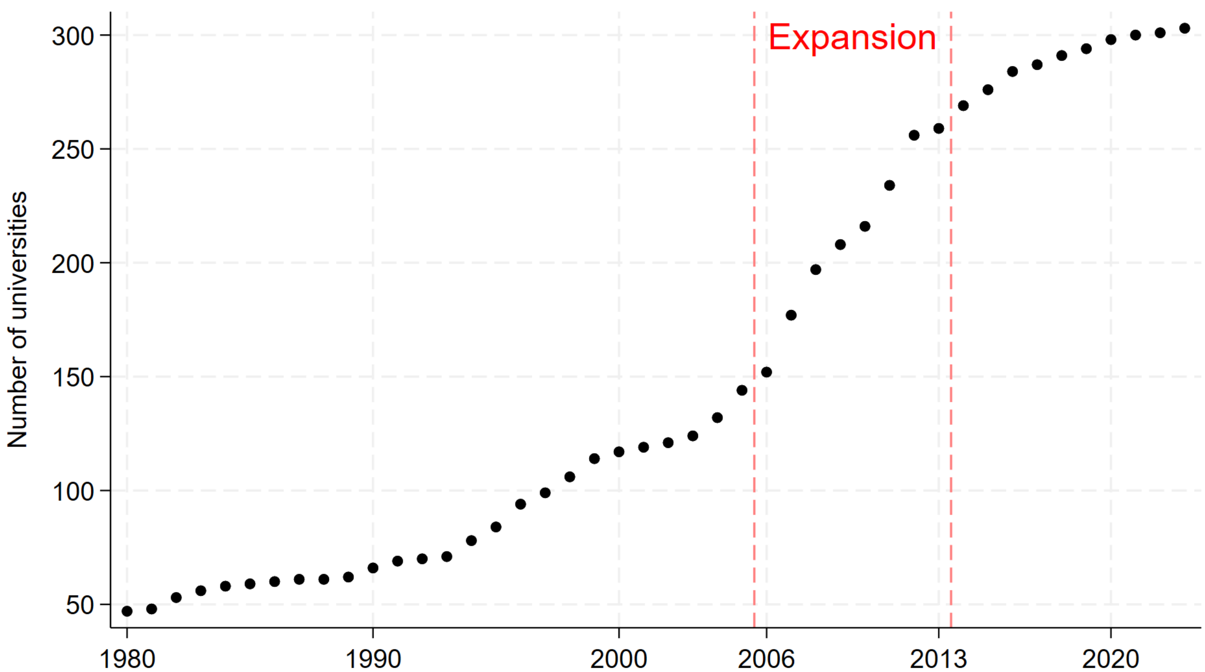
Table 6: General Equilibrium Effects of the Expansion on Local Labor Markets - By Education

	College Completion Rate	Employment: Non-College	Employment: College	Employment: All	Hourly Wage: Non-College	Hourly Wage: College	Hourly Wage: All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Old Cohort							
Year since university opening							
5 to 9	-0.001	0.036***	-0.011	0.030**	0.143	-0.061***	0.096
	(0.006)	(0.011)	(0.008)	(0.012)	(0.095)	(0.020)	(0.080)
10 to 13	-0.008	0.070***	-0.036***	0.052***	0.290**	-0.083***	0.198*
	(0.009)	(0.019)	(0.013)	(0.017)	(0.126)	(0.028)	(0.099)
N	3,996	3,996	3,996	3,996	3,996	3,996	3,996
Panel B: Young Cohort							
Year since university opening							
5 to 9	0.010*	0.052***	0.041***	0.051***	0.128*	-0.031	0.098
	(0.006)	(0.016)	(0.011)	(0.014)	(0.066)	(0.033)	(0.059)
10 to 13	0.014	0.068**	0.057***	0.065***	0.227**	-0.029	0.179**
	(0.009)	(0.026)	(0.017)	(0.023)	(0.097)	(0.039)	(0.083)
N	2,997	2,997	2,997	2,997	2,997	2,988	2,997
Panel C: Young - Old							
Year since university opening							
5 to 9	0.016***	0.023	0.040***	0.027	-0.022	0.007	-0.016
	(0.006)	(0.017)	(0.013)	(0.017)	(0.040)	(0.029)	(0.033)
10 to 13	0.029***	0.015	0.065***	0.025	-0.052	0.038	-0.011
	(0.009)	(0.021)	(0.017)	(0.022)	(0.047)	(0.034)	(0.042)
N	6,993	6,993	6,993	6,993	6,993	6,984	6,993

Note: This table presents the results of the market-level DiD analysis, in which each observation is at the province-by-year-by age cohort-by education level. Relative supply is the ratio of log college-educated workers to log non-college workers. Treatment group is provinces with the first university ever. Control group is provinces that never had a university. Panels A and B are for Old and Young Cohorts, respectively. Panel C test the difference between the two age cohorts.

