College Student Elasticity: The Impact of Job Openings on College Major Decisions

Jodie Bhattacharya*
Department of Economics
Stanford University
jodieab@stanford.edu

Under the direction of
Prof. Luigi Pistaferri

May 3, 2022

Abstract

Over time, there has been a shift in the demand for different college majors and skills, particularly engineering and science majors. Understanding how college students decide their major is important for future employment outcomes. This paper studies the effect of shifts in job opening availability on college major choice. I use a novel job openings dataset and a record of degrees conferred in the United States from 2007 to 2019. I use lags and fixed effects to estimate exposure to change in job openings by region and identify the effect on college major choice. I find that a 25 percentage point increase in job openings for a given industry leads to a 1 percentage point increase in the fraction of degrees in that industry. The results are similar for public colleges and community colleges. Students who attend small colleges and majority undergraduate colleges are more responsive to changes in job openings. In addition, students in the fields of business and humanities are more responsive to changes in job openings than other industries.

Keywords: higher education, field of study, labor market information, vacancy data, expected demand, postsecondary school, mismatch employment, human capital investments

***Acknowledgements**: I am grateful to my advisor, Prof. Luigi Pistaferri, for his mentorship and guidance on this project. I am thankful to Prof. Frank Wolak and Trevor Davis at the Program on Energy and Sustainable Development for allowing me to use their servers. I would also like to thank LinkUp for generously providing their job openings data. Finally, I would like to thank Marcelo Clerici-Arias, Cauê Dobbin, Nina Buchmann, and Andrés Rodríguez for their support. All mistakes are my own.

1 Introduction

This paper seeks to understand how responsive college students are to shifts in job openings and whether they respond by changing their college major. Many studies have explored other determinants for college major choice, such as wages, expected enjoyment, and perceived ability. However, there have been no studies on how job openings, an important labor market outcome, interact with college major choice. Job openings also serve as a measure of the status of the labor market, demonstrating how sensitive college students are to national and local trends.

Understanding how students make this decision is important since major choice impacts future earnings and employment. According to Rosenbaum, Deil-Amen and Person (2006), students feel anxious and confused about how to best choose a major, and most colleges offer little to no guidance about how to choose. Beffy, Fougère and Maurel (2012), Wiswall and Zafar (2015), and Stinebrickner and Stinebrickner (2014) find that expected salary plays a small, significant role in college major decision. Arguably, job openings are an even more important indicator of labor market outcomes, because the expected salary does not matter if there are no jobs available in that field. I am interested in how aware students are of the current job market situation and how that affects their major decisions. 4-year university students are only able to observe the labor market when they are making their major decision, roughly 3 years before they graduate, leading to a lag between job openings and college major choices. Further, different commuting zones experience differing shifts in job openings over time, which, when compared to the national status of the labor market, acts as a relatively exogenous source of change. Combining these two factors serves as a mechanism to identify a causal effect between job openings and college major choice.

This allows us to gain more insight into what students value as they choose their major. We are able to see what types of students are more responsive to the labor market and what types of fields are more responsive. If students are responsive (or, elastic), in some ways, this is considered "rational," since students are practical about what is available to them. Responsiveness to the labor market leads to less unemployment since there is less geographical mismatch between what students are majoring in and the jobs that are available. However, there are still many other reasons to choose a major, such as talent and enjoyment, which may still make the decision "rational." Regardless, this information provides valuable insight into modeling the frictions on the demand and supply side for labor.

To answer this question, I use data from U.S. job-search engine LinkUp, which aggregates job listings from employer websites. Each job posting is matched to at least one college major with relevant skills using job codes from the U.S. Department of Labor. In order to quantify the number of degrees conferred by major over time, I use data from the National Center for Education Statistics (NCES), specifically the Integrated Postsecondary Education Data System (IPEDS). I run fixed-effects regressions with lags in order to quantify students' responsiveness to changes in job openings by industry.

The rest of the paper is organized as follows. In section 2, I provide background and literature review. In section 3, I describe the data. Section 4 explains the empirical strategy and section 5 presents the results. Section 6 concludes.

2 Literature Review

My research builds on prior research analyzing college students' major choices by introducing online job vacancy data, a novel data source with relatively little analysis done on it. While other studies have used job vacancy data in the past, none have made use of the job postings in order to learn about the state of the labor market for college students — specifically, I use LinkUp data to match job posting descriptions to college majors in order to measure job market opportunities for college students at the time they are making their major decision.

I add to the existing literature on college students' major choices by examining the extent to which job availability affects student decisions, a factor that has not been evaluated before. In order to identify the factors that students use to choose a major, most economists use human capital theory, which suggests that students will make educational decisions based on a calculation of the perceived costs relative to the perceived discounted benefits (Becker, 1962). A few factors that have been evaluated are expected enjoyment (Beffy, Fougère and Maurel, 2012; Wiswall and Zafar, 2015; Stinebrickner and Stinebrickner, 2014), a student's perception of his/her ability in a given field (Arcidiacono, Hotz and Kang, 2012; Stinebrickner and Stinebrickner, 2014), and expected labor market outcomes such as salary (Wiswall and Zafar, 2015). However, there are many studies that have found that students do not have access to accurate information about estimated labor market returns. Most of these college major choice papers survey students in order to find what they value, what their expectations about various majors are, and see how telling students the real values of salaries for various majors affects their decision. Rather than running a survey, I examine students' actual major decisions, which is a much larger sample of data and fully captures all college students in the United States. This introduces more noise for what factors go into the major decision, but still quantifies how much students collectively are unaware of and do not care about labor market trends. It allows for more aggregate trend analysis, similar to Beffy, Fougère and Maurel (2012), where they examine the French post-secondary education system in order to find the effect of expected labor market income on choice of post-secondary field of study. Further, I investigate geographical mismatch between college majors and jobs available in regional areas, since major choice may solve short term shortages in regional labor markets.

2.1 College Major Choice

There have been many studies done on factors that affect college major choice. Beffy, Fougère and Maurel (2012) examine the French post-secondary education system in order to find the effect of expected labor market income on choice of post-secondary field of study (major). They use data on over 26,000 French individuals who left the French educational system in 1992 and were interviewed 5 years later. The dataset has information on educational and labor market histories. They divide postsecondary studies into three broad majors: (1) sciences, (2) humanities and social sciences (including art studies), and (3) law, economics, and management. Then, they use a multinomial probit model for the choice of major. As a result, they are able to measure the elasticity of choice of major with respect to expected earnings by taking advantage of variations in wage returns. Over time, they control for the change in the distribution of educational levels and inflation, and they also create a model to simulate a 10% variation in expected earnings. I am able to build on these models to see how well job openings predict college majors.

Baker et al. (2018) studies whether community college students are aware of the labor market outcomes for different majors. In this paper, the authors administered a survey to community college students in the Bay Area. They find that students believe salaries are 13% higher than they are and underestimate the probability of being employed by 25%. They also find that a 10% increase in salary is associated with a 14-18% increase in the probability of choosing a specific category of majors. They construct an error metric based on what % higher or lower they think salary is than it actually is for a given major. They then use regression to examine student errors: asking first if certain groups of majors are associated with larger or smaller errors, and asking second if certain groups of students have particularly large errors. Finally, they run an experiment by presenting fictional labor market outcomes to students, and then asked students to write down their beliefs about the probability of being employed and choosing that major. Since they find that students are relatively unaware of salaries associated with various majors, students may also be unaware of job opening availability. However, job openings may also be indicative of other trends in fields of study, so even if students are unaware of the exact quantity of jobs available, they may have a sense of this anyways. My study quantifies how aware students are of this labor market outcome, and how much a shock to job openings affects student decisions, at a more macro level.

Zafar (2013) uses data from Northwestern University sophomores and their subjective expectations about outcomes from various college major choices. It focuses on the gender gap in how college majors are chosen. For both genders, enjoying the coursework and gaining parents' approval are the most important determinants. Zafar finds that the gender gap is mainly due to gender differences in preferences and tastes. This provides another helpful theoretical model for predicting major choices. Zafar creates a utility model where an individual derives utility from choosing a college major based on factors from (1) outcomes realized in college: successfully graduating in 4 years, graduating with a GPA of at least 3.5 in the field of study, enjoying the coursework, hours per week spent on the coursework, and parents approve the major; and (2) outcomes realized after graduating college: getting an acceptable job immediately after graduation, enjoying working at jobs available after graduation, able to reconcile work and family in jobs, hours spent per week working in available jobs, social status of available jobs, and income of available jobs. Each outcome is chosen based on papers that have found that these are important indicators for the utility function. Based on this model, Zafar estimates the change in the utility from the occurrence of an outcome (based on characteristics). He then uses a choice model to estimate these parameters. Interestingly, he finds that income is insignificant in the decision, which he hypothesizes is because of the large variation in subjective responses (expectation of earnings). He also finds that females value the nonpecuniary aspects of the job (reconciling work and family and enjoying work) relatively more than males. My paper does not capture differences between males and females, but it does capture other aspects of heterogeneity, such as differences between different industries and college types, and finds how much the status of the labor market affects choice.

Weinstein (2017) explores whether students respond to sector-specific shocks by changing their major. In particular, he looks at the dot-com crash, the 2008 financial crisis, and a shock that transformed Delaware into an international finance center. This paper differs from mine because the data on shocks comes from looking at the Quarterly Census of Employment and Wages, rather than job openings data. Weinstein finds that universities in areas more exposed to sectoral shocks experience greater changes in sector-relevant majors. By comparing geographically mobile to immobile students, he demonstrates that this effect is driven by migration frictions (i.e. geographical mismatch between college students' majors and job openings) and not information frictions (i.e. college students being unaware of labor market demands). As a result of Weinstein's paper, there is promising evidence that college students may also respond to changes in job openings, but the effect may be much smaller than when a sector-specific shock occurs. I am able to quantify this in my paper. This paper demonstrates that the effect is not driven by information frictions, which is helpful for the assumptions made in my paper – that students are aware of their local labor markets.

