ISSN 0956-8549-773

Superstar Firms and College Major Choice

By

Darwin Choi

Dong Lou

Abhiroop Mukherjee

FINANCIAL MARKETS GROUP DISCUSSION PAPER NO 772

PAUL WOOLLEY CENTRE WORKING PAPER NO 56

December 2024

Superstar Firms and College Major Choice*

Darwin Choi
Hong Kong University of Science and Technology
dchoi@ust.hk

Dong Lou
Hong Kong University of Science and Technology
London School of Economics and CEPR
dlou@ust.hk; d.lou@lse.ac.uk

Abhiroop Mukherjee
Hong Kong University of Science and Technology
amukherjee@ust.hk

This Draft: December 2024

*We thank Ashwini Agrawal, Li An, Josh Angrist, Tania Babina, Nick Barberis, Pedro Bordalo, James Choi, Andrew Ellul, Daniel Ferreira, Xavier Gabaix, Nicola Gennaioli, Robin Greenwood, David Hirshleifer, Caroline Hoxby, Dirk Jenter, Theresa Kuchler, Daniel Metzger, Per Ostberg, Paige Ouimet, Daniel Paravisini, Steve Pischke, Christopher Polk, Wenlan Qian, Antoinette Schoar, Kelly Shue, David Solomon, Johannes Stroebel, Geoff Tate, Constantine Yannelis, and seminar participants at Australian National University, CEMFI Madrid, Chinese University of Hong Kong, Deakin University, Durham Business School, Hong Kong Polytechnic University, Hong Kong University of Science and Technology, KAIST, London School of Economics, Loughborough Business School, Monash University, New York University, Peking University, Tsinghua PBC School of Finance, Tsinghua SEM, Toulouse School of Economics, University of Hong Kong, University of Oxford, University of Toronto, Yale University, 2017 Adam Smith Finance Conference, 2017 European Finance Association Annual Meeting, 2017 Stanford Institute for Theoretical Economics (SITE) Labor and Finance Meeting, 2017 Labor and Finance Group Meeting, 2018 American Economic Association Annual Meeting, 2018 CEPR Spring Symposium in Financial Economics, 2018 NBER Summer Institute Education/Labor Studies, 2018 Miami Behavioral Finance Conference, 2019 American Finance Association Annual Meeting for helpful comments and suggestions. All remaining errors are our own. We are grateful for funding from the Paul Woolley Center at the London School of Economics.

Superstar Firms and College Major Choice

Abstract

We study the relation between the presence of superstar firms and college students' major choice. Occurrences of superstar performers in an industry are followed by a sharp rise in the number of college students choosing to major in related fields. This cohort effect remains significant after controlling for lagged industry returns and wages. Students' tendency to follow superstars, however, is met with lower real wages earned by entry-level employees when these students enter the job market. Further evidence from two college-graduate surveys shows that such adverse career outcomes can last for decades.

JEL Classification: D81, D91, G40, I23, J24

Keywords: College Major Choice, Human Capital Investment, Stock Return Skewness, News Tone Skewness, Salience

1 Introduction

Education choice and human capital investment are of great importance to both individual well-being and social-economic development (Hoxby 2003, 2004). An integral part of this decision is college major choice.[1] In this paper, we study a potentially important but understudied determinant of major choice that is frequently referred to in the popular press: the link between college major enrollment and the presence of superstar performers — that is, a small number of firms that have done exceptionally well — in related industries. For example, Stanford Daily reports that the number of Stanford undergraduate students who declared a Computer Science major in 2013 was nearly four times that in 2006, possibly due to the high-profile successes of a handful of mobile-app and social-media companies such as Facebook. A *New York Times* article argues that "students are flocking to computer science because they dream of being the next Mark Zuckerberg."

One possible account for this casual observation is that college students' attention is often drawn to—and their expectations and decisions shaped by—occurrences of superstar performers.[2] For one, superstar firms and entrepreneurs attract disproportionate media coverage and social attention (Hirshleifer, 2019): the story of Mark Zuckerberg, one of the youngest self-made billionaires, is a constant talking point in the popular press and on college campuses. In addition, the occurrences of superstar firms and entrepreneurs are often accompanied by extreme payoffs: Mr. Zuckerberg is consistently named one of the world's wealthiest and most influential individuals. A combination of these two salient features—disproportionate social attention coupled with extreme payoffs—make superstar firms particularly influential in college students' major choice.

The importance of salience in driving human decisions, given limited cognitive resources, has been extensively studied in the psychology literature. As Taylor and Thompson (1982) put it, "salience refers to the phenomenon that when one's attention is differentially directed to one portion of the environment rather than to others, the information contained in that portion will receive disproportionate weighting in subsequent judgments." Kahneman (2011) goes on to argue that "our mind has a useful capability to focus on whatever is odd, different or unusual." Applying this insight to the setting of college major choice, we argue that

[1] Prior literature links college major choice (Computer Science vs. Economics, for example) to individuals' perceptions of lifetime income of each major, personal interest in the subject, and innate ability.

[2] Note that "superstar firms/performers" in our context refer to a handful of firms with exceptional recent performance. This definition differs from the one often used in the industrial organization literature, where "superstar firms" refer to the largest firms (in terms of market value, employment or sales) in an industry (e.g., Autor et al., 2020). For example, a superstar performer in the tech sector in the early 2010s was Facebook, while the largest firm in the tech sector in the same period was Microsoft, which was struggling in its competition against Apple and Google. We argue that salient superstar performers like Facebook, rather than large firms like Microsoft in the 2010s, attract student attention.

salient occurrences of superstar firms and entrepreneurs—albeit non-representative of the whole industry—can play a substantial role in shaping students' expectations and decisions. This force is particularly relevant given the substantial search costs faced by college students in their major choice (see, e.g., Hoxby, 2004; Altonji, Blom and Meghir, 2012; Hastings et al., 2016; Huntington-Klein, 2016).

To analyze the empirical relation between superstar firms and college major choice, we take the following steps. First, as defined in Rosen (1981) and Malmendier and Tate (2009), a superstar system is characterized by a highly skewed distribution of payoffs and public attention. We measure the presence of superstar performers in each industry by the *cross-sectional* stock return skewness in that industry. By definition, positive (negative) cross-sectional skewness in an industry reflects a small number of firms performing exceptionally well (poorly) relative to the industry peers.[3] For example, Pan American World Airways (Pan Am) outperformed its peers by over 250% in the 1960s, leading to large positive cross-sectional skewness in the Air Transportation industry; Lockheed and Chrysler outperformed their peers by more than 300% in the early 1980s, giving rise to high return skewness in the Manufacturing industry. In our baseline results, we use employment-weighted skewness to give more weight to more important/visible firms in the industry.[4]

Second, we focus on a set of science and engineering majors (computer science vs. chemical engineering, for example) that can be easily mapped to one or more industry sectors (information technology vs. pharmaceutical). Third, since college students usually declare their majors by the end of the sophomore year, we focus on industry return skewness measured in years t-7 to t-3 prior to the graduation year (from their junior year in high school to sophomore year in college) to explain college major distribution at graduation in year t.

Our empirical results reveal a strong positive relation between the presence of superstar firms and enrollment in related college majors. Using cohort-level college degree data from the National Science Foundation (NSF) and controlling for both year- and major-fixed effects, we find that a one-standard-deviation increase in within-industry return skewness in years t-7 to t-3 is associated with a 11.36% (t-statistic = 3.94) increase in the number of graduates in related major fields in year t. For reference, a one-standard-deviation increase in the average industry return (wage) in years t-7 to t-3 forecasts an increase in graduates in year t by 11.19% (4.15%). This relation between industry skewness and major choice extends well beyond any particular industry or time-period. Going back to our earlier examples, the large

[3]We do not claim that high school or college students regularly follow the stock market. Instead, we think of return skewness as a proxy for important events taking place in related industries that draw students' attention, and shape their expectations and decisions. We also use a more direct measure of important events – skewness in media coverage and tones – and find very similar results.

[4]Our results are robust to a host of alternative measures of skewness.

positive cross-sectional return skewness in the Air Transportation industry in the 1960s was indeed followed by the popularity of the Aeronautical and Astronautical Engineering major; likewise, the positive skewness in the manufacturing sector in the early 1980s was followed by rising popularity of the Industrial and Manufacturing Engineering major.

A positive relation between college major enrollment and the presence of superstar performers is consistent with two general mechanisms. On the one hand, occurrences of superstar firms may indicate brighter industry prospects, so students choose related major fields in anticipation of more and better job opportunities (labor-demand based).[5] On the other hand, it could be that the presence of superstar performers is uninformative about future industry prospects, and students are attracted by these non-representative observations (labor-supply based).[6] To empirically evaluate the relative importance of labor-demand- vs. supply-based channels, we examine the wage and net hiring (the price-quantity pair) at the time the cohort enters the job market. A relatively larger outward shift in labor supply (demand) should result in lower (higher) entry-level wages.

Our results are consistent with a supply-based mechanism. After controlling for year- and major-fixed effects, a one-standard-deviation increase in industry return skewness in years t-7 to t-3 is associated with a 1.65% (t-statistic = 5.32) *drop* in the real wage earned by entry-level employees in related majors in year t. In contrast, within-industry return skewness in years t-7 to t-3 is uncorrelated with the average wage of advanced positions in year t, consistent with recent college graduates not competing for these types of jobs.[7]

Further, we find that the relation between lagged industry skewness and net hiring of entry-level employees in year t is indistinguishable from zero. This is consistent with the view that labor demand is relatively inelastic in the short run, as it takes time for firms to expand operations and production. An increase in labor supply (in the form of a larger number of college graduates in related fields) thus lowers the average wage earned by entry-level employees, without affecting the size of employment.

Next, we examine changes in the composition of labor supply, that is, the type of students

[5] Inconsistent with this demand-based mechanism, we show in later analyses that industry skewness is unrelated to a host of measures for future industry prospects (such as the industry average Return on Assets, Return of Equity, Net Profit Margins, and Sales Growth), as well as to future stock returns, which incorporate forward-looking information.

[6] A related possibility is that within-industry return skewness is informative about future industry prospects, but students overreact to this information; consequently, shifts in labor supply outweigh shifts in labor demand.

[7] Our empirical design is to compare the average market-adjusted entry-level wage of the same major across *different cohorts*. An alternative design would be to compare a student with her counterfactual-self, had she chosen a different major. An obvious issue with this second approach is that individual major choice crucially depends on *unobservable* personal characteristics, such as ability and interest in the subject-matter. This is less of a concern at the cohort level, as long as the distribution of personal characteristics in each cohort does not vary systematically with the distribution of industry skewness over time.

more likely to be attracted by occurrences of superstar performers. First, we find that the positive effect of industry return skewness on college major choice is significantly stronger among elite universities in states where the superstar's industry has a substantial presence. Second, the negative effect of industry skewness on subsequent entry-level wages is stronger for occupation codes that pay above-median wages within a major (e.g., a larger effect on the wages of software developers than on that of database administrators, for Computer Science majors). Both results suggest that superstar performers attract more able students to related fields, potentially because better students are more likely to associate themselves with extreme success. Moreover, when we directly examine individual characteristics or proxies for socio-economic status (SES) from surveys, we do not find any evidence negative selection based on SES accompanying the choice of high skewness majors.

After providing evidence at the cohort level, we next examine granular, individual-level data from the National Survey of College Graduates (NSCG). The average respondent in this survey is 43.7 years old, roughly 20 years out of college, allowing us to examine career outcomes over a long horizon. The survey also contains information on total earnings, which includes wage as well as other sources of income, allowing a more complete measurement of job prospects. Moreover, the survey reports how closely a respondent's current job is related to her main field of study, allowing us to study whether a part of the excess supply of graduates related to superstars gets absorbed in unrelated fields.

Our analysis reveals that after controlling for a host of fixed effects for different demographic characteristics, cohort, survey-year, industry, and major, a one-standard-deviation increase in industry return skewness of the respondent's declared major forecasts 1.9% lower real annual earnings (including wages and bonuses) at the time of the survey. We also find that higher skewness is associated with a 4.5–5.7% higher propensity that the respondent works in a job outside her field of study, typically associated with worse outcomes. Working outside one's field of study is associated with significantly worse outcomes—respondents who work in unrelated fields earn 17.3–20.5% less, and have 75–84.2% higher odds of reporting job dissatisfaction, compared to their peers. These results suggest that an outward shift in labor supply not accompanied by a similar shift in labor demand can have a long-lasting, adverse impact on individuals' career outcomes.