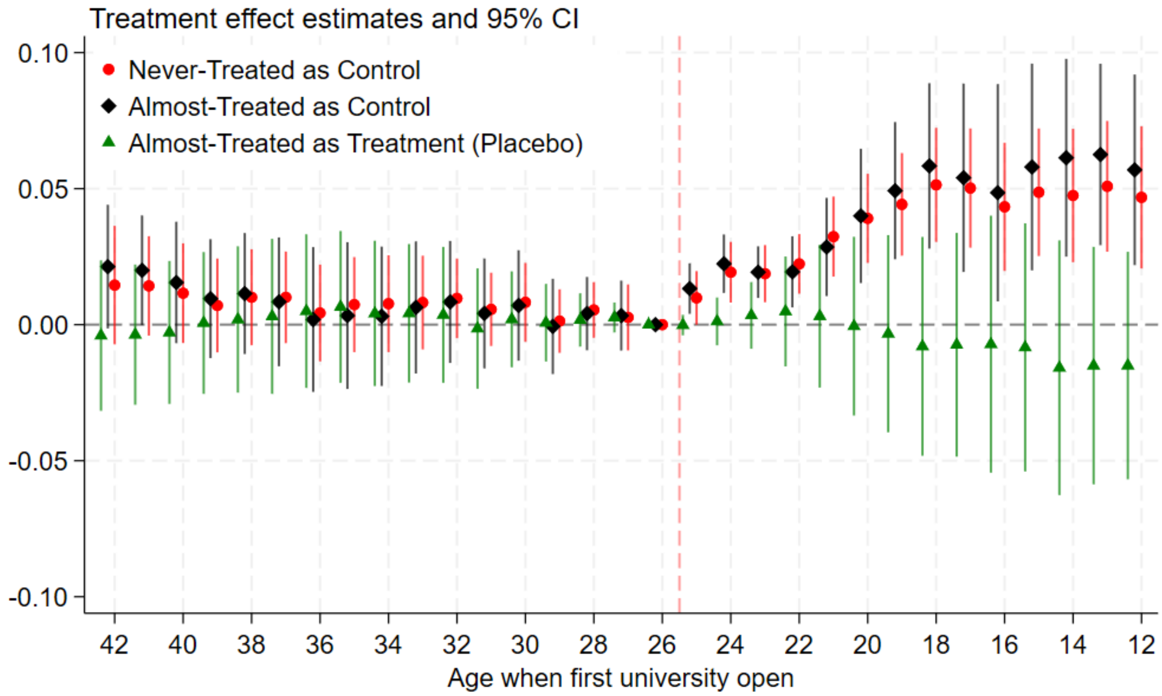
Figures

Figure 2: Number of universities by year



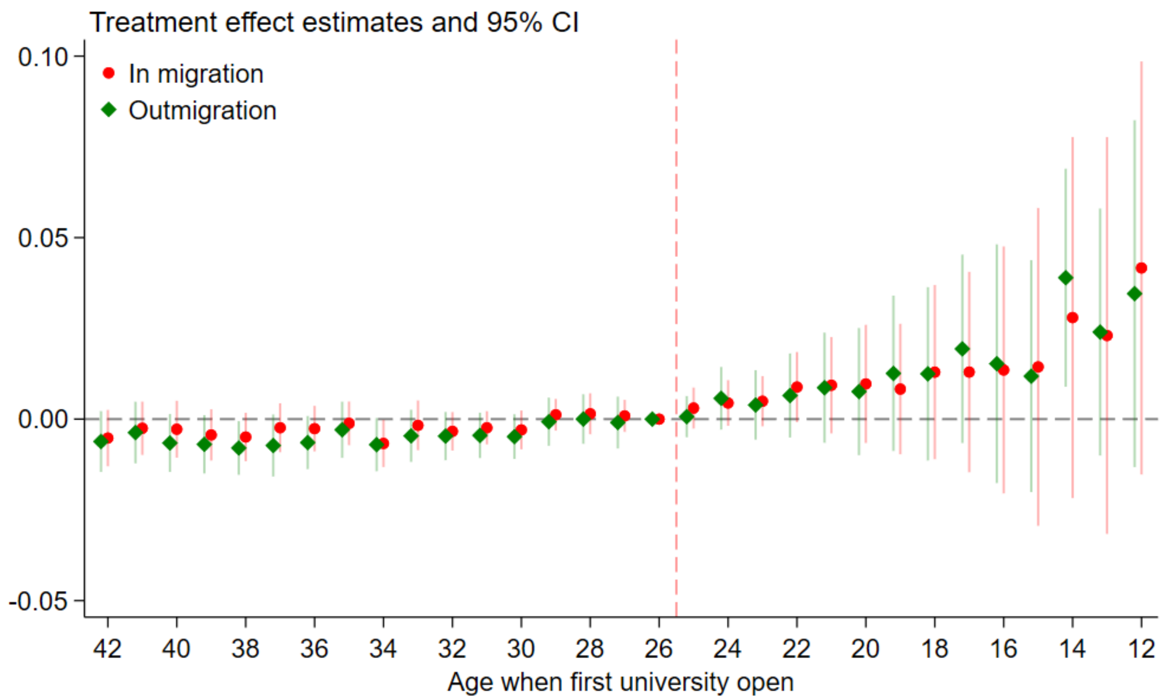
Note: The figure shows the number of universities in Vietnam by year based on data collected from official documents.

Figure 3: Event Study for the Effect on College Completion



Note: The figure shows the event study estimates from different difference-in-differences models for the effect on college completion outcome. Red dots show the estimates from using never-treated provinces as the control group. Black diamonds show the the estimates from using provinces that intended to open a university but missed the window (almost-treated) as the control group. Green triangles show the estimates from using the almost-treated group as the treatment group; treatment year is assigned from the actual treatment group as a placebo test. Standard errors are clustered at the province level.

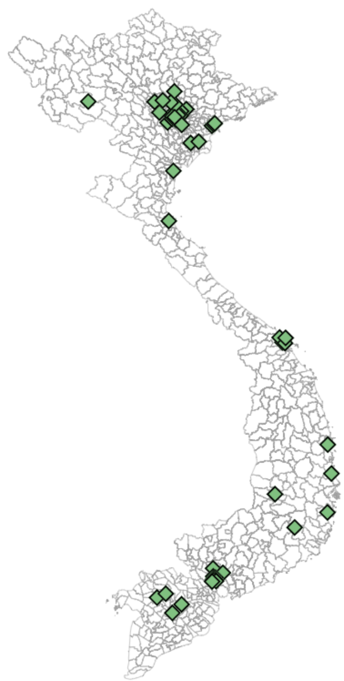
Figure 4: Event Study for the Effect on In-Migration and Out-Migration



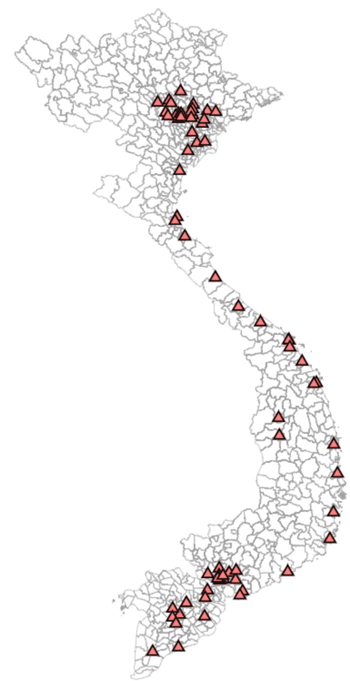
Note: The figure shows the event study estimates from different difference-in-differences models for the effect on college completion outcome. Red dots show the estimates from using never-treated provinces as the control group. Black diamonds show the the estimates from using provinces that intended to open a university but missed the window (almost-treated) as the control group. Green triangles show the estimates from using the almost-treated group as the treatment group; treatment year is assigned from the actual treatment group as a placebo test. Standard errors are clustered at the province level.

Figure 5: Locations of Existing and New Universities

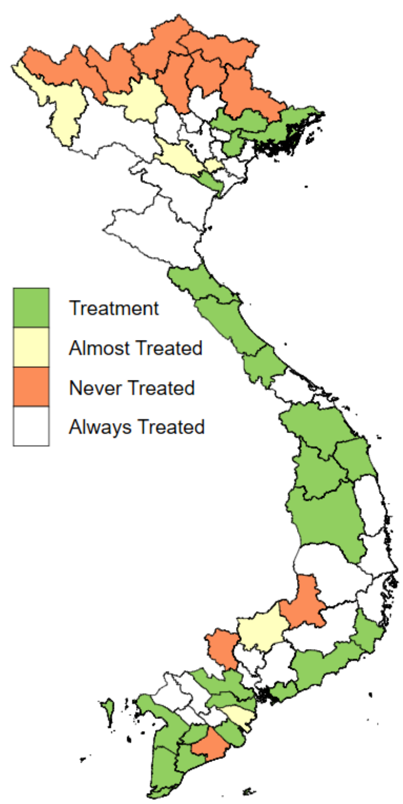
(a) Existing Universities



(b) New Universities



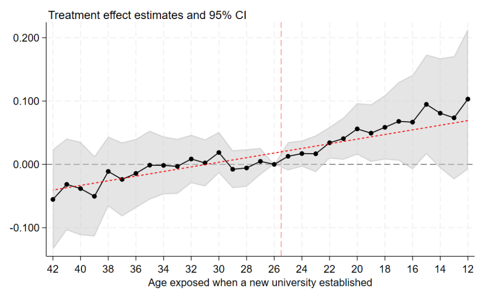
(c) New Universities



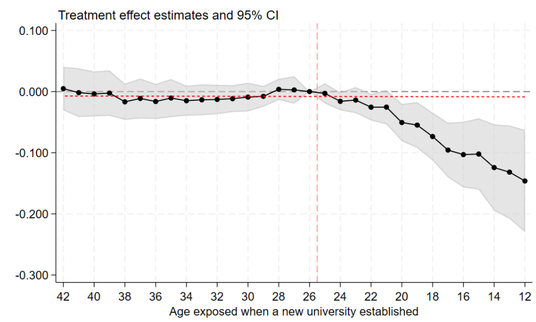
Note: Figure(a) shows the distribution of universities prior to the expansion. Figure (b) shows the distribution of new universities established during the expansion. Figure (c) shows the treatment status of provinces based on whether they had the first university due to the expansion (Treatment), announced new university but missed the expansion (Almost Treated), never had a university during the study period (Never Treated), and already had universities prior to the expansion (Always Treated).

