# Hometown Labor Markets and Degree Choice

Ryan Mather Andrew Smith*

November 18, 2024

[Click here for the most current draft of this paper]

**Abstract**

We demonstrate that conditions in a student's local hometown labor market influence their course taking choices while in high school, and also their initial and final majors when they go to attend college. As a possible mechanism to explain this, we demonstrate that growth in local relevant employment for STEM workers increases a measure of relevant wages for those workers. Next, motivated by evidence that past conditions in a student's labor market may continue to affect a students present decisions, we estimate a dynamic discrete choice model in which students can respond to labor market conditions in each period, and the decisions made in one period have implications for the next. Using this model, we find that allowing students the option to pursue STEM-intensive course loads while in high school contributes to their likelihood of pursuing a STEM degree in college and increases their earnings once in the labor market. Additionally, having this course option while in high school increases the responsiveness of students' college major choices and wages to changes in the local labor market.

*The authors appreciate valuable feedback on this paper from Ken Hendricks, Chris Taber, Jesse Gregory, and Chao Fu. The University of Wisconsin-Madison provided funding to support this project. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, the Texas Workforce Commission or the State of Texas.

## 1 Introduction

Students face difficult decisions in choosing what academic courses and majors will best prepare them for the labor market. Their choices involve trade-offs between choosing fields that fit their interests or abilities and the knowledge that academic paths have large implications for wages: The gap in expected earnings between different majors can be similar in size to the gap between college graduates and high school graduates (Altonji et al., 2012). What is more, the demand for certain skills in the labor market changes over time as different industries or production methods rise or fall in prominence. For bachelor's degrees that may require four to six years to complete, students make initial major choices based on their predictions for how valuable each major will be at the time when they graduate. Their decisions must begin even earlier than that, though, in selecting the high school courses that they will take to prepare them for life after high school.

This challenge for students creates similar challenges for policymakers. In Texas, the current *60X30TX* education plan sees postsecondary education as crucial in shaping the future labor force of Texas, and sets the goal that at least 60 percent of Texans aged 25-34 will have a certificate or degree by 2030. This plan was motivated in part by growth among minority groups like Hispanics, who made up 65 percent of state population growth in the prior census and were traditionally underrepresented in Texas higher education. Poverty among Hispanics had been increasing, particularly for those with lower educational attainment. As acknowledged in the plan, however, a "degree isn't enough... Texas students will need to match their credentials to employer needs. The role of higher education in helping students and employers coordinate their efforts is essential" (THECB, 2015). Creating educational systems that facilitate this match is difficult, because it is not clear how responsive students' academic choices are to changes in the local labor market, and how long those responses will take to occur.

In this paper, we use data from Texas to shed light on these questions. Our main goal is to assess how movements in local industry conditions influence the choices that students make between academic subjects. In doing so, we pay special attention to the fact that students do not make these decisions all at once, but work toward their final college outcome gradually over time. A main takeaway from our paper is that local labor conditions are already having significant implications for this process while a student is still in high school. The dynamics of these student decisions over time have important implications for the overall responsiveness of students to the local labor market. For example, where college seniors likely have an accurate view of what economic conditions will be like when they graduate, they have little ability to switch majors without delaying their graduation date. High school seniors, by contrast, have a greater freedom to choose between majors, but must make their decisions based on economic conditions that are further removed from the job markets that they will actually enter upon graduation from college.

We find evidence that high school graduates react strongly to their local labor markets when choosing their senior-year coursework. Importantly, however, the responsiveness of students to the labor market changes as we take the lagged difference over different lengths of time, suggesting that students' course taking choices may be influenced by past conditions in their local labor market. We interpret this as evidence that students may begin reacting to the labor market by preparing for or selecting certain senior-year courses even before entering their senior year. Our regressions also provide evidence that local labor market conditions affect the rates at which STEM course-takers earn credit for those courses.

Next, we estimate regressions looking at how students' initial and final selections between college majors, and find that STEM majors continue to be responsive to local conditions. As a likely mechanism for these effects, we test whether the wages that are relevant for a specific major increase when the relevant employment share for that major increases. We find this to be true within STEM majors, which helps to explain why STEM majors were responsive to local labor market conditions at the stage of choosing a first major in college.

Given that local labor markets can have effects on student decisions at many stages in their academic process, fully understanding these effects of the local labor market demands a richer model which can take into account student decisions over time. To address this, we estimate a dynamic discrete choice model in which students begin making decisions in high school, and are free to respond each period to changes in the local STEM employment share.

This model allows us to more fully estimate in the importance of local labor market conditions while students are in high school, both for the decisions that they make in high school and the long-term implications of those decisions as they go on to college and the labor force. The paper ends with a counterfactual estimation in which we turn off student's ability to choose their high school senior-year coursework, and observe the implications for student responses to the labor market. We find that removing this option lowers the number of students who graduate college with STEM degrees by about 4.42 percent, and decreases a measure of early career earnings by about 1.63 percent. Among students who would have otherwise pursued the STEM-Intensive senior-year track, these earnings results are even more dramatic, with early career earnings dropping by around 4.96 percent. What is more, the growth of student STEM degree shares and eventual wages after a

STEM employment share shock is diminished in the counterfactual, suggesting that the ability to choose senior year course loads has important effects on the overall responsiveness of students to the labor market.

Importantly, we also find that substantial benefits of a STEM-Intensive course load option in the senior year are present for students with the highest and lowest levels of initial STEM ability. From a policy perspective, this is encouraging evidence that providing meaningful options between career-relevant high school courses benefits the long-term earnings of students from diverse backgrounds.

### 1.1 Related Literature

This paper examines effect of a student's high school labor market on their choices between courses and eventual college majors. There are three strands of literature that seem most related to this topic. First, there are papers which estimate the effects of career-relevant high school courses on long-term outcomes (see Altonji et al. (2012) for a review). Our main contribution to this literature is to assess the impact of a student's local labor market on their high school curriculum choices. To our knowledge, we are the first to identify causal effects in this area.[1] Many of the existing papers in this literature leverage variation across high school environments to instrument for an individual student's choices between courses. For example, the number of classes taken on average in certain subjects, the dedication of full-time teachers to teach certain subjects, and school-specific graduation requirements have all been used in this way (e.g. Altonji 1995, Rose and Betts 2004, Meer 2007, and Kreisman and Stange 2017). Rather than taking aggregate conditions of the high school environment as given in this way, we are interested in asking how aggregate conditions at a high school change in response to local industry conditions. Specifically, we will estimate how the share of classes that are offered, taken, and passed in STEM and non-STEM change in response to local industry conditions.

Second, there is a large literature dedicated to understanding how students' decisions between college majors depend on conditions in the labor market. Our paper is most related to the work that looks at how student major choices respond to their local labor markets as opposed to the national labor market (e.g. Conzelmann et al. 2023, Weinstein 2022, Grosz 2022, Han and Winters 2020, Long et al. 2015). To our knowledge, this is the first paper to use a student's high school location to measure their local labor market, which we define as the commuting zone, when estimating the effects on the specific major that students pursue.[2] We argue that this offers a better measure of a student's home labor market than those used previously,[3] and that it is the most relevant local labor market for influencing a student's initial major choices.

Another contribution that we make in this literature is the use of a shift-share instrument to identify plausibly exogenous variation in local economic conditions. In doing so, we allow the "relevance" of an industry for a major to change over time, and measure it with information from recent graduates to address how the degree to which different industries are relevant may change as workers gain experience in the labor force.[4] Conzelmann et al. (2023) is one other paper that also uses a shift-share instrument.[5] The main difference between our work and theirs is that Conzelmann et al. (2023) focuses entirely on the number of people that graduate with a specific major, whereas we are more focused on the process by which a student reaches that final major. As such, our main results are about the courses that a student selects in high school and their first major when going to college.

A final contribution that we make to this literature is better describing the timing of student decisions related to major choice. Much of the evidence on student's reactions to changes in labor markets tends to focus on the degrees that students

[1] The most related paper that we have found here is Kreisman and Stange (2017), which controls for the local labor market with cohort-by-state fixed effects. However, they do not report these fixed effects or identify the portion of them that is coming from the local labor market as opposed to other things that vary at the state-by-time level. Moreover, our entire analysis is within a state, and the effects are identified using variation across commuting zones. We believe that these zones are a better proxy for a student's local labor market.

[2] Hubbard (2018) uses an even more specific measure which is the zip code of a student's home address, or, when that is not available, the location of a student's high school. However, his goal is to estimate how local labor conditions influence a student's decisions between degree levels (that is, high school diploma, associate's degree, or bachelor's degree) rather than the degree subject as measured by the major.