Our evidence, put together, points to a strong relation between superstar performers and increasing labor supply in related industries. We interpret this relation through the lens of industry salience – superstar performers increase industry salience, which helps attract students to related majors. An alternative interpretation of this relation is that the presence of superstar firms does not by itself increase an industry's attractiveness; instead, it is a reflection of an underlying industry trend/theme, which is unrelated to the industry's labor demand but attracts student attention. For instance, during the internet bubble, there was an over-hyped narrative that the information technology sector was going to transform our economy; consequently, a) a few internet companies achieved spectacular stock performance and b) students flocked to Computer Science.[8] Although we cannot rule out this possibility, it is broadly consistent with our industry-salience-based interpretation: the occurrence of superstar performers remains a *proxy* for industry salience, except that in this case salience stems from a narrative around the entire industry, rather than around a few firms in the industry.

In the last part of the paper, we provide more detailed evidence on the role of superstar firms in driving college students' major choice. Specifically, we present evidence from a survey of college graduates we conducted on the SurveyMonkey platform, to answer questions that cannot be addressed using public data. For example, neither the NSF nor NSCG data contain information on the type of industries that college students want to work in at the time of major declaration. So it is unclear whether the relation between high-skewness majors and working in unrelated fields is driven by respondents who always wanted to work in different fields (computer science students longing for a career in investment banking) or by respondents who could not get a job in their target industry due to excess supply of graduates. We ask this question directly in our survey, and the evidence clearly favors the latter possibility.

Moreover, our survey helps distinguish between the two previously-mentioned channels through which college students may be drawn to industries with superstar performers: a) belief errors—i.e., students form income expectations (or more generally, expectations of future successes) based on a small number of non-representative but highly visible observations; b) preferences for positively skewed payoffs—that is, students are happy to accept a lower average wage for a small chance of hitting the jackpot. Our survey evidence indicates that skewness-driven major choice is strongly correlated with self-reported expectation errors, but not with lottery preferences.

In sum, our set of analyses provides novel evidence that the presence of superstar performers in an industry forecasts higher college major enrollment in related fields, but lower entry-level wages and a higher likelihood of working outside target industries at the time of graduation; moreover, these adverse career outcomes last for decades. While these results are generally consistent with an outward shift in labor supply (more so than that in labor demand) in relation to superstar performers, we do not claim that this is the only mechanism driving our results. As discussed earlier, it is still possible that within-industry skewness is

[8]Moreover, it might be the parents or school administrators – rather than students themselves – whose attention is drawn by superstar firms.

correlated with unobserved industry characteristics that drive both college major choice and future industry performance. Put together, our evidence suggests an additional channel—education choice—through which superstar firms may impact social welfare and economic growth.

2 Related Literature

Our paper contributes to a large labor-finance literature on human capital investment. For example, several papers study the consequences of shocks to labor supply. Ouimet and Zarutskie (2014) show that an increase in regional supply of younger workers increases the rate of new firm creation. Agarwal et al. (2019) use the Chinese superstition of giving birth in "Dragon" years to identify the effect of labor supply shocks on earnings. Gupta and Hacamo (2018) study early career choices and subsequent long-run career outcomes of elite engineers. Bena and Simintzi (2019) show that the ability to access labor cheaply affects firm innovation. A related literature studies the impact of sectoral booms and busts on workers' choices and outcomes (Oyer, 2008; Charles, Hurst, and Notowidigdo, 2018; Hombert and Matray, 2019). We show that the presence of superstar firms in related industries has a large impact on college major choice and hence labor supply.

This paper is also related to the growing literature on the impact of salience on human decisions (Chetty, Looney, and Kroft, 2009; Bordalo, Gennaioli, and Shleifer, 2012, 2013a, 2013b; Han, Hirshleifer, and Walden, 2022; Hirshleifer and Plotkin, 2021; Cosemans and Frehen, 2021). In neuroeconomics, Fehr and Rangel (2011) show that subjects evaluate choices by aggregating information about different attributes, with decision weights influenced by attention. Recent theoretical research has also developed novel ways of incorporating limited attention in economic decisions (Gabaix, 2014). Although we are the first to examine the role of salience in education choice, we believe that it is a natural application of salience—given the complexity of the search process for information about job prospects (Stigler, 1961; 1962).

While the impact of superstar firms on current and prospective employees have been studied since at least Rosen (1981), we provide novel evidence that following superstars may lead to worse career outcomes. Our result itself can be consistent with both preference- and belief-based explanations. On the preference side, Rosen (1997) presents a model of preferences for skewness based on state-dependent utility; more generally, a preference for skewness is also a central theme in the non-standard utility literature (Kahneman and Tversky, 1979; Barberis and Huang, 2008, Kahneman, 2011). On the belief side, theory and evidence on mistaken beliefs leading to oversupply can be found as far back as in Kaldor (1934), or more recently, in Greenwood and Hanson (2015). We provide suggestive evidence that our result stems mostly from the latter (the belief-based channel).

Fluctuations in labor markets have also been studied through the lens of Cobweb theory, by Freeman (1971, 1975, 1976). Under Cobweb theory, a lower salary in a major attracts fewer freshmen and, ultimately, produces fewer graduates—but this change in supply manifests itself in the labor market only a few years later. Consequently, starting wages become higher, affecting in turn, major choice of the freshman cohort of the future year—hence producing endogenous cycles in both enrollment and wages. While our paper also a) emphasizes the time lag between major choice and graduation and b) builds on the premise that students form expectations based on stale, non-representative information, our focus is on the effect of occurrences of superstar firms in an industry on the distribution of major enrollment.

Finally, we also contribute to the broader literature on individuals' education choice and on career outcomes (Hoxby, 2003).[9] Most prior studies on college major choice (Berger, 1988) use a framework in which students form rational expectations of future earnings using Bayesian updating. Subsequent research has added various dimensions to this approach, from uncertainties (Altonji, 1993) to heterogeneity (Patnaik et al., 2020). Our paper contributes to and deviates from this literature by examining the role of salient occurrences of superstar firms in determining college students' expectations and major choice.

3 Data

3.1 College Majors and Related Industries

Our data on college degrees are obtained from the National Science Foundation (NSF) and the Integrated Postsecondary Education Data System (IPEDS). NSF uses IPEDS Completions Surveys conducted by the National Center for Education Statistics (NCES) and reports the annual number of bachelor's and master's degrees in science and engineering fields. NSF groups the 2010 Classification of Instructional Programs (CIP), a taxonomy of academic disciplines, into different major fields. The full list of the major fields is presented in Online Appendix Table A1. We use an online NSF document, which lists the number of degrees that were conferred between 1966 and 2012 by accredited institutions of higher education in the US, including the 50 states and the District of Columbia. For degrees that were conferred between 2013 and 2017, we download the number of degrees for each 6-digit CIP code from

[9]See also James, Alsalam, Conaty, and To (1989), Altonji (1993), Sacerdote (2001), Avery and Hoxby (2004), Hoxby (2004), Bhattacharya (2005), Blom (2012), Goldin (2014), Lemieux (2014), Stinebrickner and Stinebrickner (2014), Bordon and Fu (2015), Altonji, Arcidiacono, and Maurel (2016), Arcidiacono, Hotz, and Kang (2012), Fricke, Grogger, and Steinmayr (2015), Wiswall and Zafar (2015, 2022), among others.

NSF's Data Explorer and aggregate them into the corresponding major fields.[10] IPEDS provides the annual number of bachelor degrees for each CIP awarded by each institution, starting in 2001.[11]

We then map the science and engineering degrees from Appendix Table A1 to 3-digit NAICS industry codes. We focus on the set of majors that can be readily mapped to one or more industry sectors (e.g., computer science). In other words, we exclude some physical science and social science majors (e.g., physics, sociology) from our analysis, because students choosing these majors are unlikely to be targeting any specific *industry jobs*.

This mapping is carried out through merging two crosswalks, the 2010 Classification of Instructional Programs (CIP) to the 2010 Standard Occupational Classification (SOC) Crosswalk and the 2010 SOC to the 2012 NAICS map. The detailed mapping is shown in Online Appendix Table A1. Each major field can correspond to multiple industries: e.g., a degree in Health is linked to Ambulatory Health Care Services (NAICS = 621), Hospitals (NAICS = 622), Nursing and Residential Care Facilities (NAICS = 623), and Social Assistance (NAICS = 624). Each industry code can also be mapped to several major fields: for example, Petroleum and Coal Products Manufacturing (NAICS = 324) is associated with degrees in Chemical Engineering, Industrial and Manufacturing Engineering, Materials Science, and Mechanical Engineering.[12]

3.2 Return Skewness

To construct our skewness measure, we calculate the employment-weighted cross-sectional return skewness by pooling all firms that are relevant to a particular major.[13] Firms are mapped to majors based on their 3-digit NAICS industry codes and the map in Appendix Table A1. We aggregate other firm-level variables to the major level in the same way. Our results are robust to other ways to aggregate firm-level variables (calculate the industry-level

[10]The online NSF document is available from the NSF website: `https://www.nsf.gov/statistics/2015/nsf15326/pdf/nsf15326.pdf`. The number of degrees provided by NSF's Data Explorer at the 6-digit CIP level starts in 1997. While NSF's Data Explorer also provides other useful statistics such as enrollment and race composition, we opt for a longer sample period and use the NSF document, which only lists the number of degrees awarded. As shown in Figure 1, this sample period sees the rise of different popular majors and superstar firms.

[11]For the school-level analysis, we sum the number of first majors and second majors to get the number of graduates. Between 1987 and 2000, IPEDS only reports the number of first majors from each school. We do not use these years as they are not comparable to the post-2001 data.

[12]Two of the majors, Computer Sciences and Electrical Engineering, are mapped to the same set of industry codes as they are closely related.

[13]The employment-weighted return skewness is given by $\frac{\sum_i^n w_i (\frac{r_i - \bar{r}}{\hat{\sigma}})^3}{\sum_i^n w_i}$, where n is the number of firms mapped to a major, w_i is the number of employees in firm i, r_i is the return of firm i, and $\bar{r}$ and $\hat{\sigma}$ are the employment-weighted mean and standard deviation of return, respectively. The stock and firm data are obtained from CRSP and Compustat.

measures first, and then calculate the equal-weighted average across all industries associated with the major or choose the industry with the maximum absolute skewness measure), as shown in Appendix Table A2.

Figure 1 shows episodes of high major skewness, and the superstar firms in each episode that were the top contributors to the skewness measure. Specifically, we calculate the five-year residual skewness (after adjusting for the mean and standard deviation of returns in corresponding industries) for each major in each year. The solid line shows the maximum residual skewness across all majors in each year, and the shaded areas depict episodes of particularly high maximum residual skewness. The firms shown in the figure are those that contribute the most to the skewness of the related major in each episode, calculated as the difference between the skewness of the corresponding major with and without that firm.

Besides the internet bubble episode, we identify superstars in other periods as well. For example, in the late 1960s, Pan American World Airways (Pan Am) outperformed its industry peers by 250% in terms of stock returns, contributing to the high return skewness of the Air Transportation industry, and possibly the high enrolment in the Aeronautical and Astronautical Engineering major shortly after.[14] The early 1980s saw high returns to companies like Chrysler and Lockheed, and Industrial and Manufacturing Engineering was a popular major in the following years. The early 2010s saw high returns to "bulge-bracket" banks, followed by the popularity of the Economics major. In the last few years, Tesla had spectacular returns and made auto manufacturing glamorous.

3.3 Aggregate Wage and Employment

Aggregate wage and employment data between 1997 and 2017 are available from the Bureau of Labor Statistics (BLS) through the Occupational Employment Statistics (OES) program. Wage is defined as straight-time, gross pay, exclusive of premium pay. Wage and employment data are reported at the SOC code level in each industry. BLS provides projections of the job requirements (degrees and approximate number of years of experience required) for the majority of the SOC codes. We use the CIP-SOC Crosswalk and the BLS projections to define entry-level jobs for graduates of each major. These are jobs that are suitable for students of a particular major that require a bachelor's degree, but do not require prior work experience. Note that the aggregate wage and employment for each major is calculated from the employment-weighted wage and total employment across all relevant SOCs. Each SOC (e.g., Computer Programmers) can appear in multiple industries, some of which (e.g.,

[14] Although Pan Am went into bankruptcy in 1991, it was a highly successful and visible company in the 1960s. This is evidenced by the Hollywood movie, *Catch Me If You Can* (2002), which was based on the life of a con-artist who impersonated a Pan Am pilot in the 1960s.

Financial Services) can be different from our major-NAICS map in Appendix Table A1. In other words, our measures of aggregate wage and employment are derived from the actual employment data and do not rely on any major-NAICS map.

