Post-secondary Decision Making: Evidence from Large Plant Openings *

Yusuf Kulu
Johns Hopkins University

Bledi Taska
Lightcast

August, 2023

KEYWORDS: labor market shocks, postsecondary education, two-year colleges
JEL CLASSIFICATION: I23, I26, J63

*The authors thank Robert Moffitt, Nicholas Papageorge, and Yujung Hwang for their advice and guidance. We wish to thank Daniel Rock for his help and suggestions, and Jonathan Elliott for comments and discussions. Email addresses: ykulu1@jhu.edu and bledi.taska@lightcast.io

Abstract

High school graduates make an important decision with many future implications: pursuing post-secondary education or entering the labor market. This study examines the effects of large plant openings on this decision, utilizing the near universe of online vacancies in the U.S. to understand better the local labor market demand for different skill levels. Our identification strategy compares locations that narrowly secured plant investments to those that narrowly missed out, a method first used by Greenstone, Hornbeck, and Moretti (2010). Our results align with the existing literature: Upon plant openings, individuals take available jobs and move away from college enrollment. However, when the share of jobs requiring two-year college credentials increases, more high school graduates are attracted to two-year colleges, indicating that the type of labor market shocks is relevant in understanding college enrollment decisions.

The decision to attend college is a pivotal milestone for high-school graduates, with far-reaching implications for their prospects and well-being. According to the standard human capital theory, agents decide by comparing the future return from a college and the current costs. The return part mainly includes a higher wage profile over the lifecycle, and the cost side contains the expenses in terms of tuition, room and board, and the opportunity cost of foregone earnings while studying in college.

In the realm of labor market dynamics, numerous papers have explored the relationship between shocks to the labor market and college enrollment. Some studies like Betts and McFarland (1995) and Charles, Hurst, and Notowidigdo (2018), indicate that during periods of favorable labor market conditions, such as when new plants or industries open up, there tends to be a decrease in college enrollment rates among high-school graduates. It is interpreted as individuals opting for immediate employment opportunities instead of pursuing higher education. However, these studies have predominantly focused on non-college jobs and the opportunity cost of college-going. Better labor market conditions for college graduates, like an increase in college jobs, can have different effects. If a plant is mostly hiring college graduates, the expected future return of a college will increase, potentially leading to an increase in college enrollment.

This paper investigates the impact of new plant openings on high-school graduates' college decisions, utilizing a dataset containing the near universe of online vacancies to understand better the nature of local labor market demand for different skill and education groups. With this data set, the educational requirements of the jobs created by the entrance of new plants and industries can be determined, not just the total increase in labor demand used by past studies. The identification strategy is based on comparing locations that narrowly secured a million-dollar plant investment to those that narrowly missed out, first used by Greenstone, Hornbeck, and Moretti (2010), then followed by numerous studies, including Bloom et al. (2019) and Monte, Redding, and Rossi-Hansberg (2018).

We find that upon a plant opening, college enrollment rates among recent high-school graduates decline by about two percentage points when one does not distinguish between the type of vacancy. Then, we investigate the type of job opening by focusing on the vacancies posted by the plant and their educational requirements to understand whether a plant is mainly hiring higher-skilled or lower-skilled employers. It indicates that when more jobs become available for two-year college individuals, a higher share of recent high-school graduates enrolls in two-year colleges, indicating an increase in returns to two-year colleges. The same channel does not apply to four-year colleges, and it could be showing that job availability is not the main factor affecting returns to a four-year college.

Lightcast (formerly Emsi Burning Glass) provided the datasets used in this study. The first dataset contains online vacancy postings with detailed information on the job posting's occupation, location, and employer name. It is a comprehensive dataset capturing almost the universe of online vacancy postings in the U.S., collected from over 50,000 online job sites. Some jobs are not posted online, but they are a smaller fraction of the entire vacancies, so the Lightcast vacancy dataset provides a good proxy for the total labor demand in the United States. The second dataset is a resume dataset which includes approximately 36 million individuals with education and labor market information. Specifically, the educational information contains the institution name, location, and year of graduation; the labor market section includes the employer name, occupation, and years worked. This data set allows us to track individuals' transition from education to the labor market. It has a large sample size at the county level, which is the primary geographic unit of this study. However, we do not utilize the employer name information on the resume because we do not track whether an individual has taken a job in a specific plant.

Bloom et al. (2019) share the million-dollar plants' data with information including the company name and its location. We merge each plant opening with the Lightcast vacancy data by employer and county name. This study focuses on 40 new plants, and there is substantial heterogeneity of labor demanded. For example, one plant posts 81% of its vacancies to individuals with high-school or less level of education group, whereas another plant opens

92% of its vacancies to four-year college graduates or above level of education group.

Local labor demand is on the county level, and the 40 new plants are located in 35 unique "winning" counties. High-school graduates from these counties are compared to those high-school graduates from 58 runner-up (or losing) counties. The sample of individuals from the resume dataset includes individuals who have graduated from a high school in either a winning or a losing county. It covers more than 80,000 individuals from a total of 93 U.S. counties.

Our paper contributes to empirical literature investigating the effects of labor market shocks on educational outcomes by bringing detailed information on the nature of the labor market shock. The Lightcast vacancy dataset allows us to analyze the type of required education level for each job posted by a new plant. The 40 new plants located in 31 U.S. states provide a sufficient amount of heterogeneity of skills demanded by million dollar plants.

The existing literature on educational response to a labor market shock focuses on events like mass layoffs, the invention of new technology, the differential effect of regions by a national event, or a change in minimum wages. Hubbard (2018); Minaya, Moore, and Scott-Clayton (2023); and Foote and Grosz (2020) focus on how college attendance is affected by mass layoffs. Cascio and Narayan (2022) analyze the educational response to fracking technology which has increased the relative demand for less-educated males. Charles, Hurst, and Notowidigdo (2018) investigate the national housing boom and bust based on the idea that housing prices during this period derived from a speculative bubble instead of standard determinants. Schanzenbach, Turner, and Turner (2023) find that minimum wage increases has negative enrollment effects for students in community colleges.

The remaining of the paper proceeds as the following: section 1 reviews the literature, section 2 discusses the research design and presents the datasets, section 3 provides the analysis sample and how it compares to different datasets, section 4 presents the empirical strategy, section 5 shows the results, and section 6 concludes.

1 Literature Review

1.1 Labor Market Shocks and Educational Outcomes

Labor market shocks affect returns to education, but how it affects them depends on the type of labor market shock. A positive labor market shock for less skilled individuals will increase the opportunity cost of going to a college, decreasing returns to a college. On the other hand, a positive labor market shock for college graduates would increase the expected wages upon graduation, increasing returns to college.

Several early papers focus on the relationship between the unemployment rate and enrollment. Manski and Wise (1983) show a weak relationship between local labor market opportunities and four-year college applications, and Betts and McFarland (1995) find that when the unemployment rate of recent high-school graduates increases by 1 percent, it is associated with an increase in full-time community college attendance by about 0.5 percent.

The majority of the recent literature analyzes negative labor market shocks, namely mass-layoffs, like Hubbard (2018) in Michigan; Minaya, Moore and Scott-Clayton (2023) in Ohio; Foote and Grosz (2020) for the entire mass-layoffs compiled by Bureau of Labor Statistics (BLS). These studies indicate an increase in college attendance after a mass layoff, and the effect is more pronounced in two-year college margins.

Black, Mckinnish, and Sanders (2005) focus on high school enrollment rates after the coal boom and bust of the 1970s and 1980s in Kentucky and Pennsylvania. They indicate the countercyclical movement in educational outcomes as the shocks have changed the earnings of high-school dropouts. Cascio and Narayan (2022) analyze the educational response to fracking technology, which has increased relative demand for less-educated males. They find that male high-school dropout rates have increased relative to females after the technological breakthrough.

Charles, Hurst, and Notowidigdo (2018) investigate the national housing boom and bust after 2000 using several data sources, including Census and American Community Survey

(ACS). The identifying strategy exploits growth in housing demand in a Metropolitan Statistical Area (MSA) instrumented by the structural break, which is based on the idea that variation in housing prices during this period derived from a speculative "bubble" instead of standard determinants like income, population, or construction costs. The authors find that the housing boom of the 2000-2006 periods lowered college enrollment, more on the two-year college margin. This trend has reversed after the bust. The housing boom in an area generally increases opportunities for construction workers, increasing opportunity costs of college-going, thus, lowering college enrollment.

There is tangential literature in the trade where education response to the arrival of new jobs has been analyzed. Atkin (2016) find an increase in dropout rates for high-school students after an employment spike in export sectors in Mexico. On the other hand, several other papers, like Jensen (2012) find that an increase in the high-skilled IT sector in India increases enrollment rates, indicating that the type of labor market shock is relevant in understanding returns to education.