[3] The papers on the influence of local labor markets cited above use generally use the college a student goes to to construct measures of local labor market conditions. For example, Weinstein (2022) uses the area around a student's college, and Conzelmann et al. (2023) uses the locations where alumni of a certain college tend to live. An exception is Han and Winters (2020), which uses a student's state of birth. By using commuting zones, we identify a student's home location and labor market much more specifically.

[4] Other papers in this area tend to use measures of "relevance" that are constant over time and the number of years a worker has spent in the labor force. Conzelmann et al. (2023) uses a measure which can vary over time, but is constant over the number of years a worker has spent in the labor force. The differences between our method and that of other papers is discussed in more detail in Appendix A.4.

[5] Grosz (2022) uses a shift-share instrument as well, but his focus is on the supply-side responses of community colleges to state-level shocks. Also, his data is not able to match changes in industry employment shares to changes in academic majors over the same time intervals. As a result, he acknowledges that the magnitude of his coefficient estimates may be inaccurate.

have when graduating, which is the final result of this decision process.[6] Knowing the timing of student decisions will help us to understand the amount of time it takes changes in the labor market to be fully reflected in the shares of students picking certain majors. Some of the previous evidence on the degrees that students graduate with suggest that the timing of student decisions affect this lag. For example, Long et al. (2015) find that student's degree choice upon graduating is most strongly related to industry-specific wages three years earlier, which is when many of the students would have been freshman in college choosing their initial majors. To fully capture the timing of student decisions, we employ a structural model in which students are able to react to industry conditions each period, and their decisions in one period have implications for their decisions in future periods.

This brings us to the final literature we contribute to, which covers dynamic discrete choice models connecting students' degree choices and labor market conditions. Buchinsky and Leslie (2010) estimates a model in which students make a decision whether to attend college or not in response to their predictions about the labor market. However, college degrees within their model do not involve a specific "major" subject. Students do choose between science and non-science degrees in Arcidiacono et al. (2023), but the aggregate labor shocks in that model are not industry-specific, and have the same effects on the log wages of all workers regardless of their degree path. In our model, the focus is on the way that students progressively learn about and react to the broader conditions in industries that place more or less value on STEM and Non-STEM abilities. Our central goal is to describe how students make decisions between STEM and non-STEM academic options in response to changes in the relative attractiveness of those options in the labor market.

Ryoo and Rosen (2004) use a model with a similar idea, in which students can react to changes in the labor market for engineers by enrolling in engineering programs. Heckman et al. (1998) also estimates a dynamic discrete choice model in which students' educational choices react to labor market conditions, and is more similar to ours in that modeled students can choose science and non-science degrees in college.[7] The main contribution of our proposed work relative to these papers will be a greater focus on the high school and college processes. For example, neither of these papers have models that let students decide to change their majors while in college.[8]

The rest of the paper proceeds as follows: In the next section, we describe the data sets that we will use for this project and some motivating evidence on the possible connection between movements in the labor market and college majors. Section 3 describes our method for estimating the causal relationship between these two things, and section 4 shows our results from using that method. Section 5 describes our structural model model, section 6 describes how that model is identified, and section 7 shares the estimation results. Finally, section 8 concludes.

## 2 Background and Data

Our main goal in this paper is to assess how students' choices between academic subjects are influenced by the employment conditions in industries related to those subjects. Figure 1 displays the kind of relationship that we are interested in studying. In it, we look at the Dallas commuting zone,[9] and plot the changes in Health Care/Social Assistance[10] employment shares along with the share of people whose first declared major in college is in STEM.[11] The fact that the the two lines in these graphs are so similar gives suggestive evidence that students may be responding to the labor market when they are choosing

[6] Bradley (2012) is one exception, and considers the effects of recessions on the intended majors of freshman students before enrolling in classes. However, the data used there does not include the degrees students have when they graduate. Our results will show the effects of movements in specific industries on student's course taking in high school, first declared majors, and the final majors that students have upon graduation.

[7] Ryoo and Rosen (2004) also let the supply of new engineering school entrants depend on shocks to career prospects in alternative professions, but the decision paths of students who go that route are not modeled.

[8] This is a particularly limiting restriction for STEM, as we know that a large number of students initially intending to major in STEM drop out. Stinebrickner and Stinebrickner (2014) finds in their data that 19.8% of students enter college believing that a degree in science or math will be their most likely outcome, and yet only 7.4% of students actually graduate with one. There is a constant dropout rate in the model of Ryoo and Rosen (2004), but since it is constant it cannot respond to local labor market conditions.

[9] Specifically, we use commuting zone 33100 as the representative commuting zone for Dallas.

[10] Health Care/Social Assistance and Retail Trade were the two-digit NAICS codes with the highest employment shares in Texas as of 2005. To provide an initial look at the data, in these graphs we associate Health Care/Social Assistance with STEM and "Retail Trade" with Non-STEM. When we separate two-digit NAICS codes into STEM and Non-STEM intensive categories for the model "Health Care/Social Assistance" will actually fit into the Non-STEM intensive category. This is discussed in appendix section C.3. Complimentary graphs for Retail Trade and Non-STEM degrees can be found in appendix A.1.

[11] Specifically, the students included in these graph are those who attended a public high school within the commuting zone and then attended either a public university or community college within the state of Texas. We record their first major at those colleges, even if the college they attend is outside of the Dallas commuting zone. The suggestive evidence coming from this graph is more about the effects of a student's high school or "home" labor market on their decision, as opposed to the effects of the labor market of their college.

their first major. For example, if it is true Health Care hires a lot of people with STEM majors, then a student might respond to the growth in that industry by majoring in STEM.

Figure 1: Health Care/Social Assistance Employment and College Major Shares in Dallas

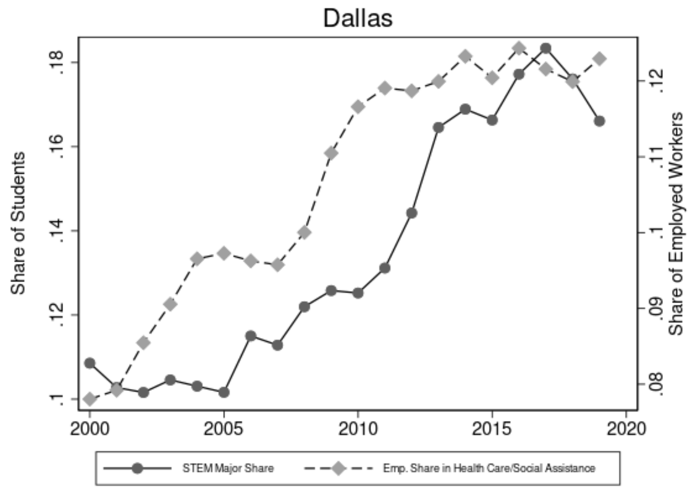
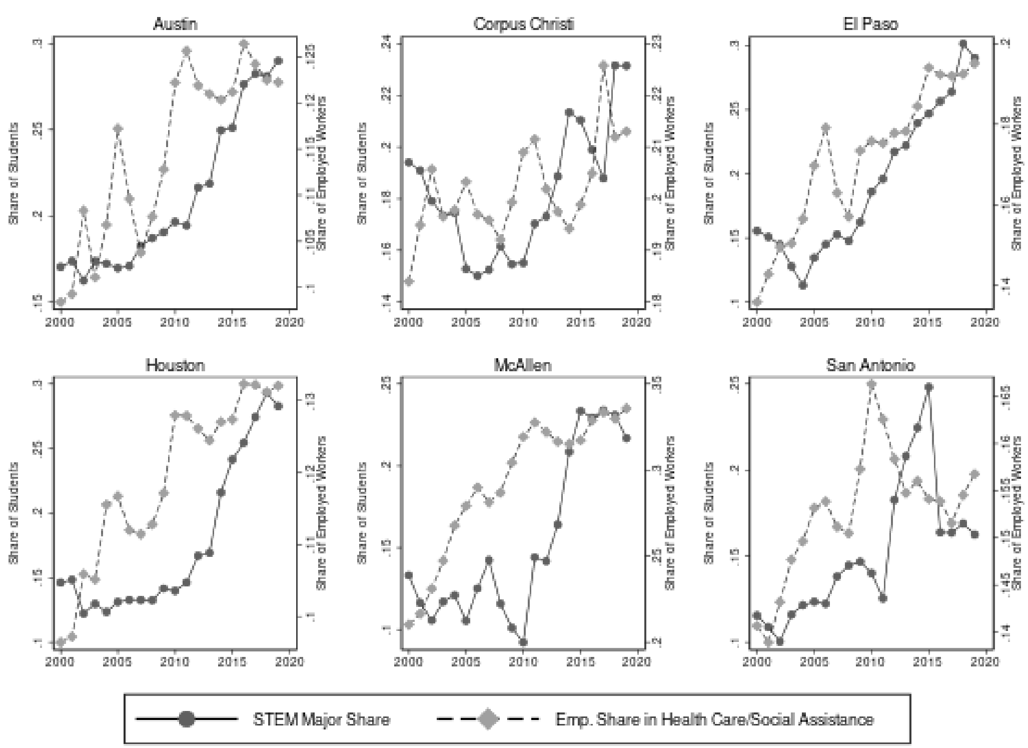



Figure 2 show this same graph for commuting zones associated with other large cities in Texas.[12] Again, the same patterns often show up. In appendix section A.1, we display similar graphs for Non-STEM degrees and the "Retail Trade" industry. This paper will try to assess whether there is any causal connection underlying this sort of relationship. Specifically, we study how the conditions of industries in labor markets that a student is exposed to at different times in their education can affect their choices between majors.