Our tests focus on engineering and applied science majors that can be readily mapped to industry sectors. In our analysis of aggregate-level data (from NSF, IPEDS, and BLS), we exclude Biological Sciences and Health majors. Although students of these majors can follow superstars in specific industries (e.g., in the Health sector), many biology- and health-related jobs require an advanced degree and students often go to graduate schools before entering the job market (health and bio majors are 26.9% more likely to have higher degrees, relative to other majors in our sample). This means that these majors are particularly subject to selection. For example, if there is an excess supply of graduates in some cohort, leading to potentially worse entry-level job opportunities, more of them could choose to go for higher education. As a result, the entry-level wage data will come from a selected sample who cannot (or choose not to) go to grad school in spite of poor job conditions. Nevertheless, our conclusion remains unchanged if we include Biological Sciences and Health majors in the aggregate-level analysis, as reported in Appendix Table A2.

Such selection issues do not affect our results in the individual-level NSCG or Survey-Monkey samples. This is because these surveys are not based on entry-level employees, but also contain respondents that graduated 20 or more years earlier. Following the graduation cohort years later implies that it contains both those that chose to get higher education, as well as those that elected to go to the job market immediately. So in these samples, we include Biological Sciences and Health majors.

3.4 Individual-Level Wage and Employment

We also use the National Survey of College Graduates, sponsored by the National Center for Science and Engineering Statistics (NCSES) and the NSF and conducted by the Census Bureau. The survey provides data on college graduates, focusing on those in the science and engineering workforce. It is conducted every two years since the 1970s and samples individuals who live in the U.S., have at least a bachelor's degree, and are younger than 76. Eligible individuals are identified by the education attainment responses to the U.S. Census long form and the American Community Survey (ACS). The data are collected through online surveys, paper questionnaires, and computer-assisted telephone interviews. From the survey, we can obtain information about individual survey respondents' graduation year, major, demographics, total earnings (including variable compensation, e.g., bonuses), and employment status. Data are available online beginning 1993. We require information on respondents' majors, which is publicly available in the years 1993, 2003, 2010, 2012, 2015 and 2017.

3.5 News Sentiment

Our news sentiment data are from RavenPack News Analytics, which quantifies positive and negative contents of news reports. We focus on the Event Sentiment Score (ESS) constructed by RavenPack. ESS is an entity-level sentiment score, determined by matching stories typically categorized by financial experts as having positive or negative financial or economic impact. It ranges between 0 and 100, where 50 represents neutral sentiment, and is available between 2000 and 2017.

We follow the methodology by Dang, Moshirian, and Zhang (2015). First, a score of -1 to 1 is assigned to each news article by re-scaling the ESS. For each calendar day, we keep the most novel story (the story with the highest Event Novelty Score, ENS). Just like our skewness measure based on annual stock returns, we pool all relevant firms to calculate news-based skewness. More specifically, we calculate a news sentiment score for every firm in every year (which is given by the sum of rescaled ESS scores); we then use these sentiment scores to calculate news-based skewness, analogous to what we do with returns.

3.6 Summary Statistics

Panel A of Table 1 summarizes the datasets, including the sample period, coverage, and how they are used in the paper. We present summary statistics for our variables of interest in Panel B. The median number of bachelors in each major is 6,921 students per year. On average, firm returns are positively skewed in the cross-section, with a median annual skewness of 1.39. The median cross-sectional skewness in news tone, measured from the Ravenpack news analytics data, is 4.85. The correlation between the return-based and news-based skewness measures is roughly 30%. The employment-weighted average entry-level wage for workers with a bachelor's degree in science and engineering has a median of $58,070 (in 1997 dollars).[15] The median annual net new hiring of these positions is 3,102, or 2.3% of the number of employees in the previous year.

[15] Note that we do not have data specifically on the first year of employment; the wage and employment figures include seasoned workers who are still in these entry-level positions and have not been promoted.

4 Main Results at the Major Level

This section presents the main results of our paper. We start by examining the relation between major-level skewness and the subsequent number of students choosing that major. Note that in contrast to standard major choice regressions typically run at the individual level using survey data, in this section we are mostly interested in variation at the cohort level, as a function of related-firm characteristics.

4.1 Number of Graduates in Different Majors

Our main hypothesis is that given the substantial cognitive costs faced by college students in figuring out job prospects, personal ability, or interest, in various industries, students' expectations—and hence major choice—are disproportionately influenced by exceptional performance by a handful of superstar employers for that major. To analyze the effect of cross-sectional return skewness on major choice decisions at the cohort level, we estimate the following regression equation:

$$log(bachelor_{i,t}) = \alpha + \beta LaggedSkew_{i,t-3} + \gamma \mathbf{X}_{i,t-3} + \mu_i + \tau_t + \epsilon_{i,t}, \tag{1}$$

where $log(bachelor_{i,t})$ is the natural logarithm of the number of graduates in major i in year t (t refers to the calendar year of graduation, all other time variables are expressed with respect to this; i.e., t-k refers to k years before graduation). $LaggedSkew_{i,t-3}$ is the cross-sectional return skewness relevant to that major, calculated using returns data from t-7 to t-3.[16] This skewness measure can be described best by considering firms related to any two majors, say, Mechanical Engineering and Computer Science. Suppose that firms that typically hire from each of these majors have similar average performance, as well as similar dispersion in performance. These respective firm-performance distributions are different – the performance of firms related to Mechanical Engineering is evenly distributed across the average, but for Computer Science, there are a handful of superstar firms that have done exceptionally well while other firms have slightly below-average performance. Then our measure will classify Computer Science as a high-skewness major.

Note that our skewness measure is lagged to reflect that exceptional performance by a handful of related firms can only affect major choice if they occur *before the major is decided*, which for most students is by their sophomore year in college.[17]

[16] Our results are also robust to other return windows, e.g., t-8 to t-3 and t-6 to t-3, as shown in Appendix Table A3.

[17] For example, in our own survey using SurveyMonkey, we asked each respondent the year she decided on her major, and find that about 80% decided their majors by the end of their sophomore year in college.

$\mathbf{X}_{i,t-3}$ is a vector of controls, suitable to our setting of analyzing major choice at the cohort level. Our vector of controls includes the average return and return volatility between t-7 and t-3, calculated using firms relevant to the major. These two controls ensure that our skewness measure picks up differences between industries that have performed similarly in the recent past, and have had similar dispersion in performance, as in our example above. Next, we include as a control the average size (market capitalization) of firms. We also account for the average firm age and industry valuation ratio (book-to-market, B/M), which are typically regarded as proxies for firm growth. μ_i and τ_t are major and time (year) fixed effects, respectively. The inclusion of major fixed effects ensures that our identification of the coefficient of interest, β, comes from changes in the number of graduates, not its level. Inclusion of time fixed effects purges out any market-wide fluctuation from our estimate. We also control for the average wage earned by each major (*Lagged Log Average Wage*). This is to examine whether the empirical relation we document above is distinct from the cycles of major enrollment, employment and wages as described in Cobweb models. Following prior literature on Cobweb theory (e.g., Freeman, 1976), we control for the average real wage of each major in the past three years.[18] We cluster the standard errors at the year level and not at the major level because of the small number of majors; our results remain highly statistically significant with bootstrapped standard errors and block-bootstrapped standard errors (see Online Appendix Table A2).

If our hypothesis—that college students' major choice is influenced by superstar firm performance—is true in the data, we expect major-level skewness to positively predict the number of graduates in the future. That is, the coefficient on *LaggedSkew*, β, should be positive. We present the results in Table 2. As we can see from Column (1), *LaggedSkew* predicts major choice strongly, even after controlling for the average return and return volatility. A one-standard-deviation increase in *LaggedSkew* is associated with an increase in the number of students majoring in related fields by 11.36% with a t-statistic of 3.94 (all explanatory variables are standardized). As a benchmark, one-standard-deviation increase in *LaggedSkew* is about the same as the episode in 1986–88 or that in 2010–13 in Figure 1. For comparison, a one-standard-deviation increase in the lagged average return is associated

Further, Table A3 in the Online Appendix conducts a test using skewness measured in years t-2 and t-1, i.e., the two years that are likely after major declaration and hence mostly reflect those who switch majors. As expected, *LaggedSkew* in these two years has a much more muted effect on major choice, and in a horse race has virtually no impact on the coefficient on our *LaggedSkew* variable in years t-7 to t-3, suggesting that switching majors based on skew is less popular than sticking to a declared major.

[18] We show in Table A4 that skewness in t-3 to t-7 is not correlated with future stock return or with future return volatility. To the extent that stock returns are forward-looking and incorporate all available future information, skewness is not proxying for future industry conditions. Also, in Table A5, we study the relations between skewness and other independent variables that are used in this regression. We regress each of these variables on skewness and do not find any significant relation with skewness.

with a 11.19% (t-statistic = 4.72) increase in the number of students choosing the major. Further, our results show that past wages indeed positively predict future major enrollment, our skewness measure retains its predictive power.

An important consideration is the robustness of our results. In the Online Appendix, Table A6 reruns regression (1) by replacing major and year fixed effects with the interaction of major fixed effects and time trends, in the forms of linear, quadratic, and logarithmic trends. These time trends capture specific trends in graduation by major. The results remain robust. We also find that the relationship between $log(bachelor)$ and $LaggedSkew$ holds even if we drop the Tech Bubble period, or use cohort-level data on Masters degrees instead of Bachelors (Appendix Table A2).

Next, we study the dynamics of the relationship between graduates in a major and skewness by adopting an event-study approach. We identify discrete jumps in skewness based on Figure 1, when the adjusted skewness of the major is higher than the sample median. Year 0 is defined as the year when adjusted skewness begins to rise. We then look at the log number of bachelor's degrees from year -5 to year 10, adjusting for year and major effects. A control group is made up of majors not in the top 5 for adjusted skewness around the event year. Figure 2 shows the average difference between the treatment and control groups for each year. Consistent with our main results, the graph shows that the response is more prominent around year 2 to year 7.

To put our discussion above in perspective, we would like to highlight two key aspects of our empirical design. First, many of our majors can be stepping stones to careers in multiple industries; so choosing to graduate with a particular major does not necessarily limit the student to work in the industry most closely related to it. For example, Computer Science graduates can also work as librarians. All we assume in our analysis is that *at the time* the student chooses to major in Computer Science, he is much more interested in a career in the Computing or Tech industry than he is interested in librarianship.

Second, we use skewness in stock returns, rather than concentration in firm size or sales (e.g., Autor et al., 2020), to reflect the fact that students are more likely to be attracted by what is "exciting" at that time. This notion of attraction is more closely captured by a handful of superstar firms recently doing exceptionally well and thereby capturing public imagination and media attention, than it is by the presence of a few dominant firms in the industry. Note that even though such exciting firms may lead a student to choose a major field, there is often a small chance of actually working for these dream companies. For example, even though Facebook's success was (and perhaps still is) capturing social attention from college campuses to movie studios, the firm accounted for a tiny fraction of all jobs for Computer Science majors. We revisit this issue in Section 5.

4.2 Entry-Level Wages and Employment

The fact that students are drawn to industries with high skewness can be consistent with both increased labor demand and increased labor supply. In other words, students are attracted to these majors either because a) they rationally anticipate improving job prospects in related industries, or b) they are simply drawn by extreme performance by a few related superstar firms; but this is in fact uninformative about future job opportunities (or less informative than they expect). To examine the relative importance of labor demand vs. supply channels, we simultaneously examine two quantities: wages (inflation-adjusted) and employment. Examining the price-quantity pair allows us to evaluate the relative magnitude of shifts in the labor supply vs. demand curves.

Note that in these wage/employment tests, while our skewness measure is based on related industries that students likely think about while making their major choice decisions, the wage data are from actual occupations that absorb graduates from various majors (from the CIP to SOC Crosswalk). Each occupation code can appear in multiple industries (e.g., computer programmers in the tech industry vs. financial services industry); we take an employment-weighted average across all industries to quantify employment opportunities more accurately. In our baseline result, we focus on entry-level employment and wages for jobs that require a bachelor's degree but no prior work experience.

We examine what happens to job opportunities at the time our year t cohort enters the job market. Specifically, we estimate the following regression equation:

$$log(annual\_wage_{i,t}) = \alpha + \beta LaggedSkew_{i,t-3} + \gamma \mathbf{X}_{i,t-3} + \mu_i + \tau_t + \epsilon_{i,t}, \qquad (2)$$

where $annual\_wage_{i,t}$ is the employment-weighted average entry-level wage for major i in year t. $LaggedSkew_{i,t-3}$ is our employment-weighted pooled skewness measured using all firms related to major i, as explained earlier. We also control for major and time fixed effects in our regressions, so one way of thinking about our empirical design is that we relate the average market-adjusted entry-level wage of the same major across different cohorts to lagged industry return skewness. Other control variables are similar to those in equation (1). In particular, we control for lagged wages of related majors to distinguish our results from predictions of a Cobweb model, and add two additional controls: a) the average number of graduates in related majors in years t-1 and t-2, to ensure that the delayed absorption of previous graduates is not driving our results, and b) the ratio of male to female graduates, to account for changes in gender balance.