1.2 Million Dollar Plant (MDP) Literature

The first paper in the literature that uses million-dollar plant opening winner and losing county as an identifying strategy is Greenstone, Hornbeck, and Moretti (2010), where the authors investigate how a million-dollar plant opening in a county impacts the total factor productivity (TFP) of the existing firms. The authors think of a spillover mechanism where they investigate productivity advantages through agglomeration. The authors use the MDP data combined with several other datasets, including Census Bureau's Standard Statistical Establishment List (SSEL), to verify the year of the plant opening in the MDP dataset. Their empirical strategy builds upon a difference-in-differences framework with plant fixed effects, two-digit industry-by-year fixed effect, and pair (they call it a case) fixed effect to ensure that the impact of the MDP opening is identified from comparisons within a winner-loser pair. The paper's main result is that an MDP opening is associated with a 12 percent relative increase in incumbent plants' TFP five years later.

The authors validate their research design by comparing preexisting observable county characteristics and trends in TFP in winning and losing counties prior to the MDP opening. They show that in the years before the MDP opening, TFP trends among incumbent plants were similar in both types of counties.

Bloom et al. (2019) investigate drivers of good management practices and their mechanisms. In particular, they use the MDP dataset to understand the impact of new plant openings on the management practices of the incumbent firms by learning spillovers from the newcomer. They find evidence of positive management practices by learning from the newcomers.

Bhardwaj et al. (2022) investigate how an MDP affects the prices of retail goods and wages using the MDP dataset combined with IRI, retail price data, March Current Population Survey (CPS), and American Time Use Survey (ATUS) to obtain labor market data like wages, and hours worked. Their empirical strategy is a form of difference-in-differences that includes pair, county, industry, and year fixed effects. They find that product level prices increase by 0.7 percent after a new establishment in the county. The authors find that wages only increased for skilled workers by 15%, and rising local prices translate into a decline in real wages for unskilled workers.

Monte, Redding, and Rossi-Hansberg (2018) use the MDP strategy as a separate empirical analysis to support the model predictions. They find that employment in winning counties is, on average, close to 4 percent larger than in runner-up counties after 10 years of the MDP announcement. They further find greater increases in employment in counties with more open local labor markets.

Qian and Tan (2021) do not exploit the MDP pair strategy; instead, they exploit variation in proximity to big-plant opening. Their results indicate that high-skill firm entry benefits high-skilled homeowners, and its impact on low-skilled people are ambiguous. Kim (2020), Patrick and Partridge (2019), Giroud et al. (2021), and Slattery and Zidar (2020) are several other papers that have used the MDP strategy to tackle questions mostly related questions including agglomeration, productivity spillovers, and firm capital structure.

2 Research Design and Data Sources

2.1 Research Design

A magazine called Site Selection had an article titled "Million Dollar Plants" where the article compiled information about the location rankings of profit-maximizing firms. Usually, firms initially consider a long list of potential locations to invest. After several months of investigations of many possible locations, firms reduce the number of potential locations to a shorter list of final candidate counties for the investment.

Firms make an investment decision based upon profit maximization motive. In the context of location preferences, two factors are essential to consider, as Greenstone, Hornbeck, and Moretti (2010) mentioned: expected future cost of production in a location and the total subsidy received.

Runner-up and winner counties have similar expected future cost functions from the perspective of a firm, so several county characteristics like the share of college-educated population, labor force participation rate, and population change over the years are similar. One caveat of the MDP research design is the second factor, where the total subsidy received by the firm is usually unknown, and it limits our understanding of the role it played during the selection process. Overall, runner-up and winner counties are a better way to compare themselves instead of the winner county with all the remaining counties and several papers in the literature like Greenstone, Hornbeck, and Moretti (2010), Bloom et al. (2019), Bhardwaj et al. (2022) have incorporated this type of a research design.

We discuss why this research design is valid by providing information on how these winner and runner-up counties are similar in observable characteristics and how they differ from the remaining U.S. counties. In the following sections, we also present evidence that these two types of counties had similar college enrollment trends prior to plant openings.

2.2 Data Sources

This paper uses the Million Dollar Plants data provided by Bloom et al. (2019), extending an earlier dataset by Greenstone, Hornbeck, and Moretti (2010). The dataset includes 132 large investments between 2000 and 2017, where winner and runner-up counties exist for an investment. We refer to a "pair" as consisting of a winning county and either a losing county or multiple losing counties competing for a plant opening.

Each new plant opening is merged into Lightcast Vacancy data. Lightcast is an employment analytics firm collecting vacancy data from over fifty thousand online job sites. Duplicate vacancies are removed, and the company claims that this vacancy data covers the entire online vacancies in the United States. The data is available for 2007 and all the years between 2010 and 2021. The vacancy data only captures new job postings rather than active postings. Thus, when a new vacancy is posted, it is included in the data in that month. It is removed from the dataset next month even if the posting remains active.[1] Information on the educational requirement of the vacancy is not available for about half the vacancies. However, almost all vacancy postings include employer name, location, sector, and occupation information. The six-digit occupation code for each vacancy posting is merged into Occupational Information Network (O*NET) dataset to obtain educational requirements for each job posting.

The O*NET dataset provides information on almost 1,000 occupations covering the entire U.S. economy. It is developed under the sponsorship of the U.S. Department of Labor/Employment and Training Administration. We use the Education, Training, and Experience section, where the dataset includes the required level of education for each occupation. The required level of education variable is provided in a probability distribution. However, we reduce it into a single required education group for each six-digit occupation by taking

[1]It is different than BLS The Job Openings and Labor Turnover Survey where they keep track of active job postings.

the biggest mass education category. Ultimately, we categorize each vacancy posting into one of three groups: ones requiring high school or less education, ones requiring two-year college credentials, and ones requiring four-year college credentials. The appendix details how O*NET data is converted into these three classifications.

Of the 132 pairs available in the original Bloom et al. (2019) dataset, we use 40 pairs according to the following criteria: (1) The loser county location should exist; (2) The year of plant opening should be 2007 or later as Lightcast Vacancy dataset starts at 2007; (3) A county should not be a winner and a loser at the same time; and (4) The county location of a new plant and the plant name should match the Lightcast Vacancy data county and employer name.

These 40 new plants are located in 35 different U.S. counties. One new plant is constructed in 31 counties, two new plants are built in 3 counties, and three new plants are constructed in 1 county. Therefore, there are a total of 35 winning counties. On the other hand, there are a total of 58 losing counties. Fifty of them were losers once, and 8 of them were runner-ups multiple times. In total, we used 93 U.S. counties for the analysis sample.

National Center for Education Statistics (NCES) Integrated Postsecondary Education Data System (IPEDS) provides college-level information. We use this dataset to understand county-level college enrollment numbers. For each college, IPEDS provides first-time Fall enrollment numbers. There are three levels of institution: four or more years, at least two but less than four years, and less than two years (below associate). We take the first one to be the four-year college category and the second one to be the two-year college category. The institutions below associate degrees are not included in the two-year college category.

The Lightcast Resume dataset is used to obtain individual-level data for the 93 counties described above.[2] This dataset has resumes of approximately 36 million individuals in the United States. The data has information on individuals' educational backgrounds and job

[2]We discuss why this dataset is used in the next section. There could be potential issues of non-representativeness of using this dataset for the entire U.S. population. However, it has many other advantages.

histories, including city-level locations. Educational background has information on the institution's name, the type of degree received (high school, associate's degree, bachelor's degree), and the graduation year. Job history provides information on the 6-digit O*NET code, employer name, and job date range. However, we only use county-level locations of individuals and their college enrollment patterns before and after a new plant opening in a county. We do not utilize information like the employer's name from the resume. We also do not use the panel nature of the data set and treat each year's observations as an independent sample. Thus, our sample consists of high-school graduates of the winner or loser counties.

The analysis sample consists of individuals with the following criteria: (1) The resume should include high-school information[3]; (2) High-school graduation year and high-school county location should exist in the resume; (3) The high school should be located in one of the 93 U.S. counties; (4) There should not be a gap of more than two years after high-school graduation in the resume. The first criterion can create a sample selection, and the next section discusses why it is used and the possible issues with having that criterion. We think the remaining criteria do not create many problems with the sample selection. The analysis sample consists of 81,281 individuals who graduated from high schools in one of the 93 counties.

3 Sample

Table 1 presents summary statistics of the million-dollar plants. Panel A presents information on the original data obtained from Bloom et a. (2019). The average size of a plant is \$505 million, and the average size of employment is around 1,100. Approximately half of the plants are in the manufacturing sector, and about three-quarters are newly established. Panel B information is obtained by merging each of these plants with the Lightcast Vacancy dataset. The educational requirement information is further obtained using the

[3]This condition is necessary. Otherwise, if the first information in the resume is a college, it is impossible to understand whether this person has worked before college enrollment and whether this person has moved to study in a college.