Figure 2: Health Care/Social Assistance Employment and STEM Major Shares Across Texas




Texas is an ideal state in which to consider these questions. First, as mentioned in the introduction, Texas is in the

[12]In that figure, we use commuting zone 32000 as the representative commuting zone for Houston.

midst a policy agenda to increase the share of Texans aged 25-34 with a certificate or degree to 60 percent by 2030.[13] This plan was motivated by future employer needs and the growth of minority populations like Hispanics—the Hispanic population aged 25 to 24 is projected to increase by 41 percent between 2015 and 2030. According to the *60x30TX* strategic plan, African Americans and Hispanics together accounted for over 60 percent of the K-12 pipeline,[14] and traditionally are underrepresented in Texas higher education (THECB, 2015). Texas may be well-positioned to increase economic equity through higher education. In Chetty et al. (2020), the authors found that Texas had three of the nation's top ten colleges for bottom-to-top quintile income mobility rates.[15] measured by moving from the bottom 20 percent to the top 20 percent of the income distribution. Achieving positive and equitable outcomes for students in the midst of such a large increase in the college-educated workforce, however, will demand educational systems that can help match student training with employers who need their skills and are hiring. We view this paper as making important contributions in that context.

Finally, among state administrative datasets to access, it is helpful that Texas is a large state both in terms of covering many commuting zones and educating many students. The many commuting zones are helpful so that we can observe meaningful variation in local labor market conditions, and having data on many students provides us with large sample sizes. In fact, Texas educates over 10 percent of all public-school students in the United States (NCES, 2023).[16] The total population in Texas is greater than that of the entire Nordic region combined, meaning that the scope of our data is similar to nationwide data from a reasonably sized country.[17]

The main data that we will use for this paper comes from the Texas Schools Project (TSP). This data set links educational data from grades K-12 and colleges to labor force outcomes in Texas. It has a few features are particularly well-suited for our research question. First, we can see the high schools that students attended, and can find the location of these high schools. This gives us reliable information of a student's location before entering college, and serves as a good proxy of a student's home location. Second, we are able to see the declared majors of college students over time while they are in public Texas colleges. We use this to to identify changes to a student's major around shifts in industry conditions. Third, we are able to see the wages, employment status, and industry classification of Texas workers each quarter. This can help us to understand the link between college majors and industries.

Our sample is comprised of students that graduated from public high schools in Texas from 2000 through 2019. We restrict the sample to only include students that graduated high school within 4 years of starting it, and that were only observed in one commuting zone over the course of high school.[18] Table 17 in appendix section A.2 displays some summary statistics for these students. For our analysis, "no college" means that the student did not enroll in any public university or community college in Texas, and "Enrolled in college" means that they did enroll in one of those options.[19]

For the share of employment in each industry, we use annual Census County Business Patterns (CBP) data. We use the 2022 Department of Homeland Security (DHS) STEM Designated Degree Program List to classify degrees as STEM or

[13] A progress report published in 2022 suggests that Texas has made progress toward meeting that goal. In 2013 the share of Texans aged 25-34 with a certificate or degree was 38.3 percent (THECB, 2015), and by 2021 that share had grown to 48.6 percent (THECB, 2022).

[14] Within our sample, which is limited to high school graduates and covers many years of data, these two groups account for 54.4 percent of the sample. Table 17 in appendix section A.2 displays some summary statistics for the students in our sample.

[15] This rate measures the share of students at a school where (i) the student came from a family in the bottom quintile of the income distribution and (ii) the student went on to themselves enter the highest quintile of the income distribution. Their analysis excludes colleges with fewer than 300 students, which excluded about 5 percent of the students in their original sample. The three Texas universities in the top ten were the "University of Texas, Pan American," "South Texas College", and the "University of Texas, El Paso" (Chetty et al., 2020). The "University of Texas, Pan American" has been merged with the University of Texas at Brownsville and is now called "the University of Texas Rio Grande Valley."

[16] In this respect, Texas is nearly identical to California, which is the state that educates the most public school students. Texas educates 11.2% of all public school students, and California educates 11.7%. These statistics are taken from Fall 2023 and cover enrollment in public elementary and secondary schools (NCES, 2023).

[17] Specifically, the Nordic Statistics database records a population of 28.0 million in 2023 across Denmark, Greenland, Finland, Åland, Iceland, Norway, Sweden, and the Faroe Islands. The Census estimates the 2023 population in Texas at 30.5 million.

[18] We will consider a student's local labor market to be the commuting zone associated with their school district. This second condition of requiring that we only see a student in a single commuting zone simplifies things because students that attended high schools from multiple commuting zones may have been influenced each of their local labor markets. With this restriction, we will only consider the effects of a single labor market on each student. Restricting to students that stayed in the same commuting zone also helps to lower the risk of considering students that were influenced to attend a school in a certain commuting zone because of recent changes in the labor market. For example, if a student's parent is an engineer, and they move school districts during high school because the engineering labor market in another location greatly improved that year, then this could potentially bias our results. In order to determine the school commuting zones a student was in for this restriction, we use graduation data from a student's first observed high school graduation and high school enrollment data taken at or before the school year of that graduation.

[19] The TSP data does provide information on student majors at independent colleges for certain years, and also is linked to National Student Clearinghouse data for certain years, which makes it possible to track the colleges that these students attend even if they leave the state of Texas. In the time periods where these data sets are available, they can help us to better observe whether a student went to any college. We could do this in the future to check the robustness of our results.

non-STEM.[20] Appendix A includes additional details on other data sources and on how the data were prepared.

## 3 Estimation Strategy

Our goal is to determine how students' choices between different academic subjects is affected by changes in industries that are relevant for those subjects. Understanding which industries are relevant, and how relevant they are, is difficult for two reasons. First, bachelor's degrees are often generally applicable to many different jobs. In the case of community college degrees studied by Grosz (2022), the connection between a particular degree and job in the labor market is likely to be much more stable. Something like a bachelor's degree in English, by contrast, could be applied to work in many different contexts. This makes it more difficult to pin down exactly what industries are relevant for a particular major. Second, the relevance of a subject to certain industries could change over time and over the life cycle of workers. This is a particularly relevant for technology-intensive subjects, where the skills needed to complete a job may change rapidly over time. Deming and Noray (2020) show that workers tend to shift away from occupations like this as they gain experience in the labor market.

### 3.1 Measuring Relevant Employment by Major

In view of these challenges for measuring the share of local employment that is relevant to major subject $m$, we propose a measure that incorporates both the applicability of a major across multiple industries, and the way that that applicability can change over time.[21] Specifically, we will multiply the share of employment in each industry $k$ by a measure of the relevance of that industry to students with major $m$ in time $t$. Let $RE_{m\ell t}$ represent the share of employment that is relevant for major $m$ in location $\ell$ and time $t$, where

$$RE_{m\ell t} = \sum_{k=1}^{K} r_{mkt} s_{k\ell t} \tag{1}$$

$$r_{mkt} = \frac{(\#\text{ new graduates in major } m \text{ and industry } k)_t}{(\#\text{ new graduates in industry } k)_t}$$

$$s_{k\ell t} = \frac{(\#\text{ workers employed in } k)_{\ell t}}{(\#\text{ workers employed})_{\ell t}}$$

Here, $r_{mkt}$ gives the relevance of industry $k$ to graduates with major $m$, which we measure as the share of jobs for new graduates in industry $k$ going to those with major $m$. Notice that $r_{mkt}$ does not differ across locations $\ell$—we are assuming here that the relevance of a major to a particular industry does not vary geographically.[22] Then, $s_{k\ell t}$ gives the share of employed workers in area $\ell$ that work in industry $k$.