Columns (2)-(5) of Table 2 reports results on wages and employment. As shown in Column (2), *LaggedSkew* is significantly and negatively associated with future entry-level wages. A one-standard-deviation increase in lagged industry skewness is associated with a 1.65% (t-statistic = 5.32) lower real wage for entry-level jobs requiring a bachelor's degree. In Column (3), we find an insignificant relation between $LaggedSkew$ and future net new hires. This suggests that even though extreme performance drives more students to related major fields, entry-level job positions do not immediately expand to absorb these extra graduates. This is consistent with labor demand being relatively inelastic in the short-run. In the Online Appendix Table A7, we show that industry turnover, defined as total separations minus hires (as a percentage of employment), is also unrelated to $LaggedSkew$.

In Columns (4) and (5) we conduct an additional test by repeating our analysis in Columns (2) and (3) using data on *advanced* job positions that require a Doctoral or Professional degree or substantial prior work experience. If our results are driven by $Skew$ reflecting changes in demand for labor, we should see a similar pattern with these advanced positions. If, instead, our results are driven by increased supply of fresh graduates, this should mostly affect entry-level jobs, and not advanced positions. Our evidence supports the latter: $LaggedSkew$ is unrelated to wages and net new hires of advanced positions.

Combined, the evidence presented in Table 2 suggests that the presence of a few exceptional firms forecasts larger enrollment in related major fields, and yet lower future wages when these students enter the job market. Such skewness, however, is uncorrelated with the number of entry-level jobs. Put differently, students' decision to enroll in high skewness majors is subsequently met with worse job opportunities.

4.3 Heterogeneity in the Cross Section

4.3.1 Composition Changes in Labor Supply

Our focus so far has been on the shifts in the entire labor supply curve, without explicitly considering the composition of the supply; in other words, we implicitly assume that the *quality/type of students* choosing each major are unrelated to the presence of superstar firms. In this section, we discuss two related variants of the labor-supply channel that allow for changes in supply composition.

The first variant is that less capable students disproportionately select into high-skewness majors. If these students drive up labor supply, then the observed lower entry-level wage partly reflects their lower marginal productivity. We take this possibility to the data and examine whether our results are stronger among students of lower versus higher overall quality.

Our first test looks at college reputation. In Table 3, Panel A we examine dis-aggregated school-level data from four-year universities. We focus on 336 schools that offer at least 5 out of our 10 majors and study whether the effect of superstar firms is any different among top schools, especially those located in states with a significant presence of related industries (e.g., California for Tech jobs). To this end, we construct a dummy variable, *TopSchool*, which equals one if a school is in the top 50 in the *US News Best Colleges* list four years prior to graduation and is located in the state that hires the most people in related industries (e.g., Stanford University for Computer Science), and zero otherwise. Our results show that top school students are especially drawn by local superstar performers. Specifically, the effect of lagged industry skewness on subsequent major enrollment among top schools is more than thrice as large as that for lower-ranked schools.

Our second test exploits wage differences for jobs within a major (e.g., software developers vs. database administrators, for Computer Science majors). Specifically, for each major-year, we separate combinations of occupation codes and industries into those offering wages above and below the sample median in year t-4. The former (latter) group represents higher-(lower-) skilled jobs within the major. We calculate the average wage and change in net hiring in each of these major-wage groups, and add an interaction term between *LaggedSkew* and *Lowskilled*, where *Lowskilled* indicates the lower-wage group. (We further include in the regression specification $Major\text{-}Wage\text{-}group$ fixed effects to subsume (time-invariant) differences between high- vs. low-skilled jobs within each major.) Column (1) of Table 3 Panel B shows that *LaggedSkew* negatively predicts wages of high-skilled occupations but not wages of low-skilled occupations (-1.64% vs. -0.08%). Column (2) shows that *LaggedSkew* is not systematically related to the net new hires in either group.

In sum, the evidence shown above—a) that top school students react more to local superstar firms and b) that the wage decline in relation to superstar firms is present only in high-skilled occupations—suggests that the documented drop in average entry-level wage in relation to superstar performers is unlikely to be a reflection of an influx of less capable students.

Another way to check for selection driving negative effects on wages is to examine whether students drawn to high-skewness majors tend to come from specific backgrounds, typically associated with differential future earnings profiles. We use available information from the *National Survey of College Graduates* (NSCG) on various proxies of SES to assess this possibility. These include information on gender, race, parents' education, besides how survey respondents financed their undergraduate information (for example, student loan amounts or tuition waivers obtained). In Table A8 (in the Internet Appendix), we do not find any evidence of negative selection on SES accompanying the choice of high skewness majors.

A second variant of the labor-supply mechanism is that although students of good overall quality are attracted to majors with superstar firms, the additional students who choose a major attracted by superstar firms have lower major-specific skill (and/or are poorly matched to the field) than the typical student majoring in that field in other years without the superstar effect. One example of this could be where a student who would make a good mechanical engineer now chooses to study computer science. If companies' entry-level hiring technology (how they pick entry-level employees from graduates) is noisy, then they may end up picking some of these lower quality graduates. Since these graduates' have lower major-specific skills, their productivity is lower. Therefore they get lower wages, depressing the overall average wage for that major-year cohort. In other words, the lower average wage we see could be a reflection of lower marginal productivity. (Note that if the hiring technology is perfect, then companies could simply screen out these additional low-quality graduates; and hence this would not be a concern.)

We present a test to examine this hypothesis. Our test is based on the following idea. If the average company's entry-level hiring technology is indeed noisy, and they end up hiring some of these low-quality graduates from high-skewness majors, entry-level wages would be low. But over time these companies should learn about the ability (or match) of these employees, and should start paying more to the high-quality (or better matched) employees and less to the low-quality (or worse matched) ones. This process should increase wage dispersion within the cohort. In sum, if one looks at cohorts that have excess supply of graduates attracted by skewness (the superstar effect), then they should see more wage dispersion in that cohort going forward.

We examine wage dispersion in Table A9 using NSCG data. While we see that the effect of skewness on average earnings is moderated over time, there is no increase over time in wage dispersion for cohorts of high-skewness majors. Our evidence, therefore, does not support this second variant of the compositional change hypothesis either.

Finally, a related possibility is that salient occurrences of superstar firms lead to an overall increase in enrollment in science and engineering. In other words, students that would not go to college or would not choose science and engineering majors are now attracted to do so. In Online Appendix Table A10, we show that neither the max skewness nor the average skewness across all majors predicts total enrollment in science and engineering majors, so this is an unlikely possibility.

4.3.2 Major Versatility

Next, we investigate the notion that the fungibility of employment opportunities is different across majors. Specifically, if graduates from a particular major have employment opportunities in a variety of industries, then part of the excess labor supply can be shared among those industries, leading to less downward pressure on wages in each industry. We test this hypothesis using an interaction term between $LaggedSkew$ and $Versatility$, defined as the Herfindahl index of employment for graduates from a particular major in different industries. Our evidence in the first column of Table 3, Panel C reveals that return skewness negatively affects real wages mostly for majors that have concentrated job opportunities in a small number of industries. (We do not find statistically significant differences in entry-level employment size in Column (2) of the same panel.)

5 Granular Data from Surveys

5.1 National Survey of College Graduates

One important concern about our finding that following superstar firms seems to hurt college students is that while this is true at graduation, job prospects for high-skew majors may actually improve in the longer term. Another concern is that our wage measure does not adequately capture total earnings, which should also include bonuses and other payments, and these payments may be higher for high-skew majors. A third issue is that while we show some results consistent with employment not expanding in the short-term to keep pace with increased labor supply, we have not shown how this excess supply is eventually absorbed by the labor market. Finally, the evidence we have presented is based on aggregate, macro-level data; it is useful to show that our wage results also hold up at the individual level, after controlling for age, gender, and other well-known determinants whose distributions might also shift across cohorts.

We explore all these issues using granular survey data from the NSCG. First, the average age of respondents (63.3% male, 70.1% married) in the NSCG surveys is 43.7 years—roughly 20 years out of college; thus the survey provides useful information on long-term outcomes. Second, NSCG respondents report their total earnings including bonuses and stock grants. Third, NSCG data allow us to explore whether college graduates from high-skew majors have to accept jobs unrelated to their majors. Finally, this dataset is at the *individual* level, allowing us to add various non-parametric controls, e.g., a host of fixed effects for Age, Gender, Marital status, Minority status, Region, Major, Second major (if any), Survey-year, and NSCG Job code for respondent's principal job (e.g., Aerospace, aeronautical or astronautical engineers vs. Electrical or computer hardware engineers).

Table 4 shows our regression results. This table reports results from the National Survey of College Graduates. In Panel A, the first four columns report results from fixed effects panel regressions with log(Earnings) as the dependent variable, while the last 4 columns report results from Ordered Logit regressions of a variable indicating whether the graduate's current job area is from her field of study (which is measured in 3 levels: "closely related", "somewhat related" and "unrelated", with higher values indicating decreasing relatedness). Job unrelatedness to UG major is defined directly for those who are either only graduates (no higher degrees), or those who got higher degrees but in the same field as their UG major. We exclude respondents working outside their fields of study if the respondent switched industry due to changes in interest or family-related reasons or retirement (6307 respondents). We also report below each column the impact of our variables of interest on the Odds Ratios for all ordered logit regressions. Standard errors are clustered by Job code for the respondent's principal job, to account for the fact that opportunities could be correlated within similar job functions across years.

As can be seen from Column (1), one-standard-deviation higher *LaggedSkew* is associated with 1.9% lower total earnings (comparable to the effect on starting wages in Table 2). This is consistent with adverse initial labor market conditions affecting long-term income of college graduates (also see Oyer, 2006, 2008; Oreopolous et al. 2012). In sum, the wage effect we document earlier is neither short-lived, nor is it an artifact of leaving out non-wage compensation. Columns (2)–(4) show that this long-term earnings effect is robust to excluding health and bio majors, looking only at those with bachelors degrees, and to high dimensional fixed effects controlling for Highest degree attained and Highest degree field (see, e.g., Macpherson and Hirsch, 1995; Johnson and Neal, 1997; Dickson, 2010).

The last four columns of Table 4, Panel A report the odds ratios from Ordered Logit regressions where the dependent variable reflects how closely the graduate's current job is related to her field of study in college, with higher values indicating decreasing relatedness. Holding other variables constant, we find that a one-standard-deviation higher LaggedSkew is associated with a 4.5–5.7% higher odds of working in a job outside the field of study (versus combined odds of working in a related or somewhat related field), indicating the absorption of excess supply of graduates by other industries.

In Panel B, we examine the consequences of having a job unrelated to field of study. The first four columns examine log(Earnings) as the dependent variable, while the next four columns examine Job satisfaction. The sample and fixed effect specifications in each column of Panel B are analogous to those in Panel A. We find significant negative career outcomes associated with having to take a job outside one's field: earnings are lower by 17.3–20.5%, and the odds of reporting one-notch lower job satisfaction are higher by 75–84.2% for those working in less related sectors.

5.2 Evidence from Our Own Survey

In this section, we provide evidence from a survey of college graduates we conducted on the SurveyMonkey platform, specifically designed for the purpose of this study. While the NSCG survey shows that students influenced by superstar performers are more likely to work in industries outside their fields of study, that evidence may not necessarily reflect oversupply of graduates. For example, some engineering students may have hoped to work for investment banks when choosing their major. Further, we do not know from NSCG whether a graduate's first job was in a different industry, or whether she changed her jobs later due to a change in interest or learning about job contents.[19]

Moreover, the survey allows us to contrast two behavioral mechanisms underlying student attraction to superstar-related majors: a) belief errors—students form income expectations (or more generally, expectations of future successes) based on a small number of non-representative but highly visible observations; and b) skewness preference—students are happy to accept a lower average wage for a small chance of hitting a jackpot.