There are numerous other papers that examine factors that affect college major choice that are similar but not the same as mine, such as: the effect of local shocks on high school completion and college enrollment (Cascio and Narayan, 2015; Charles, Hurst and Notowidigdo, 2015), reallocation of college majors in response to unemployment rates (Blom, Cadena and Keys, 2021), and college major responses to changes in occupation-specific wages (Long, Goldhaber and Huntington-Klein, 2015).

2.2 Online Job Vacancy Data

Online job vacancy data has been used in the past, to offer alternate perspectives of the labor market. Azar et al. (2020) uses data on online job vacancies from Burning Glass Technologies (BGT) to calculate labor market concentration using the Herfindahl-Hirschman index (HHI) for each commuting zone by 6-digit SOC occupation. In order to validate their data, they check that BGT data is fairly similar in terms of industry composition when compared to all vacancies recorded in the Job Openings and Labor Turnover Survey (JOLTS), which is nationally representative of employers. The job advertisements that are not online now are usually in small businesses (the classic example being the "help wanted" sign in the restaurant window) and union hiring halls. Overall, however, research shows the online job market has consistently expanded over the last few years. They also explore what drives vacancy-posting as an economic behavior, and additionally, whether the concentration of vacancy-posting is a good metric for the underlying economic characteristic they seek to measure: the market power of employers in a labor market.

In order to investigate job availability, I seek to measure employee demand for college graduates from various majors, so online job vacancy data is likely the best estimate of what opportunities are available to college graduates. Further, if college students are aware of the status of the labor market, it would be through their knowledge of job postings and friends getting hired. College students are more likely to use the Internet than older age groups, and more likely to use Internet advertising rather than to search for mom-and-pop stores or union jobs. According to a survey of 2,500 undergraduates by College Pulse, 88% of students say they are most likely to visit the company website when they're interested in learning more about a company or are considering applying for a job, and 48% use LinkedIn to search for job opportunities (Pulse, 2020). As a result, using online job vacancy data may be a slight underestimate of overall job opportunities available. However, most job postings that are not available online are typically not marketed for recent college graduates, and would be evenly spread across industries. Thus, job vacancy data is likely an unbiased way to measure the labor market that college students are aware of.

2.3 Impact of College Major

I am also interested in observing the mismatch between major choice and job openings, which serves to quantify the tangible impact of college major decisions. Other papers have investigated labor market mismatch, geographically and through workers searching for jobs in the wrong fields. However, there has been no research done on the mismatch between skills attained in college and jobs available. Marinescu and Rathelot (2018) investigate geographical mismatch using data at the job seeker level from CareerBuilder.com. Using Poisson regression, they estimate the probability that a job seeker in a given zip code applies to a job in a different zip code, as a function of the distance between the zip codes. They find that job seekers are 35% less likely to apply to a job 10 miles away from their zip code of residence. In the paper, they create a search and matching model which predicts that relocating job seekers to minimize unemployment would only decrease unemployment by 5.3%, so they conclude that geographic mismatch is a minor driver of aggregate unemployment. My research also examines mismatch for college graduates at the geographical level.

Ayşegül Şahin, Joseph Song, Giorgio Topa and Giovanni L. Violante (2012) seek to quantify mismatch unemployment, or an unemployed person searching in the "wrong" sector. They develop a theoretical framework where there are a large number of distinct labor markets, and each one is frictional. Then, they compare the actual allocation of unemployed workers across sectors to an ideal allocation. Using the Help Wanted Online database, they perform their analysis at the 2 and 3 digits occupational level, state and county level, and define labor markets as a combination of occupation and location. They find that mismatch unemployment accounts for 0.75 out of 5.4 percent of US unemployment. They also analyze mismatch within education groups. They find that the contribution of occupational mismatch to the rise in unemployment between 2006 and 2010 grows as you move from the highest to lowest education category. In other words, for college graduates, mismatch explains 24% of the rise in unemployment. The implication is that college graduates may be looking in the wrong sectors for work. This problem may be compounded by college graduates majoring in a field where there are no jobs, an issue that I quantify.

3 Data

In order to construct the data for analysis, I combine data on job openings from LinkUp, data on degrees conferred from IPEDS, and multiple other data sources to make the data panel at the commuting zone and industry level. In the section that follows, I describe in detail the choices I make in merging and cleaning the data, and provide validation that the data is representative of the metrics I am interested in. In addition, I perform Granger Causality tests as a means of exploratory data analysis.

3.1 Job openings data

I use proprietary data from LinkUp, which provides a historical database of job openings from 2007 to 2021 (LinkUp, 2007-2019). LinkUp, a leading provider of job market data and analytics, is the only job search engine that indexes jobs exclusively from company websites. As a result, the jobs in their database have no duplicate listings, expired jobs, or fraudulent jobs. LinkUp scrapes the web daily for job postings, and they have a database with job openings from over 50,000 employers. LinkUp scrapes 100% of publicly traded U.S. companies that have jobs on their website. Since around 15% of public companies currently don't post jobs on their website, this means the LinkUp database covers 85% of publicly traded companies. About 75% of companies in the LinkUp database are private. For the purposes of replication, interested researchers can obtain LinkUp's data by contacting LinkUp.[1]

An alternative source of job openings data comes from the U.S. Bureau of Labor Statistics. The survey is called the Job Openings and Labor Turnover Survey (JOLTS), collected from sampled establishments on a voluntary basis. The main reason that I chose to use LinkUp data instead is that JOLTS provides aggregate data by industry. In contrast, LinkUp data contains individual job postings data, which can be matched more closely to college majors. For example, computer scientists are required in most industries, from finance to manu-

[1] `https://www.linkup.com/contact-us/`

facturing. As a result, job openings by industry are hard to map back to college majors. Further, due to the more granular nature of the LinkUp data, we also have location-level data, allowing us to investigate mismatch in local labor markets, rather than at the country aggregate level.

LinkUp's data has been used by several other researchers for academic papers, providing validation that the data is representative of the market for job openings. Zhang, Yu and Hu (2011) studies the correlation between the LinkUp engine and the Bureau of Labor Statistics Job Reports. They use numbers for total U.S. employment from the nonfarm payroll report (NFP). Notably, they find a positive correlation of about 75% between the LinkUp Index (number of jobs per month divided by the number of companies per month) and 1-month lagged NFP. They also try Help Wanted Online (HWOL), the largest dataset of job listings available online. However, even with a time lag, they find no correlation between HWOL and NFP, indicating that HWOL has much lower quality data due to duplicate and fake/garbage listings. As a result, LinkUp has meaningful predictive attributes for the NFP report, and it is far more predictive than HWOL.

Campello, Kankanhalli and Muthukrishnan (2020) find that the total number of job postings in LinkUp consistently captures around 50% of the total private sector hires in JOLTS. Jordan (2018) compares the LinkUp job duration with the DHI-DHF vacancy duration measure, finding that the LinkUp data is more volatile, but still provides a very close match. They also compare the share of vacancies by occupational cluster between JOLTS and LinkUp, finding similar magnitudes overall, with differences that can be explained by differences in compositions or labelings of occupation types.

For each job posting, I have the job title, company name, city, state, ZIP code, created and deleted date, and O*NET occupation code. O*NET codes are a classification developed by the U.S. Department of Labor to describe occupations in terms of the knowledge, skills, and abilities required. LinkUp assigns each job posting an O*NET code using natural language processing.

In order to clean the data, first, I restrict the sample to U.S. job openings. Since we are interested in entry-level jobs available to college students, I subset the data based on the Job Zone from the O*Net database (U.S. Department of Labor and Administration, 2016). For students graduating from 2-year institutions, I subset to Job Zone Three or "Medium Preparation Needed": "Most occupations in this zone require training in vocational schools, related on-the-job experience, or an associate's degree." This dataset has 27,496,113 observations, where each observation is a job posting.

For students graduating from 4-year institutions, I subset to Job Zone Four or "Considerable Preparation Needed": "Most of these occupations require a four-year bachelor's degree, but some do not" (U.S. Department of Labor and Administration, 2016). This method is likely not perfectly accurate at capturing entry-level jobs but is more reliable than scraping degree requirements from the text of the job posting. This dataset has 29,854,276 observations.

Data Validation Since I am interested in using the LinkUp data in order to estimate trends in job openings by industry over time, it is important to investigate how representative this variable is in relation to true domestic dynamics. Specifically, I want the trends by industry to be representative and ensure that the LinkUp algorithm for scraping job postings is not scraping from one sector more than another. By looking at the share of job openings by industry, I eliminate the issue with LinkUp's job openings consistently growing over time due to adding more companies to the index. We can also note that we are interested in whether the distributions between sources are the same, not whether the levels are the same. The empirical strategy relies on changes in job openings over time, not absolute numbers of job openings.

I compare LinkUp with a reliable source of estimates on job openings from JOLTS. The LinkUp data comes with O*NET codes, which can be aggregated to O*NET Career Clusters and loosely mapped to JOLTS Industry categorizations. The O*NET cluster *Education* *and training* maps to *Educational services*, *Finance* maps to *Finance and insurance*, *Health science* maps to *Health care and social assistance,* and *Hospitality and tourism* maps to *Accommodation and food services.* I select these industries because these are the JOLT industry categories that are most relevant to college graduates.

From Figure 1, we see that shares from both data sources have similar magnitudes and trends over time. In order to measure this quantitatively, I use the distance correlation metric ($dCor$) to measure the similarity between the two time series. More information on calculating this metric is available in Székely, Rizzo and Bakirov (2007). I choose this metric because it generalizes classical linear correlation, and $dCor = 0$ implies true independence. The distance correlation statistics for education, finance, health services, and hospitality are 0.209, 0.200, 0.343, and 0.504, respectively. In all cases, performing a nonparametric t-test of independence yields $p < 0.02$, indicating that the time series are dependent and significantly correlated.