When we use a shift-share instrument below, we will break $s_{k\ell t}$ into two terms: One that gives the share of employment in industry $k$ for some base year 0, and another term $g_{k\ell t}$ that gives the growth rate in the employment share since that baseline year. First, define

$$g_{k\ell t} = \frac{(\#\text{ workers employed in } k)_{\ell t}}{(\#\text{ workers employed in } k)_{\ell 0}}$$

as the growth rate in the the number of workers in industry $k$ since the base year 0. Then, appendix A.3 shows that

$$s_{k\ell t} = \frac{s_{k\ell 0} g_{k\ell t}}{\sum_{k'} s_{k'\ell 0} g_{k'\ell t}} \tag{2}$$

[20] Where the DHS classifies degrees as STEM or non-STEM at the six-digit CIP code level, Texas has a degree classification that adds on on two additional digits for an eight-digit CIP code. When matching these eight-digit codes in with the DHS list, then, we ignore the last two digits added by Texas.

[21] In Appendix A.4, we discuss some of the methods other papers have used to link degrees to opportunities in the labor market, and why we think our method shown here is preferable.

[22] In our data, we do not observe the locations of people with certain majors after they leave college, and so we cannot directly calculate a relevance measure that differs by location. Even if we could, this may not be a good decision because local shares of this kind are likely to be noisy, and may not accurately reflect the share of jobs in industry $k$ and location $\ell$ that are available to people with the major $m$. A local $r_{mk\ell t}$ term could be greatly effected by the chosen degrees of local graduates, raising concerns about reverse causality and also omitted variable bias if unknown local factors could affect both $Y_{m\ell t}$ and the $r_{mk\ell t}$ term. Instead, then, we assume that $r_{mkt}$ does not vary with location $\ell$, and measure it using data from all of Texas. There may be a concern with this assumption for industries where the activities involved in production differ widely across areas. In the future, this concern could be addressed more directly by identifying supplemental datasets in which $r_{mk\ell t}$ terms can be estimated in different locations $\ell$, and viewing the extent to which those terms vary across $\ell$.

With this, (1) can be rewritten as

$$RE_{m\ell t} = \sum_{k=1}^{K} r_{mkt} \frac{s_{k\ell 0} g_{k\ell t}}{\sum_{k'} s_{k'\ell 0} g_{k'\ell t}} \tag{3}$$

### 3.2 Estimating the Response of Majors to Industry Conditions

Let $m$ represent academic subject, and $\ell$ be their location. We will use commuting zones for the location $\ell$, as this provides a natural representation of the local labor market.[23] For the results relating to high school course taking, we allow $m$ to be either "STEM" or "non-STEM," but for the results at the college level we are able to use a more disaggregated measure and define $m$ as the 2-digit CIP code of a degree. The equations we estimate will take the form

$$\Delta_d Y_{m\ell t} = \gamma_{mt} + \beta_1 \Delta_d RE_{m\ell t} + X_{\ell t}\beta_2 + \varepsilon_{m\ell t}. \tag{4}$$

where $X_{\ell t}$ includes a number of controls for students in location $\ell$ and time $t$.[24] $Y_{mlt}$ is our outcome of interest, which in our baseline results is the share of classes taken by students from location $\ell$ in subject $m$ at time $t$. The $\Delta_d$ operator means taking the $d$-year difference, so $\Delta_d Y_{m\ell t} = Y_{m\ell t} - Y_{m\ell t-d}$. We will show results for multiple different values of this difference, which will help to pin down the dynamics of student decision making.

$\beta_1$ is the key parameter that we want to estimate. It says that when there is a 1 percentage point increase in the share of jobs that are relevant for major $m$, the share of students majoring in $m$ increases by $\beta_1$ percentage points. There are two main concerns for identifying this coefficient. The first is that unobserved changes in specific labor markets could have an effect on both employment shares and major shares, creating omitted variable bias. For example, imagine that one of the enjoyable things about being a financial adviser is getting to golf with your clients, and a new golf course opens up in location $\ell$. This has the potential to increase both the share of jobs held by people with finance degrees and the share of students majoring in finance, but the response of the students could be more related to the golf course than the movement in employment shares.

Second, we might be worried about reverse causality, especially when $Y_{m\ell t}$ is measuring the share of people that actually graduate with a particular degree. For example, if $\beta_1$ is positive, it may be that a higher share of people in location $\ell$ are holding degrees in major $m$ because a local college is graduating more students with that degree, and not that students are reacting to changes in the local labor market with the majors they choose.

To mitigate these concerns, we use a shift-share instrument. As shown above, we can rewrite $RE_{m\ell t}$ as

$$RE_{m\ell t} = \sum_{k=1}^{K} r_{mkt} \frac{s_{k\ell 0} g_{k\ell t}}{\sum_{k'} s_{k'\ell 0} g_{k'\ell t}}$$

To make the shift-share instrument $B_{mlt}$, we make two changes: First, our concern about omitted variable bias comes from the fact that the local growth rate $g_{k\ell t}$ may be correlated with changes specific to that location like new golf courses. However, some of the movements in $g_{k\ell t}$ are likely also driven by things that are not specific to location $\ell$, and that would affect all locations in Texas. Examples of this include the development of new production methods in a certain industry or increases in demand from other countries following some kind of trade deal. Therefore, rather than using the location-specific industry growth rate $g_{k\ell t}$, we use the growth rate in industry $k$ across all of Texas, $g_{kt}^{TX}$. Second, rather than allowing $r_{mkt}$ to vary over time, we fix it at what it was in a base year, $r_{mk0}$.[25] This will mitigate concerns about reverse causality, because we are fixing the relevance of a particular industry $k$ to a major $m$, and not allowing it to vary as the share of people majoring in $m$ varies. Then, the instrument is given by[26]

$$B_{m\ell t} = \sum_{k=1}^{K} r_{mk0} \frac{s_{k\ell 0} g_{kt}^{TX}}{\sum_{k'} s_{k'\ell 0} g_{k't}^{TX}} \tag{6}$$

[23] In using commuting zones, we align with other papers in the literature like Autor et al. (2013).

[24] We link data from each school year to the employment data associated with the fall semester. For example, for the 2010-2011 school year, the $RE_{m\ell t}$ variable is measured in 2010. Note that because equation (4) is estimated in differences, an implicit "location" fixed effect has been differenced out.

[25] Currently, we are using 1998 as the base year for $r_{mk0}$ and $s_{k\ell 0}$.

[26] Since (4) is in differences, we use the differenced version of this shift-share instrument in estimation:

$$\Delta_d B_{m\ell t} = \sum_{k=1}^{K} r_{mk0} s_{k\ell 0} \left( \frac{g_{kt}^{TX}}{\sum_{k'} s_{k'\ell 0} g_{k't}^{TX}} - \frac{g_{kt-d}^{TX}}{\sum_{k'} s_{k'\ell 0} g_{k't-d}^{TX}} \right). \tag{5}$$

# 4 Regression Results

In this section, we present estimates for (4) using $Y_{m\ell t}$ taken from different stages in a student's education. First, we will look at the effect of a student's high school labor market on the classes that they take in their senior year of high school. Then, we will look at the effect of labor markets on the first majors that these students declare when going to college.

## 4.1 The Courses Taken by High School Seniors

Table 1 shows the results of estimating equation (4),[27] where the $Y_{m\ell t}$ variable is the share of classes taken by high school graduates in their senior year that are in subject $m$, where $m$ could be STEM or non-STEM.[28] To sort classes into these categories, we label classes in the subjects Mathematics, Science, and Computer Science as "STEM," as well as any "Career and Technical Education" courses with the subject "Science, Technology, Engineering, and Math." Then, we label all remaining classes as "non-STEM," with the exception of Physical Education courses which do not count as STEM or non-STEM.

It is known that when clusters are of substantially different sizes, standard-cluster robust estimation will over-reject the null hypothesis of a null effect (MacKinnon and Webb, 2017). To address this when clustering standard errors at the commuting zone level, below each table we report p-values from a wild cluster bootstrap method, which performs better in these circumstances (MacKinnon and Webb 2017, Cameron et al. 2008).