Figure 6: Event Study Estimates for Effects on Labor Market Outcomes

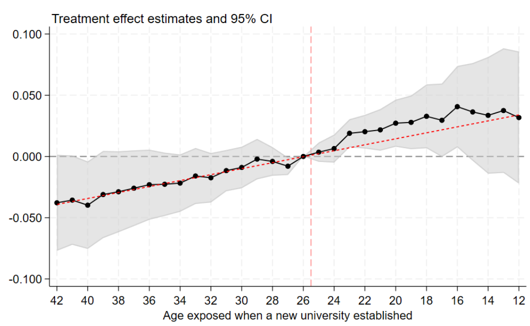
(a) Hourly Wage



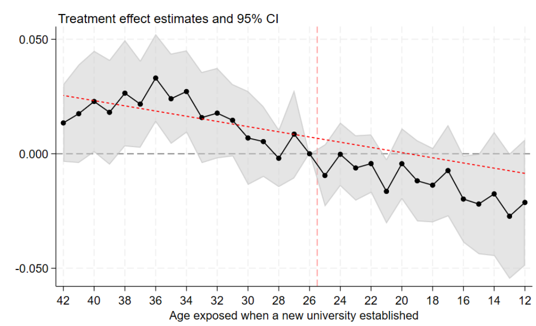
(b) Ag. Sector



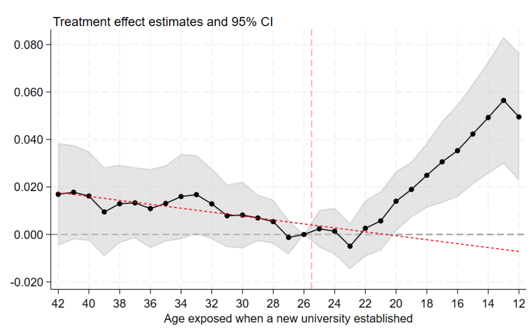
(c) Manu. Sector



(d) Basic Service

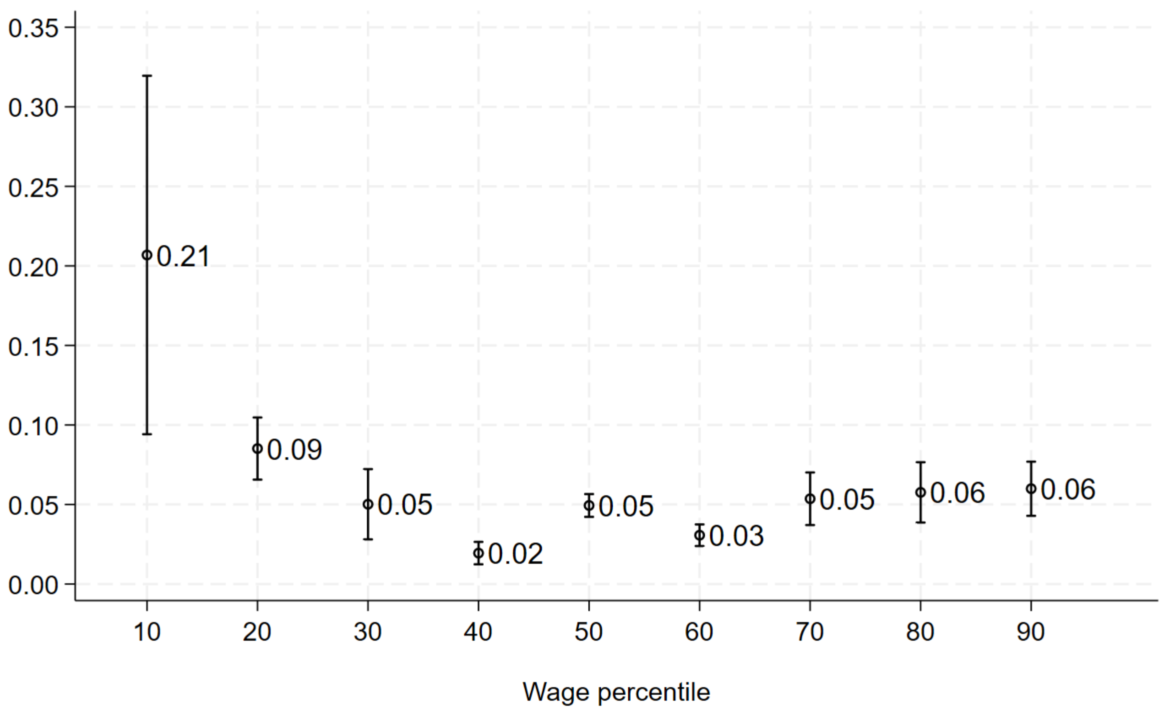


(e) Skill-Intensive Service



Note: The graphs display event study estimation for the effects of the higher education expansion on labor market outcomes. All models control for age, age squared, and gender. Standard errors are clustered at the province-level and 95% confidence intervals are displayed.

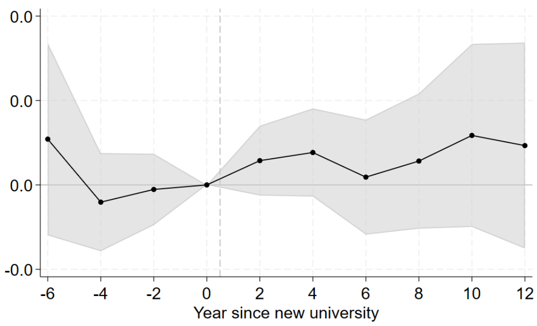
Figure 7: Change-in-Changes Estimated Effects on Monthly Wage



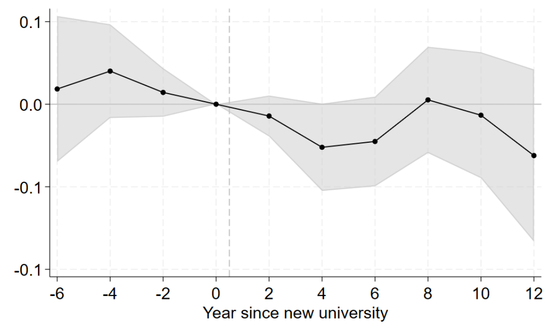
Note: The figure shows the wage effects of the expansion on different wage quantiles estimated using Athey and Imbens (2006) Change-in-Changes approach.