We used SurveyMonkey to conduct this survey.[20] We selected College graduates between the ages of 22 to 65, employed full time in the United States, with a major in our list of NSF majors, and who thought that the availability of jobs or future income prospects in related industries was at least "Somewhat Important" in their major choice decision. SurveyMonkey sent the survey to 1200 people enlisted on their platform; out of which we ended up with 394 respondents that met our selection criteria.[21] In our sample, 49.2% of respondents are male and the median age group of survey respondents is 30–44. The median household income is in the \$75,000–\$99,999 range (demographic information on respondents– such as gender, age, region and income brackets – was provided to us by SurveyMonkey, as background information on our sample). The median income in our sample is in the same range as that in other comprehensive national statistics of Science and Engineering graduates (e.g., surveys by NSF). Still, given that SurveyMonkey respondents have clearly chosen to join the platform and to answer our survey for a few dollars, we do not think that they are representative of the general graduate population; instead, we only compare one respondent to another *within* this dataset. SurveyMonkey also asked respondents about the geographic region they were in; our sample contains people from all of the nine regions of the US (East North Central, East South Central, Middle Atlantic, Mountain, New England, Pacific, South

[19] In our sample, 26.2% of respondents change industries after graduation, similar to the figure reported by Ellul, Pagano, and Scognamiglio (2020).

[20] This survey was designed in May-June 2018 and conducted in July 2018, and there was no pre-analysis plan for it at the time.

[21] This sample size was primarily limited by research budget constraints.SurveyMonkey charged approximately 3 GBP per respondent, and it took about one week to get the responses back.

Atlantic, West North Central, and West South Central). The exact sampling distribution of these statistics are presented in Panel A of Table A14 in the internet appendix.[22]

Note that we did not directly ask our survey respondents the exact superstar firms that triggered their major choice decisions. This was a conscious choice on our part, because many of these respondents chose their major years back. So when asked to name superstar firms that motivated their major choice, they might not mention the firms that were superstars at the time of major choice but failed subsequently. For example, one could have been motivated by Nokia or Blackberry in the early 2000s, or Kodak in the 1990s, but when asked to name their dream firms years later, they might name Apple. To avoid such issues, we only asked respondents about their majors, and the years (e.g., sophomore vs freshman vs high school) when they made the major-choice decision. Using that information, we back out the relevant major skewness for each respondent. For example, if the respondent graduated in 2014 and chose her major in the freshman year (i.e., in 2010-11), we examine cross-sectional return skewness for that major measured from returns to related firms in 2010 and 2011.

In all regressions we control for Major, Graduation-year, Importance of student debt repayment (measured in four levels, including one for not having any such debt; see, e.g., Chakrabarti et al., 2020), Gender, Region, When major was chosen (e.g., freshman vs. sophomore year), and the importance of career prospects when choosing major (measured in two levels) through fixed effects. We also control for Industry, appropriately defined depending on context (that is, as current industry for current household income regressions, and target or first-job industry for outcomes pertaining to the time of graduation, e.g., for the dependent variable measuring the propensity of first job being in respondent's target industry). We report marginal effects for all logit regressions, and also report below each column the impact of our variables of interest on the Odds Ratios for all ordered logit regressions. Standard errors are clustered by industry.

Table 5 reports results from this survey. In Panel A, Columns (1) and (4) report results from an ordered logit regression with Household income (in 8 buckets: 7 buckets of size $25,000, starting from $25,000, and the last bucket for income above $200,000) as the dependent variable. Column (2) reports results from a logit regression with a dummy dependent variable indicating whether the graduate started her post-College career in an industry she was expecting to work in when she chose her major. In Column (3) we examine a dummy dependent variable, reflecting regret about major choice: it takes a value of 1 if respondents thought that they could have chosen a more suitable major had they done more research, and thereby formed more accurate expectations regarding future job/income outcomes. In

[22] The survey instrument is available at `https://personal.lse.ac.uk/loud/ChoiLouMuk_SurveyQuestions.pdf`.

Column (5) we examine a dummy dependent variable, which is 1 if the graduate currently works in an industry that is different from the industry she started in after graduation. Starting job not in target industry is a dummy variable indicating whether the graduate's first job was not in an industry she was targeting.

We find that the choice of a high-skewness major is associated with lower household income, as reported to SurveyMonkey (column (1)), a 6.9% lower probability of the respondent's first job being in the industry that they had targeted in college (column (2)), as well as a 4.6% higher probability of the graduate admitting sub-optimal major choice due to them having incomplete information about job opportunities while selecting their major.

Columns (4) and (5) examine consequences of the respondent's first job not being in their target industry. We find that this is associated with both lower household income, as well as a 13.6% higher likelihood of changing their job to yet another industry.

In Panel B, we examine two behavioral channels that could potentially drive respondents' decisions to follow superstar firms, as reflected by their choice of a high-skewness major. The dependent variable here is Skew, the return skewness associated with the major chosen by each respondent. The key explanatory variables are: (i) *Decision based on small number of observations*: takes a value of 1 if, before choosing her major, the respondent gathered information about a small number of firms that were doing well from the news media or friends & relatives and used that to form overall expectations about job prospects in that industry; zero otherwise. (ii)*Lottery_Preference*: takes a value of 1 if the graduate answered that she would have chosen to play a fair lottery over a stable, future income stream; zero otherwise.

First, we find that respondents indeed confirm making errors in expectations. Panel B of Table A14 in the internet appendix presents summary information on key survey answers. We find that 44.8% of respondents say that they had chosen their major based on information only about a small number of firms employing graduates from that major – rather than finding out job prospects from a broader set. Similarly, 56.4% of respondents say that their peers' actual performance did not conform to what they had expected about prospects from their major. When we ask about Expectation Errors due to incomplete information collection – that is whether their perception of career prospects could have been more accurate with a bit more research – 64.9% say yes. On the other hand, only 29.6% of respondents say that they would have picked a lottery-type payoff over a stable, average income stream.

Next, we examine the association between the choice of high-skewness majors and making such expectations errors. Column (1) of Panel B shows that respondents who decided about their major based on a small number of firms were significantly more likely to have chosen a high skewness major. However, when we examine *Lottery_Preference* in Column (2), we do not find any significant relation to high-skewness major choice. The last column shows that including both these variables in the same regression does not change these conclusions.

Overall, this evidence suggests that belief errors, rather than inherent preferences for skewness, are associated with individuals' decisions to choose high-skewness majors in college.

6 Discussions of Our Results

6.1 Further Discussions of Labor Demand

As discussed earlier, the increase in major enrollment associated with high skewness majors may also be consistent with increased labor demand; that is, occurrences of superstar performers may be indicative of brighter future industry prospects, and students choose related major fields in anticipation of more and better job opportunities. This labor-demand mechanism, however, is unable to account for our findings a) that return skewness negatively forecasts entry-level wages when students enter the job market, b) that high skewness is not followed by an increase in entry-level hiring, and c) survey evidence that a disproportionate number of graduates in superstar-related majors have to accept jobs in industries different from where they wanted to join, which is costly in terms of both lower wages and job dissatisfaction.

Further, we show in Online Appendix Table A7 that return skewness is not correlated with past operating performance of the industry, nor does it predict future performance. In that table, we find that industry skewness is unrelated to Return on Assets, Return of Equity, Net Profit Margins, or Sales Growth in years t-3, t, t+3 or t+6 (with t being the year of graduation).

One could be still concerned that our skewness measure is proxying for industry conditions that are not captured by operating performance. For example, it could be that high industry skewness is associated with better growth prospects for that industry years or even decades later, and this is not reflected in operating performance six years out. However, stock prices, and hence returns, are forward-looking, and should incorporate all available future information. So one way to address this concern would be to replace the dependent variable of our baseline regressions in Table 2 with average stock returns and return volatility. As we show in Table A4, skewness in either t-1 to t-2 or in t-3 to t-7 is not correlated with future stock return or with future return volatility.

Changes in the Composition of Labor Demand A more nuanced version of the labor-demand mechanism is that occurrences of superstar performers are associated with changes in the composition of labor demand. More specifically, skewness may forecast higher demand for low-quality employees but lower demand for high-quality employees, thus keeping the total employment constant while driving down entry-level wages. First, this mechanism is inconsistent with our earlier finding that both entry-level net hiring and wages for low-skilled jobs are unrelated to major skewness. Moreover, it cannot explain why a larger number of students (especially those from elite universities) choose to major in related fields in response to industry return skewness, knowing that future job opportunities are worse.

Changes in Bargaining Power Another related possibility is that return skewness reflects changes in the industry structure, which impacts firm bargaining power vis-a-vis employees. Specifically, superstar performers gain market power as they grow, which they then exploit to negotiate wages down. First, this mechanism does not explain the finding that industry return skewness negatively predicts entry-level wages, but not wages earned by occupations in the same industry that require prior experience. Second, and more importantly, this mechanism also fails to explain why more students are attracted to related major fields by superstar firms, knowing that their job prospects are going to be worse.

In sum, our evidence points to a strong relation between high skewness majors and a subsequent increase in labor supply in related industries. One way to interpret this relation is through the lens of salience: superstar performers are salient, which attracts students to related majors. Alternatively, it could be that "public interest" (e.g., during the internet bubble and amidst the ongoing climate crisis) leads to the creation of superstar firms in related industries, and at the same time draws students to related majors. Similarly, it might be parents or school administrators – rather than students themselves – whose attention is drawn by superstar firms. The common element across these different mechanisms is that industries with superstar firms attract attention in a way that affects students' major choice, and that the resultant increase in major popularity, is not fully justified by labor market conditions at graduation (or even a couple of decades later), as evidenced by our results on lower wages/incomes.

6.2 Alternative Measures of Superstar Performers

6.2.1 An Intuitive Measure of Skewness

The main skewness variable we use follows the standard definition in the literature, designed to capture the presence of outliers in a smooth and continuous fashion exploiting the entire distribution. To provide more intuition for skewness, we use an alternative, more discrete definition of distribution asymmetry: the distance between the right tail of a distribution and its median, minus the distance between the left tail and the median. More formally, we define $tail_N = (|top_N - median| - |bottom_N - median|)/(90^{th}percentile - 10^{th}percentile))$.[23] Take $N = 1$ for example, a large and positive $tail_1$ indicates that the best performing firm in the industry does spectacularly better than the median firm while the worst performing firm only mildly worse—which intuitively captures our definition of an industry with superstar performers.

The usefulness of this alternative measure is its ease of interpretation (e.g., Green and Hwang, 2012). The tail measure also allows us to study the top (firms that did exceptionally well) vs. bottom (firms that did exceptionally poorly) separately. The correlation between this alternative tail measure and our baseline skewness measure is 0.44.

We repeat our analyses in Table 2 using $tail_N$ in place of cross-sectional skewness. Again, the measure is constructed using return data from t-7 to t-3. The results are shown in Table 6, Panel A.[24] Column (1) shows that this alternative measure positively predicts the number of bachelors. A one-standard-deviation increase in $tail_N$ forecasts a 14.6% (t-statistic = 2.73) higher number of graduates in related major fields. Column (2) shows that the effect of $tail_N$ mainly operates through superstar firms ($top_N - median$) attracting more students, and less so through super losers ($bottom_N - median$) repelling students (the difference between the coefficients on $top_N - median$ and $bottom_N - median$ is significant at the 1% level).[25] The evidence in Column (3) suggests that a one-standard-deviation increase in our return asymmetry measure predicts 2.79% (t-statistic = 3.24) lower entry-level wages in related industries when the cohort graduates. Column (4) shows that this wage effect—in line with the effect on the supply of graduates in Column (2)—is also coming mostly from superstars, and not super losers. As before, $tail_N$ does not predict future changes in entry-level employment size in related industries (Columns (5) and (6)).

6.2.2 Skewness in Media Tones

Until now, we measure salient occurrences of superstar firms using cross-sectional return skewness. As mentioned earlier, this is not to suggest that high school students, or first and second year college students, follow the stock performance of all firms on a regular basis.

[23] We exclude firms below the 50th percentile of the size distribution when selecting the top and bottom N firms in each industry so that the measure is not dominated by small firms.

[24] In our baseline result, we pick $N = 10$ to reduce noise in the measure, but as shown in Online Appendix Table A2, our results are similar but statistically weaker for other choices of N (e.g., 1, 3, 5).

[25] The Lagged $bottom_{10}$ measure is more negative in industries with more extreme losers; so the positive coefficient means that super losers drive students away. While our results here indicate that the bottom firms also have an effect on major choice, these results are sensitive to the specification choice (unlike the results for top firms).

Indeed, we think of *LaggedSkew* as a proxy for extreme performance by related firms that draw students' attention and shape their expectations and decisions.

Here, we utilize the idea that such exceptional performance is also likely to be accompanied by significant media attention. Thus, an alternative way of capturing superstar performance by a handful of firms is to exploit variation in media coverage and tones. Here, we use a media-coverage score supplied by RavenPack, as discussed in Section 3.5.