Table 1: Change in the Share of Classes High School Seniors take in subject $m$

| | (2)<br>2-year | (4)<br>4-year | (6)<br>6-year | (8)<br>8-year | (10)<br>10-year |
|---|---|---|---|---|---|
| 2SLS | -0.0450 | 0.5640 | 1.0525$^{**}$ | 0.9471$^{*}$ | 0.9549$^{*}$ |
| | (0.3986) | (0.4110) | (0.4041) | (0.3914) | (0.4540) |
| Observations | 2,412 | 2,144 | 1,876 | 1,608 | 1,340 |
| Bootstrapped P Value | 0.864 | 0.003 | 0.011 | 0.002 | 0.010 |
| F-Stat | 22.654 | 21.810 | 20.166 | 20.389 | 18.323 |

Standard errors in parentheses, and are clustered at the commuting zone level.

The STEM and Non-STEM rows only used classes from those subjects, so they have half as many observations.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Each column of the table shows the results of using a different period of time for the difference $d$. For example, the estimate from the "6-year" column shows that raising the employment share for a subject by 1 percentage point over 6 years is expected to raise the share of classes high school seniors take in that subject by 1.0525 percentage points. The point estimates generally increase with the number of years used difference $d$ until leveling off at an effect around 1. This pattern also holds true for the other differences from one through ten years not shown here. A larger version of table 1 with all of these differences, as well as various other regressions included for robustness checks, is present in appendix section B.3. That appendix holds similar tables for all of the regression results shown here in section 4 of the paper.

The fact that the estimates do not remain constant as different lag lengths are used is interesting. This provides evidence against a model in which local industry conditions only affect the contemporaneous decisions of students, and past values of relevant employment shares have no current effects. If that were true, the only relevant distinction between year $t$ and year $t-d$ would be the employment shares from those specific years, and we would expect the estimated coefficients to be the same regardless of the lagged difference $d$. Instead, these results fit better with a more general distributed lag model. These points are discussed in more detail in appendix B.1. We take these results, along with similar results presented later in other regressions,[29] as motivation for estimating a structural model which more flexibly allows past and present labor market

[27] Here and throughout section 4, we partial out the $X_{\ell t}$ controls and fixed effects before estimating the regressions.

[28] We limit the course taking data to courses that are eligible for high school credit. For 2012 and later, this can be done using an explicit variable in the ERC which indicates whether a course is eligible for high school credit. Where this variable does not exist, we make a guess of whether a course counts for high school credit using information on the campuses that teach the course and on the course itself.

[29] It is also worth mentioning that unless the classes that a student takes in high school have no effect on their future course taking, the fact that local labor markets affect high school classes implies that the local labor market will have non-contemporaneous affects once a student enters college. We would expect, for example, that taking STEM classes either provides students with a greater amount of human capital specific to STEM, a greater

conditions to affect a student's present choices. The estimates shown in table 1, as well as versions of the other estimates shown in this section, will serve as moments for estimating that structural model.

One possible explanation for why past labor market conditions could continue to affect the present decisions of students is that students begin preparing for their senior-year coursework before entering their senior year. For example, a student intending on taking upper-level STEM courses in their senior year may want to be taking the necessary preparatory math courses in their sophomore or junior years. This could be done informally, by simply selecting those courses, or more formally by opting into certain "tracks" of high school coursework meant to funnel students toward certain classes as upper-classmen.[30] If there are improved conditions in the STEM labor market, current seniors may have little ability to adjust to these changes by switching the classes they are taking, either because of administrative difficulties in switching tracks or the challenges of "catching up" to other students who had already been preparing.

Regardless of why past labor market conditions continue to affect present decisions, an implication of this is that the full effect of a change in the local labor market could take many years to appear, with each grade level of affected students completing their schooling and entering the labor market at different times. A positive shock in the oil industry, for example, may in the short run convince college junior studying chemical engineering to switch their major to petroleum engineering before graduating next year. The long run effect is potentially much larger, however, because the same shock to the oil industry can also affect present day high school students. Whereas perhaps only chemical engineers were prepared to switch to petroleum engineering in the short run, a larger contingent of current high school students may be able to switch since they can change their preparation before even entering college. These are the sorts of dynamics we intend to explore in the structural model.

Given that local labor market conditions affect student's high school course taking, this could be driven by one of two things: Either students (and perhaps their parents) are themselves reacting to local labor market conditions when they choose courses, or else school systems are effectively forcing students to respond by somehow changing their course options. We argue that the dominant effect of these two is student decisions. While we cannot directly measure the internal pressures school systems might place on students, we can measure the share of courses offered by a schools in a commuting zone that are either STEM or Non-STEM.[31] Table 2 shows the results of estimating equation (4) with this share as the $Y$ variable. The results are largely null and small in magnitude compared to those shown in the 2SLS line of 1. This provides evidence that the effects shown in table 1 are driven largely by student reactions to the local labor market as opposed to decisions made by the local schooling system.

Table 2: Change in the Share of Course Offerings that are in subject $m$

| | (2)<br>2-year | (4)<br>4-year | (6)<br>6-year | (8)<br>8-year | (10)<br>10-year |
|---|---|---|---|---|---|
| 2SLS | -0.0915 | 0.0045 | 0.0989 | 0.1115 | 0.1592 |
| | (0.2204) | (0.1963) | (0.1754) | (0.1691) | (0.1850) |
| Observations | 2,412 | 2,144 | 1,876 | 1,608 | 1,340 |
| Bootstrapped P Value | 0.428 | 0.007 | 0.105 | 0.102 | 0.158 |
| F-Stat | 22.654 | 21.810 | 20.166 | 20.389 | 18.323 |

Standard errors in parentheses

The STEM and Non-STEM rows only used classes from those subjects, so they have half as many observations.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Finally, having shown that the local labor market can influence students to take more STEM or Non-STEM classes, we

understanding of whether they are interested in STEM topics, or both. Later on in table 10, we will demonstrate the first of these channels, that increasing the share of classes taken in STEM increases the accumulation of STEM human capital. By increasing the share of classes that students take in STEM, local labor conditions from high school would thereby also affect a student's preparation or willingness to take similar courses later on in college.

[30] More formal "tracks" of this kind were available to students toward the end of our sample. Specifically, in the Foundations High School Program (FHSP) went into affect for students entering high school in the 2014-2015 school year. In it, these students could elect early on in high school to pursue a STEM "endorsement," which involves choosing from a certain set of courses.

[31] Technically, we are only observing course offerings by high schools if at least one student decides to take the class. It is possible, then, that schools do react to local labor market conditions by increasing course offerings, but that this does not show up in the data because no students take advantage of those additional offerings. Even if this were happening, though, it would only serve to underline the point that it is student reactions to local labor market conditions rather than changes to course offerings that are driving the results shown in table 1.

can ask what effect local labor market conditions have on how well students do in these classes. Table 3 shows the results of estimating (4) when the $Y_{m\ell t}$ variable is the share of senior-year classes in subject $m$ where the student received credit. The "2SLS STEM" and "Non-STEM" rows run these regressions separately for STEM and Non-STEM courses. They are each included here because the results are dramatically different between the two. For the Non-STEM row, higher labor shares for Non-STEM workers have positive effects on the share of students in Non-STEM courses that earn credit. Among STEM classes, the effect is much stronger and moves in the opposite direction. This is what you would expect to find if STEM classes generally attract students with higher levels of academic ability. In that case, an improvement in the Non-STEM labor market moves marginal students out of STEM field, and they find themselves to be above-average in the STEM field, raising the share of Non-STEM classes taken where credit is earned.[32] On the other hand, an improvement in the STEM labor market would move marginal students into STEM classes who were less prepared than their peers, and the overall share earning credit goes down.[33] The overall effect in the 2SLS row tends to be small and negative, reflecting how the negative effect on STEM passing rates is stronger than the positive effect on Non-STEM passing rates.