Figure 8: Event Study Estimates for Effects on Firm-Level Outcomes

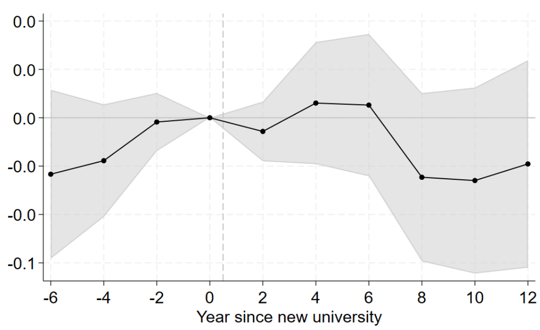
(a) Agriculture



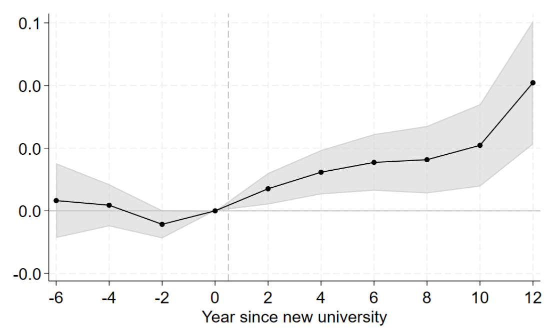
(b) Manufacturing



(c) Basic Service

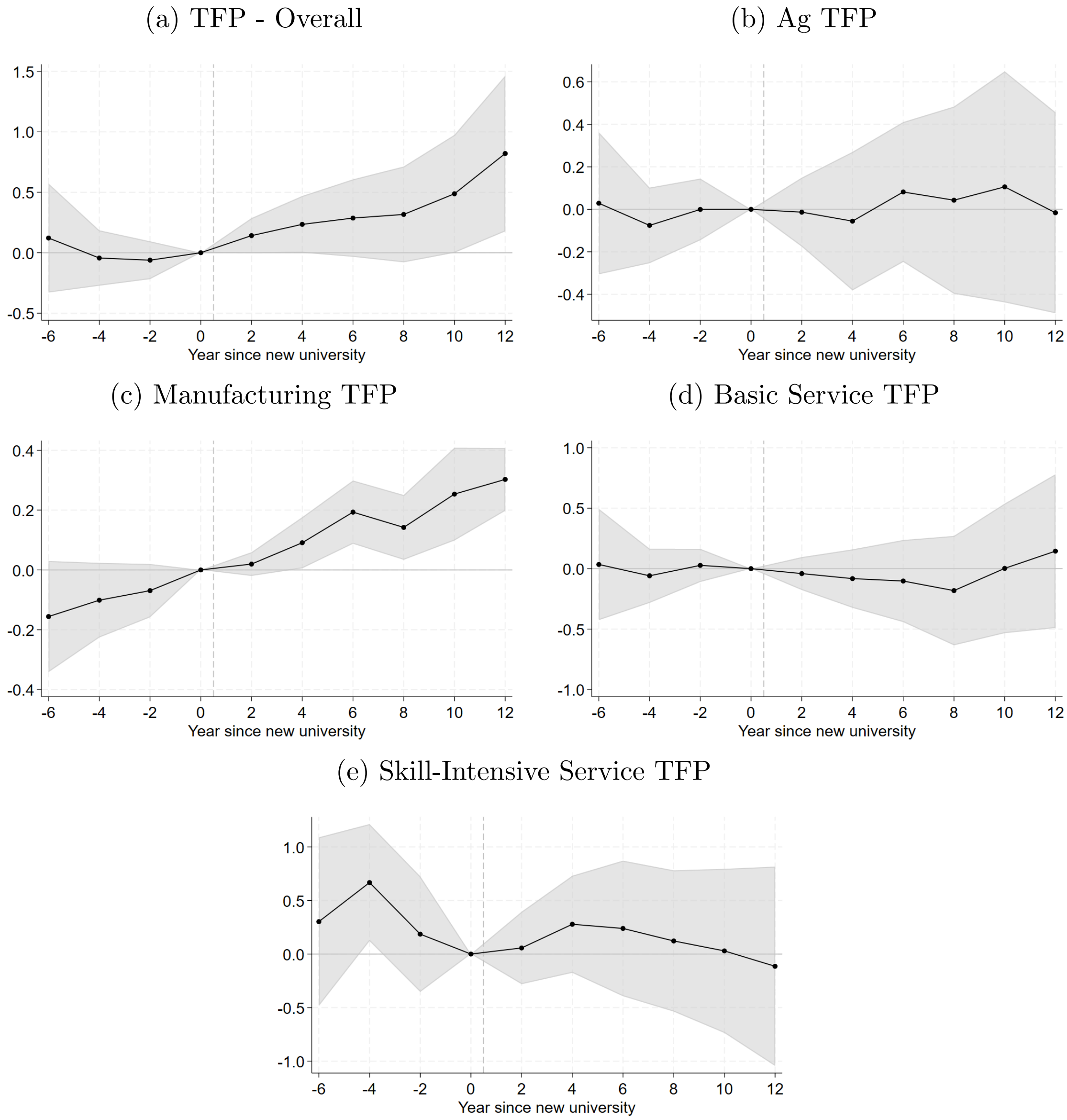


(d) Skill-Intensive Service



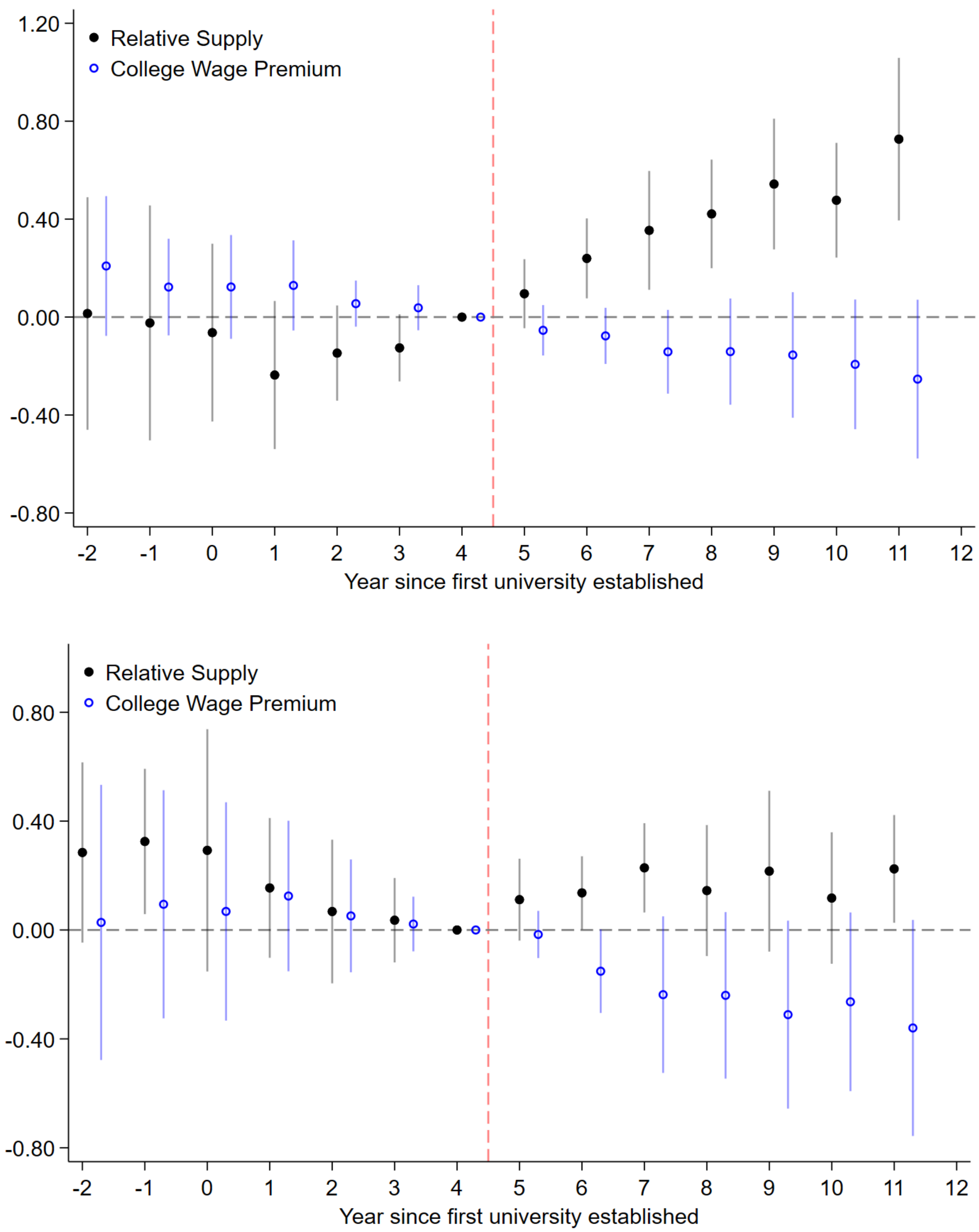
Note: The graphs display event study estimation for the effects of the higher education expansion on firm-level outcomes. Standard errors are clustered at the province-level and 95% confidence intervals are displayed.

Figure 9: Event Study Estimates for Effects on Firm-Level TFP



Note: The graphs display event study estimation for the effects of the higher education expansion on firm-level outcomes. Standard errors are clustered at the province-level and 95% confidence intervals are displayed.

Figure 10: Event Study Estimates for Effects on Market-Level Outcomes - By Age Cohort



Note: The graphs display event study estimation for the effects of the higher education expansion on market-level labor market outcomes.

Appendices

Table A1: Effects of Higher Education Expansion on Individual Labor Market Outcomes - Highschool or Above

				Sectoral Employment			
	Complete College	Hourly Wage	Employed	Agri-culture	Manu-facturing	Basic Service	Skill-Intensive Service
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Main Specification (Never-Treated as Control)							
Age exposed:							
19 to 25	0.031*	0.061***	0.071***	-0.061***	0.043**	0.009	-0.019
	(0.016)	(0.020)	(0.020)	(0.017)	(0.016)	(0.010)	(0.015)
14 to 18	0.065**	0.124***	0.197***	-0.173***	0.060**	-0.006	0.038**
	(0.025)	(0.035)	(0.040)	(0.034)	(0.025)	(0.016)	(0.015)
N	1,403,021	1,088,242	1,315,114	1,315,949	1,315,949	1,403,021	1,399,680
Panel B: Alternative Specification (Almost-Treated as Control)							
Age exposed:							
19 to 25	0.022*	0.062***	0.052***	-0.051**	0.035*	0.016	-0.015
	(0.011)	(0.015)	(0.015)	(0.020)	(0.018)	(0.013)	(0.021)
14 to 18	0.073**	0.088*	0.164***	-0.156***	0.042	0.006	0.037**
	(0.033)	(0.044)	(0.052)	(0.052)	(0.031)	(0.015)	(0.015)
N	636,637	503,009	594,063	594,332	594,332	636,637	634,280
Panel C: Falsification Tests (Almost-Treated as Treatment)							
Age exposed:							
19 to 25	0.011	0.008	0.035	-0.021	0.017	-0.010	-0.004
	(0.021)	(0.027)	(0.026)	(0.029)	(0.019)	(0.012)	(0.019)
14 to 18	-0.018	0.066	0.059	-0.036	0.031	-0.016	0.002
	(0.041)	(0.062)	(0.066)	(0.066)	(0.034)	(0.018)	(0.019)
N	1,182,943	915,123	1,114,876	1,115,591	1,115,591	1,182,943	1,180,480

The table reports DiD estimates for impacts on individual outcomes using stacked regression, where treated provinces are compared to never-treated provinces across birth cohorts. I control for age, age squared, gender, province, birth, and cohort fixed effects. All standard errors are clustered at the province level. Data is drawn from LFS 2010-2018, but the sample is limited to 5 years after treatment or more.