We then run regressions similar to equation (1) but replace return skewness with media skewness, and report these results in Table 6, Panel B. In Column (1), we find that media skewness positively predicts major choice; a one-standard-deviation increase in *Lagged News Skew* is associated with 14.9% (t-statistic = 3.86) more graduates in related majors in year t. In Columns (2) and (3), we show that a one-standard-deviation increase in news skewness is associated with a 1.12% (t-statistic = 2.49) lower entry-level wage, but is not significantly associated with net new hires.

6.2.3 Regional Skewness

In our baseline specification, we do not use regional variation of skewness because students might also be motivated by firms outside their local regions in choosing their majors. For example, MIT students might be affected by the extreme success of a few tech firms in the Silicon Valley (e.g., Facebook) in their major choice.

In Appendix Table A11, we conduct an analysis exploiting major-region variation, using the school-level data provided by IPEDS (note that neither the NSF aggregate graduate data nor the BLS wage data contain major-regional information). Regional Skew is cross-sectional return skewness calculated using firms in the school's region. The region is one of nine census regions: East North Central, East South Central, Middle Atlantic, Mountain, New England, Pacific and US Territories, South Atlantic, West North Central, and West South Central. We control for major-year fixed effects and clustered standard errors at the major-region level. Top school is an indicator that the school is in the top 50 in the *US News Best Colleges* list 4 years prior to graduation. The results remain similar to our baseline specification in Table 3 Panel A.

6.2.4 An Alternative Map Based on SurveyMonkey Responses

Our major-industry map in Table A1 is based on ex-ante target industries that are likely to motivate a student to choose that major. It is possible to use an alternative map based on the mapping used to assign wages (CIP-SOC crosswalk), and the other based on the current job industries as reported in the NSCG survey. Both, however, have severe limitations precisely because they are ex-post maps. For example, a lot of computer science majors go to the financial industry (ex-post), but when students choose computer science as a major, they are more likely to target the tech industry than the financial industry.

Fortunately, we had asked about the respondent's college major and his/her target industry in our own survey on SurveyMonkey. Using that data, we find a high degree of concordance between our original major-industry map and what respondents report on SurveyMonkey. For example, we find that over 70% of the target industries reported by survey respondents match our original major-industry map.

We also show a robustness table (Table A12) in the appendix where we use an alternative major-industry map based on this SurveyMonkey data to calculate return measures (skewness, mean return, and standard deviation of return) and other controls. In that table, we only keep the major-industry links from our baseline map that correspond to the top-3 *target* industries of students from that major based on our survey. We find that our results are robust to this alternative map.

6.3 More Evidence on Industry Salience

In this section, we provide further evidence for the role of industry salience in driving student major choice. Our test is motivated by the analysis in Charles, Hurst, and Notowidigdo (2018). We exploit structural breaks in industry valuation during the NASDAQ bubble in the late 1990s to identify superstar industries. The results are reported in Online Appendix Table A13. Not surprisingly, Computer Science-related industries experience the largest structural break in industry valuation among all science-engineering majors in our sample; moreover, the size of the structural break is significantly and positively associated with subsequent changes in major enrollment.

7 Conclusion

This paper studies the relation between superstar firms and college major choice. Using cross-sectional skewness in stock returns, as well as that in favorable news coverage, to capture the occurrences of superstar performers in each industry, we show that such occurrences are associated with larger college enrollment in related fields. However, students attracted by superstar performers earn lower real wages upon entering the job market. Coupled with the finding that entry-level hiring does not vary with the presence of superstar performers, the wage result is consistent with the view that an increase in labor supply, without an accompanying shift in the labor demand curve, lowers the average wage earned by entry-level employees without affecting employment size.

Further evidence from both the National Survey of College Graduates and our own survey indicates that many graduates from these high-skewness majors have to take up jobs in fields outside their target industries at graduation. Moreover, these adverse career outcomes last for decades: cohorts drawn into major fields by superstar performers earn lower wages 20 years after graduation, and have a lower probability of working in fields related to their college majors. In sum, our paper is the first to provide evidence on the role played by superstar firms in driving one of the most important and irreversible decisions in life—human capital investment. Given the rise of such superstars across a diverse array of sectors, and across various countries worldwide, our results bring to attention another way in which they have a crucial impact on people's lives.

References

Agarwal, Sumit, Qian, Wenlan, Sing, Tien Foo, and Tan, Poh Lin, 2019, "Dragon Babies," National University of Singapore Working Paper.

Altonji, J.G., 1993, "The demand for and return to education when education outcomes are uncertain," Journal of Labor Economics, 11(1), pp.48-83.

Altonji, J. G., Blom, E. and Meghir, C., 2012, "Heterogeneity in human capital investments: High school curriculum, college major, and careers," Annual Review of Economics 4, 185-223.

Altonji, J.G., Arcidiacono, P. and Maurel, A., 2016, "The analysis of field choice in college and graduate school: Determinants and wage effects," in Handbook of the Economics of Education, Elsevier.

Arcidiacono, P., Hotz, V. J. and Kang, S., 2012, "Modeling college major choices using elicited measures of expectations and counterfactuals," Journal of Econometrics 166, 3–16.

Autor, D., Dorn, D., Katz, L.F., Patterson, C.,, Van Reenen, J., 2020, "The fall of the labor share and the rise of superstar firms," Quarterly Journal of Economics 135, 645-709.

Avery, Christopher and Caroline M. Hoxby, 2004, "Do and Should Financial Aid Decisions Affect Students' College Choices?" in Caroline Hoxby, ed. College Choices: The New Economics of Choosing, Attending, and Completing College. University of Chicago Press.

Barberis, N. and M. Huang, 2008, "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices," American Economic Review 98, 2066–2100.

Bena, J. and Simintzi, E., 2019, "Machines Could Not Compete with Chinese Labor: Evidence from U.S. Firms' Innovation," Working Paper.

Berger, M.C., 1988, "Predicted future earnings and choice of college major," Industrial & Labor Relations Review, 41, 418–429.

Bhattacharya, J., 2005, "Specialty selection and lifetime returns to specialization," Journal of Human Resources 15, 115–143.

Blom, E., 2012, "Labor market determinants of college major," Yale University Working Paper.

Bordon, P. and Fu, C., 2015, "College-major choice to college-then-major choice," Review of Economic Studies, 82, 1247-1288.

Bordalo, P., Gennaioli, N. and Shleifer, A., 2012, "Salience theory of choice under risk," Quarterly Journal of Economics 127, 1243–1285.

Bordalo, P., Gennaioli, N. and Shleifer, A., 2013a, "Salience and consumer choice," Journal of Political Economy 121, 803–843.

Bordalo, P., Gennaioli, N. and Shleifer, A., 2013b, "Salience and asset prices," American Economic Review 103, 623–628.

Chakrabarti, Rajashri, Vyacheslav Fos, Andres Liberman and Constantine Yannelis, 2020, "Tuition, Debt, and Human Capital," Working paper.

Charles, Kerwin Kofi, Erik Hurst and Matthew J. Notowidigdo, 2018, "Housing Booms and Busts, Labor Market Opportunities, and College Attendance," American Economic Review, 108, 2947–94.

Chetty, Raj, Adam Looney, and Kory Kroft, 2009, "Salience and Taxation: Theory and Evidence," American Economic Review 99(4): 1145–77.

Cosemans, Mathijs, and Rik Frehen, 2021, "Salience theory and stock prices: Empirical evidence," Journal of Financial Economics 140, no. 2, 460-483.

Dang, T.L., Moshirian, F. and Zhang, B., 2015, "Commonality in News Around the World," Journal of Financial Economics 116(1), 82–110.

Dickson, L., 2010, "Race and gender differences in college major choice," The Annals of the American Academy of Political and Social Sciences, 627, 108–124.

Ellul, Andrew, Marco Pagano, and Annalisa Scognamiglio, 2020, " Careers in Finance," Working Paper.

Fehr, Ernst, and Antonio Rangel, 2011, "Neuroeconomic Foundations of Economic Choice—Recent Advances," Journal of Economic Perspectives 25(4): 3–30.

Freeman, R.B., 1971, "The Labor Market for College-Trained Man-power" (Cambridge: Harvard University Press).

Freeman, R.B., 1975, "A Recursive Model of the Market for New Lawyers," The Review of Economics and Statistics, 57, 171–179.

Freeman, R.B., 1976, "A cobweb model of the supply and starting salary of new engineers," Industrial & Labor Relations Review, 29, 236–248.

Fricke, H., Grogger, J., Steinmayr, A., 2015, "Does exposure to economics bring new majors to the field? Evidence from a natural experiment," NBER Working Paper.

Gabaix, Xavier, 2014, "A sparsity-based model of bounded rationality", The Quarterly Journal of Economics, 129(4), pp.1661-1710.

Goldin, C., 2014, "A grand gender convergence: Its last chapter," American Economic Review 104, 1091–1119.

Gupta, N. and Hacamo, I., 2018, "(Mis)matching Superstar Engineers to Finance Jobs," Working Paper.

Greenwood, Robin, and Hanson, Samuel, 2015, "Waves in ship prices and investment," The Quarterly Journal of Economics, 130, 55–109.

Han, B., Hirshleifer, D., and Walden, J., 2022, "Social Transmission Bias and Investor Behavior," Journal of Financial and Quantitative Analysis, 57(1), pp.390-412.

Hastings, J.S., Neilson, C.A., Ramirez, A. and Zimmerman, S.D., 2016, "(Un) informed college and major choice: Evidence from linked survey and administrative data," Economics of Education Review, 51, 136-151.

Hirshleifer, David, 2019, "Attention, Psychological Bias, and Social Interactions," Finance Theory Working Group Paper, Wharton.

Hirshleifer, D. and Plotkin, J.B., 2021, "Moonshots, investment booms, and selection bias in the transmission of cultural traits," Proceedings of the National Academy of Sciences, 118(26).

Hombert, J. and Matray, A., 2019, "Technology Boom, Labor Reallocation, and Human Capital Depreciation," HEC Paris Working Paper.

Hoxby, Caroline M., ed., 2003, "The Economics of School Choice," Chicago: University of Chicago Press.

Hoxby, Caroline M., ed., 2004, "College Choices: The Economics of Where to Go, When to Go, and How to Pay for It," Chicago: University of Chicago Press.

Huntington-Klein, Nick, 2016 "(Un) informed College and Major Choice: Verification in an alternate setting." Economics of Education Review 53: 159–163.

James, E., Alsalam, N., Conaty, J. C., To, D. L., 1989, "College quality and future earnings: Where should you send your child to college?" American Economic Review 79, 247–252.

Johnson, W.R. and Neal, D.A., 1997, "Basic skills and the black-white earnings gap," University of Virginia, Thomas Jefferson Center for Political Economy.

Kahneman, D., 2011, "Thinking, fast and slow," Macmillan.

Kahneman, D. and A. Tversky, 1979, "Prospect Theory: An Analysis of Decision Under Risk, Econometrica 47, 263–291.

Kaldor, Nicholas, 1934, "A Classificatory Note on the Determination of Equilibrium," Review of Economic Studies, 1, 122–136.

Lemieux, T., 2014, "Occupations, fields of study and returns to education," Canadian Journal of Economics 47(4), 1047–1077.

Macpherson, D.A. and Hirsch, B.T., 1995, "Wages and gender composition: why do women's jobs pay less?" Journal of labor Economics, 13(3), pp.426-471.

Malmendier, U. and Tate, G., 2009, "Superstar CEOs," Quarterly Journal of Economics, 124(4), pp.1593-1638.

Oreopoulos, Philip, Till von Wachter, and Andrew Heisz, 2012, "The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates," American Economic Journal: Applied Economics 4, 1–29.

Ouimet, Paige and Rebecca Zarutskie, 2014, "Who Works for Startups? The Relation between Firm Age, Employee Age and Growth," Journal of Financial Economics 112, 386-407.

Oyer, Paul, 2006, "The Macro-Foundations of Microeconomics: Initial Labor Market Conditions and Long-Term Outcomes for Economists," National Bureau of Economics Working Paper 12157.

Oyer, Paul, 2008, "The Making of an Investment Banker: Stock Market Shocks, Career Choice, and Lifetime Income," Journal of Finance, 63, 2601–28.

Patnaik, A., Venator, J., Wiswall, M. and Zafar, B., 2020, "The role of heterogeneous risk preferences, discount rates, and earnings expectations in college major choice," Journal of Econometrics.

Rosen, S., 1981, "The economics of superstars," American Economic Review, 71(5), pp.845-858.

Rosen, S, 1997, "Manufactured Inequality," Journal of Labor Economics 15, 189–196.