6 digit occupation code of each vacancy linked to the O*NET dataset. On average, these plants have posted 1,549 vacancies. More than half of these vacancies are posted for individuals with at least a bachelor's degree and about a third are posted for those with at most high-school degrees. It shows the heterogeneity between plants with respect to the type of labor demanded.

Table 2 Panel A compares winning county characteristics with the losing county characteristics and the remaining U.S. county characteristics using the U.S. Department of Agriculture dataset. The variables are population, labor force participation rate, unemployment rate, and educational level before the plants were established. Winning and losing counties have many similarities, like the size of their population and their population characteristics. They are much bigger compared to the remaining U.S. counties. They have higher labor force participation rates and lower unemployment rates than the remaining U.S. counties. Their population is more educated than the remaining U.S. population. None of the observable differences between winning and losing counties are statistically significantly different from each other, and it provides credibility to the research design. Panel B uses the Lightcast Vacancy dataset. Approximately, 600,000 vacancies are posted in each winning county over 2007 and 2010-2021, which corresponds to aroung 45,000 job postings per year. About half of these vacancies are posted for individuals with high-school or less education level, 10% are posted for two-year college graduates, and the remaining are posted for four-year college graduates.

Table 3 presents the analysis sample consisting of 81,281 individuals. Column 2 provides information on people in the winner counties before the plant opening and column 3 after the plant opening. Column 4 presents statistics on individuals in the runner-up counties prior to the investment and column 5 is the post-investment. Approximately 7% of the sample has continued their education in two-year colleges after high school, 11% has moved to a four-year college, and 92% in the labor market upon high-school graduation. Many college students are working while they are studying, so these three numbers do not add up to one.

66% of the sample has high school as the highest level of schooling, 15% of the sample has a two-year college degree or diploma, and 19% of the sample has a bachelor's degree or higher level of education. Post-MDP columns have higher ratio of labor market, and a lower ratio of college enrollments compared to Pre-MDP years.

3.1 Representativeness of the Sample

As mentioned in the previous section, the sample consists of individuals with high-school information on their resumes. These people are slightly lower educated compared to the whole population. When we compare Table 2 and Table 3, 49% of the population in winning counties have a high-school diploma or less level of education, but this ratio for the analysis sample is 66%.

There are some advantages of using a Resume dataset: (1) It is a panel-like dataset where no panel dataset in the U.S. has a sample size as big as the Lightcast Resume dataset. Although this paper does not utilize the panel nature, it has potential for future use. (2) Employer and postsecondary institution names allow leveraging more information by merging firm-level and institutional-level characteristics. We only differentiate a four-year college from a two-year college, but postsecondary institution names can be merged with IPEDS dataset to leverage college-level information further. (3) City-level location information is also unique to track individual location choices over time. This study focuses on county-level location as the labor market is defined at that level.

American Community Survey (ACS) and Current Population Survey (CPS) might be the two alternative datasets, but they have some disadvantages preventing us from using them[4]. Two-year and four-year college enrollment patterns can be identified using the variable indicating the highest achieved education level as it has distinct categories for associate's and bachelor's degrees. However, this variable do not indicate the year of graduation. Therefore, these datasets can only be used to cross-validate the general pattern between labor market

[4]We replicate some of the results using the CPS ASEC sample in the Appendix.

shocks and college enrollment, not the enrollment timing for different colleges.

Another related concern is about the collection of the Lightcast Resume dataset. If there was an increase in the number of resumes in winning counties right after a plant opening, then an increase in the number of high-school graduates entering the labor market might be due to a rise in the number of resumes. It might not have anything to do with a change in choices for high-school graduates towards the labor market in winning county. We think this is not the case. We compare the populations in the winning and losing counties with the number of resumes in each county. They are almost identical at the aggregate level and also the same for different years. Therefore, we conclude that the number of available resumes in the Lightcast dataset is unaffected by labor market conditions.

4 Empirical Strategy

We estimate the difference-in-differences type of equations where a plant opening is the onset of an event. The first estimating equation has the following form:

$$Y_{ipct} = \beta_1 Winner_{ct} + \rho_p + \tau_t + \epsilon_{ipct} \tag{1}$$

where Y is a binary outcome variable for categories representing which option an individual has moved right after high school: labor market, two-year college, and four-year college. Results estimate three separate linear probability models and a multinomial logistic regression for comparison. i stands for the individual, p is pair of counties running for a plant, c is a county, and t is a calendar year. The winner is a dummy variable taking the value one after a million-dollar plant starts to operate, and it remains one after that.[5] ρ_p is pair fixed effects, and it allows identifying the parameter of interest by comparing winner and loser counties. τ_t is year fixed effects accounting for time-variant variables between students in

[5]Bloom et al. (2019) million dollar data provide the decision year of a plant. However, plants start to operate several years after they are decided to build. The vacancy data allows us to determine a plant's opening year. It is identified by observing a plant's first peak of vacancy postings.

different counties. All standard errors are clustered at the county level.

We further investigate heterogeneities in the labor market shock and their impact on college decisions by focusing on vacancies posted by million-dollar plants. Each plant seeks to employ individuals with different skill levels. For example, one new plant posts 81% of its vacancies to individuals with high school or lower level of education, whereas another plant posts only 7% of its vacancies to the same group. We estimate the following equation to understand these differences:

$$Y_{ipct} = \beta_2 Winner_{ct} + \sum_{k=1}^{3} \delta_k PlantEducShare_{pck} + \rho_p + \tau_t + \epsilon_{ipct} \tag{2}$$

where PlantEducShare stands for the share of vacancies posted by a new plant for different education levels. The shares are nonzero in winning counties, and they are zero in losing counties as they do not have a plant opening. We partition these shares, k, into three groups: (1) jobs requiring high-school or less level or education group, (2) jobs requiring two-year college credentials, (3) jobs requiring four-year college or above the level of education. The shares add up to one. We estimate separate linear probability models and a multinomial logistic regression for comparison.

Finally, we estimate the nature of the vacancies at the county level. A new plant opening changes the characteristics of vacancies in the county, so we utilize the entire job postings in winning and losing counties with the following equation:

$$Y_{pct} = \beta_3 Winner_{ct} + \sum_{k=1}^{3} \eta_k EducShare_{pctk} + \sum_{k=1}^{3} \alpha_k Winner_{ct} EducShare_{pctk} + \rho_p + \tau_t + \epsilon_{ipct} \tag{3}$$

where EducShare means the share of job openings for different education levels in the county. The variable of interest is α's because they capture the changes in the educational shares created by a new plant opening. The η coefficients estimate how different shares affect college enrollment rates. Winner variable proxies the total number of vacancies posted in a plant.

We also estimate a different version of this equation where the winner variable is replaced by the total number of vacancies posted. In this case, the interaction variables indicate the total number of vacancies posted for each education group.

There are 40 pair fixed effects, and the year fixed effects represent each individual's high-school graduation year. Year-fixed effects are available for every between 1955 to 2020. Only 13% of individuals have high-school graduation years before 2000, so the results are not affected by these earlier cohorts, as shown in the Appendix.

Estimating equations above are similar to a difference-in-differences framework, and they have the following assumptions:

Assumption 1: This assumption is the parallel trends assumption where the counties in the same pair should have a similar trend. Formally, it is the following:

$$E(Y_{ipcpost}(0) - Y_{ipcpre}(0)|Winner_{ct} = 1) = E(Y_{ipcpost}(0) - Y_{ipcpre}(0)|Winner_{ct} = 0) \quad \forall \rho = \bar{\rho}$$

Assumption 2: This assumption is the no anticipation effect. It can also be called an exclusion restriction where the treatment would be unknown a priori. It is written formally as:

$$Y_{ipcpre}(1) = Y_{ipcpre}(0)$$

The most common strategy to empirically show these assumptions is to present an event study plot of the treatment. Although insignificant pre-period coefficients do not prove that these assumptions hold for sure, as Roth (2022) discussed in detail, it is a suggested strategy by Lee and Wooldridge (2023) to present multiple coefficients over time instead of providing a single coefficient. The following section shows event-study plots to indicate that the assumptions above are likely to hold, and it also presents a single coefficient estimate of winning a new plant on high-school graduates' college decisions to understand the combined effect.

5 Results

We first investigate how a million-dollar plant opening impact the total number of vacancies posted in a county. Table 4 presents the coefficient estimates from a regression including the Winner variable with its leads and lags. The coefficient estimates for the years before the plant opening are small in magnitude and statistically insignificant. During the plant opening year and the preceding year, there is an increase in the total number of vacancies posted in the county. Then, the number of vacancies in winning counties returns to similar values to those in losing counties.