Qualitatively, education has a slightly larger share in the LinkUp dataset but follows the same general trend over time. Finance and Health Services appear to decrease more in the LinkUp dataset, but otherwise, follow trends fairly well. The hospitality group has a differing fluctuation from the LinkUp dataset in 2010. Definitions of the industries can likely explain any differences between the LinkUp and JOLTS data. For example, hospitality and tourism jobs may differ from JOLTS' definition of accommodation and food services. Regardless, the shares seem to align well enough to support the assumption that the LinkUp data has representative sampling by industry. Non-online job postings may explain further differences that LinkUp is unable to capture. Especially in the hospitality industry, some job postings may not be done online. I am using job openings to measure what college students are aware of for their labor market opportunities, so an online sample may be even more representative of this, since college students may not see non-online job postings. Still, the LinkUp data acts as an imperfect proxy for measuring job openings over time. I am primarily interested in what college students are aware of. They may not be aware of job openings at all, so all

Figure 1: Comparing JOLTS and LinkUp Trends

we can measure is how responsive they are to job openings measured in the LinkUp data.

LinkUp Index Since LinkUp is constantly adding more companies to scrape the web for and index, their job openings counts are continually growing over time. To account for this, I introduce the LinkUp Index, a statistic used in many other papers such as Zhang, Yu and Hu (2011).

$$\text{LinkUp Index} = \frac{\#\ \text{active job openings}}{\#\ \text{companies}} \tag{1}$$

By accounting for the number of companies in the dataset, I now have a more accurate measure of the status of the labor market, with the relative number of job openings available per company.

One issue with this strategy is that I don't account for the fact that some companies are smaller than others and would necessarily have fewer jobs available. This metric also doesn't capture the creation of new companies over time. However, introducing this metric can serve as a robustness check on the original outcome of "Share of job openings by industry."

3.2 College majors data

In order to observe trends in college majors over time, I use the Integrated Postsecondary Education Data System (IPEDS) Survey, which provides a variety of data from public and private institutions in the U.S. and outlying areas (U.S. Department of Education, 2007-2019). Reporting is mandatory for institutions that are eligible to receive Title IV funding. As a result, the IPEDS data covers the majority of universities in the U.S. since Title IV provides funding for federal student financial aid programs. The remaining institutions may voluntarily respond to the survey. The survey is conducted annually by the National Center for Education Statistics (NCES).

This data is very reliable, as it represents nearly all universities in the United States. The web-based survey instrument contains edit checks to detect major reporting errors. In addition, IPEDS help desk staff manually reviewed the data for additional errors. When necessary, the help desk staff contacted keyholders to verify the accuracy of the data.

Specifically, I take the datasets for Completions, or degrees conferred, from 2007 to 2019. I merge this with IPEDS datasets on institutional characteristics so that we have degrees conferred by Classification of Instructional Program (CIP) code, year, and zip code. This results in a dataset with 9,590 unique institutions. Each institution has data on characteristics such as whether they are public or private, their size, majority undergraduate, and residential. We have the CIP code, a taxonomy created by the NCES to correspond to academic and occupational instructional programs offered for credit at the postsecondary level. I use the CIP code as the measure of college majors, so the two terms are interchangeable. In the final dataset, there are 3,615,773 observations with the institution name, year, location, CIP code, award level, and counts of degrees granted.

3.3 Constructing the panel data

In order to construct panel data by year, industry, and commuting zone (C.Z.), I start by matching industries between the LinkUp and IPEDS data. Since the LinkUp data contains 2010 O*NET codes, I use a crosswalk to map these to 2019 O*NET codes, then another crosswalk to map these to 2020 CIP codes. The CIP codes are then aggregated to 7 sub-categories of majors, which I will refer to as industries: Education, Health, Legal Services, Social Sciences, Engineering, Business, and Humanities. The crosswalk for this is manually determined and is available in the appendix (Table A1). Since the IPEDS data contains 2000 and 2010 CIP codes, this is crosswalked to 2020 CIP codes. Then, these are aggregated to the 7 industries using the same crosswalk as in Table A1.

The next step is to put the data at the commuting zone level. A commuting zone is a geographic area defined by the Economic Research Service and U.S. Department of Agriculture, intended to be a spatial measure of the local labor market. I choose this level so that we can still isolate geographic variation under the assumption that students would be most aware of the area that is most accessible to them. I use the HUD-USPS ZIP Code Crosswalk from the U.S. Department of Housing and Urban Development to map zip codes to FIPS codes at the state and county level. In the rare case where a zip code was assigned to multiple counties, I match it to the one with the highest ratio of zip codes in that county. This matches job posting zip codes to FIPS codes and college address zip codes to FIPS codes. Then, I use data from Fowler and Jensen (2020) to map the FIPS codes to commuting zones. Fowler and Jensen (2020) use a consistent methodology from the Census Bureau to determine commuting zones (referred to as ERS delineations), but they use more recent data from 2010, rather than the data from the Census Bureau, which was last updated in 2000. Fowler, Rhubart and Jensen (2016) identified some discrepancies between the method described by ERS and the results reported for previous years, so a second delineation for

2010 (OUT10) is included. As such, I use OUT10 for my analysis.

For the college data, I group by industry, C.Z., and year and sum the degree counts. For the LinkUp data, the data is at the job posting level, so I transform it in order to count the number of active job postings and the number of active companies by industry, C.Z., and year. Finally, I merge the two datasets together by industry, C.Z., and year for panel-level data. In order to investigate the causal effect of job postings on college majors, I merge the data together with a 3 year lag on college majors for bachelor's degrees, and 2 year lag for associate's degrees.

As a result, the final dataset for bachelor's level data has the years 2010 to 2019 for degree data, 2007 to 2016 for job openings data, 7 industries, and 624 commuting zones, yielding 43,680 observations. The descriptive statistics are available in Table 1. We can note that each of these variables is at the year, commuting zone, and industry level. As such, there is large variation in statistics such as number of degrees in a year, since some commuting zones are much larger than others and have more universities in them. We see that the LinkUp Index also has a wide range, from a minimum of 0 to a maximum of 72.5, demonstrating the large spread of job openings in this measurement.

Similarly, the final dataset for associate's level data has a lag of 2 years, so it has the years 2009 to 2019 for degree data, 2007 to 2017 for job openings data, 7 industries, and 625 commuting zones, yielding 48,125 observations. The descriptive statistics are available in Table 2. The statistics look similar to Table 1, but a lower number of degrees granted overall and smaller job market.

I plot the data geographically, so Figure 2 shows the variation in fraction of degrees in health by commuting zone. We see that there is significant variation across commuting zones, and that the fraction of degrees changes from 2011 to 2019. Similarly, there is significant variation in the engineering industry in Figure 3. Similar plots using the LinkUp Index instead demonstrate far less variation across commuting zones, indicating that using the LinkUp index may not be helpful measurement of the labor market.

Figure 4 plots job openings and degrees over time by industry, averaged across commuting zones. We see in Figure 4a that overall degree counts have been increasing over time, except for in the fields of Education and Legal Services. Figure 4b demonstrates relatively flat shares of jobs over time, with some fluctuations in the fields of Health, Business, and Social Sciences. Ideally, the empirical strategy works best when there is significant variation over time, across industries, and across commuting zones. It seems that by averaging across commuting zones, much of the variation is masked, but there still may be enough to generate powerful results. We still see quite large differences in the levels of industries overall.

Figure 5 plots alternative ways of measuring job openings and degrees over time. Figure 5a plots share of degrees in a given industry, rather than the absolute number of degrees, in order to capture trends over time. We see that the lines look relatively flat but some industries have declined or grown slightly. This data subsets only to non-profit institutions, a characteristic that conforms more closely with the typical idea of a college. Figure 5b plots the LinkUp index over time. The LinkUp index has a spike in jobs in 2015, which seems to be driven by a spike in hospitality and tourism relative to other industries. This may due to the way that the LinkUp data is scraped. There is also a large spike in 2007, which is due to the LinkUp database not having a very large sample size at this point, leading to large fluctuations in the numerator of the statistic. This measure of job openings does not seem very accurate, which is why I instead focus on share of jobs in given industries.

3.4 Granger Causality

As an exploratory data analysis method, I perform Granger Causality tests of lags 1, 2, and 3 years. While a typically outdated method, this serves to identify and confirm what lag duration makes sense and highlight any patterns that would align with my hypothesis that college students respond to the labor market. I use the fact that U.S. university students must choose their college major typically about 3 years before they enter the job market. As a result, there is a lag between them observing the status of the labor market and graduating

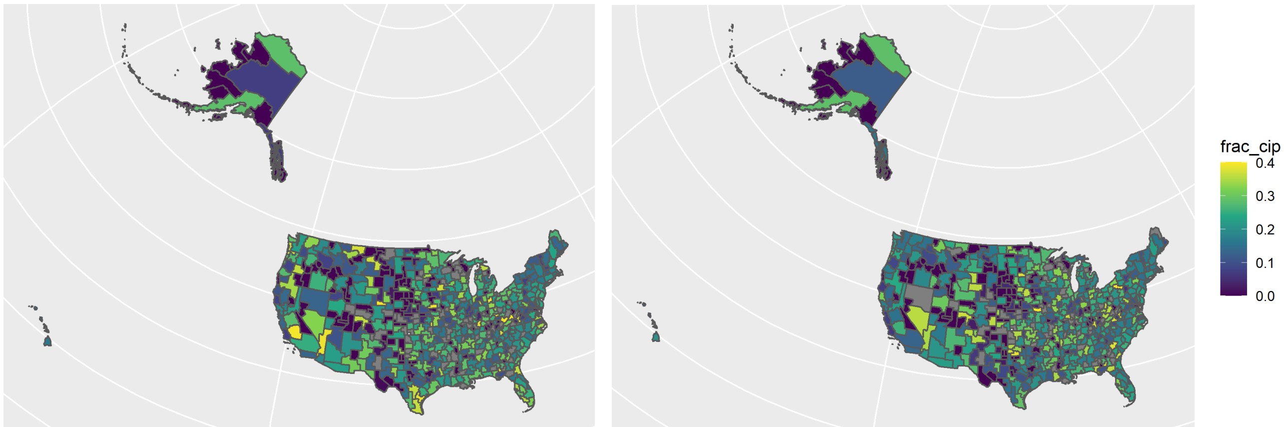
Table 1: Descriptive statistics (bachelor's level)