Table A2: Effects of Higher Education Expansion on Individual Labor Market Outcomes - Non-Migrant Sample

				Sectoral Employment			
	Complete College	Hourly Wage	Employed	Agri-culture	Manu-facturing	Basic Service	Skill-Intensive Service
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Main Specification (Never-Treated as Control)							
Age exposed:							
19 to 25	0.019**	0.035	0.055***	-0.023	0.042**	-0.020***	-0.004
	(0.008)	(0.021)	(0.016)	(0.016)	(0.016)	(0.007)	(0.006)
14 to 18	0.043***	0.077**	0.139***	-0.098***	0.053*	-0.018	0.028***
	(0.011)	(0.033)	(0.036)	(0.033)	(0.027)	(0.012)	(0.007)
N	3,048,142	2,254,729	2,854,404	2,855,838	2,855,838	3,048,142	3,045,731
Panel B: Alternative Specification (Almost-Treated as Control)							
Age exposed:							
19 to 25	0.022**	0.047**	0.047**	-0.025	0.032	-0.016	0.002
	(0.008)	(0.021)	(0.022)	(0.022)	(0.020)	(0.010)	(0.010)
14 to 18	0.052***	0.048	0.122**	-0.092*	0.042	-0.016	0.031***
	(0.016)	(0.046)	(0.056)	(0.053)	(0.035)	(0.016)	(0.011)
N	1,347,399	1,025,138	1,253,347	1,253,862	1,253,862	1,347,399	1,345,760
Panel C: Falsification Tests (Almost-Treated as Treatment)							
Age exposed:							
19 to 25	-0.003	-0.008	0.016	-0.003	0.016	-0.007	-0.005
	(0.010)	(0.024)	(0.028)	(0.025)	(0.023)	(0.010)	(0.009)
14 to 18	-0.013	0.051	0.035	-0.020	0.022	-0.004	-0.002
	(0.019)	(0.054)	(0.067)	(0.061)	(0.041)	(0.018)	(0.013)
N	2,574,564	1,876,843	2,425,219	2,426,416	2,426,416	2,574,564	2,572,751

The table reports DiD estimates for impacts on individual outcomes using stacked regression, where treated provinces are compared to never-treated provinces across birth cohorts. I control for age, age squared, gender, province, birth, and cohort fixed effects. All standard errors are clustered at the province level. Data is drawn from LFS 2010-2018, but the sample is limited to 5 years after treatment or more.

Table A3: General Equilibrium Effects of the Expansion on Local Labor Markets - Robustness Check

	Relative Supply	College Premium	Sectoral Employment: Agriculture	Sectoral Employment: Manu-facturing	Sectoral Employment: Basic Service	Sectoral Employment: Skill-Intensive Service
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Old Cohort						
Year since university opening						
5 to 9	0.075	-0.195	-0.020	0.007	0.004	-0.001
	(0.093)	(0.130)	(0.030)	(0.007)	(0.017)	(0.005)
10 to 13	0.092	-0.311	-0.030	0.028*	-0.005	-0.001
	(0.111)	(0.188)	(0.046)	(0.013)	(0.026)	(0.009)
N	1,692	1,692	1,692	1,692	1,692	1,692
Panel B: Young Cohort						
Year since university opening						
5 to 9	0.370***	-0.154*	-0.027	0.020	-0.017	0.012*
	(0.082)	(0.079)	(0.032)	(0.023)	(0.023)	(0.006)
10 to 13	0.632***	-0.253*	-0.028	0.041	-0.048	0.018*
	(0.116)	(0.140)	(0.060)	(0.035)	(0.040)	(0.009)
N	1,269	1,268	1,269	1,269	1,269	1,269
Panel C: Young - Old						
Year since university opening						
5 to 9	0.244***	0.006	-0.004	0.022	-0.033***	0.013***
	(0.080)	(0.058)	(0.022)	(0.018)	(0.010)	(0.004)
10 to 13	0.537***	0.030	-0.001	0.036	-0.061***	0.016**
	(0.120)	(0.066)	(0.032)	(0.032)	(0.016)	(0.006)
N	2,961	2,960	2,961	2,961	2,961	2,961

Note: This table presents the results of the market-level DiD analysis, in which each observation is at the province-by-year level. Relative supply is the ratio of log college-educated workers to log non-college workers. Treatment group is provinces with the first university ever. Control group is provinces that almost had a university. College premium is measured as the ratio of log college wage to log non-college wage. Sectoral employment outcomes are the share of workers in a given sector. Basic service includes food, hospitality, entertainment, domestic, and other services that do not typically require a college degree. Skill-intensive service includes sectors that typically require a college degree, including health, education, finance, insurance, and science. Panels A and B are for Old and Young Cohorts, respectively. Panel C test the difference between the two age cohorts.

Table A4: General Equilibrium Effects of the Expansion on Sectoral Employment

	Agriculture		Manufacturing		Low-Skilled Service		High-Skilled Service	
	Non-College	College	Non-College	College	Non-College	College	Non-College	College
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Old Cohort								
Year since university opening								
5 to 9	-0.039*	0.008**	0.010	0.014***	0.018	0.001	0.002	-0.024
	(0.019)	(0.003)	(0.007)	(0.004)	(0.014)	(0.008)	(0.003)	(0.018)
10 to 13	-0.064**	0.007	0.036**	0.009	0.016	0.014	0.005	-0.034
	(0.024)	(0.006)	(0.014)	(0.010)	(0.020)	(0.015)	(0.005)	(0.030)
N	3,996	3,996	3,996	3,996	3,996	3,996	3,996	3,996
Panel B: Young Cohort								
Year since university opening								
5 to 9	-0.038*	-0.020***	0.032*	0.037**	-0.007	-0.003	0.009**	-0.003
	(0.020)	(0.006)	(0.018)	(0.018)	(0.015)	(0.013)	(0.004)	(0.013)
10 to 13	-0.051*	-0.028**	0.061**	0.059**	-0.032	-0.034*	0.012**	0.029
	(0.029)	(0.013)	(0.026)	(0.024)	(0.029)	(0.019)	(0.005)	(0.020)
N	2,997	2,997	2,997	2,997	2,997	2,997	2,997	2,997
Panel C: Young - Old								
Year since university opening								
5 to 9	-0.001	-0.016**	0.031*	0.028**	-0.034***	-0.014	0.008***	0.008
	(0.021)	(0.008)	(0.017)	(0.013)	(0.007)	(0.018)	(0.003)	(0.021)
10 to 13	0.004	-0.016	0.047	0.060**	-0.064***	-0.039*	0.008**	0.018
	(0.027)	(0.011)	(0.030)	(0.026)	(0.013)	(0.023)	(0.004)	(0.030)
N	6,993	6,993	6,993	6,993	6,993	6,993	6,993	6,993

Note: This table presents the results of the market-level DiD analysis, in which each observation is at the province-by-year-by age cohort-by education level. Relative supply is the ratio of log college-educated workers to log non-college workers. Treatment group is provinces with the first university ever. Control group is provinces that never had a university. Panels A and B are for Old and Young Cohorts, respectively. Panel C test the difference between the two age cohorts.

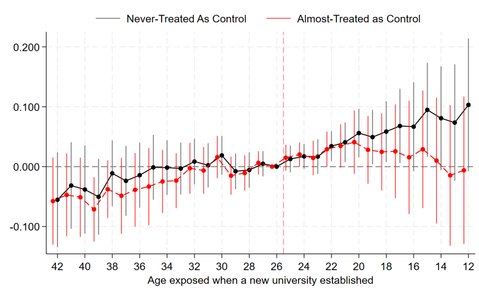
Table A5: Effects of Higher Education Expansion on Household Expenditure

	Share of Total Expenditure			
	Basic Service	Skill- Intensive Service	All Service	Expenditure per capita
	(1)	(2)	(3)	(4)
Panel A: Main Specification (Never-Treated as Control)				
Years since university opening				
5 to 9	0.017***	-0.004	0.012	26.246
	(0.005)	(0.005)	(0.008)	(23.121)
10 to 13	0.021***	0.001	0.022*	17.461
	(0.007)	(0.007)	(0.012)	(31.834)
N	161,880	161,880	161,880	161,880
Panel B: Alternative Specification (Almost-Treated as Control)				
Years since university opening				
5 to 9	0.020***	-0.003	0.017***	33.461
	(0.005)	(0.004)	(0.005)	(32.651)
10 to 13	0.027***	0.006	0.032***	23.727
	(0.008)	(0.005)	(0.010)	(44.913)
N	91,230	91,230	91,230	91,230
Panel C: Falsification Tests (Almost-Treated as Treatment)				
Years since university opening				
5 to 9	-0.004	-0.002	-0.007	-4.935
	(0.005)	(0.007)	(0.009)	(26.130)
10 to 13	-0.007	-0.005	-0.011	-6.885
	(0.007)	(0.009)	(0.012)	(37.233)
N	107,409	107,409	107,409	107,409