Table 5 reports coefficient estimates from the equation (1). Three outcomes (labor market, two-year college, and four-year college) are regressed on the pair fixed effects, year fixed effects, and a dummy variable indicating a county's million-dollar plant-winning status. Results are not sensitive to these fixed effects, as Appendix Table A2 compares different estimating equations.

Figures 1 and 2 show event study designs for two-year and four-year college enrollment differences between winning and losing counties. Both figures indicate that two-year and four-year college enrollment rates were similar between winning and runner-up counties before the plant. It provides credibility to the research design, indicating that these two types of counties were comparable before the opening of a million-dollar plant. Table 6 provides estimation results from an equation similar to (1) with interactions to capture the lagged effects of the event. These results are inline with the figures.

There is a reduction in college enrollment rates after a plant opening. Four-year college enrollment rates are lower two years after a plant opening, and two-year college enrollment rates are declined six years after the plant opening. Both results indicate that when a new plant is opened in a county, the college enrollment rates are declined by about two percentage points compared to a similar county without a new plant. These results align with the literature where a positive local labor market shock reduces college enrollment rates.

Table 7 presents a multinomial logistic regression with the three outcomes. The labor market outcome is the base outcome, so its coefficients are omitted. The results indicate that winning counties' two-year college enrollment rates declined after a plant opening. When we compare Table 7 to Table 5, there are slight differences. Table 5 indicates a new plant opening reduces four-year college enrollment more, but Table 7 shows two-year college enrollment declines more than the four-year one. Both results indicate a decline in college enrollment rates after a plant opening.

The subsequent results investigate the type of available jobs for each education group. Million-dollar plants differ widely in terms of the labor demanded. We analyze the labor demand at two-different levels: plant and county. The first one is on the plant level, where we investigate the share of vacancies for high-school or less education level, for individuals with two-year college credentials, and for those individuals with at least a bachelor's degree for each plant. The second part focuses on county-level labor demand, where we look at the same shares by education level in winning counties and runner-up counties for 2007 and the years between 2010 and 2021.

Table 8 presents the results with plant-level vacancy variables. The first column indicates that fewer high-school graduates are attracted to the labor market when the share of jobs available for two-year college credentials increases. Table 8 column 2 shows that when the percentage of jobs requiring two-year college credentials increases in a county, it increases returns to a two-year college. More students are enrolling in a two-year college after such a plant opening. We observe different patterns for four-year college enrollment rates, and it might indicate that returns to four-year colleges are independent of the type of available jobs.

Table 9 presents the same results from a multinomial logistic regression. It shows reverse patterns for two-year and four-year college enrollment rates, but Tables 8 and 9 both indicate heterogeneities by skill level for college enrollment rates. A higher share of jobs for two-year college credentials increases college enrollment rates.

Table 10 and 11 show the results at the county level using IPEDS data. All these variables are available for both winner and runner-up counties for 2007 and between 2010 and 2021. The variables also interact with share of jobs by education levels to observe extra labor demand created by the new plants. We use winner dummy variable, proxies the total number of vacancies posted in a plant, in Table 10, and the total number of posted vacancies in a county in Table 11.

Both results indicate that when the share of jobs (or the number of jobs) for two-year college graduates increases due to a new plant opening, college enrollments increase in a county. The same channel does not work when the number of jobs available for high-school graduates increases. Thus, it is crucial to distinguish whether the new openings are for low-skilled or high-skilled individuals. Surprisingly, the results indicate that more two-year college jobs increase enrollment in two-year colleges as well as four-year colleges. Four-year college channel might be related to the federal data classification of a four-year college. If a college is offering a four-year degree, it is classified as a four-year college. Some two-year colleges started to offer a small number of four-year degrees, so they might be driving these results.

Table 10 Winner coefficient is representing share of jobs available for four-year college graduates, as it is omitted due to perfect multicollinearity. Four-year college enrollment does not increase when the share of jobs available for four-year college graduates increases in a county. It could have many reasons related to the uncertainty of vacancies in the future, wages, and migration. Four-year college is a long process, and those available jobs might not be there in four years. Therefore, recent high-school graduates might not be enrolling in four-year colleges after observing an increase in the share of jobs available for four-year college graduates. Secondly, the number of available jobs might not affect the returns to four-year colleges; instead, returns might be affected by wages for four-year college graduates. The risk of unemployment might be a minor concern for four-year college graduates, so returns to four-year colleges might be a function of wages instead of available jobs. Finally, it has been documented by many studies, including Conzelmann et al. (2022), that four-year college alumni disperse widely compared to two-year college graduates. Thus, county-level labor market shock might not be the correct geographical unit for four-year college graduates, but it might be relevant for two-year college students.

6 Conclusion

This paper investigates the college decisions of high-school graduates in response to positive labor market shocks. The identification strategy is based on a magazine article titled "Million Dollar Plants" which compiled location rankings of firms for a million-dollar investment. In the end, there is one winner location of a million-dollar plant, and a few other areas have competed for the investment but could not get it by a small margin. Winner counties have many observable similarities to runner-up counties, including their population characteristics and college enrollment rates before the plant opening. This empirical design was first used by Greenstone, Hornbeck, and Moretti (2010), and many other studies, including Bloom et al. (2019) and Bhardwaj et al. (2022), have followed a similar identification strategy.

The utilization of rich information on the Lightcast Vacancy dataset has multiple advantages. It allows us to understand the type of labor market shocks at different locations. Each vacancy posting is suitable for an individual, so leveraging this information by education groups indicates a substantial variation in labor demand by each plant. Some plants mostly hire high-skilled individuals, and some others hire less-skilled individuals. Thus, returns to each college are differently affected by the type of each plant opening.

The paper's main results indicate that when the share of jobs suitable for two-year college individuals increases, more high-school graduates enroll in two-year colleges, indicating that these types of plants might be increasing returns to a two-year college degree. Thus, whether a new plant opening would reduce the college enrollment rates of recent high-school graduates depends on the nature of available jobs created by a new plant.

The study's main limitation is the sample which is obtained from the Lightcast Resume dataset. It is not representative of the U.S. population, but we perform robustness checks using many other datasets like CPS and IPEDS. CPS is a representative dataset, and we compare college enrollment trends with the Lightcast Resume sample. IPEDS provides information at the college level, and we utilize college enrollment numbers at the county level to double-check our results.

There are two main areas for future research. First, utilizing the employer name in resumes and merging them with the actual plant opening would be helpful in understanding the type of individuals working in new plants. Another area for future research is using wage data in the Lightcast vacancy dataset. The wage information is only available for a small fraction of the vacancy postings. However, it is worth investigating how wages affect returns to different colleges to complement this paper.

7 References

Atkin, David. "Endogenous skill acquisition and export manufacturing in Mexico." American Economic Review 106, no. 8 (2016): 2046-85.

Betts, Julian R., and Laurel L. McFarland. "Safe port in a storm: The impact of labor market conditions on community college enrollments." Journal of Human resources (1995): 741-765.

Black, Dan A., Terra G. McKinnish, and Seth G. Sanders. "Tight labor markets and the demand for education: Evidence from the coal boom and bust." ILR Review 59, no. 1 (2005): 3-16.

Bhardwaj, Abhishek, Devaki Ghose, Saptarshi Mukherjee, and Manpreet Singh. "Million Dollar Plants and Retail Prices." (2022).

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen. "What drives differences in management practices?." American Economic Review 109, no. 5 (2019): 1648-83.

Cascio, Elizabeth U., and Ayushi Narayan. "Who needs a fracking education? The educational response to low-skill-biased technological change." ILR Review 75, no. 1 (2022): 56-89.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. "Housing Booms and Busts, Labor Market Opportunities, and College Attendance." American Economic Review 108, no. 10 (2018): 2947-94.

Conzelmann, Johnathan G., Steven W. Hemelt, Brad Hershbein, Shawn M. Martin, Andrew Simon, and Kevin M. Stange. Grads on the go: Measuring college-specific labor markets for graduates. No. w30088. National Bureau of Economic Research, 2022.

Foote, Andrew, and Michel Grosz. "The effect of local labor market downturns on postsecondary enrollment and program choice." Education Finance and Policy 15, no. 4 (2020): 593-622.

Giroud, Xavier, Simone Lenzu, Quinn Maingi, and Holger Mueller. Propagation and amplification of local productivity spillovers. No. w29084. National Bureau of Economic Research, 2021.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings." Journal of Political Economy 118, no. 3 (2010): 536-598.

Hubbard, Daniel. "The impact of local labor market shocks on college choice: Evidence from plant closings in Michigan." In Unpublished manuscript. Retrieved from https://www. aeaweb. org/conference/2019/preliminary/paper/Q67dtN77. 2018.