Table 3: Change in the Share of Senior-year Classes in subject $m$ where the student earns credit

| | (2)<br>2-year | (4)<br>4-year | (6)<br>6-year | (8)<br>8-year | (10)<br>10-year |
|---|---|---|---|---|---|
| 2SLS | 0.0313 | -0.0718 | -0.1665 | -0.1072 | -0.0802 |
| | (0.0810) | (0.0758) | (0.0934) | (0.0822) | (0.0966) |
| 2SLS STEM | -0.2011 | -0.5634*** | -0.8684*** | -0.7042*** | -0.6694*** |
| | (0.2022) | (0.1684) | (0.2045) | (0.1677) | (0.1937) |
| 2SLS Non-STEM | 0.1424 | 0.2181* | 0.2702* | 0.2114 | 0.1933 |
| | (0.1128) | (0.0935) | (0.1137) | (0.1152) | (0.1138) |
| Observations | 2,412 | 2,144 | 1,876 | 1,608 | 1,340 |
| Boot. P Val. for 2SLS | 0.931 | 0.623 | 0.443 | 0.738 | 0.204 |
| Boot. P Val. for STEM | 0.410 | 0.045 | 0.001 | 0.001 | 0.002 |
| Boot. P Val. for Non-STEM | 0.441 | 0.042 | 0.066 | 0.141 | 0.106 |
| F-Stat for 2SLS | 22.654 | 21.810 | 20.166 | 20.389 | 18.323 |

Standard errors in parentheses

The STEM and Non-STEM rows only used classes from those subjects, so they have half as many observations.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

### 4.2 The First College Majors Chosen By High School Graduates

In addition to affecting student's high school decisions, local labor conditions can also affect their first major choices in college. Table 4 shows the results of estimating equation (4), where the $Y_{m\ell t}$ variable is the share of college-bound[34] students graduating from high school in location $\ell$ at time $t$ that choose major $m$ as their first major in college. Here, we are able to disaggregate the list of majors beyond the simple "STEM/Non-STEM" distinction, and define each major $m$ at the two-digit CIP-code level. In the "2SLS STEM" and "2SLS Non-STEM" rows of table 4, we present the separate estimates of equation (4) using just two-digit CIP codes $m$ that are "STEM" and then only using those that are "Non-STEM."[35] As before, we use a

[32] Another possible explanation here is that students have greater motivation to do well in non-STEM courses when the labor market for non-STEM workers is better. However, this does not help to explain why STEM students would do worse when the labor market for STEM workers improves.

[33] Another possible explanation, along the lines of what Ahn et al. (2022) find in a college setting, is that professors increase their grading standards when there is higher demand for their classes. This would explain why students are less likely to earn credit in good times when the STEM labor market, but does not do as good of a job explaining why improvements in the Non-STEM labor market would increase the share earning credit in Non-STEM courses. Given the hypothesis of Ahn et al. (2022), you might expect the professors of Non-STEM classes to raise their grading standards in that case, and then the effects would be negative for both Non-STEM and STEM.

[34] As noted earlier, the only college students that we are including are those who attend public universities or community colleges in the state of Texas covered by the ERC data. In the future, we will test the robustness of our conclusions to including students that attend independent universities and colleges outside of Texas. Since we are only including students who attend college here, this is a measure of the "intensive" margin of what a student majors in conditional on them attending college. Later on, we will also present results that incorporate the "extensive" margin of whether a student attends one of the college options we consider.

[35] We describe how we divide two-digit CIP codes into the "STEM" and "Non-STEM" categories in appendix section A.6.

wild cluster bootstrap to test the statistical significance of the 2SLS estimates, and place the resulting p-values at the bottom of the table.

The results show that the effects of local labor market conditions among STEM majors are much more precisely estimated, and show consistent positive effects. In appendix section B.3, we present similar robustness checks for these regressions as we did for the high school course taking regressions above.[36]

Table 4: Change in the Share of College-Bound High School Graduates Choosing $m$ as their First Major in College

| | (2)<br>2-year | (4)<br>4-year | (6)<br>6-year | (8)<br>8-year | (10)<br>10-year |
|---|---|---|---|---|---|
| 2SLS | 1.5376 | 1.3149 | 1.5956 | 1.8343 | 2.4063 |
| | (0.8706) | (1.3821) | (2.9268) | (3.1406) | (4.1204) |
| 2SLS STEM | 0.7077** | 0.5621** | 0.4929** | 0.4770** | 0.4790* |
| | (0.2699) | (0.2131) | (0.1711) | (0.1740) | (0.1949) |
| 2SLS NonSTEM | 2.0733 | 2.2386 | -9.7656 | -15.9694 | -13.1304 |
| | (1.4753) | (3.7166) | (95.7088) | (181.2749) | (103.7415) |
| Observations | 41,004 | 36,448 | 31,892 | 27,336 | 22,780 |
| Boot. P Value for 2SLS | 0.023 | 0.238 | 0.403 | 0.376 | 0.360 |
| Boot. P Value for 2SLS STEM | 0.012 | 0.019 | 0.009 | 0.016 | 0.027 |
| Boot. P Value for 2SLS NonSTEM | 0.060 | 0.334 | 0.858 | 0.879 | 0.837 |
| F-Stat for 2SLS | 6.034 | 2.396 | 0.700 | 0.732 | 0.542 |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the NonSTEM row used the other 76.471%.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Another possible way of preparing these results would be to define $Y_{m\ell t}$ as the share of all students graduating from high school in location $\ell$ at time $t$ that choose major $m$ as their first major in college. The results shown in Table 4 computed this share only over students who attended college, but this new share also includes students that graduated high school and did not attend a Texas public or community college.[37] Therefore, in addition to the "intensive" margin of deciding what to major in once at college, it also measures the "extensive" margin of deciding to attend college at all. It is a unique advantage of our data set that we are able to measure this share, because data sets which are specific to college can only observe students conditional on them having made the decision to attend college. The results of using this "extensive" method to calculate the $Y_{m\ell t}$ share are shown in table 25 in appendix B.3. The results are similar to those showed at the intensive margin, with consistent positive effects of local labor markets among STEM classes and statistically insignificant effects for Non-STEM classes.

### 4.3 The Majors that Students have when they Graduate College

We now turn to estimating the effects of local labor market conditions on the majors that students have when they graduate college. There are many potential years in which local labor market conditions could influence this outcome, from a student's high school years or before all the way up until the year they graduate college. For fitting the structural model, we will use moments taken at two of these time periods. First, we will look at the effects of local labor market conditions measured in the year a student graduates high school on the final major that they have when graduating college. To estimate this, we set

[36] We also include additional tables in that appendix section which present these estimates separately for students attending community college and for students attending public universities, as well as a table which shows these results using the college's location rather than the student's high school location.

[37] Since our results shown here are only for students that attend public universities and community colleges in Texas covered by the TSP, we cannot say for certain that the additional students included in this share did not attend college anywhere. However, for certain years we can pair our data with data from independent universities in Texas and all colleges covered by the NSC database, which together comprise a much more thorough list of the colleges that students could have attended. In the future, we can use these additional data sets to check the results that we show here.

$Y_{m\ell t}$ in equation (4) equal to the share of all students graduating from high school in location $\ell$ at time $t$ that graduate from college with major $m$ and do so within 6 years of their high school graduation year.[38] Table 5 shows the results of estimating this equation, and the results are similar to those shown for first major choices in the previous section. "STEM" majors are the only ones with statistically significant effects, and those effects are positive.

Table 5: Change in the Share of College Graduates Choosing $m$ as their Final Major in College, with labor market conditions measured in the Senior Year of High School

| | (2)<br>2-year | (4)<br>4-year | (6)<br>6-year | (8)<br>8-year | (10)<br>10-year |
|---|---|---|---|---|---|
| 2SLS | -0.1949 | -0.7927 | -0.9348 | 0.0561 | 0.0332 |
| | (0.6499) | (1.1593) | (1.8148) | (0.5678) | (0.6927) |
| 2SLS STEM | 0.6342 | 0.3523* | 0.2962* | 0.2683 | 0.3337* |
| | (0.4082) | (0.1689) | (0.1485) | (0.1377) | (0.1661) |
| 2SLS NonSTEM | -0.8122 | -3.2027 | 12.3443 | -0.8434 | -1.3981 |
| | (1.2271) | (7.1581) | (113.3491) | (4.2624) | (8.0831) |
| Observations | 31,824 | 27,268 | 22,746 | 18,190 | 13,634 |
| Boot. P Value for 2SLS | 0.736 | 0.408 | 0.425 | 0.884 | 0.936 |
| Boot. P Value for 2SLS STEM | 0.117 | 0.068 | 0.085 | 0.081 | 0.096 |
| Boot. P Value for 2SLS NonSTEM | 0.384 | 0.503 | 0.842 | 0.721 | 0.762 |
| F-Stat for 2SLS | 3.497 | 1.513 | 0.593 | 1.523 | 0.977 |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the NonSTEM row used the other 76.471%.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Second, we will look at the effects of local labor market conditions measured in the year a student graduates college on the final major that they have when in that year. Taking all students who graduated high school in location $\ell$ and then went on to graduate college in year $t$, here $Y_{m\ell t}$ is the share of those students that graduated with major $m$. Table 6 shows the results of those regressions. The differences between tables 5 and 6 encode important information about the relative importance of local labor market conditions at different times in a student's progression, and will be useful in fitting the structural model.[39] Qualitatively, though, the effects are again broadly similar to what has been seen before, remaining largely positive for STEM majors. One difference for the Non-STEM majors is also showing a statistically significant positive effect.