Statistic	Mean	St. Dev.	Max
Num degrees in industry	394.048	1,208.004	29,378
Num degrees in year	2,646.841	6,627.169	95,755
Frac degrees in industry	0.093	0.124	1.000
Num job postings created	632.410	4,003.271	303,056
Num job postings deleted	620.036	3,971.365	302,111
Cumulative job postings created	2,937.450	20,747.350	859,070
Cumulative job postings deleted	2,849.589	20,358.380	850,930
Num active job postings	87.861	455.303	11,984
Num active companies	56.803	179.418	3,070
LinkUp index	0.612	1.168	72.500
LinkUp index (scaled)	615.026	2,591.799	40,902
Total jobs by industry and year	0.123	0.150	1.000

Table 2: Descriptive statistics (associate's level)

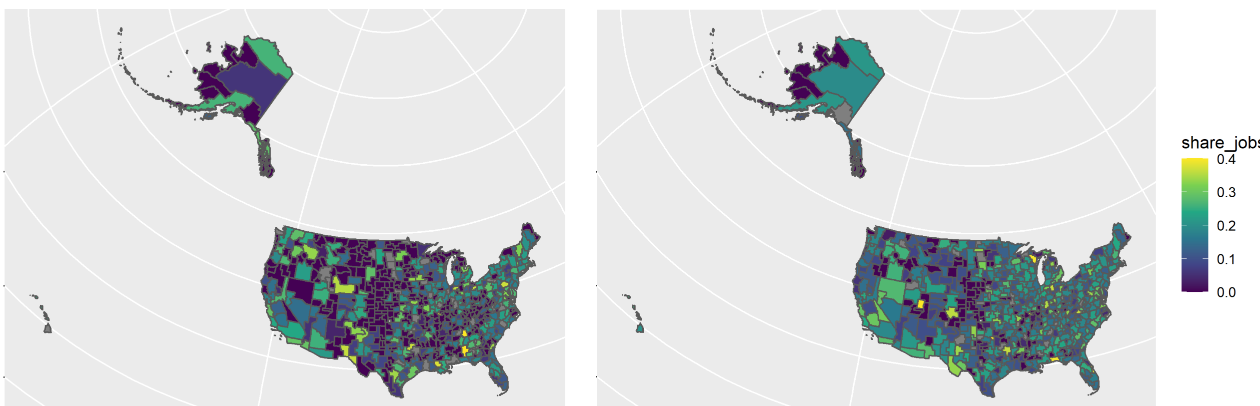
Statistic	Mean	St. Dev.	Max
Num degrees in industry	202.069	909.795	55,283
Num degrees in year	1,380.657	3,860.882	98,301
Frac degrees in industry	0.113	0.178	1.000
Num job postings created	404.112	2,152.219	206,067
Num job postings deleted	392.590	2,123.552	205,455
Cumulative job postings created	1,714.420	10,255.430	511,105
Cumulative job postings deleted	1,648.674	9,986.915	502,468
Num active job postings	65.746	295.334	13,180
Num active companies	37.517	111.505	2,245
LinkUp index	0.899	2.011	104.000
Total jobs by industry and year	460.223	1,427.748	27,459
Share jobs in industry	0.128	0.181	1.000

Figure 2: Fraction of degrees in health by commuting zone



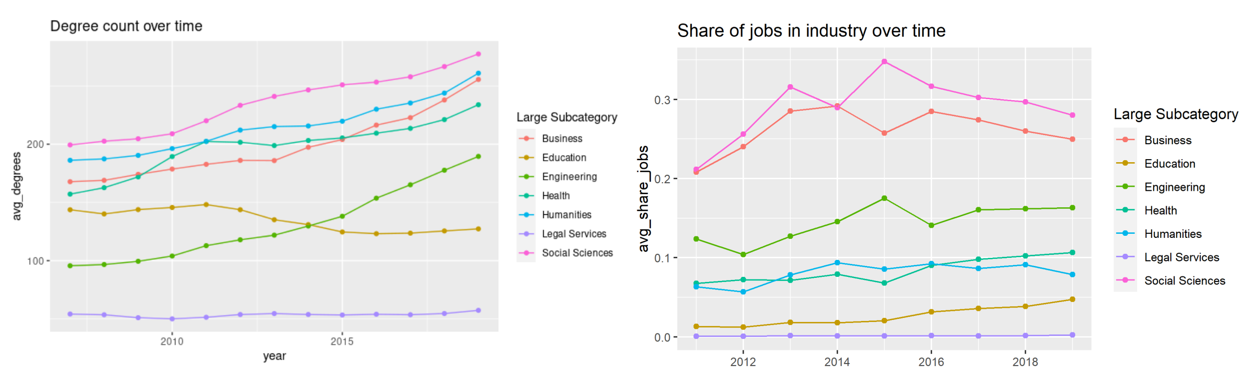
(a) 2011 (b) 2019

Figure 3: Share of job openings in engineering by commuting zone



(a) 2011 (b) 2019

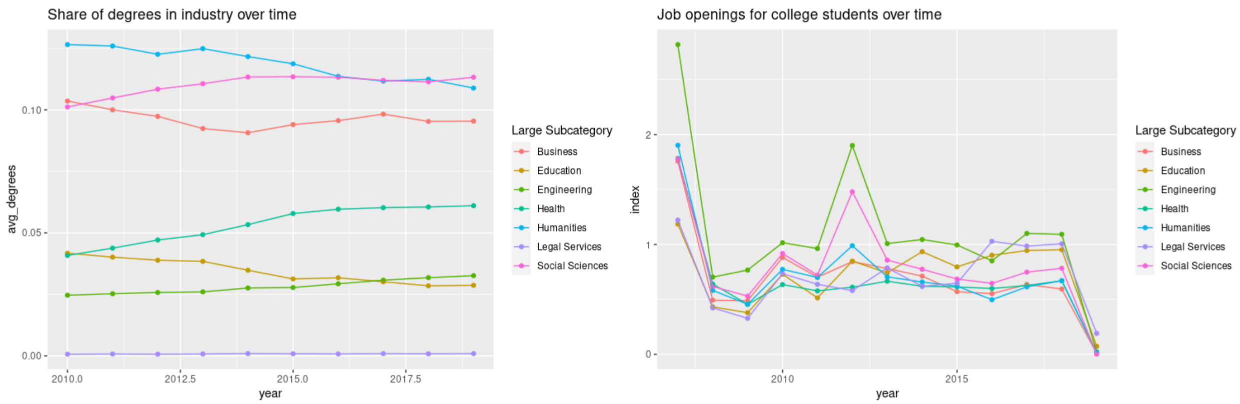
Figure 4: Job openings and degrees over time by industry



(a) Average number of degrees per institution by year

(b) Share of jobs

Figure 5: Alternate measures of job openings and degrees over time by industry



(a) Share of degrees in non-profit institutions (b) LinkUp Index by year

with their chosen degree. As such, I would expect a lag of 3 years to have the lowest p-value. I also explore, as a secondary question, whether employers respond to shifts in college majors by changing their job openings offered. In this case, there should similarly be a lag between employers observing what degrees college students are graduating with, and responding to possible new employees.

The univariate time series process $\{z_t\}$ is said to Granger-cause univariate time process $\{x_t\}$ if

$$P^t_{x,t+h}(\Omega_t) < P^t_{x,t+h}(\Omega_t \backslash \{z_s \mid s \leq t\}) \tag{2}$$

where Ω_t represents all relevant information in the universe available and up to time t, $x^t_{t+h}(\Omega_t)$ is the optimal h-step-ahead predictor of process $\{x_t\}$, and $P^t_{x,t+h}(\Omega_t)$ is the Mean Squared Error (MSE). In other words, does including the history of z_t reduce the variance of the optimal prediction error of x_{t+h}? The Granger Causality tests do not necessarily imply a causal effect, but they indicate that one time series has predictive validity for forecasting another. This has important implications on its own, to answer the research questions and determine which is more plausible (or both). Do college students decide their major based on the status of the job market? Do companies shift their job openings based on college majors?

I use aggregate univariate time series for job openings over time and college majors over time. Since the time series are univariate, I consider 3 aggregations: all majors/industries, STEM majors/jobs, and non-STEM majors/jobs. STEM majors are of particular interest given that STEM is a field that has been growing in popularity lately, and anecdotally, it seems that many students may be responding to this. In order to create STEM-level statistics, I subset to STEM-related CIP codes, using the STEM-designated degree program list from the Department of Homeland Security.

For the question of whether students choose their majors based on job opportunities, either significant or non-significant results could be justified. I would expect that students tend to follow industry trends, but existing literature suggests that students tend to be unaware of what options are available to them. Regardless, if there were a significant result, it would make the most sense if it appeared with a lag of 3 years. For the question of whether employers respond to student supply, I wouldn't expect this to be the case, although that would certainly be an interesting result. In this case, the assumption of a 1-year lag may not make sense because employers could also forecast how many students would graduate with a particular major ahead of time – for example, by looking at internship applications. Regardless, they may respond in some way to the supply because business priorities may reflect what students are majoring in.