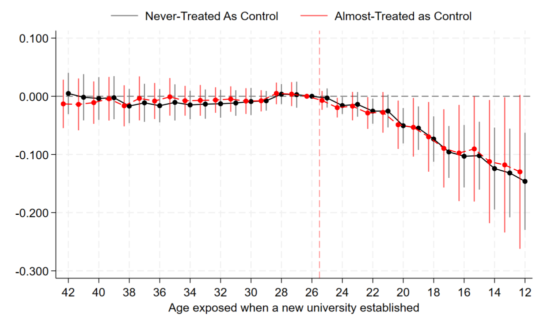
The table reports DiD estimates for impacts on individual outcomes using stacked regression, where treated provinces are compared to never-treated provinces across birth cohorts. I control for age, age squared, gender, province, birth, and cohort fixed effects. All standard errors are clustered at the province level. Data is drawn from LFS 2010-2018, but the sample is limited to 5 years after treeatment or more.

Figure A1: Event Study Estimates for Effects on Labor Market Outcomes

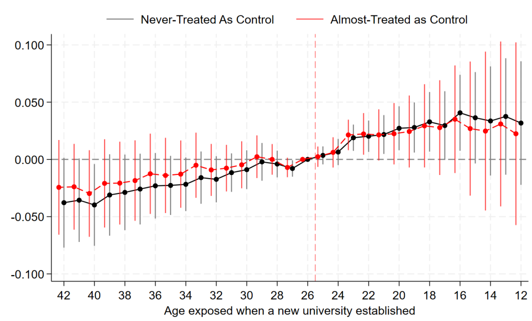
(a) Hourly Wage



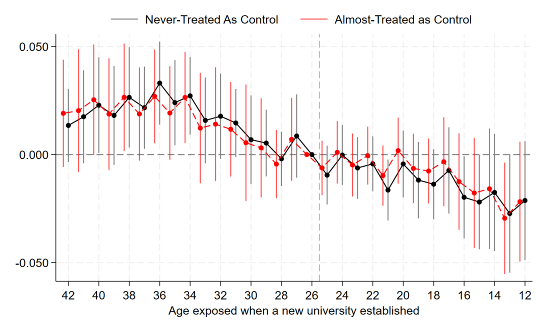
(b) Agriculture



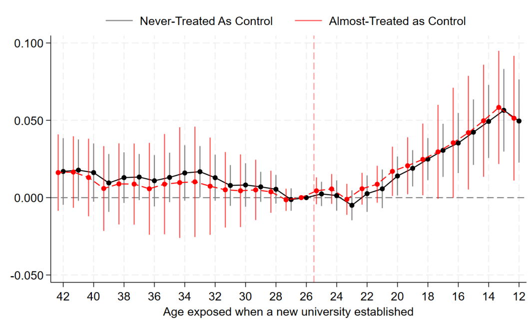
(c) Manufacturing



(d) Basic Service

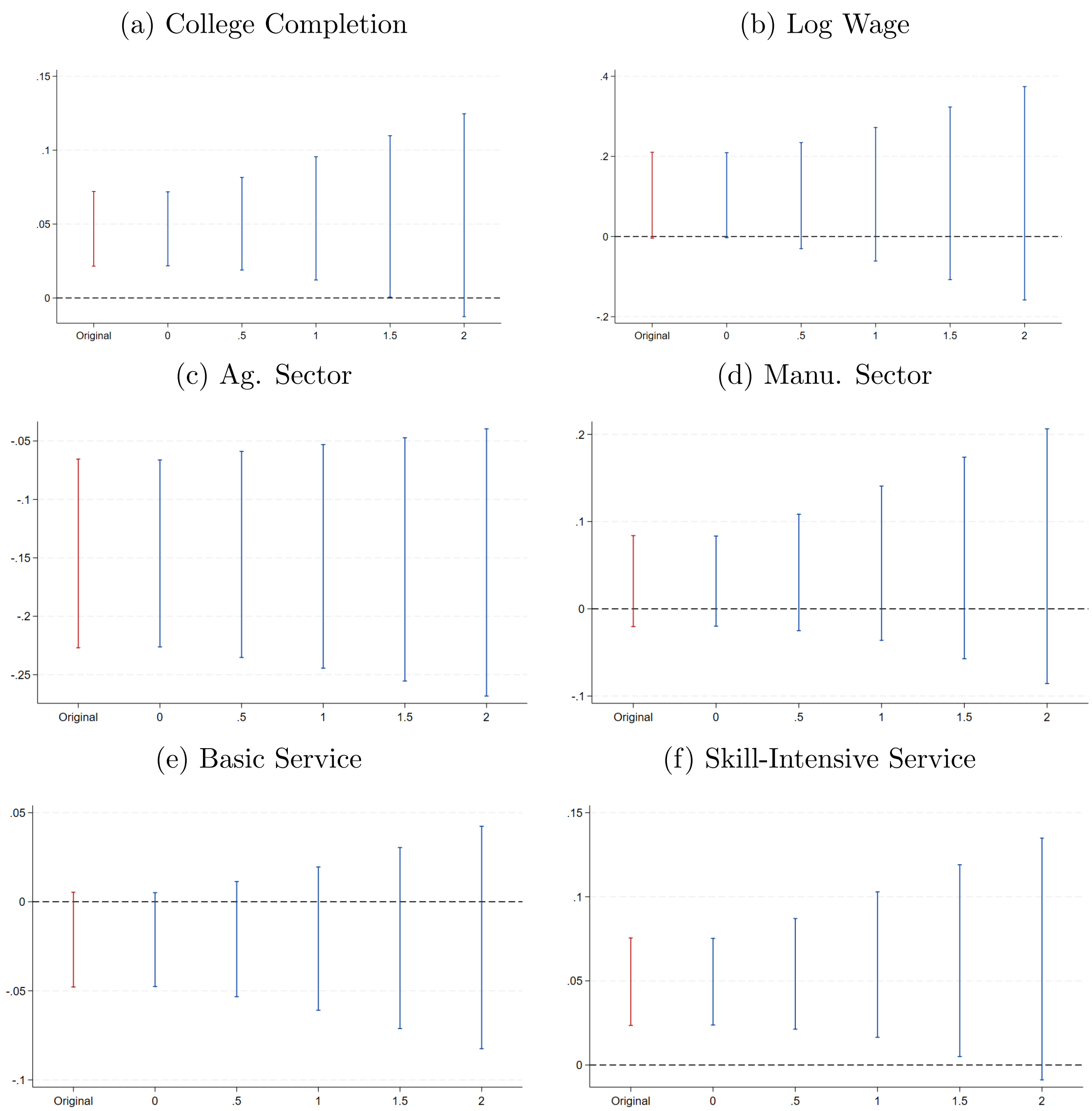


(e) Skill-Intensive Service



Note: The graphs display event study estimation for the effects of the higher education expansion on college completion and labor market outcomes. All models control for age, age squared, and gender. Standard errors are clustered at the province-level and 95% confidence intervals are displayed.

Figure A2: Honest DiD estimates



(a) College Completion

(b) Log Wage

(c) Ag. Sector

(d) Manu. Sector

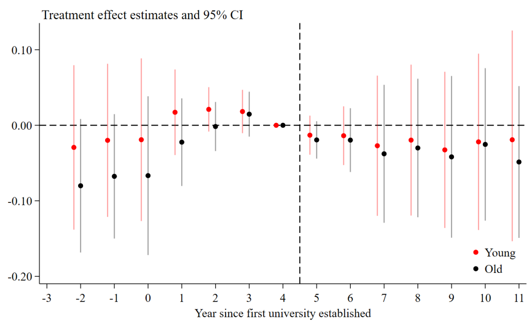
(e) Basic Service

(f) Skill-Intensive Service

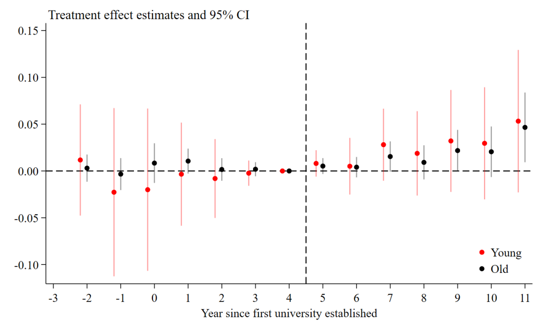
Note: The graphs display the robust confidence set for treatment effects using Rambachan and Roth (2023) for a given M threshold. M represents the maximum violation of the parallel trends for the post-treatment period as much as the violation in the pre-treatment period. For example, $M = 1$ means that the post-treatment period parallel trends are no more violated than the violation of pretrends.