### 4.4 The Effects of Relevant Employment on Local Wages

Having established that students' choices between academic subjects in high school and college respond to local employment levels in industries where those subjects are relevant, there remains a question about why students respond in this way. In this section, we will examine the channel of wages. One possible intuition here is that industries may need to raise their wages in order to capture a larger share of employed workers with a certain skill set in a local area. To the extent that this occurs, the resulting connection between the relevant employment share and wages may help to explain why students are responding to the relevant employment rates: By studying subjects that are more highly demanded locally, they are able to secure higher wages for themselves in the local economy.

We create a measure of relevant wages $RW_{m\ell t}$ for each major group $m$, location $\ell$, and year $t$. Specifically, we define $RW_{m\ell t}$ as the expected average industry wage for employed workers with major $m$, and construct it similarly to our measure of relevant employment $RE_{m\ell t}$. More details are provided in appendix section A.5.

Table 7 shows the results of estimating equation (4) with the percentage change in $RW_{m\ell t}$ as the $Y$ variable.[40] To take

[38] Placing a time limit like "within 6 years" is important, because we cannot rule out that the students without degrees by the end of our sample period will not ever earn a degree with a public or community college in Texas.

[39] Tables 28 and 29 in appendix section B.3 present similar robustness checks for these results as were prepared elsewhere in this section, and table 30 in that section provides a version of table 6 in which labor market conditions are measured around the college location.

[40] That is, we set $(RW_{m\ell t} - RW_{m\ell t-d})/RW_{m\ell t-d}$ as the $\Delta_d Y_{m\ell t}$ variable. Estimating this regression in percentage changes like this will hopefully make it easier to use the estimates for fitting our structural model.

Table 6: Change in the Share of College Graduates Choosing *m* as their Final Major in College, with labor market conditions measured at the year of college graduation

| | (2) 2-year | (4) 4-year | (6) 6-year | (8) 8-year | (10) 10-year |
|---|---|---|---|---|---|
| 2SLS | 0.2176 | 0.7759 | 1.4049 | -0.6419 | 5.0134 |
| | (0.2515) | (0.5142) | (2.0767) | (2.7116) | (9.4967) |
| 2SLS STEM | -0.0772 | 1.0337*** | 1.0643** | 0.4844 | 0.2694 |
| | (0.2314) | (0.3060) | (0.3994) | (0.2477) | (0.3470) |
| 2SLS NonSTEM | 0.3344 | 0.4698 | 40.2345 | 2.6526 | 2.2475* |
| | (0.3849) | (0.8293) | (8927.6760) | (5.7570) | (1.1193) |
| Observations | 29,614 | 25,058 | 20,502 | 15,946 | 11,390 |
| Boot. P Value for 2SLS | 0.370 | 0.072 | 0.310 | 0.676 | 0.427 |
| Boot. P Value for 2SLS STEM | 0.753 | 0.004 | 0.016 | 0.094 | 0.529 |
| Boot. P Value for 2SLS NonSTEM | 0.366 | 0.528 | 0.994 | 0.446 | 0.059 |
| F-Stat for 2SLS | 17.565 | 4.971 | 0.517 | 0.316 | 0.220 |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the NonSTEM row used the other 76.471%.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

the two-year change estimated in the "2SLS STEM" row as an example, the way to interpret these coefficients is that a 1 percentage point increase in the relevant employment share $RE_{m\ell t}$ over two years increases relevant wages $RW_{m\ell t}$ by 5.2375 percent.

Here, the "2SLS STEM" row shows that among STEM majors, movement in the relevant employment share has a large effect on wages which generally grows in magnitude as the lags increase. The "2SLS Non-STEM" row, by contrast, shows effects that are smaller and less precise. This may help to explain why student's first major responses seem more responsive to local labor market conditions for STEM majors as opposed to Non-STEM majors.

# 5 Model

In this section, we outline a structural model that we intend to estimate using a set of the shift-share estimates like those described above as moments. The central goal of this model is to describe how students make decisions over time, and particularly how lagged values of local relevant employment shares could influence a student's present decisions. In our model, students can potentially begin to prepare for their eventual college major by altering the courses that they take while in high school.

There are two primary channels by which a student's past decisions can affect their present decisions within the model. First, decisions between courses in high school affect student's accumulation of subject-specific human capital. In this way, past employment shares can influence a student's present decisions if they led the student to invest in different kinds of human capital. Second, there may be transition costs between different academic tracks and options. This is largely captured here in the utility costs associated with switching tracks, which may reflect things like administrative hurdles to switching majors, psychological costs associated with switching one's plans, or the difficulty of "catching up" to peers that have already been pursuing a certain track.

Here is a general outline of the model timing: A model period represents two years. At the beginning of each period, students observe the local labor market.

1. Period 1 (Freshman and Sophomore Years of High School): Students enter the model with initial STEM and non-STEM abilities. At the end of the period, they choose between a STEM-intensive senior year course-load and non-STEM intensive course load.

2. Period 2 (Junior and Senior Years of High School): After accumulating human capital from their courses, students choose whether to go to college or directly into the labor market. If they choose college, they also pick their initial

Table 7: Percentage Change in Relevant Wages

| | (2) 2-year | (4) 4-year | (6) 6-year | (8) 8-year | (10) 10-year |
|---|---|---|---|---|---|
| 2SLS | 4.4736 | 6.1818 | 20.8868 | 17.5381 | 20.3676 |
| | (3.1330) | (5.0250) | (33.6415) | (22.2037) | (31.1140) |
| 2SLS STEM | 5.2375* | 7.2805*** | 14.5189*** | 16.3991*** | 15.9682*** |
| | (2.2587) | (1.9690) | (2.1652) | (2.0423) | (2.1182) |
| 2SLS NonSTEM | 3.4740 | 2.6421 | 3.8840 | 14.6670 | 8.7188 |
| | (4.4242) | (9.8271) | (23.8056) | (76.8673) | (39.5854) |
| Observations | 45,492 | 40,936 | 36,414 | 31,858 | 27,302 |
| Boot. P Value for 2SLS | 0.149 | 0.120 | 0.300 | 0.197 | 0.277 |
| Boot. P Value for 2SLS STEM | 0.052 | 0.006 | 0.000 | 0.000 | 0.000 |
| Boot. P Value for 2SLS NonSTEM | 0.336 | 0.661 | 0.779 | 0.746 | 0.712 |
| F-Stat for 2SLS | 6.675 | 2.709 | 0.392 | 0.624 | 0.419 |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the NonSTEM row used the other 76.471%.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

major in this period.

3. Period 3 (Freshman and Sophomore Years of College): Students choose to whether to continue in college or drop out. If they decide to continue, they also decide whether to switch majors for period 4.

4. Period 4 (Junior and Senior Years of College): Students choose whether to drop out or graduate with their chosen major.

5. Labor Force: Once in the labor force, students choose which industry to work in each period. Workers retire $T$ periods after graduating high school, so students who graduated from college work for $T-2$ periods.

This information on the model timing is presented graphically in figure 3. For easier exposition below, we will present these periods backwards from the time that a student is in the labor market.

## 5.1 Labor Market

A student's college outcome defines their college degree "$m$" in the labor market. This is the two-digit CIP code of their degree for students that graduate college, and a "non-college" type for those that do not. A worker is able to supply an amount of "effective labor" that depends on their type, experience $e_{it}$, and STEM and non-STEM abilities:

$$l_{imt} = A_{is13}^{\tau_m^s} A_{in13}^{\tau_m^n} e_{it}^{\tau_m^e} \tag{7}$$

$A_{is13}$ and $A_{in13}$ are the STEM and non-STEM abilities from the end of twelfth grade.[41] We measure experience $e_{it}$ as the number of periods since an individual has graduated high school.