I perform Granger Causality tests on the univariate time-series data, using the LinkUp index and degree counts for Bachelor's degrees. The results are available in Table 3. Even though some results are significant, the exact values should not be taken too seriously, since they are only done with 13 data points of averaged data.

We see that only two results are significant, both of which are in Panel A, where the data is subset to STEM fields. As expected, one significant result occurs for the model where job openings predict college degrees with a 3-year lag (similar to Equation 3). This makes sense because students must decide their major based on the status of the job market about 3 years before they actually graduate. One could argue that a 4- or 5-year lag may also make sense, but in the U.S., most colleges allow students to change their major up to 3 years before they graduate.

The other significant result occurs for the reverse causal effect, where college degrees predict job openings with a 1-year lag (similar to equation 5). If firms were to respond to students' degrees, it makes sense that this would occur with a 1-year lag, because any longer than that would be too long of a response time for a firm to change the way they operate or their job infrastructure.

Interestingly, the only significant results appear in the STEM panel, indicating that non-STEM field students may not be as responsive to the job market, and non-STEM industries may not be as responsive to students. One possible explanation for this is that there is a baseline number of students who would major in non-STEM fields regardless of the demand, due to their own skills and interests. STEM students may be more responsive to job opportunities due to salary differences.

Table 3: Granger Causality Test Results

Lag	degrees ∼ jobs p-value	jobs ∼ degrees p-value
Panel A: STEM fields		
1	0.6987	0.04046**
2	0.2499	0.1202
3	0.08716*	0.6091
Panel B: Non-STEM fields		
1	0.6713	0.4324
2	0.4391	0.1904
3	0.2604	0.1578
Panel C: All fields of study		
1	0.7768	0.3735
2	0.4186	0.4808
3	0.2244	0.3218
Note:	*p<0.1; **p<0.05; ***p<0.01	

4 Empirical Strategy

My paper contains several analyses to answer the question: When people are choosing their college majors, do they decide based on relevant job opportunities in their local labor market? And, the secondary question, which I refer to as "reverse causal": do shifts in college majors cause shifts in job openings?

4.1 Panel Regression with Fixed Effects

In order to answer these questions, I perform regressions with the lagged variables and introduce fixed effects by commuting zone and industry. I make use of the fact that each granular location and industry experiences its own unique fluctuations in job openings and college majors, so they are exposed to different amounts of change. As a result, I can measure any response in job openings or college majors depending on level of exposure.

Since I am looking at data at the local labor market level by using commuting zones, defined by Fowler and Jensen (2020), it is important to establish that this is a valid geographical level of granularity. In particular, I argue that the majority of students stay in the area where they went to college, making their job search far more local than a national labor market. And even if they don't stay, students are most aware of job openings in their local labor market. The only case where this does not seem to hold is for elite universities, and it holds most strongly for community college graduates.

A study by EMSI demonstrates the fact that the majority of students stay in the area where they went to college. The study tracked college alumni from 445 prominent research universities and liberal arts colleges to see where they moved after they graduated (Sentz et al., 2018). EMSI found that, "on average, a student who attends a community college will stay within 300 miles of the college and 61% live within 50 miles of the college. State university grads generally stay within state lines with an average distance of 330 miles from their alma mater, and 40% are within 50 miles of the university. Graduates of elite schools flock to the big cities and tend to move nearly 700 miles away from their universities."

As a result, the study demonstrates that it's valid to assume that students are aware of their local job openings, and this assumption holds more strongly in some places than others. Further, I am not interested in whether a student actually stays in their local area, but rather whether they are most *aware* of the job options in their local labor market. This seems plausible given that job fairs usually have local companies come, so a students' perception of job opportunities is likely driven by their college location.

For my empirical strategy, the panel specification controls for industry, commuting zone, and year fixed effects, which are unobservable variables that could bias the estimation of the effect of job openings. Year fixed effects capture variables that change across time which are constant across industries and commuting zones, such as macroeconomic domestic trends. By adding fixed effects, the regression measures job openings relative to a baseline (the average in the economy), which identifies sectors that are growing faster than average. In other words, the regression can be thought of as identifying how shares of degrees shift in response to a jump in job openings in a given commuting zone, relative to others. Since panel data generates autocorrelation between unit observations, I use commuting zone × industry clustered standard errors.

There is one primary specification:

$$Share\_Degrees_{ijt} = \beta_0 + \beta_1 \cdot Share\_Jobs_{i,j,t-3} + \gamma_i + \alpha_j + \phi_{t-3} + \varepsilon \tag{3}$$

where i is the industry, j is the commuting zone, and t is the year. *Share_Degrees* is the fraction of degrees awarded for the given industry and commuting zone in year t, and *Share_Jobs* is the fraction of job openings in the given industry among the LinkUp data. This aligns with the variable that I compared to JOLTS in the data validation section. γ, α, ϕ represent fixed effects for the given variables. The job openings are lagged by 3 years to represent the lag between observing the job openings and graduating with a chosen major. This is chosen based on the experience of a typical college student. While lags of 4 and 5 years could also be included, this would introduce autocorrelations, and the Granger Causality tests demonstrate that a lag of 3 years makes the most sense.

I perform this specification using different college characteristics: 1) a baseline model of all 4-year colleges in the IPEDS database, 2) a model subset to non-profit colleges that are title IV eligible and primarily baccalaureate and above, 3) a model subset to 4-year public universities, and finally, 4) a model with community colleges and corresponding job openings for those with an associate's degree.

For the second subset, essentially, this corresponds to our typical idea of a 4-year school. Non-title IV eligible schools are typically not traditional since they do not offer aid. Colleges must be "baccalaureate" in order to offer undergraduate degrees, so this removes some outliers. It makes sense to focus on non-profit colleges, since about 11 percent of students are enrolled in for-profit colleges, which tend to offer short-degree programs with a strong vocational focus (Darolia et al., 2015). In fact, Gilpin, Saunders and Stoddard (2015) find that for-profit colleges are more responsive than public colleges to local employment and wage growth. The primary difference between non-profit and for-profit colleges is that the average annual tuition at for-profit colleges is nearly five times higher than at public community colleges. As a result, for-profit colleges are sometimes viewed as predatory and non-traditional. As such, I would expect that the second specification may lead to a lower estimate of college student responsiveness to shifts in the job market, due to removing for-profit colleges, but this estimate may more accurately represent the experience of the typical student in the US.

The third and fourth subsets follow from the EMSI study discussed above, and they both serve to identify a potentially higher level of college student responsiveness, due to the fact that these students may be more aware of their local labor markets and more likely to live in the same area after they graduate.

One common concern with fixed effects is reverse causality and simultaneity bias. I am able to address this by the lagged nature of the data. Another concern is that it doesn't address time-variant unobserved heterogeneity. This is addressed in the robustness section.

In addition to answering questions about the general trend of students in the US, I am also interested in observing heterogeneity in response by industry and college type. This serves to identify whether certain college types are more responsive to changes in the job market and whether certain industries within colleges are more responsive. This can be observed through the following regressions:

$$\begin{aligned} Share\_Degrees_{ijtk} &= \beta_0 + \beta_1 \cdot Share\_Jobs_{i,j,t-3,k} + \beta_2 \cdot \mathbb{1}\{CollegeChar\}_{ijtk} \\ &+ \beta_3 \cdot Share\_Jobs_{i,j,t-3,k} \cdot \mathbb{1}\{CollegeChar\}_{ijtk} + \gamma_i + \alpha_j + \phi_{t-3} + \varepsilon \quad (4) \end{aligned}$$

where *CollegeChar* represents the college characteristic of interest, always as an indicator variable, and k is college characteristic level. As a result, the data is still at the i, j, t level, with an additional aggregated level of the k college characteristic. In other words, the data is not at the college level, so including college fixed effects would not make sense.

I consider whether the college is majority undergrad, whether it's a small university, and whether the college is residential. Majority undergraduate versus graduate is typically a proxy for whether or not a college is a research university. I would expect that non-research universities may be more career-oriented and be more responsive to the job market. Small universities may be more flexible and able to change their program offerings faster, making them more responsive to the job market. If a college is not residential, students may be more likely to live at home and want to stay in the area long-term, making them more aware of their local labor market and thus more responsive to it.

In order to identify heterogeneity by industry, I follow Equation 4, except *CollegeChar* is instead a dummy for the industry category. In order to avoid collinearity, the Education category is chosen to be dropped and acts as a baseline for comparison.

Reverse Causality The reverse causal question is inspired by Freeman (1975), where Freeman creates a recursive model of the market for lawyers. Due to the time delay between the decision to study in a field and graduation and entrance into the market, he finds that the supply of lawyers fluctuates over time in response to the current income, which is a function of the demand for lawyers. Then, cyclically, income changes based on the current supply of lawyers. Freeman refers to this as a "cobweb" model. I perform a similar specification to investigate the reverse causal effect of degrees on job openings, where degrees are lagged by 1 year:

$$Share\_Jobs_{ijt} = \beta_0 + \beta_1 \cdot Share\_Degrees_{i,j,t-1} + \gamma_i + \alpha_j + \phi_{t-1} + \varepsilon \tag{5}$$

This serves to answer the question of whether employers respond to changes in college majors by shifting their job postings. This seems unlikely since college cohorts are a small share of employment in any given region, but it is possible that firms may take shifts in college majors as a sign that this is a good time to upskill since there are a lot of talented majors. For example, if more people major in economics, Uber may decide to create an economics research group, since economists are generally cheaper so the marginal cost may now be equal to their marginal benefit. I will call this reverse causality story "theory 1." The lag is 1 year since it would seem feasible that it would take employers no longer than 1 year to respond, and the Granger Causality test results were significant for a lag of 1 year.