Jensen, Robert. "Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India." The Quarterly Journal of Economics 127, no. 2 (2012): 753-792.

Juszkiewicz, Jolanta. "Trends in Community College Enrollment and Completion Data, Issue 6." American Association of Community Colleges (2020).

Kim, Hyunseob. "How does labor market size affect firm capital structure? Evidence from large plant openings." Journal of Financial Economics 138, no. 1 (2020): 277-294.

Lee, Soo Jeong, and Jeffrey Wooldridge. "A Simple Transformation Approach to Difference-in-Differences Estimation for Panel Data" Unpublished

Manski, Charles F., and David A. Wise. College choice in America. Harvard University Press, 1983.

Minaya, Veronica, Brendan Moore, and Judith Scott-Clayton. "The effect of job displacement on public college enrollment: Evidence from Ohio." Economics of Education Review 92 (2023): 102327.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. "Commuting, migration, and local employment elasticities." American Economic Review 108, no. 12 (2018): 3855-90.

Patrick, Carlianne, and Mark Partridge. "Identifying Agglomeration Spillovers: New Evidence from Large Plant Openings." Unpublished. https://www. dropbox. com/sh/w9ilyyom5te4hgf/AAB1

(2019).

Roth, Jonathan. "Pretest with caution: Event-study estimates after testing for parallel trends." American Economic Review: Insights 4, no. 3 (2022): 305-22.

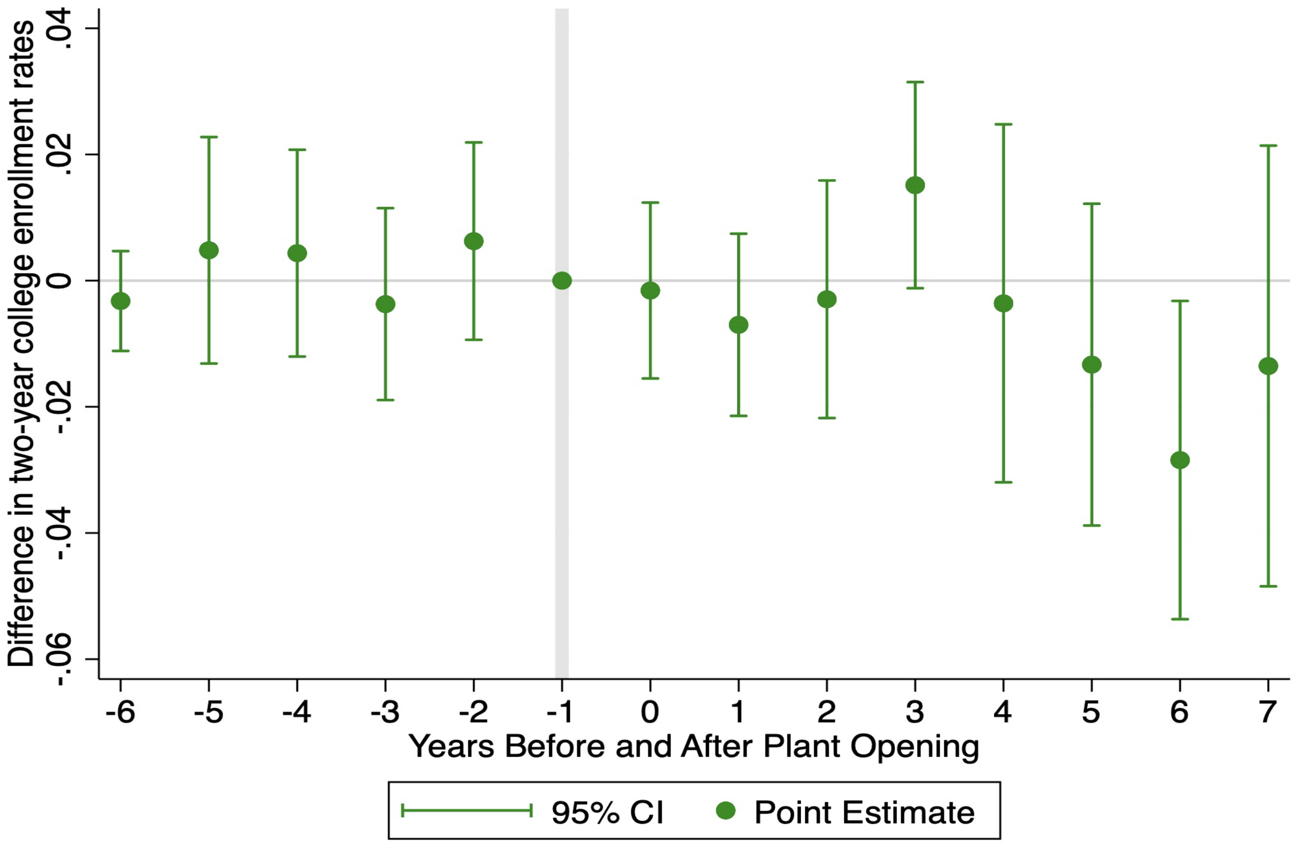
Schanzenbach, Diane Whitmore, Julia A. Turner, and Sarah Turner. "Raising State Minimum Wages, Lowering Community College Enrollment." National Bureau of Economic Research (NBER) Working Paper No. 31540, 2023.

Slattery, Cailin, and Owen Zidar. "Evaluating state and local business incentives." Journal of Economic Perspectives 34, no. 2 (2020): 90-118.

Qian, Franklin, and Rose Tan. The effects of high-skilled firm entry on incumbent residents. Stanford Institute for Economic Policy Research (SIEPR), 2021.

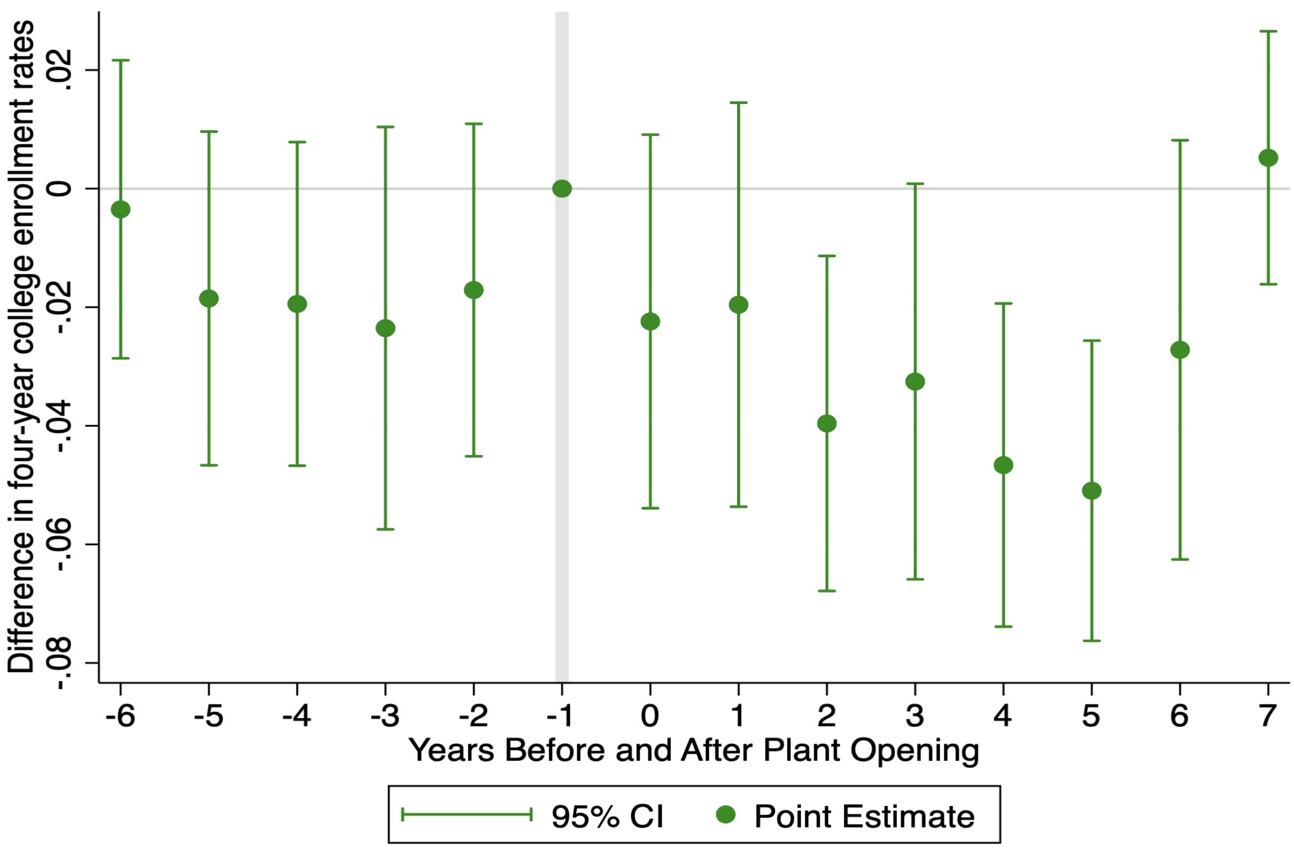
Figures and Tables

Figure 1. Difference in Two-Year College Enrollment Rates between Winner and Loser Counties



Notes: This figure shows the difference in the ratio of two-year college enrollment rates between the counties that a new plant has opened and those counties that were runner-ups, but could not get the plant. Time -1 is the base year, and time 0 is the year of plant opening.