Sacerdote, B., 2001, "Peer effects with random assignment: Results for Dartmouth roommates," Quarterly Journal of Economics, 116, 681–704.

Stigler, George J., 1961, "The economics of information," Journal of Political Economy, 69, 213–225.

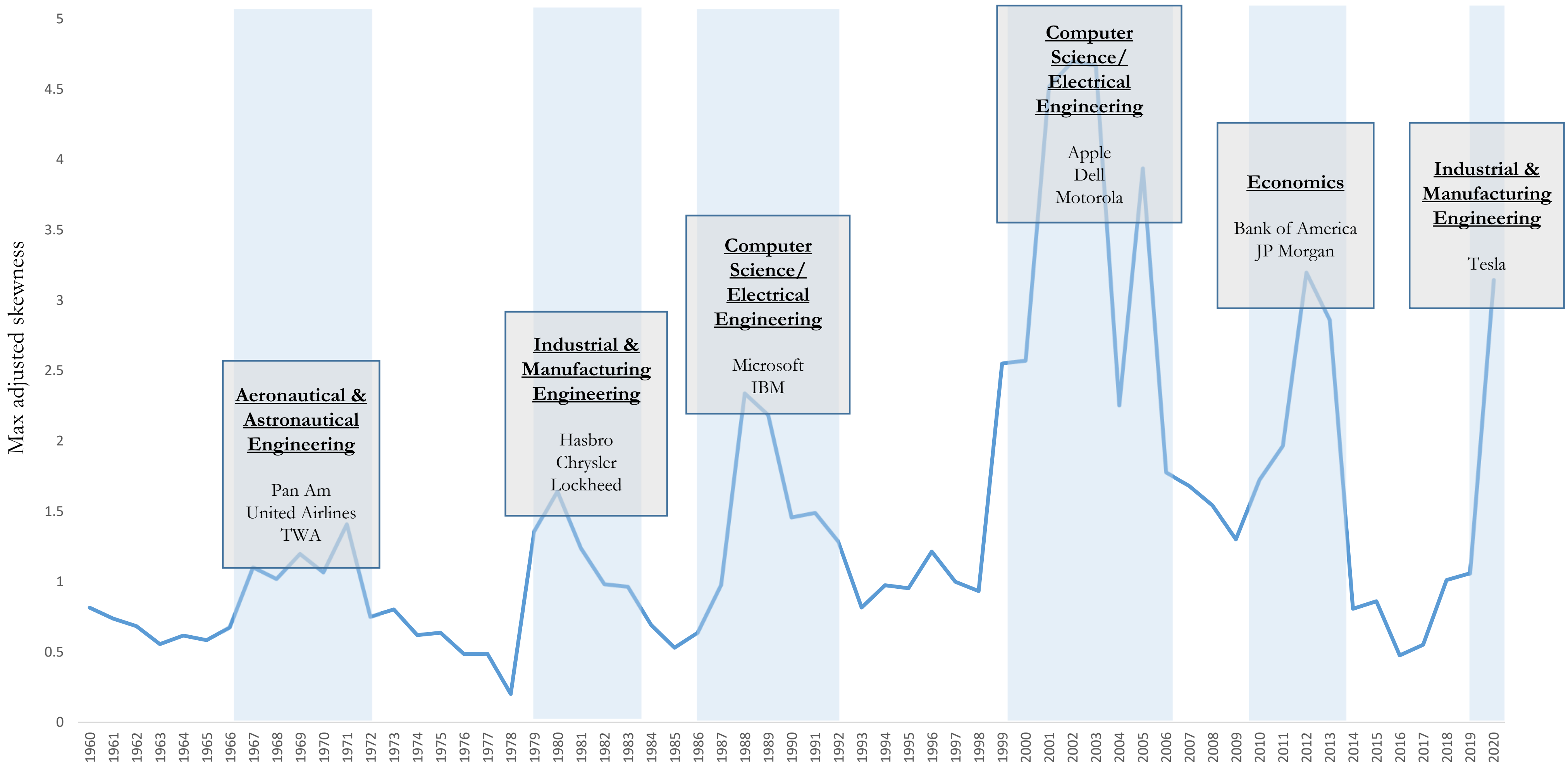
Stigler, George J., 1962, "Information in the labor market," Journal of Political Economy, 70, 94–105.

Stinebrickner, T. and Stinebrickner, R., 2014, "A major in science? Initial beliefs and final outcomes for college major and dropout," Review of Economic Studies 81, 426–472.

Taylor, S.E. and Thompson, S.C., 1982, "Stalking the elusive "vividness" effect," Psychological review, 89, 155.

Wiswall, M. and Zafar, B., 2015, "Determinants of college major choice: identification using an information experiment," Review of Economic Studies 82, 791-824.

Wiswall, M. and Zafar, B., 2022, "Human Capital Investments and Expectations about Career and Family," Journal of Political Economy, forthcoming.



This figure shows majors with the maximum five-year skewness after adjusting for year fixed effects, and the mean and standard deviation of returns. Shaded areas reflect episodes of high relative skewness. The firms indicated are among the top-3 contributors to the skewness of the related major during that period, calculated as the difference between the skewness of its major with and without that firm.

Figure 1. Superstar firms, and corresponding high skewness majors

Figure 2
Number of Bachelors Around High Skewness Events

This figure shows the adjusted Log Number of Bachelors around high skewness events. The treatment group includes majors that are popular, if the adjusted skewness of the major in Figure 1 is higher than the sample median. The event year 0 is the year when the adjusted skewness starts increasing. Log Number of Bachelors are adjusted for year and major fixed effects. A control group of majors is constructed by using majors that are not in the top 5 adjusted skewness around the event year. The graph shows the average difference between the treatment group and the control group in each year. The dotted lines refer to the 95% confidence intervals.

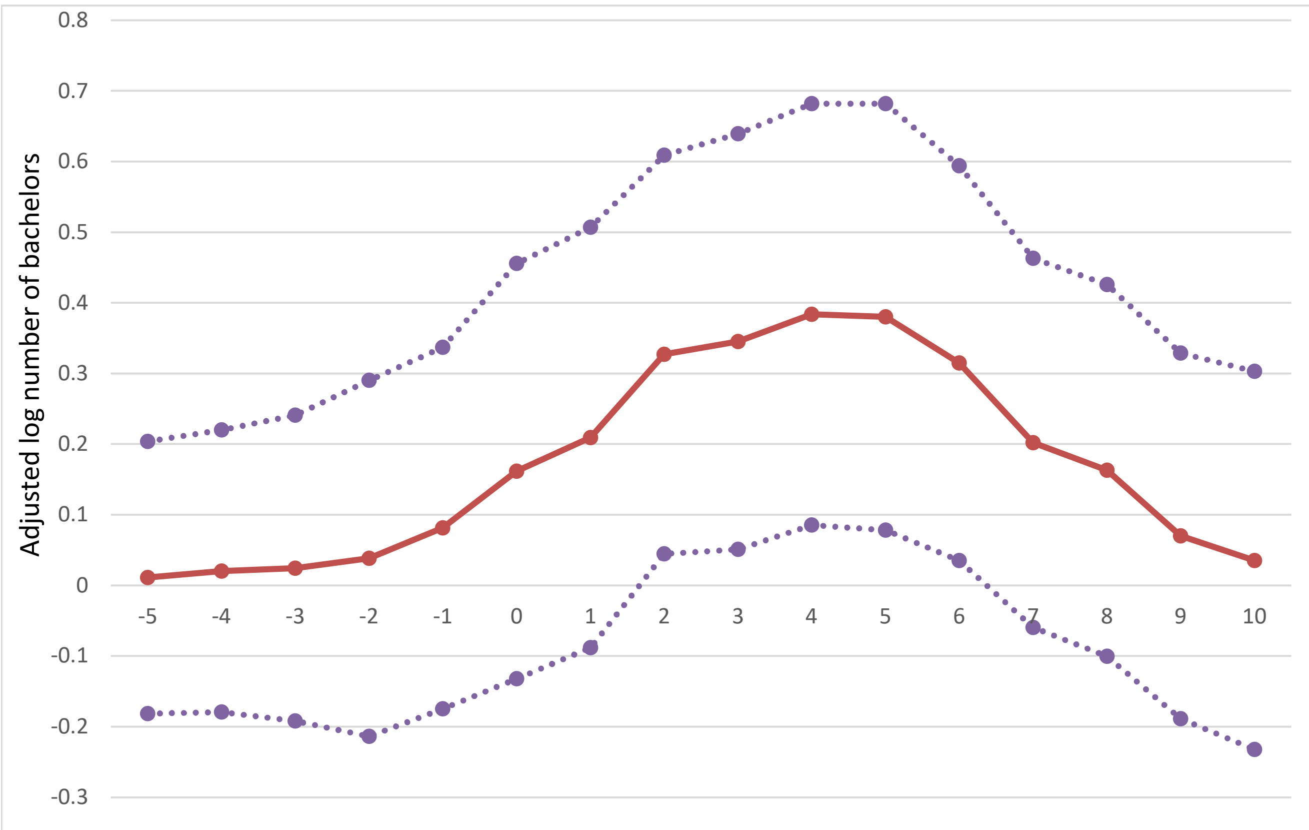


Table 1
Data and Summary Statistics

This table describes the data in Panel A and provides summary statistics of our major variables in Panel B. Number of Bachelors is the annual number of bachelor degrees awarded for a major. Skew is the cross-sectional skewness of annual returns in all firms that are mapped to a major, weighted by the number of employees of firms. $Tail_{10}$ is the sum of Top_{10} and $Bottom_{10}$. Top_{10} ($Bottom_{10}$) is the average return of the top (bottom) 10 firms among all firms that are mapped to a major minus the median return, divided by the difference between the 90th and 10th return percentile, after dropping firms in the lowest 50th size percentile. News Skew is the employment-weighted cross-sectional skewness of annual sum of RavenPack ESS scores (the ESS scores are rescaled to -1 to 1). Mean Return and Standard Deviation of Return (based on monthly returns), Average Market Cap, Average Book-to-Market, and Average Firm Age are weighted by the number of employees of firms. Firms are mapped to majors using the map in Appendix Table A1. Annual Wage is the employee-weighted average wage across occupation codes that are mapped to the major and that require bachelor s degree and do not require prior experience, inflation-adjusted (1997 level). Net new hires is the net new hires in these positions.

Panel A: Data

Dataset	Sample Period	Coverage	Data Used
NSF	1966-2017	US accredited institutions of higher education	Number of bachelor's degrees in science and engineering fields
CRSP, Compustat	1959-2017	US public firms	Stock and firm information
RavenPack	2000-2017	US public firms	Firm-level news sentiment
BLS	1997-2017	US employees	Aggregate wage and employment
National Survey of College Graduates	1993-2017	College graduates in the US	Individual-level wage and employment

Panel B: Summary Statistics

	Sample Period	Data Source(s)	Median	25th Pctl	75th Pctl	Std Dev	# Obs
Number of Bachelors	1966-2017	NSF	6,921	3,298	16,196	11,314	520
Skew	1959-2017	CRSP, Compustat	1.386	0.553	2.488	2.383	520
$Tail_{10}$ (Top_{10}+$Bottom_{10}$)	1959-2017	CRSP	0.567	0.221	0.951	0.588	520
Top_{10}	1959-2017	CRSP	1.383	0.839	1.735	0.690	520
$Bottom_{10}$	1959-2017	CRSP	-0.711	-0.856	-0.502	0.267	520
News Skew	2000-2017	RavenPack	4.845	1.701	8.073	3.612	160
Mean Return	1959-2017	CRSP, Compustat	0.013	0.009	0.017	0.007	520
Standard Deviation of Return	1959-2017	CRSP, Compustat	0.058	0.048	0.071	0.022	520
Average Market Cap ($ mil)	1959-2017	CRSP	2,019	775	6,025	5,601	520
Average Book-to-Market	1959-2017	CRSP, Compustat	0.574	0.409	0.835	0.361	520
Average Firm Age	1959-2017	CRSP	26	20	34	9	520
Annual Wage (1997 Dollars)	1997-2017	BLS	58,070	53,346	64,994	7,818	210
Net New Hires	1998-2017	BLS	0.0231	-0.0130	0.0491	0.2917	200
Number of Net New Hires	1998-2017	BLS	3,102	-658	12,070	256,172	200

Table 2
Aggregate Number of Graduates and Employment

This table reports the results of regressions of aggregate number of graduates and employment on skewness measures (measured in years t-7 to t-3, relative to the graduation year t). Number of Bachelors is the annual number of bachelor degrees awarded for a major. Annual Wage is the employment-weighted average wage across occupation codes that are mapped to the major and that require bachelor s degree and do not require prior experience, inflation-adjusted (1997 level). Net New Hires is the log net new hires in these positions. In Columns (4) and (5), we examine advances positions that typically require a Doctoral or Professional degree or substantial prior work experience. Skew is the employment-weighted cross-sectional skewness of annual returns in all firms that are mapped to the major. Our control variables are all measured at year t-3 and include Log Average Wage, the 3-year average wage obtained from Compustat (up to 1998) or from BLS (1999 and onward); Mean Return and Standard Deviation of Return, both are employment-weighted. Other controls are Log Average Market Cap, Log Average Book-to-Market, and Log Average Firm Age, weighted by employment. In Columns (2) through (5), we also control for Log Number of Bachelors, the log annual number of bachelor degrees awarded for a major, averaged across years t-1 to t-2; and Lagged Male/Female Ratio, the ratio of male to female graduates in the major in years t-1 to t-2. Standard errors are clustered at the year level. All independent variables are standardized with zero mean and unit standard deviation. ** $p < .05$; *** $p < .01$.