We can note that if we assume that both Equations 3 and 5 hold, Equation 5 can be simplified to a reduced form by plugging in the expression for $Share\_Degrees_{i,j,t-1}$. For the sake of clarity, the coefficients are changed to different variable names:

$$\begin{aligned} Share\_Degrees_{ijt} &= \alpha + \beta \cdot Share\_Jobs_{i,j,t-3} + ... + \varepsilon \\ Share\_Jobs_{ijt} &= \gamma + \delta \cdot Share\_Degrees_{i,j,t-1} + ... + \nu \\ &= \gamma + \delta[\alpha + \beta \cdot Share\_Jobs_{i,j,t-4} + ... + \varepsilon] + ... + \nu \end{aligned} \tag{6}$$

Clearly, from Equation 6, the reduced form demonstrates that job openings are a function of job openings four years ago. This presents an alternate possibility where, rather than theory 1 holding, it's possible that companies who boom continue to post jobs in the same area. I refer to this as "theory 2." In order to test the two theories, I perform the regression:

$$Share\_Jobs_{ijt} = \beta_0 + \beta_1 \cdot Share\_Degrees_{i,j,t-1} + \beta_2 \cdot Share\_Degrees_{i,j,t+3} + \varepsilon \qquad (7)$$

If β_1 is significant, this provides evidence for theory 1, since degrees predict job openings with a 1-year lag, while if β_2 is significant, this provides evidence for theory 2, since degrees 3 years in the future predict current job openings. This allows me to test the two theories in the same regression to see which one holds.

Robustness checks A few other specifications can serve as robustness checks. First, rather than using $Share\_Jobs$, I can use Job_Index, which is defined in Equation 1. This may serve as a helpful way of controlling for the fact that LinkUp is always adding jobs to their index, and is used as a statistic by many other authors.

Second, I consider a quadratic specification where the independent variable $Share\_Jobs$ also has a coefficient on its squared form, $Share\_Jobs^2$, to test whether the model may be non-linear. I also perform a F-test to compare the linear and non-linear models.

Next, there is also another alternative form of the regression in Equation 3:

$$Share\_Degrees_{ijt} = \beta_0 + \beta_1 \cdot Share\_Jobs_{i,j,t-3} + \gamma_{ij} + \phi_{t-3} + \varepsilon \qquad (8)$$

The difference between the specifications is that Equation 8 includes fixed effects for each commuting zone interacted with each industry, while Equation 3 includes separate fixed effects by industry and by commuting zone. The difference between these models relies on the assumption that, in the first case, jobs within an industry are a local labor market, so people tend to only be aware of job opportunities near them. In the second case, given industries and given commuting zones may have heterogeneity, but the demand effects are separable between these two dimensions. This is a relatively stable period, so using the fixed effects for commuting zone interacted with each industry will only leave time, commuting zone, and industry variation. There may not be enough time variation to identify a relationship, since I only use 9 or 10 years of data once the lag is taken into account. The specification in Equation 3 is more common, and used in papers such as Aizer (2010) with a similar empirical setup. Regardless, the alternative specification serves as a robustness check.

Future versions of the paper will try a Bartik instrument strategy similar to Autor, Dorn and Hanson (2013), which may serve to identify a stronger causal effect. However, this strategy requires additional data which was not possible to get under the time constraints.

One additional concern with using fixed effects is that it doesn't address time-variant unobserved heterogeneity. The most plausible pathway for this would be through an aggregate shock to the labor market from 2007 to 2019 that enables sorting. The main aggregate shock is the financial crisis, so for robustness, I also consider the same specification with the years of the financial crisis dropped (2007 to 2010). This is the main shock discussed in Weinstein (2017), which is the primary shock that occurs during this time period. The fact that there are not many other shocks makes the Bartik instrument strategy difficult to implement, but it makes this strategy valid. Another possible pathway would be via a shock to degree-granting programs. However, we need not be concerned about this, since there are permanent industry and geographic effects due to the fact that it is hard for universities to enter and exit the market. Light (2020) finds that institutional response to change in demand for skills is slow, so universities do not change their degree-granting programs often or create or remove spots.

5 Results

5.1 Predicting College Majors

First, I perform regressions to quantify the effect of job market shifts on college majors, following Equation 3. The results are presented in Table 4. We see that the effect of job openings on college majors drops significantly when we add both commuting zone and industry fixed effects, from 0.215 and 0.138 (respectively) to 0.043. This makes sense, because there is a lot of specificity to certain commuting zones and industries. For example, there are certain geographical areas that always produce a lot of degrees and jobs in given fields, such as the Bay Area, Boston, North Carolina, Chicago, etc. In terms of industries, there are permanent characteristics about the way that colleges work so that certain fields tend to be more popular than others. Controlling for both commuting zones and industries explains a lot of variation in the model, increasing the R^2 from 0.356 and 0.202 (respectively) to 0.481. The further addition of CZ $\times$ industry fixed effects is discussed in the robustness section.

Most notably, in column 3, which is my preferred specification, the model can be interpreted as: an increase in the fraction of job openings in the industry by 1 percentage point leads to a 0.043 percentage point increase in the fraction of degrees in that industry. More suggestively, a 25 percentage point increase in job openings leads to a 1 percentage point increase in the fraction of degrees in that industry. In addition, the coefficient is significant at the $p < 0.01$ level. In other words, it appears that college students do respond to shifts in availability of job openings, but only in a very minor way – while the result is significant, it is extremely small. As a result, only a major shock to the labor market would elicit a notable shift in the distribution of college majors. This is corroborated by Weinstein (2017), who finds that after the dot-com crash, there was an 8% decline in computer science degrees at research universities in high computer employment areas. In other words, one must look at specific degrees, in specific colleges, in specific areas, after a major shock, in order to find a notable shift in college majors. As such, the estimate of 0.043 seems reasonable.

Next, in Tables 5, 6, and 7, I look at subsets of university types. In Table 5, I look at 4-year nonprofit colleges that are title IV eligible and primarily baccalaureate and above. We see that the estimate for the impact of the job market on degrees drops, as I predicted in the Empirical Strategy section. This is due to eliminating major choice from for-profit colleges, which have been empirically found to be more responsive to the labor market. This indicates that an estimate more representative of the typical undergraduate student in the U.S. would be closer to a 0.031 percentage point increase.

When we look at public universities in Table 6, the estimate drops further, to about a 0.027 percentage point increase. Public universities are one setting where students tend not to move after college, so the assumption that they are only concerned with their local labor market holds more strongly here. As such, the baseline estimate may be a slight overestimate. Finally, we look to community college degrees and job openings in Table 7, where this assumption is most strongly met. The job openings are now lagged by 2 years instead of 3 since an associate's degree typically requires 2 years to complete, and one must declare a major before or upon starting. The point estimates are not too different in this setting, with an estimate from the preferred specification (column 3) of a 0.031 percentage point effect. This indicates robustly that the result is significant and likely around this range.

I would expect that the point estimate would have been the highest for community colleges, but it seems that both private and public universities are still responding similarly. This indicates that there are a lot of frictions on the supply side of the labor market. Despite the fact that labor demand changes, it takes longer for students to respond to these changes in demand. This corresponds with the findings of Light (2020), that it typically takes a long time for colleges to update their degree program or course offerings. In addition, between years, colleges have had the same capacity for various majors, which is why it is now more competitive to apply to colleges in growing fields like computer science.

In order to investigate heterogeneity in sensitivity to the labor market, Table 8 includes results according to Equation 4 for different college characteristics. We see that the effect is much stronger in colleges that are majority undergrad (compared to majority grad), colleges that are small (compared to large), and colleges that are residential (compared to non-residential). These all confirm my hypotheses and explanations described in Section 4.

Next, I investigate heterogeneity by industry in Table 9. The baseline industry which is dropped from the model is Education. We see that there is significant heterogeneity across industries, indicating that the 0.043 estimate in Table 4 masks this variation across industries. We can identify the marginal effect of job openings by taking the derivative with respect to job openings. Then, we have that:

$$\begin{aligned}\frac{\partial Degrees}{\partial Jobs} = &-0.105 + 0.188 \cdot Business + 0.098 \cdot Engineering + 0.1 \cdot Health \\ &+ 0.19 \cdot Humanities - 0.063 \cdot LegalServices + 0.169 \cdot SocialSciences\end{aligned}$$

As a result, the marginal effect for students with degrees in business is $-0.105 + 0.188 = 0.083$. Similarly, we have that the effect for degrees in engineering is -0.007, health is -0.092, humanities is 0.085, legal services is insignificant, and social sciences is 0.064. We can consider the baseline industry of education as set to 0, so all of these estimates are relative to education, where the effect for education is not identifiable. Clearly, the most responsive industries are humanities and business. I find this surprising because my initial hypothesis was that engineering would be the most responsive, since the field tends to be more high-paying. This also conflicts with the exploratory results found from Granger Causality tests in Section 3. However, this can be explained by the fact that engineering is a difficult field, so students may not be responsive since it would be hard for them to gain the skills to succeed. It may be easier to switch to a humanities or business degree since these fields typically require more soft skills rather than technical skills.