Figure 2. Difference in Four-Year College Enrollment Rates between Winner and Loser Counties



Notes: This figure shows the difference in the ratio of four-year college enrollment rates between the counties that a new plant has opened and those counties that were runner-ups, but could not get the plant. Time -1 is the base year, and time 0 is the year of plant opening.

Table 1: Summary Statistics of Million Dollar Plants

Variable	Mean	SD	Min	Max
Panel A				
Size (million $)	505	679	19	3700
Size of employment	1,102	770	40	3300
Sector				
Manufacturing	0.49			
Non-Manufacturing	0.38			
Headquarters and Mix	0.13			
Type				
New	0.75			
Expansion	0.18			
Relocation	0.07			
Panel B				
Number of Posted Vacancies	1,549	2247	21	9,847
Share vacancies for high-school	0.33	0.21	0.07	0.81
Share of vacancies for two-year college	0.09	0.08	0	0.34
Share of vacancies for four-year college	0.58	0.22	0.08	0.92
N	40			

Notes: This table reports summary statistics for the Million Dollar Plants. Panel A variables are included in the original Bloom et al. (2019) data. Panel B information is obtained using the Lightcast Vacancy dataset where each of the forty plants are identified. Share of vacancies by education level are obtained by merging 6-digit occupation codes available in the Lightcast Vacancy data to the Occupation Information Network (O*NET) Education dataset. The O*NET Education dataset provides required level of education for each 6-digit occupation code.

Table 2: Summary Statistics of Million Dollar Plant Winning and Losing Counties

Variable	**Winning Counties** (1) Mean	**Losing Counties** (2) Mean	**Remaining Counties** (3) Mean	**p-Value Col.1 - Col.2** (4) Mean
Panel A				
Population (2000)	430,771	492,109	77,242	0.788
Labor Force Participation Rate (2000)	0.50	0.51	0.48	0.714
Unemployment Rate (2000)	3.83	4.02	4.57	0.524
Percentage of adults less than HS	0.20	0.19	0.23	0.363
Percentage of adults with HS diploma	0.29	0.29	0.35	0.982
Percentage of adults some college or AA	0.28	0.27	0.26	0.824
Percentage of adults with BA or higher	0.23	0.24	0.16	0.429
Panel B				
Total number of posted vacancies	602,086	709,714	-	0.709
Share vacancies for high-school	0.50	0.48	-	0.377
Share of vacancies for two-year college	0.10	0.10	-	0.455
Share of vacancies for four-year college	0.40	0.41	-	0.513
N	35	58	3,128	

Notes: This table reports summary statistics for the Million Dollar Plants Winning Counties, Losing Counties, and the Remaining US Counties. p-Value is from a t-test of the mean difference between the Winning and Losing Counties. All the variables in Panel A are as of the year 2000, before the Million Dollar Plants data have started. Panel A data is obtained from the U.S. Department of Agriculture's website (https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/). Panel B information is obtained using the Lightcast Vacancy dataset where each vacancy postings in these counties are included for the years 2007 and 2010 to 2021. Share of vacancies by education level are obtained by merging 6-digit occupation codes available in the Lightcast Vacancy data to the Occupation Information Network (O*NET) Education dataset. The O*NET Education dataset provides required level of education for each 6-digit occupation code.

Table 3: Summary Statistics of Lightcast Resume Dataset

	All Sample	**Winner Counties**		**Runner-up Counties**	
		Pre-MDP	Post-MDP	Pre-MDP	Post-MDP
	(1)	(2)	(3)	(4)	(5)
Variable	Mean	Mean	Mean	Mean	Mean
Highest level of schooling					
high-school	0.661	0.642	0.713	0.655	0.733
associate's degree or less	0.150	0.150	0.137	0.156	0.120
bachelor's degree or more	0.189	0.209	0.150	0.189	0.146
Total	1.000	1.000	1.000	1.000	1.000
Decision after high-school					
labor-market	0.920	0.907	0.973	0.912	0.978
two-year college	0.073	0.081	0.048	0.078	0.034
four-year college	0.106	0.124	0.066	0.108	0.063
N	81,281	23,363	5,256	46,084	6,578

Notes: This table reports summary statistics for the sample. The sample consists of individuals graduated from a high-school located in million dollar plants winning or runner-up counties. Column two and three are for those individuals graduated from a winner county high-school and column four and five are for runner-up county high-school graduates. Pre-MDP means before the plant opening and Post-MDP means after the plant openings. Runner-up counties did not experience a plant opening, so their Post-MDP is whenever their winning counterpart county wins the plant.

Table 4: The Effect of MDP on the Total Number of Vacancies

Panel A	
Dependent Variable :	$\text{TotalVacancies}_{pct}$
Event Year	
$\tau = -4$	-13242.19
	(16045.57)
$\tau = -3$	2249.562
	(5059.481)
$\tau = -2$	-5639.218
	(4443.592)
$\tau = -1$	725.6458
	(3200.809)
$\tau = 0$	10006.38**
	(4934.65)
$\tau = 1$	15798.24**
	(7583.614)
$\tau = 2$	-1356.919
	(4765.843)
$\tau = 3$	8774.928*
	(5261.622)
$\tau = 4$	3644.551
	(13761.78)
Pair Fixed Effect	✓
Year Fixed Effect	✓
Dependent Variable Mean	58,157
R^2	0.5138
N	1,053

Notes: This table presents coefficient estimates from a linear regression model with leads and lags of the Winner variable. Winner is a dummy variable taking the value one after a plant opening. Total-Vacancies is the total number of vacancies posted in a county. Event year 0 is the time of the million dollar plant opening. Standard errors are clustered at the county level and they are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: The Effect of MDP on Labor Market Entrance and College Enrollment

	Lightcast Resume Sample		
Panel A			
Dependent Variable :	$\text{LaborMarket}_{ipct}$	$\text{Two-Year College}_{ipct}$	$\text{Four-Year College}_{ipct}$
Winner	0.0106**	-0.0054	-0.0209***
	(0.0051)	(0.0059)	(0.0072)
Pair Fixed Effect	✓	✓	✓
Year Fixed Effect	✓	✓	✓
Dependent Variable Mean	0.9200	0.0730	0.1064
R-squared	0.0472	0.0162	0.0267
N	81,281	81,281	81,281

Notes: This table presents coefficient estimates from three separate linear probability models where the dependent variables are dummy variables representing decision after finishing high school. Winner is also a dummy variable taking the value 1 after a plant opened. Standard errors are clustered at the county level and they are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: The Effect of MDP on Labor Market Entrance and College Enrollment

	Lightcast Resume Sample		
Panel A			
Dependent Variable :	LaborMarket$_{ipct}$	Two-Year College$_{ipct}$	Four-Year College$_{ipct}$
Winner × 1.YearsSince	0.0121*	0.00210	-0.0225
	(0.00649)	(0.00959)	(0.0157)
Winner × 2.YearsSince	0.00208	-0.0146	-0.00815
	(0.0101)	(0.0107)	(0.0216)
Winner × 3.YearsSince	0.0183**	-0.00553	-0.0269**
	(0.00771)	(0.0126)	(0.0125)
Winner × 4.YearsSince	0.0101	0.00693	-0.0195
	(0.00982)	(0.0145)	(0.0239)
Winner × 5.YearsSince	0.0150	-0.00622	-0.0385**
	(0.0102)	(0.0171)	(0.0191)
Winner × 6.YearsSince	0.0110	-0.0106	-0.0454**
	(0.0141)	(0.0145)	(0.0220)
Winner × 7.YearsSince	0.00741	-0.0269**	-0.0161
	(0.00914)	(0.0132)	(0.0208)
Pair Fixed Effect	✓	✓	✓
Year Fixed Effect	✓	✓	✓
Dependent Variable Mean	0.9200	0.0730	0.1064
R-squared	0.0473	0.0163	0.0269
N	81,281	81,281	81,281