Figure A3: Event Study Estimates for Effects on Market-Level Outcomes

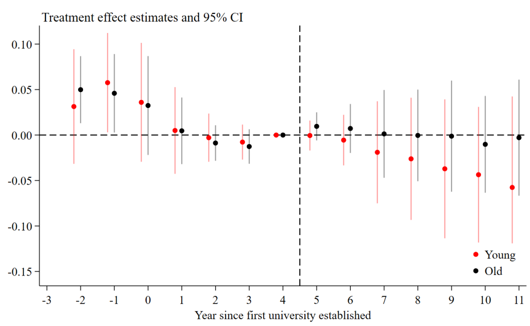
(a) Agriculture



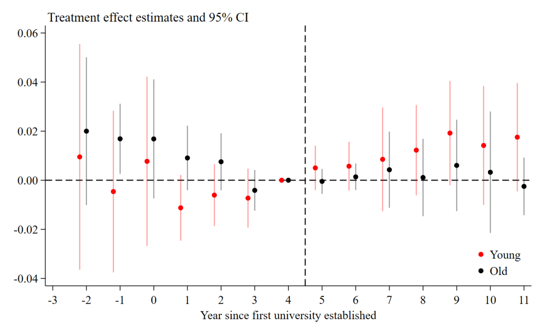
(b) Manufacturing



(c) Basic Service

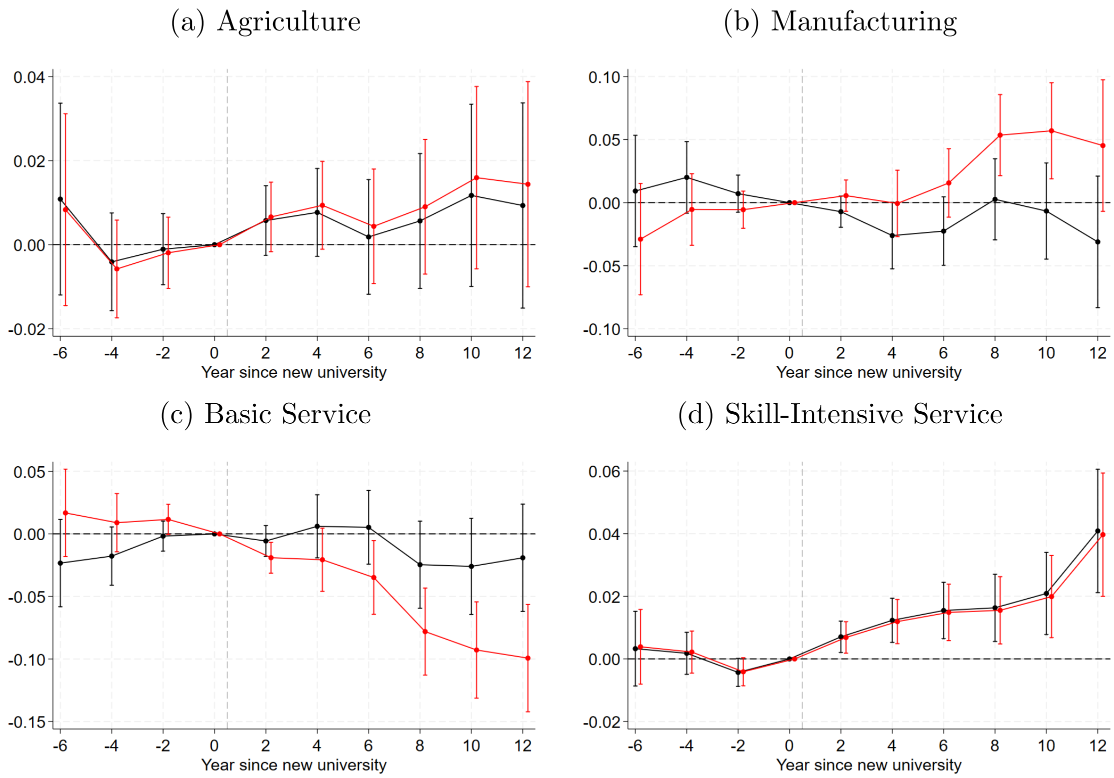


(d) Skill-Intensive Service



Note: The graphs display event study estimation for the effects of the higher education expansion on market-level labor market outcomes.

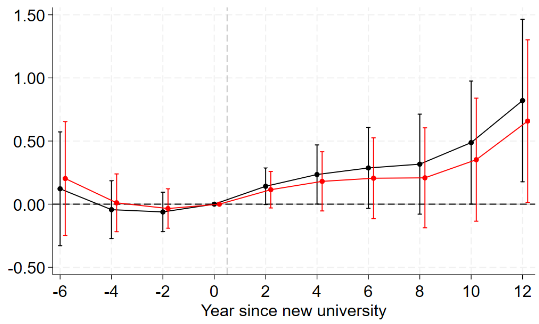
Figure A4: Event Study Estimates for Effects on Firm-Level Outcomes - With and Without Detrending



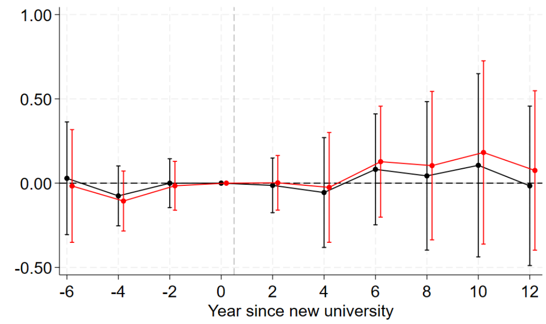
Note: The graphs display event study estimation for the effects of the higher education expansion on firm-level outcomes. The baseline results come from estimating the stacked regression. The detrended results come from detrending the outcome with a linear trend interacted with treatment status. Standard errors are clustered at the province-level and 95% confidence intervals are displayed.

Figure A5: Event Study Estimates for Effects on Firm-Level TFP - With and Without Detrending

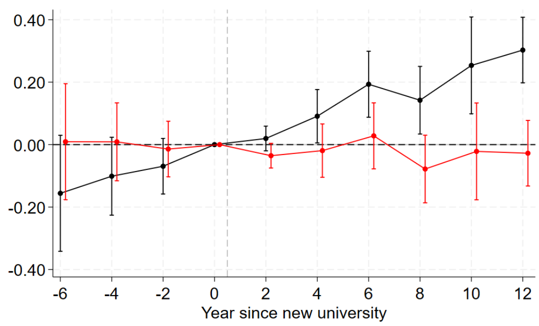
(a) Overall TFP



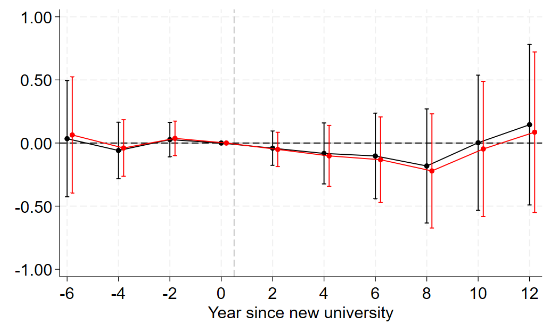
(b) Agriculture TFP



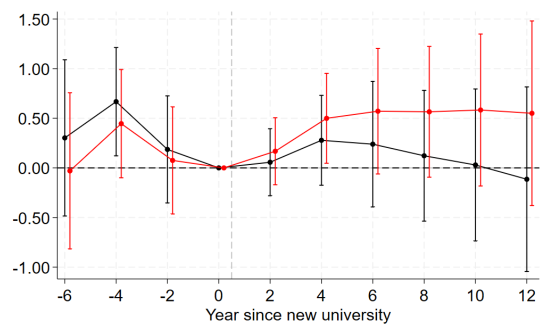
(c) Manufacturing TFP



(d) Basic Service TFP



(e) Skill-Intensive Service TFP



Note: The graphs display event study estimation for the effects of the higher education expansion on firm-level outcomes. The baseline results come from estimating the stacked regression. The detrended results come from detrending the outcome with a linear trend interacted with treatment status. Standard errors are clustered at the province-level and 95% confidence intervals are displayed.