Table 4: OLS with fraction of job openings

	Dependent variable:		
	Frac degrees in industry		
	(1)	(2)	(3)
Frac job openings in industry, 3 year lag	0.215***	0.138***	0.043***
	(0.008)	(0.009)	(0.005)
CZ FE	X		X
Industry FE		X	X
Year FE	X	X	X
Observations	43,680	43,680	43,680
R^2	0.356	0.202	0.481
Adjusted R^2	0.347	0.202	0.474
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 5: OLS (4-year nonprofits)

	Dependent variable:		
	Frac degrees in industry		
	(1)	(2)	(3)
Frac job openings in industry, 3 year lag	0.147***	0.107***	0.031***
	(0.008)	(0.008)	(0.005)
CZ FE	X		X
Industry FE		X	X
Year FE	X	X	X
Observations	43,680	43,680	43,680
R^2	0.357	0.125	0.444
Adjusted R^2	0.348	0.125	0.435
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 6: OLS (public universities)

	Dependent variable:		
	Frac degrees in industry		
	(1)	(2)	(3)
Frac job openings in industry, 3 year lag	0.164***	0.110***	0.027***
	(0.007)	(0.008)	(0.004)
CZ FE	X		X
Industry FE		X	X
Year FE	X	X	X
Observations	43,680	43,680	43,680
R^2	0.485	0.155	0.580
Adjusted R^2	0.477	0.155	0.573
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 7: OLS (community colleges)

	Dependent variable:		
	Frac degrees in industry		
	(1)	(2)	(3)
Frac job openings in industry, 2 year lag	0.185***	0.078***	0.031***
	(0.009)	(0.010)	(0.008)
CZ FE	X		X
Industry FE		X	X
Year FE	X	X	X
Observations	48,125	48,125	48,125
R^2	0.138	0.406	0.506
Adjusted R^2	0.126	0.406	0.499
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 8: OLS (heterogeneity by college type)

	Dependent variable:		
	Frac degrees in industry		
	(1)	(2)	(3)
Jobs	-0.180^{***}	-0.014	-0.079^{***}
	(0.010)	(0.011)	(0.011)
Majority undergrad	0.048^{***}		
	(0.002)		
Jobs * Undergrad	0.446^{***}		
	(0.021)		
Small University		0.024^{***}	
		(0.002)	
Jobs * Small		0.124^{***}	
		(0.022)	
Residential			0.033^{***}
			(0.002)
Jobs * Residential			0.260^{***}
			(0.022)
CZ FE	X	X	X
Industry FE	X	X	X
Observations	87,360	87,360	87,360
R^2	0.374	0.376	0.390
Adjusted R^2	0.369	0.371	0.386
Residual Std. Error (df = 86718)	0.087	0.087	0.086

Note: *p<0.1; **p<0.05; ***p<0.01
Jobs represents frac job openings in industry, 3 year lag

Table 9: OLS (heterogeneity by industry)

	Dependent variable:
	Frac degrees in industry
Jobs	-0.105^{***}
	(0.032)
Jobs * Business	0.188^{***}
	(0.034)
Jobs * Engineering	0.098^{***}
	(0.033)
Jobs * Health	0.100^{***}
	(0.035)
Jobs * Humanities	0.190^{***}
	(0.044)
Jobs * Legal Services	-0.063
	(0.165)
Jobs * Social Sciences	0.169^{***}
	(0.034)
Business	0.053^{***}
	(0.005)
Engineering	0.001
	(0.005)
Health	0.013^{***}
	(0.004)
Humanities	0.078^{***}
	(0.006)
Legal Services	-0.062^{***}
	(0.004)
Social Sciences	0.080^{***}
	(0.005)
CZ FE	X
Industry FE	X
Year FE	X
Observations	43,680
R^2	0.483
Adjusted R^2	0.475
Note:	*p<0.1; **p<0.05; ***p<0.01
	Jobs represents frac job openings in industry, 3 year lag

5.2 Predicting Job Openings

Next, in Table 10, I perform the regression corresponding to Equation 5. Column 3 demonstrates significant results at the $p < 0.01$ level, indicating that college majors can predict job openings with a 1-year lag. The coefficient in the model can be interpreted as: a 1 percentage point increase in the fraction of degrees in an industry leads to a 0.043 percentage point increase in the jobs in that industry. As discussed in Section 4, I use the model in Equation 7 to test whether this result is indicative of either, 1) firms responding to college majors, or 2) persistence in job openings 4 years later (based on the reduced-form Equation 6). Theory 1 is my baseline original model. The results of testing these theories is in Table 11.

We see that the coefficient for degrees at time $t - 1$ (β_1 in the model) is significant, providing support for theory 1. In other words, this provides evidence that firms may respond to trends in college majors by modifying their business model and creating more job openings in that industry. However, the coefficient is still small and thus would not have a meaningful effect unless there was a large shock to the market for a particular college major. In addition, the difference between the two coefficients $(0.036 - 0.021)$ is small and due to lack of power, we cannot conclude that the difference between the coefficients is significant. Regardless, this is still a novel result which would be interesting to investigate further with more data. It is likely that this theory holds in some industries but not others, since for certain majors, it does not seem plausible that companies would ever need to expand to create job openings for those majors.

It serves as a manifestation of Say's law, "Supply creates its own demand." In this case, supply of college students with a particular major creates its own demand for college students with that major. One possible explanation for this is that a higher supply of students with that major makes the labor cheaper. Another explanation comes from sunspot theory (Woodford, 1990): changing everyone's state of mind at once, through something as arbitrary as a sunspot, leads to a self-fulfilling prophecy where the economy improves. In this case, perhaps all firms need to expand is an arbitrary sign that there are ample college students available to hire, leading them to grow and create more job postings.

Table 10: Cobweb model with fraction of job openings

	Dependent variable:		
	Frac job openings in industry		
	(1)	(2)	(3)
Frac degrees in industry, 1 year lag	0.383***	0.113***	0.043***
	(0.021)	(0.010)	(0.008)
CZ FE	X		X
Industry FE		X	X
Year FE	X	X	X
Observations	52,416	52,416	52,416
R^2	0.119	0.469	0.500
Adjusted R^2	0.108	0.469	0.494
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 11: Reverse Causality Theory Test

	Dependent variable:
	Frac job openings, time t
Frac degrees, time $t-1$	0.036***
	(0.013)
Frac degrees, time $t+3$	0.021
	(0.013)
CZ FE	X
Industry FE	X
Year FE	X
Observations	39,312
R^2	0.528
Adjusted R^2	0.520
Residual Std. Error	0.104 (df = 38672)
Note:	*p<0.1; **p<0.05; ***p<0.01

5.3 Robustness checks

Table A2 presents the results from Equation 3, using the LinkUp index rather than share of job openings. We see similar results to Table 4. Note that in this table, the LinkUp index is scaled so that it is divided by 100. As a result, the coefficient in column 3 can be interpreted as: a 1-point increase in the LinkUp index (1 more job per 100 companies) leads to a 0.1 percentage point increase in the fraction of degrees in that industry. These results are similarly suggestive that college students are responsive to the status of the job market, albeit in a small way.

In Table A3, we see the specification with interacted commuting zone and industry fixed effects according to Equation 8. Column 1 shows the baseline, and we see that the model captures the majority of the variation: the R^2 increases to 0.916 from 0.481 in Table 4. However, the result is not significant. Similarly, columns 3 and 4 have insignificant results for this specification, and column 2 only has $p < 0.1$. This indicates that the CZ $\times$ industry fixed effect may absorb too much of the variation for there to be any meaningful variation left for us to analyze. Clearly, there is not enough time variation to identify a relationship, since once we account for the lag, we are essentially only looking at 9 years worth of data. This is corroborated by the lack of variation in the plots in Section 3. We would need many more years of data in order for this specification to be useful. Take, for example, the Silicon Valley. It was created over more than 50 years, not 9, so we do not have enough data to identify these trends (the term was coined and Atari, Apple, Microsoft, and Oracle were founded in the 1970s).

Table A4 presents the results from dropping the years 2007 to 2010, in order to prevent time-variant unobserved heterogeneity. We see that in my preferred specification in column 3, the coefficient is 0.022, indicating that college student response may be slightly smaller when there are only small shifts in the job market and not large shocks. The number is similar to the specification with public universities. However, it is still significant and positive, demonstrating that the baseline result is still robust but may be a slight overestimate.

Table A5 presents the quadratic specification of the baseline model. Column 3 provides suggestive results that perhaps college student responsiveness to jobs is positive up to a certain point and then becomes negative if the share of job openings in a given industry is too high (above 0.4). In addition, the F-test comparing the linear and quadratic models is significant at the $p < 0.001$ level, indicating that the true model may be nonlinear. Since it is very uncommon for share of job openings to exceed 0.4, the main result that college students are elastic to job openings still holds, and may even be larger than the linear estimates when share of job openings is less than 0.4.

Overall, the robustness checks demonstrate that the model is robust to various specifications and that the key result still holds, though more data may be needed to investigate certain specifications with more power.

6 Conclusion

In this paper, I seek to understand how shifts in job openings lead to a shift in college majors. I contribute to the growing literature to explain the determinants of college major choice and am the first to explore job openings. In order to analyze this, I use LinkUp job postings data in the US from 2007 to 2019 and degrees conferred over time from the IPEDS system. I use regression with lags and fixed effects in order to quantify college student responses.

I find that a 25 percentage point increase in the fraction of job openings in a given industry leads to a 1 percentage point increase in the fraction of degrees in that industry. Essentially, the college student response to job openings is small but significant. The response levels are similar for public universities and community colleges, settings where students are most likely to be aware of their local labor market. College students who attend majority undergraduate schools and small colleges are more responsive to shifts in job openings. This may indicate that majority undergraduate schools, which are less research-focused, are more focused on the vocational aspect of college. Smaller colleges could be more flexible and able to change their program offerings faster. In addition, I find that the students in the industries of business and humanities are more responsive to changes in job openings, demonstrating a larger willingness to switch to and from these fields.

The small student response to job openings has two potential implications: one, students lack information on job openings for given industries or do not care about this; and two, there is a lot of friction for the supply of labor. There may be a limit to the number of degrees available in a given field at a university, and though labor demand changes quickly, it takes longer for universities to respond on the supply side.

The mismatch between job availability and student skills can be partially explained by students' slow responses to labor market demands. However, the lack of responsiveness by students (and universities) is not necessarily negative. Other reasons for students choosing majors, such as enjoyment and talent, are valid and lead to more skilled workers. The inability of universities or students to quickly adapt demonstrates a larger question about the role of universities to educate students, and whether degrees serve as a signal for any job, or whether courses provide useful skills for jobs. If degrees simply serve as a signal, students need not change their major to find a job. If college provides relevant job skills, policies providing information to students about their job market opportunities and encouraging universities to adapt their curriculum to currently demanded skills will be valuable.