	Log Number of Bachelors	Log Annual Wage	Net New Hires	Log Annual Wage (Advanced Positions)	Net New Hires (Advanced Positions)
	(1)	(2)	(3)	(4)	(5)
Lagged Skew	0.1136***	-0.0165***	0.0426	-0.0056	-0.0094
	(0.0288)	(0.0031)	(0.0320)	(0.0049)	(0.0283)
Lagged Mean Return	0.1119***	0.0058	0.0294	0.0149**	0.0188
	(0.0237)	(0.0044)	(0.0258)	(0.0068)	(0.0520)
Lagged Standard Deviation of Return	-0.0627**	0.0005	0.0202	-0.0248***	0.0400
	(0.0304)	(0.0056)	(0.0856)	(0.0085)	(0.0578)
Lagged Log Average Wage	0.0415**	0.0066**	-0.0128	0.0059**	-0.0331
	(0.0203)	(0.0029)	(0.0165)	(0.0027)	(0.0171)
Lagged Log Number of Bachelors and Lagged Male/Female Ratio	No	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes
Year and Major Fixed Effects	Yes	Yes	Yes	Yes	Yes
# Observations	513	210	200	204	192
Adj. R-Squared	0.85	0.96	0.29	0.91	0.13

Table 3

Composition Changes in Labor Supply and Concentration of Employment

In Panel A, Number of Bachelors is the annual number of bachelor degrees awarded for a major at the school-level. Top School is a dummy variable, indicating that the school is in the top 50 in US News Rankings 4 years prior to graduation and is located in the state that hires the most employees in the major-related industries. In Panel B, the dependent variables are Log Annual Wage and Net New Hires of high- and low-skilled occupations. Low Skilled is a dummy variable indicating the low-skilled occupations, which are the occupation codes and industries that offer below median wage within a major in year t-4. In Panel C, Versatility is a dummy variable indicating that the concentration of employment in various industries is low for the major. The concentration is measured by the Herfindahl index of employment in different industries. Other variables are the same as those in the corresponding regressions in Table 2. Standard errors are clustered at the year level. All independent variables are standardized with zero mean and unit standard deviation. ** $p < .05$; *** $p < .01$.

Panel A: Number of Bachelors		
	School-level Log Number of Bachelors	
	(1)	(2)
Lagged Skew	0.0706***	0.0695***
	(0.0155)	(0.0155)
Lagged Skew * Top School		0.1459**
		(0.0695)
Lagged Mean Return and Standard Deviation of Return	Yes	Yes
Other Controls	Yes	Yes
Various Fixed Effects	Yes	Yes
# Observations	33,301	33,301
Adj. R-Squared	0.21	0.21

Panel B: Wage and Net New Hires		
	Log Annual Wage	Net New Hires
	(1)	(2)
Lagged Skew	-0.0164	0.0492
	(0.0088)	(0.0889)
Lagged Skew * Low Skilled	0.0156**	-0.0193
	(0.0066)	(0.0646)
Lagged Mean Return and Standard Deviation of Return	Yes	Yes
Other Controls	Yes	Yes
Various Fixed Effects	Yes	Yes
# Observations	330	307
Adj. R-Squared	0.90	-0.03

Panel C: Major Versatility		
	Log Annual Wage	Net New Hires
	(1)	(2)
Lagged Skew	-0.0333***	0.0336
	(0.0088)	(0.0548)
Lagged Skew * Versatility	0.0221**	0.0145
	(0.0103)	(0.0808)
Lagged Mean Return and Standard Deviation of Return	Yes	Yes
Other Controls	Yes	Yes
Year and Major Fixed Effects	Yes	Yes
# Observations	210	200
Adj. R-Squared	0.97	0.29

Table 4
Individual-Level Analysis

This table reports results from the National Survey of College Graduates. In Panel A, Columns (1) to (4) report results from fixed effects panel regressions with log(Earnings) as the dependent variable, while Columns (5) to (8) report results from Ordered Logit regressions of a variable indicating the graduate s current job area is from her field of study (measured in 3 levels: "closely related", "somewhat related" and "unrelated", with higher values indicating decreasing relatedness). Job unrelatedness to UG major is defined directly for those who are either only graduates (no higher degrees), or those who got higher degrees but in the same field as their UG major. We exclude respondents working outside their fields of study if the switch of industry is due to changes in interest or family-related reasons or retirement. In Panel B, we examine the consequences of having a job unrelated to field of study. Throughout the table, the following fixed effects are included: Age, Gender, Marital status, Minority status, Region, Major, Second major (if any), and Survey-year. In Panel A, industry fixed effects are also included. In Panel B, fixed effects for Job code for principle job are also included. We report the impact of our variables of interest on the Odds Ratios for all ordered logit regressions. Standard errors are clustered by job code for principal job. ** $p < .05$; *** $p < .01$.

Panel A: Earnings and Job Unrelatedness								
	Log Earnings				Odds of job unrelated to UG major (Ordered logit)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lagged Skew	-0.019***	-0.025***	-0.008	-0.014**	0.047	0.055***	0.043	0.044
	(0.005)	(0.006)	(0.004)	(0.006)	(0.027)	(0.020)	(0.026)	(0.026)
Lagged Mean Return and Standard Deviation of Return	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Fixed Effects	No	No	No	Highest degree attained, Highest degree major	No	No	No	Highest degree attained, Highest degree major
Impact on Odds Ratios	-	-	-	-	1.048	1.057***	1.044	1.045
Sample	Full	Exclude Health and bio majors	Highest degree is bachelors	Full	Full	Exclude Health and bio majors	Highest degree is bachelors	Full
# Observations	134,174	85,432	77,309	134,174	117,011	78,186	86,151	117,011
Adj./ Pseudo R-Squared	0.31	0.29	0.34	0.35	0.19	0.15	0.19	0.19

Panel B: Earnings and Job Dis-satisfaction								
	Log Earnings				Odds of Job Dis-satisfaction (Ordered logit)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Job unrelatedness to UG major	-0.205***	-0.173***	-0.197***	-0.192***	0.577***	0.611***	0.559***	0.568***
	(0.029)	(0.028)	(0.031)	(0.029)	(0.05)	(0.049)	(0.048)	(0.047)
Additional Fixed Effects	No	No	No	Highest degree attained, Highest degree major	No	No	No	Highest degree attained, Highest degree major
Impact on Odds Ratios	-	-	-	-	1.781***	1.842***	1.749***	1.765***
Sample	Full	Exclude Health and bio majors	Highest degree is bachelors	Full	Full	Exclude Health and bio majors	Highest degree is bachelors	Full
# Observations	99,650	64,945	71,570	99,650	100,436	65,442	72,156	100,436
Adj./Pseudo R-Squared	0.31	0.31	0.32	0.32	0.04	0.04	0.04	0.05

Table 5
Survey of College Graduates Using *SurveyMonkey*

In Panel A, Columns (1) and (4) report results from an ordered logit regression of Household income (in 8 buckets: 7 buckets of size $25,000, starting from $25,000, and the last bucket for income above $200,000). Column (2) reports results from a logit regression with a dummy dependent variable indicating whether the graduate started her post-College career in an industry she was expecting to work in when she chose her major. In Column (3) we examine a dummy dependent variable indicating if respondents thought that they could have chosen a more suitable major had they done more research. In Column (5) we examine a dummy dependent variable, indicating if the graduate currently works in an industry that is different from the industry she started in after graduation. Starting job not in target industry is a dummy variable indicating whether the graduate s first job was not in an industry she was targeting. In Panel B, the dependent variable is Skew, the return skewness associated with the chosen major of each respondent. The key explanatory variables are: (i) Decision based on small number of observations, indicating if a respondent formed expectations only through information on a small number of firms. (ii) Lottery Preference, indicating if the graduate answered that she would have chosen to play a fair lottery over a stable, future income stream. Throughout the table, the following fixed effects are included: Major, Graduation-year, Student debt, Gender, Region, and When major was chosen. In Panel A, fixed effects for the importance of career prospects when choosing major are also included. We report marginal effects for all logit regressions, and also report below each column the impact of our variables of interest on the Odds Ratios for all ordered logit regressions. Standard errors are clustered by industry in the earnings and job change regressions, and by target industry for starting job elsewhere. ** $p < .05$; *** $p < .01$.

Panel A: Income and Employment

	Odds of higher HH Income (Ordered Logit)	First job in target industry (Logit)	Better with more research (Logit)	Odds of higher HH Income (Ordered Logit)	Job change to yet another industry (Logit)
	(1)	(2)	(3)	(4)	(5)
Lagged Skew	-0.194**	-0.069***	0.046***		
	(0.094)	(0.020)	(0.009)		
Starting Job Not in Target Industry				-0.514**	0.136**
				(0.251)	(0.063)
Lagged Mean Return and Standard Deviation of Return	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Current industry	Target industry, Actual industry of first job	Target industry, Actual industry of first job	Current industry	Current industry
Impact on Odds Ratio	0.824**	-	-	0.598**	-
# Observations	355	390	390	355	388
Pseudo R-Squared	0.19	0.55	0.36	0.19	0.33

Panel B: The Choice of a High-Skew Major			
		Skew	
	(1)	(2)	(3)
Decision Based on Small Number of Observations	0.455***		0.455***
	(0.153)		(0.153)
Lottery Preference		-0.015	-0.011
		(0.178)	(0.209)
Lagged Mean Return and Standard Deviation of Return	Yes	Yes	Yes
Industry Fixed Effects	Target industry, Actual industry of first job	Target industry, Actual industry of first job	Target industry, Actual industry of first job
# Observations	390	390	390
Adj. R-Squared	0.38	0.37	0.38

Table 6
Alternative Return Skewness Measures

This table reports the results of regressions of aggregate number of graduates and employment on alternative skewness measures (measured in years t-7 to t-3, relative to the graduation year t). $Tail_{10}$ is the sum of Top_{10} and $Bottom_{10}$. Top_{10} ($Bottom_{10}$) is the average return of the top (bottom) 10 firms among all firms that are mapped to the major minus the median return, divided by the difference between the 90th and 10th return percentile, after dropping firms in the lowest 50th size percentile. News Skew is the employment-weighted cross-sectional skewness of annual sum of RavenPack ESS scores (the ESS scores are rescaled to -1 to 1). Other variables are the same as those in the corresponding regressions in Table 2. Standard errors are clustered at the year level. All independent variables are standardized with zero mean and unit standard deviation. ** $p < .05$; *** $p < .01$.

Panel A: Tail Measures

	Log Number of Bachelors		Log Annual Wage		Net New Hires	
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged $Tail_{10}$ (Top_{10}+$Bottom_{10}$)	0.1459***		-0.0279***		-0.0068	
	(0.0535)		(0.0086)		(0.0483)	
Lagged Top_{10}		0.1938**		-0.0374***		-0.0185
		(0.0741)		(0.0126)		(0.0731)
Lagged $Bottom_{10}$		0.1066***		-0.0101		0.0153
		(0.0380)		(0.0061)		(0.0815)
Lagged Mean Return and Standard Deviation of Return	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year and Major Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	513	513	210	210	200	200
Adj. R-Squared	0.85	0.85	0.96	0.96	0.28	0.28

Panel B: News Skew

	Log Number of Bachelors	Log Annual Wage	Net New Hires
	(1)	(2)	(3)
Lagged News Skew	0.1486***	-0.0112**	0.0537
	(0.0385)	(0.0045)	(0.1012)
Lagged Mean Return and Standard Deviation of Return	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes
Year and Major Fixed Effects	Yes	Yes	Yes
# Observations	150	150	150
Adj. R-Squared	0.99	0.98	0.07