A Theoretical Model

A.1 Model Setup

There are two levels of education: non-college and college, denoted as $i = \{N, C\}$ and two sectors: goods and services, denoted as $j = \{g, s\}$. Assume a non-homothetic, CES preference for consumptions of goods and services, denoted as c_g and c_s:

$$U(c_g, c_s) = \left[a_g c_g^{\frac{\epsilon-1}{\epsilon}} + (1 - a_g)(c_s + \gamma)^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}}$$

Workers of education level i and with a wage level w_i maximize their utility function $U(c_{gi}, c_{si})$ subject to $p_g c_{gi} + p_s c_{si} = w_i$.

Let Y_j be the CES production function only using non-college and college-educated labors as inputs, N_j and C_j. Firms in sector j solves the following profit maximization problem:

$$\max_{\{N_j, C_j\}} p_j A_j \left[\alpha_j C_j^{\frac{\rho-1}{\rho}} + (1 - \alpha_j) N_j^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}} - w_N C_N - w_C C_j$$

To illustrate the impacts of the exogenous shock in the supply of college-educated workers, I assume that $f(u)$ is the share of college-educated workers, which is a function of whether there is a university within the area, and is independent of the college wage premium. I also normalize non-college wage to be one, so the college wage premium is w_C.

A.2 Equilibrium

Two factor-markets and two output-markets have prices w_N, w_C, p_g, p_s. Solve for the first-order conditions of the consumer problem yields the aggregate expenditure share of service consumption as

$$e_s = \frac{p_s[(1-f)c_{s,N} + fc_{s,C}]}{1-f+fw_C} = \frac{1}{\left(\frac{1-a_g}{a_g}\right)^{\epsilon} + \left(\frac{p_g}{p_s}\right)^{1-\epsilon}} \left[\left(\frac{1-a_g}{a_g}\right)^{\epsilon} - \frac{p_s \bar{c}_s \left(\frac{p_g}{p_s}\right)^{1-\epsilon}}{1-f+fw_C}\right] \tag{3}$$

where f is the share of college-educated workers. This equation captures the essence of standard structural transformation theory: As the relative price of goods, $\frac{p_g}{p_s}$, falls, the expenditure share of service increases. Alternatively, as the overall income, measured as $(1-f+fw_C)$, increases, the expenditure share of service also increases because of non-homotheticity.

Solve for the first order conditions of the firm problem yields the relative demand for college-educated workers $\frac{C_j}{N_j} = \left[\frac{\alpha_j}{(1-\alpha_j)} \frac{1}{w_C}\right]^{\rho}$ and the prices of sector j in the equilibrium as a function of the relative wage of college-educated workers: $p_j(w_C) = \frac{1}{A_j}\left[\frac{\alpha_j^{\rho}}{w_C} + (1-\alpha_j)^{\rho}\right]^{\frac{1}{1-\rho}}$. The demands for non-college workers and college-educated workers in sector j are

$$N_j = \left[(1-\alpha_j)\hat{p}_j(w_H)A_j\right]^{\rho} \frac{Y_j}{A_j}; \quad C_j = \left[\frac{\alpha_j \hat{p}_j(w_H)A_j}{w_H}\right]^{\rho} \frac{Y_j}{A_j}$$

The market-clearing condition for college-educated workers can be written as supply equal to the equilibrium demands for college-educated workers in the goods and service markets:

$$f = \sum_j \hat{C}_j = \sum_j \left[\frac{\alpha_j \hat{p}_j(w_H)A_j}{w_H}\right]^{\rho} \frac{\hat{Y}_j}{A_j} \tag{4}$$

where $\hat{Y}_j = f\hat{c}_{j,C} + (1-f)\hat{c}_{j,N}$ are the market-clearing conditions for the goods and service markets, and $\hat{c}_{j,C}$ and $\hat{c}_{j,N}$ are equilibrium demands for goods and service of college-educated and non-college workers.

One can then derive the effect of an increase in the relative supply of college-educated workers, f, on the college wage premium is derived as

$$d \log w_C = -\frac{1}{\tilde{\rho}} d \log \frac{f}{1-f} \tag{5}$$

where $\tilde{\rho}$ is a function of e_s

The general equilibrium effects of the higher education expansion are captured in Equations 3, 4, and 5. As f increases, w_C falls due to imperfect substitution between C and N. Higher f also means higher overall income $(1 - f + f w_C)$ and, thus, higher aggregate expenditure share towards service consumption, e_s, due to the non-homothetic preference assumption. This structural transformation raises demand for college-educated workers in the service sector, thus raising w_C.

B Estimating total factor productivity at the firm level

Consider the following Cobb-Douglas production function for firm i in year t:

$$VA_{it} = \beta_l L_{it} + \beta_k K_{it} + \omega_{it} + u_{it}$$

where VA_{it} is the annual value-added, L_{it} is total labor, K_{it} is capital, measured as the value of assets at the beginning of the year (Newman et al., 2015), and ω_{it} is the unobserved productivity shock. Given that OLS is typically biased as both L_{it} and K_{it} are likely affected by the unobserved productivity shock, we first assume that firms' investment decision is a function of labor, capital, and productivity shock, i.e., $I_{it} = f_t(L_{it}, K_{it}, \omega_{it})$,[16] which makes ω_{it} observable in the production function (by inverting f_t):

$$VA_{it} = \beta_l L_{it} + \beta_k K_{it} + f_t^{-1}(L_{it}, K_{it}, I_{it}) + u_{it}$$

This approach by Ackerberg et al. (2015) (ACF) is different from two other convention approaches to estimate production functions, namely Olley and Pakes (1996) (OP) and Levinsohn and Petrin (2003) (LP), who do not include labor input as part of firms' investment decision. Not allowing labor input to enter the investment function means that L_{it} is a deterministic function of capital and investment and, hence, would be functionally dependent on the inverse function of investment; in other words, the coefficient of labor would not be identified.

Assuming that the productivity shock follows a first order Markov process, we can write $\omega_{it-1} = g(\omega_{it-1}) + \zeta_{it}$ where $g(\omega_{it-1})$ is the predictable component and ζ_{it} is the unpredictable/innovation component of productivity (Olley and Pakes, 1996). We also assume the following capital formation process: $K_{it} = (1 - \delta)K_{it-1} + I_{it-1}$. These assumptions give us

[16] Investment is measured as annual change in value of fixed and long-term assets plus accumulated depreciation (Newman et al., 2015).

$E[\zeta_{it}|I_{it-1}] = 0$ and $E[\zeta_{it}|K_{it}] = 0$ (since K_{it} is determined at $t-1$). Lastly, we assume that $E[\zeta_{it}|L_{it-1}] = 0$ (Ackerberg et al., 2015). Given this set of moment conditions, we can estimate β_l and β_k.

Given that this approach requires panel data, we aggregate the firm-level variables to the district-by-year-by-industry level. We then estimate the production function by sectors and present the estimation results in Table A6. Once we obtain the estimates for β_l and β_k, we can use these estimates to calculate ω_{it} for each firm in each year, which is also our measure of total factor productivity (TFP).

Table A6: Production function estimation results

	Capital		Labor	
Agriculture				
	0.271***	(0.098)	1.143***	(0.392)
Mining				
	0.222***	(0.031)	1.063***	(0.048)
Manufacturing				
	0.342***	(0.031)	0.919***	(0.041)
Waste and electricity				
	0.380***	(0.065)	0.980***	(0.126)
Construction				
	0.198***	(0.028)	0.948***	(0.028)
Wholesale and retail				
	0.201***	(0.036)	1.386***	(0.055)
Transportation				
	0.057	(0.038)	1.103***	(0.055)
Hospitality				
	0.222***	(0.071)	0.682***	(0.186)
Information and communication				
	0.314***	(0.062)	1.035***	(0.105)
Finance, banking, and real estate				
	0.175***	(0.024)	1.348***	(0.088)
Science and technology				
	0.123***	(0.035)	1.384***	(0.050)
Administrative and support				
	0.156***	(0.041)	1.144***	(0.046)
Education, health, and social support				
	-0.116	(0.191)	1.535***	(0.296)
Entertainment				
	-0.112	(0.166)	1.677***	(0.268)
Other services				
	0.236	(0.250)	1.220**	(0.584)