Notes: This table presents coefficient estimates from three separate linear probability models where the dependent variables are dummy variables representing decision after finishing high school. Winner is also a dummy variable taking the value 1 after a plant opened. YearsSince variable indicates how many years have passed after a plant opening. For example, 1.YearsSince is the year of a plant opening. Standard errors are clustered at the county level and they are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: The Effect of MDP on Labor Market Entrance and College Enrollment

	Lightcast Resume Sample	
Panel A		
Dependent Variable :	Two-Year College$_{ipct}$	Four-year College$_{ipct}$
Winner	-0.3613**	-0.1134
	(0.1633)	(0.1258)
dy/dx	-0.0115	-0.0040
Constant	-3.6010***	18.8498
	(1.1866)	(12430.54)
Pair Fixed Effect	✓	
Year Fixed Effect	✓	
Pseudo R-squared	0.0751	
N	81,281	

Notes: This table presents coefficient estimates from a multinomial logistic regression where there are three outcomes: labor market, two-year college, and four-year college. Labor market outcome is the base outcome, so the coefficients are omitted. dy/dx represents the marginal effect of a variable. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: The Effect of MDP on Labor Market and College Enrollment by Skill Level

	Lightcast Resume Sample		
Panel A			
Dependent Variable :	$\text{LaborMarket}_{ipct}$	$\text{Two-Year College}_{ipct}$	$\text{Four-Year College}_{ipct}$
Winner	0.0199***	-0.0154	-0.0255**
	(0.0063)	(0.0103)	(0.0129)
ShareHSJobs	-0.0071	0.0321	0.0132
	(0.0170)	(0.0309)	(0.0298)
ShareTYCJobs	-0.1157***	0.0530*	0.0291
	(0.0316)	(0.0316)	(0.0419)
Constant	0.0034	-0.0095**	1.0001***
	(0.0254)	(0.0048)	(0.0446)
Pair Fixed Effect	✓	✓	✓
Year Fixed Effect	✓	✓	✓
Dependent Variable Mean	0.9200	0.0730	0.1064
R^2	0.0473	0.0162	0.0267
N	81,281	81,281	81,281

Notes: This table presents coefficient estimates from three separate linear probability models where the dependent variables are dummy variables representing decision after finishing high school. Winner is also a dummy variable taking the value 1 after a plant opened. ShareHSJobs stands for the fraction of vacancies posted by a plant that are suitable for people with high-school or less level of education. ShareTYCJobs is the same for people with two-year college level of education. ShareFYCJobs is the omitted category because the three shares add up to one. Standard errors are clustered at the county level and they are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: The Effect of MDP on Labor Market and College Enrollment by Skill Level

	Lightcast Resume Sample	
Panel A		
Dependent Variable :	Two-Year College$_{ipct}$	Four-Year College$_{ipct}$
Winner	-0.8783***	-0.1419
	(0.3319)	(0.3489)
dy/dx	-0.0282	-0.0040
ShareHSJobs	2.8691*	-0.9544
	(1.5498)	(1.7725)
dy/dx	0.0952	-0.0468
ShareTYCJobs	0.3473	2.7858**
	(1.9034)	(1.1314)
dy/dx	0.0049	0.1168
Constant	-3.5454***	18.8548
	(1.1864)	(12325.17)
Pair Fixed Effect	✓	
Year Fixed Effect	✓	
Pseudo R-squared	0.0753	
N	81,281	

Notes: This table presents coefficient estimates from a multinomial logistic regression where there are three outcomes: labor market, two-year college, and four-year college. Labor market outcome is the base outcome, so the coefficients are omitted. dy/dx represents the marginal effect of a variable. Winner is also a dummy variable taking the value 1 after a plant opened. ShareHSJobs stands for the fraction of vacancies posted by a plant that are suitable for people with high-school or less level of education. ShareTYCJobs is the same for people with two-year college level of education. ShareFYCJobs is the omitted category because the three shares add up to one. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: The Effect of MDP on College Enrollment by Skill Level

	NCES IPEDS Data	
Panel A		
Dependent Variable :	$\text{TwoYearCollege}_{pct}$	$\text{FourYearCollege}_{pct}$
Winner	-9242.988	-6398.276
	(7620.198)	(4208.92)
ShareHSJobs	-6564.143***	-6285.706*
	(2522.436)	(3246.387)
ShareTYCJobs	-46034.07*	-46608.53**
	(27390.72)	(18906.29)
Winner × ShareHSJobs	4225.613	473.4853
	(7340.802)	(5588.335)
Winner × ShareTYCJobs	75227.33	62453.32**
	(46637.26)	(30726.17)
Constant	9652.319***	7805.02***
	(2971.518)	(2360.97)
Pair Fixed Effect	✓	✓
Year Fixed Effect	✓	✓
Dependent Variable Mean	2,458	3,097
R^2	0.4742	0.5653
N	1,053	1,053

Notes: This table presents coefficient estimates from two separate linear probability models. The dependent variables are dummy variables representing for decision after finishing high school. Winner is also a dummy variable taking the value 1 after a plant opened. ShareHSJobs stands for the fraction of vacancies posted by a plant that are suitable for people with high-school or less level of education. ShareTYCJobs is the same for people with two-year college level of education. ShareFYCJobs is the omitted category because the three shares add up to one. Standard errors are clustered at the county level and they are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: The Effect of MDP on College Enrollment by Skill Level

	NCES IPEDS Data	
Panel A		
Dependent Variable :	$TwoYearCollege_{pct}$	$FourYearCollege_{pct}$
TotalVacancies	0.142^{*}	0.0851^{***}
	(0.0775)	(0.0296)
ShareHSJobs	14878.3^{*}	7947.4^{**}
	(7759.1)	(3164.4)
ShareTYCJobs	-5409.7	-19032.3^{*}
	(11411.4)	(9909.8)
TotalVacancies × ShareHSJobs	-0.282^{*}	-0.190^{***}
	(0.158)	(0.0550)
TotalVacancies × ShareTYCJobs	0.295	0.359^{***}
	(0.223)	(0.116)
Constant	-5069.7	-2369.4
	(3780.0)	(1590.0)
Pair Fixed Effect	✓	✓
Year Fixed Effect	✓	✓
Dependent Variable Mean	2,458	3,097
R^2	0.746	0.855
N	1,053	1,053

Notes: This table presents coefficient estimates from two separate linear probability models. The dependent variables are dummy variables representing for decision after finishing high school. TotalVacancies is the total number of posted vacancies in winning and losing counties for 2007 and between 2010 and 2021. ShareHSJobs stands for the fraction of vacancies posted by a plant that are suitable for people with high-school or less level of education. ShareTYCJobs is the same for people with two-year college level of education. ShareFYCJobs is the omitted category because the three shares add up to one. Standard errors are clustered at the county level and they are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: The Lagged Effect of MDP on College Enrollment by Skill Level

	NCES IPEDS Data	
Panel A		
Dependent Variable :	$TwoYearCollege_{pct}$	$FourYearCollege_{pct}$
Winner	-9340.6	-5173.5
	(7434.9)	(3802.0)
ShareHSJobs	-8003.6**	-8159.8**
	(3047.6)	(3603.3)
ShareTYCJobs	-37935.2	-36049.0*
	(28411.7)	(19036.7)
Winner × ShareHSJobs	5523.1	373.2
	(5224.0)	(4010.8)
Winner × ShareTYCJobs	64004.3	46124.0
	(56438.2)	(34205.4)
Winner × ShareHSJobs × 1.YearsSince	-8441.0	-1601.2
	(8194.5)	(6267.6)
Winner × ShareHSJobs × 2.YearsSince	4855.3	8879.4
	(12520.6)	(9385.3)
Winner × ShareHSJobs × 3.YearsSince	-2873.8	2828.9
	(6645.2)	(9314.6)
Winner × ShareHSJobs × 4.YearsSince	3632.3	-302.0
	(17366.2)	(13749.9)
Winner × ShareHSJobs × 5.YearsSince	-13665.5**	-3106.3
	(5536.8)	(8227.6)
Winner × ShareHSJobs × 6.YearsSince	-13983.7	-10432.4
	(10497.1)	(13447.9)

Winner × ShareTYCJobs × 1.YearsSince	121067.4	86335.4
	(76899.7)	(57223.7)
Winner × ShareTYCJobs × 2.YearsSince	77361.3	65749.2
	(53139.5)	(43764.3)
Winner × ShareTYCJobs × 3.YearsSince	119937.5*	140798.8**
	(68980.2)	(56922.0)
Winner × ShareTYCJobs × 4.YearsSince	-41901.2	22011.0
	(99716.5)	(58532.6)
Winner × ShareTYCJobs × 5.YearsSince	-14517.6	65614.9*
	(51542.6)	(36822.7)
Winner × ShareTYCJobs × 6.YearsSince	-69025.3	10967.8
	(66545.9)	(49462.6)
Constant	9511.0***	7490.2***
	(3272.8)	(2431.3)
Pair Fixed Effect	✓	✓
Year Fixed Effect	✓	✓
Dependent Variable Mean	2,458	3,097
R^2	0.410	0.519
N	1,053	1,053

Notes: This table presents coefficient estimates from two separate linear probability models. The dependent variables are dummy variables representing for decision after finishing high school. Winner is also a dummy variable taking the value 1 after a plant opened. ShareHSJobs stands for the fraction of vacancies posted by a plant that are suitable for people with high-school or less level of education. ShareTYCJobs is the same for people with two-year college level of education. ShareFYCJobs is the omitted category because the three shares add up to one. YearsSince variable indicates how many years have passed after a plant opening. For example, 1.YearsSince is the next year after a plant opened. Standard errors are clustered at the county level and they are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8 Appendix

8.1 Sample Creating Criteria

This section details the construction of the sample. The following groups have been removed to create the analysis sample.