References

Aizer, Anna. 2010. "The Gender Wage Gap and Domestic Violence." *American Economic Review*, 100(4): 1847–1859.

Arcidiacono, Peter, V. Joseph Hotz, and Songman Kang. 2012. "Modeling college major choices using elicited measures of expectations and counterfactuals." *Journal of Econometrics*, 166(1): 3–16.

Autor, David H, David Dorn, and Gordon H Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review*, 103(6): 2121–2168.

Azar, José, Ioana Marinescu, Marshall Steinbaum, and Bledi Taska. 2020. "Concentration in US labor markets: Evidence from online vacancy data." *Labour Economics*, 66: 101886.

Baker, Rachel, Eric Bettinger, Brian Jacob, and Ioana Marinescu. 2018. "The Effect of Labor Market Information on Community College Students' Major Choice." *Economics of Education Review*, 65: 18–30.

Becker, Gary S. 1962. "Investment in Human Capital: A Theoretical Analysis." *Journal of Political Economy*, 70(5, Part 2): 9–49. Publisher: The University of Chicago Press.

Beffy, Magali, Denis Fougère, and Arnaud Maurel. 2012. "Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter?" *The Review of Economics and Statistics*, 94(1): 334–347.

Blom, Erica, Brian C. Cadena, and Benjamin J. Keys. 2021. "Investment over the Business Cycle: Insights from College Major Choice." *Journal of Labor Economics*, 39(4): 1043–1082.

Campello, Murillo, Gaurav Kankanhalli, and Pradeep Muthukrishnan. 2020. "Corporate Hiring under COVID-19: Labor Market Concentration, Downskilling, and Income Inequality." 45.

Cascio, Elizabeth U., and Ayushi Narayan. 2015. "Who Needs a Fracking Education? The Educational Response to Low-Skill Biased Technological Change." National Bureau of Economic Research Working Paper 21359. Series: Working Paper Series.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2015. "Housing Booms and Busts, Labor Market Opportunities, and College Attendance." National Bureau of Economic Research Working Paper 21587. Series: Working Paper Series.

Darolia, Rajeev, Cory Koedel, Paco Martorell, Katie Wilson, and Francisco Perez-Arce. 2015. "Do Employers Prefer Workers Who Attend For-Profit Colleges? Evidence from a Field Experiment: Do Employers Prefer Workers." *Journal of Policy Analysis and Management*, 34(4): 881–903.

Fowler, Christopher S, and Leif Jensen. 2020. "Bridging the gap between geographic concept and the data we have: The case of labor markets in the USA." *Environment and Planning A: Economy and Space*, 52(7): 1395–1414. Publisher: SAGE Publications Ltd.

Fowler, Christopher S., Danielle C. Rhubart, and Leif Jensen. 2016. "Reassessing and Revising Commuting Zones for 2010: History, Assessment, and Updates for U.S. 'Labor-Sheds' 1990–2010." *Population Research and Policy Review*, 35(2): 263–286.

Freeman, Richard. 1975. "Legal Cobwebs: A Recursive Model of the Market for Lawyers." *Review of Economics and Statistics.*

Gilpin, Gregory A., Joseph Saunders, and Christiana Stoddard. 2015. "Why has for-profit colleges' share of higher education expanded so rapidly? Estimating the responsiveness to labor market changes." *Economics of Education Review*, 45: 53–63.

Jordan, Guillermo Carranza. 2018. "Vacancy Duration and Demand for H-1B Labor in the United States." 55.

Light, Jacob. 2020. "How Have the Skills College Graduates Develop Changed Over Time? Evidence from 3 Million Course Descriptions." Social Science Research Network SSRN Scholarly Paper 3751682, Rochester, NY.

LinkUp. 2007-2019. "Job Market Data Raw." `https://pages.linkup.com/DataSolutions.html#Raw` (accessed Sep 15, 2021).

Long, Mark C., Dan Goldhaber, and Nick Huntington-Klein. 2015. "Do completed college majors respond to changes in wages?" *Economics of Education Review*, 49: 1–14.

Marinescu, Ioana, and Roland Rathelot. 2018. "Mismatch Unemployment and the Geography of Job Search." *American Economic Journal: Macroeconomics*, 10(3): 42–70.

Pulse, College. 2020. "College Students Rely Heavily on LinkedIn During the Job Search."

Rosenbaum, James E., Regina Deil-Amen, and Ann E. Person. 2006. *After admission: from college access to college success.* New York:R. Sage Foundation.

Sentz, Rob, Meredith Metsker, Paul Linares, and Josh Clemans. 2018. "How Your School Affects Where You Live."

Stinebrickner, Ralph, and Todd Stinebrickner. 2014. "Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model." *Journal of Labor Economics*, 32(3): 601–644. Publisher: The University of Chicago Press.

Székely, Gábor J., Maria L. Rizzo, and Nail K. Bakirov. 2007. "Measuring and testing dependence by correlation of distances." *The Annals of Statistics*, 35(6).

U.S. Department of Education. 2007-2019. "Integrated Postsecondary Education System." Institute of Education Sciences, National Center for Education Statistics. `https://nces.ed.gov/ipeds/use-the-data` (accessed Sep 10, 2021).

U.S. Department of Labor, Employment, and Training Administration. 2016. "O*NET 21.0 Database." `https://www.onetcenter.org/db_releases.html` (accessed Sep 15, 2021).

Weinstein, Russell. 2017. "Local Labor Markets and Human Capital Investments." *IZA Discussion Papers*, 82.

Wiswall, Matthew, and Basit Zafar. 2015. "How Do College Students Respond to Public Information about Earnings?" *Journal of Human Capital*, 9(2): 117–169.

Woodford, Michael. 1990. "Learning to Believe in Sunspots." *Econometrica*, 58(2): 277.

Zafar, Basit. 2013. "College Major Choice and the Gender Gap." *Journal of Human Resources*, 48(3): 545–595. Publisher: University of Wisconsin Press.

Zhang, Qi, Bowen Yu, and Junlin Hu. 2011. "Statistical Correlation Between The LinkUp Job Search Engine and Bureau of Labor Statistics Jobs Reports." 17.

ΩŞahin et al.

Şahin, Ayşegül, Joseph Song, Giorgio Topa, and Giovanni L. Violante. 2012. "Mismatch Unemployment." National Bureau of Economic Research w18265.

A Appendix: Figures and Tables

Table A1: Aggregated CIP codes

Code Prefix	Large Subcategory
01	Social Sciences
03	Social Sciences
04	Engineering
05	Humanities
09	Humanities
10	Humanities
11	Engineering
12	Humanities
13	Education
14	Engineering
15	Engineering
16	Humanities
19	Social Sciences
21	N/A
22	Legal Services
23	Humanities
24	Humanities
25	Social Sciences
26	Social Sciences
27	Social Sciences
28	Social Sciences
29	Social Sciences
30	Humanities
31	N/A
32	N/A
33	N/A
34	Health
35	N/A
36	N/A
37	N/A
38	Humanities
39	Humanities
40	Engineering
41	Engineering
42	Social Sciences
43	N/A
44	Social Sciences
45	Social Sciences
46	N/A
47	N/A
48	N/A
49	N/A
50	Humanities
51	Health
52	Business
53	N/A
54	Humanities
55	N/A
60	Health
61	Health

Table A2: OLS with LinkUp Index

	Dependent variable:			
	Frac degrees in industry			
	(1)	(2)	(3)	(4)
LinkUp index (scaled), 3 year lag	0.713***	1.104***	0.100**	−0.026*
	(0.137)	(0.178)	(0.049)	(0.015)
CZ FE	X		X	
Industry FE		X	X	
CZ x Industry FE				X
Year FE	X	X	X	X
Observations	43,680	43,680	43,680	43,680
R^2	0.297	0.197	0.480	0.916
Adjusted R^2	0.286	0.197	0.472	0.906
Note:				*p<0.1; **p<0.05; ***p<0.01

Table A3: OLS with CZ x Industry FE

	Dependent variable:			
	Frac degrees in industry			
	Baseline	Nonprofit	Public	CC
	(1)	(2)	(3)	(4)
Frac job openings in industry, 3 year lag	0.003	0.002*	−0.001	0.001
	(0.003)	(0.001)	(0.001)	(0.003)
CZ x Industry FE	X	X	X	X
Year FE	X	X	X	X
Observations	43,680	43,680	43,680	48,125
R^2	0.916	0.959	0.966	0.922
Adjusted R^2	0.906	0.954	0.963	0.914
Note:				*p<0.1; **p<0.05; ***p<0.01

Table A4: OLS (financial crisis years dropped)

	Dependent variable:		
	Frac degrees in industry		
	(1)	(2)	(3)
Frac job openings in industry, 3 year lag	0.229***	0.086***	0.022***
	(0.010)	(0.012)	(0.008)
CZ FE	X		X
Industry FE		X	X
Year FE	X	X	X
Observations	26,208	26,208	26,208
R^2	0.363	0.189	0.477
Adjusted R^2	0.347	0.188	0.464
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table A5: OLS (quadratic)

	Dependent variable:		
	Frac degrees in industry		
	(1)	(2)	(3)
Frac job openings	0.472***	0.481***	0.158***
	(0.015)	(0.017)	(0.012)
Frac job openings squared	−0.524***	−0.627***	−0.197***
	(0.023)	(0.028)	(0.018)
CZ FE	X		X
Industry FE		X	X
Year FE	X	X	X
Observations	43,680	43,680	43,680
R^2	0.380	0.236	0.484
Adjusted R^2	0.371	0.236	0.476
Note:	*p<0.1; **p<0.05; ***p<0.01		
	Frac job openings in industry, 3 year lag		