The first criterion is that individuals with high-school information on their resumes are used. A little more than 10% of people have this information on their resume.

The second main criterion is that individuals with more than a two-year gap between high-school graduation year and the next step decision are dropped from the sample. For example, if an individual graduated from a high school in 2010, then the gap in the resume should not be wide enough so that the next information in the resume starts in 2013. This criterion does not necessarily create an important limitation, but it is needed to analyze the decision right after graduation.

Finally, if an individual high-school graduation year or high-school location is not included in the resume, they are also removed from the sample.

All the remaining individuals who graduated from either a winner or a runner-up county high school have been included in the analysis sample.

8.2 Million Dollar Plants Data

Bloom et al. (2019) raw data has 132 pairs. 27 of the pairs do not include the location information for the runner-ups. 33 pairs have decided to built before 2005. Since the vacancy data starts at the year 2007. It is not possible to capture early plant vacancies. Of the remaining 72 pairs, 32 of the plants do not have any vacancy postings or they are in a county that was a winner and runner-up at the same time at different years. 40 pairs remain in the analysis sample, but pairs do not have to consist on two counties. Some pairs consist of three or four counties where one of them is a winner, and the others are runner-ups for the million dollar plant.

There are a total of 93 counties. 35 of them are winner counties, and 58 of them are loser counties.

8.3 O*NET Required Level of Education

O*NET categorizes required level of education into the following twelve categories: less than a high-school diploma, high-school diploma or equivalent, post-secondary certificate, some college courses, associate's degree, bachelor's degree, post-baccalaureate certificate, master's degree, post-master's certificate, first professional degree, doctoral degree, post-doctoral training. We reduce this classification into three groups as: jobs requiring high-school or less level of education, requiring two-year college credentials, requiring four-year college credentials. O*NET's first two category belongs to our first category, O*NET's next three category is included into our second classification, and the remaining ones are included into four-year college credentials category.

The O*NET dataset is provided in a distribution. Each occupation has a distribution of required level of education. We take the biggest mass for each occupation and make it the required level of education for that one. The results are very similar if we take the mean of the distribution. The biggest mass is preferred because an occupation requires an interpretable level of education. For example, captains, mates, and pilots of water vessels require a high-school diploma or equivalent level of education if we take the biggest mass of the distribution. However, if we take the mean of the distribution, this occupation requires an education level between a high-school diploma and a post-secondary certificate, which is hard to interpret.

8.4 Different Specifications

Table A1 compares the results from the Lightcast Resume sample and the CPS ASEC. It is impossible to differentiate the two-year and four-year college separately using the CPS ASEC sample, so only college enrollment vs. labor market entrance after high-school graduation are compared. The results indicate similar coefficient estimates between these two different datasets. Coefficient estimates from the CPS ASEC sample are insignificant and it might be due to the smaller sample size.

Table A2 presents the empirical specification with pair, year, state, and county fixed effects. Four different specifications are presented and all of them are similar to one another.

Table A3 drops earlier cohorts to show that none of the results are dependent on them. There are not many individuals with high-school graduation year before 2000, so they do not affect the results.

Table A1: The Effect of MDP on Labor Market Entrance and College Enrollment

	Lightcast Resume Sample	CPS ASEC Sample
	(1)	(2)
Panel A		
Dependent Variable :	$\text{LaborMarket}_{ifpct}$	
Winner	0.0106**	0.0107
	(0.0051)	(0.0142)
Constant	0.0038	0.6531***
	(0.0253)	(0.0348)
R-squared	0.0472	0.0304
Dep. Vrb. Mean	0.92	0.47
Panel B		
Dependent Variable :	College_{ifpct}	
Winner	-0.0235***	-0.0207
	(0.0057)	(0.0141)
Constant	0.9900***	0.3302***
	(0.0478)	(0.0347)
R-squared	0.0358	0.0301
Dep. Vrb. Mean	0.17	0.45
Pair Fixed Effect	✓	✓
Year Fixed Effect	✓	✓
N	81,281	23,228

Notes: This table presents coefficient estimates from a linear probability model where the dependent variables are dummy variables representing college decision after high-school graduation. Winner is also a dummy variable taking the value 1 after a plant opened. The sample of the first column is from the Lightcast Resume sample and the sample of the second column is from the Current Population Survey Annual Social and Economic Supplement. The sample of the second column consist of college age population group, ages 18 to 21. Standard errors are clustered at the county level and they are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: The Effect of MDP on Labor Market Entrance and College Enrollment

		Lightcast Resume Sample		
	(1)	(2)	(3)	(4)
Panel A				
Dependent Variable :		$\text{LaborMarket}_{ifpct}$		
Winner	0.0106**	0.0084	0.0056	0.0056
	(0.0051)	(0.0054)	(0.0054)	(0.0054)
Constant	0.0038	0.1463***	0.0741***	0.1654***
	(0.0253)	(0.0147)	(0.0007)	(0.0038)
R-squared	0.0472	0.0489	0.0497	0.0497
Dep. Vrb. Mean		0.9200		
Panel B				
Dependent Variable :		$\text{Two-Year College}_{ifpct}$		
Winner	-0.0054	-0.0034	-0.0077	-0.0077
	(0.0059)	(0.0065)	(0.0064)	(0.0064)
Constant	-0.0098**	-0.0178***	-0.0226***	-0.0627***
	(0.0046)	(0.0178)	(0.0006)	(0.0019)
R-squared	0.0162	0.0168	0.0176	0.0176
Dep. Vrb. Mean		0.0730		
Panel C				
Dependent Variable :		$\text{Four-Year College}_{ifpct}$		
Winner	-0.0209***	-0.0219***	-0.0144*	-0.0144*
	(0.0072)	(0.0075)	(0.0076)	(0.0076)
Constant	1.0000***	0.8226***	0.8714***	0.7806***
	(0.0445)	(0.0237)	(0.0006)	(0.0024)
R-squared	0.0267	0.0304	0.0328	0.0376
Dep. Vrb. Mean		0.1064		
Pair Fixed Effect	✓	✓	✓	✓
Year Fixed Effect	✓	✓	✓	✓
State Fixed Effect	X	✓	X	✓
County Fixed Effect	X	X	✓	✓
N		81,281		

Notes: This table presents coefficient estimates from three separate linear probability models where the dependent variables are dummy variables representing decision after finishing high school. Winner is also a dummy variable taking the value 1 after a plant opened. Standard errors are clustered at the county level and they are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: The Effect of MDP on Labor Market Entrance and College Enrollment

	Full Resume Sample	Resume Sample excluding early cohorts
	(1)	(2)
Panel A		
Dependent Variable :		$\text{LaborMarket}_{ifpct}$
Winner	0.0106**	0.0077
	(0.0051)	(0.0050)
Constant	0.0038	0.8700***
	(0.0253)	(0.0213)
R-squared	0.0472	0.0205
Dep. Vrb. Mean	0.92	0.94
Panel B		
Dependent Variable :		College_{ifpct}
Winner	-0.0235***	-0.0217***
	(0.0057)	(0.0060)
Constant	0.9900***	0.1908***
	(0.0478)	(0.0311)
R-squared	0.0358	0.0333
Dep. Vrb. Mean	0.17	0.16
Pair Fixed Effect	✓	✓
Year Fixed Effect	✓	✓
N	81,281	70,372

Notes: This table presents coefficient estimates from a linear probability model where the dependent variables are dummy variables representing college decision after high-school graduation. Winner is also a dummy variable taking the value 1 after a plant opened. The sample of the first column is from the entire Lightcast Resume sample and the sample of the second column excludes individuals who have graduated from a high-school before 2000. Standard errors are clustered at the county level and they are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$