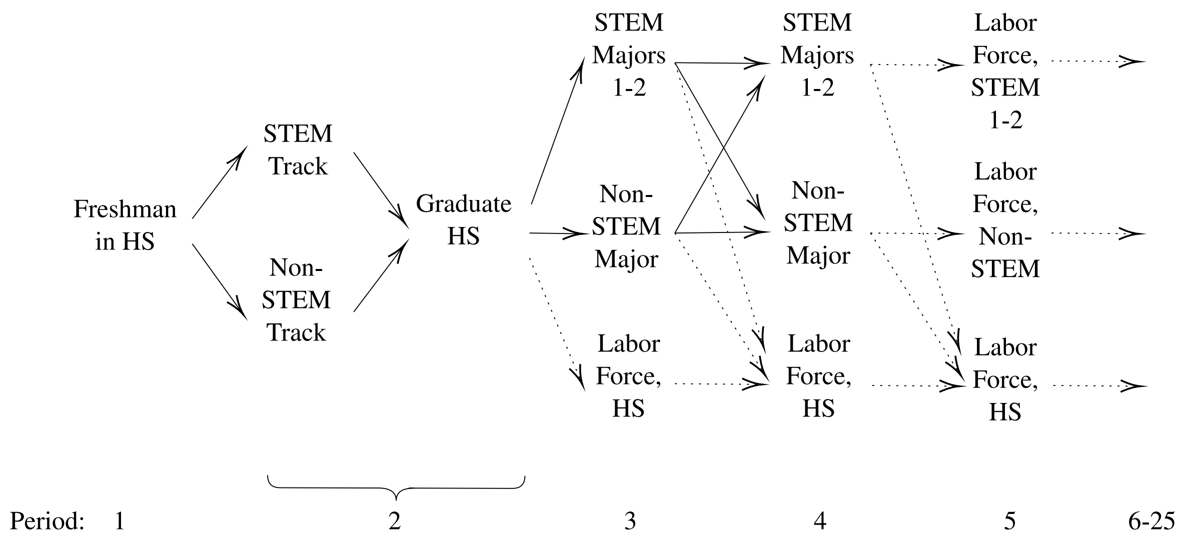
Each type of worker receives a different wage for their effective labor $w_{mt}$. The total amount of wages a person receives in the labor market is then given by $w_{imt} = l_{imt} w_{mt}$, and this determines the amount of utility that they receive each period from the labor market:

$$U_{imt} = \delta_{0m}^{lab} + \delta_1^{lab} \ln(w_{imt}) + \delta_3^{lab} \mathbb{1}\{\text{Dropped out last period}\}$$

Here, $\delta_3^{lab}$ is a one-time utility cost or benefit a student receives if and when they drop out of college without earning a college degree.

[41] The idea here is to get a measure of their human capital when entering college. The $\tau$ terms in equation (7) capture both the growth in human capital that occurs through pursuing degree type $m$, and also the productivities of STEM or non-STEM ability in generating effective labor. We will not differentiate between these two functions of the $\tau$ terms.

Figure 3: Graphical Depiction of the Model Timing




Each period, the share of employment in STEM-intensive industries $SE_t$ is realized. We assume that this share follows the following AR-1 process over time:

$$SE_{t+1} = \rho_0^{SE} + \rho_1^{SE} SE_t + \omega_t^{SE} \tag{8}$$

This share plays an important role within the model because it is predictive of the effective wages for each degree type in the next period:

$$\log\left(w_{mt+1}\right) = \rho_{0m}^{w} + \rho_{1m}^{w} SE_t + \omega_{mt}^{w} \tag{9}$$

From the moment that they enter the model onwards, students are able to use information on the current $SE_t$ to forecast the future path of $SE_t$ and their eventual labor market wages using formulas (8) and (9). Putting these two equations together allows students to predict the future path of $SE_t$ and $w_{mt}$. Specifically, a student curious about their wages at some time $\tilde{t}$ in the future would use equation (8) to project the path of the STEM-Intensive employment share up until $\tilde{t}-1$, and then they would use the expected $SE_{\tilde{t}-1}$ to forecast the wages of interest $w_{m\tilde{t}}$ using equation (9).[42] The $\omega$ terms create uncertainty in these predictions, and students will factor that uncertainty into the decisions that they make. We assume $E\left[\omega_t^{SE}\right] = E\left[\omega_{mt}^{w}\right] = 0$, that $\omega_t^{SE}$ is normally distributed, and that $\omega_{mt}^{w}$ and $\omega_t^{SE}$ are uncorrelated.

It would be possible to assume that equations (8) and (9) truly reflect the aggregate movements in STEM employment shares and wages, in which case the students in our model would have fully rational expectations. Within the current estimation of the model, we do use equation (8) to develop $SE_t$ shares over time.[43] However, when comparing to the real world, it seems more reasonable to view students' use of equations (8) and (9) as a restriction that we impose on their ability to forecast future values. This is discussed in more detail in appendix section C.1.

From the perspective of a student in high school or a student in college, the value function upon entering the labor force is given by their utility from having a degree and their earnings. These earnings depends on a student's ability, which will never change while they are in the labor force, and the future values of wages for their type, which could change. To make the notation less cluttered, let $\Omega^{SE} = \left\{\omega_t^{SE}\right\}_t$ represent the set of all future values of $\omega_t^{SE}$ shocks, and $t'$ represent the number of periods a person has spent in the labor force from 1 to $T_{im}$.[44] Ignoring the potential one-time dropout cost $\delta_3^{lab}$, a student's

[42] In matrix form, the student would create the following expectation about the $SE_t$ and $w_{mt+1}$ values $d$ periods from now:

$$E\begin{bmatrix} 1 \\ SE_{t+d} \\ \log\left(w_{mt+d}\right) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \rho_0^{SE} & \rho_1^{SE} & 0 \\ \rho_0^{w} & \rho_1^{w} & 0 \end{bmatrix}^d \begin{bmatrix} 1 \\ SE_t \\ 0 \end{bmatrix}$$

[43] In the future, we intend to change this so that the true development of $SE_t$ shares over time is more realistic to how this process occurs in the real world. Once we make this adjustment, students' use of equations (8) and (9) will not be perfectly rational within the context of the model.

[44] This is indexed by $m$ because different worker types will have different amounts of time to work in the labor force. Students that begin working right after high school will work for two additional periods in the labor force (representing 4 years), since they do not need to spend that time in college.

value function when entering the labor market is given by

$$\begin{aligned} v_{it}^{labor}\left(A_{ist}, A_{int}, m, SE_{t-1}\right) &= \int_{\Omega^{SE}} \sum_{t'=1}^{T_{mi}} \beta^{t'-1} \left(\delta_{0m}^{lab} + \delta_1 \log\left(w_{ikmt}\right)\right) dG\left(\Omega^{SE}\right) \\ &= \left(\frac{1-\beta^{T_{mi}}}{1-\beta}\right) \delta_{0m}^{lab} + \delta_1^{lab} \int_{\Omega^{SE}} \sum_{t'=1}^{T_{mi}} \beta^{t'-1} \left(\log\left(l_{imt}\right) + \log\left(w_{mt}\right)\right) dG\left(\Omega^{SE}\right) \\ &= \left(\frac{1-\beta^{T_{mi}}}{1-\beta}\right) \delta_{0m}^{lab} + \delta_1^{lab} \left(\sum_{t'=1}^{T_{mi}} \beta^{t'-1} \log\left(l_{imt}\right)\right) + \delta_1^{lab} \sum_{t'=1}^{T_{mi}} \int_{\Omega^{SE}} \beta^{t'-1} \log\left(w_{mt}\right) dG\left(\Omega^{SE}\right) \end{aligned} \quad (10)$$

where $G\left(\Omega^{SE}\right)$ is the distribution of future $\omega_t^{SE}$ shocks. The first two terms term above do not depend on wages in the labor market. We can simplify them as

$$\left(\frac{1-\beta^{T_{mi}}}{1-\beta}\right) \delta_{0m}^{lab} + \delta_1^{lab} \left(\frac{1-\beta^{T_{mi}}}{1-\beta}\right) \left(\log\left(\tau_m^c\right) + \tau_m^n \log\left(A_{in}\right) + \tau_m^s \log\left(A_{is}\right)\right) + \delta_1^{lab} \tau_m^e \sum_{t'=1}^{T_{mi}} \beta^{t'-1} log\left(e_{it}\right)$$

The second term of equation (10) involves taking the expectation of log $w_{mt}$ terms each period using equations (8) and (9), which is also written in the equivalent matrix notation in footnote 42. Since the student's future utility depends upon log wages and $E\left[\omega_{mt}^w\right] = 0$, the $\omega_{mt}^w$ terms drop out of the expectation here. An implication of this is that increasing or decreasing the variance of the $\omega_{mt}^w$ terms will not affect student decisions. The variance of the $\omega_t^{SE}$ terms, however, may still affect student decisions, because students integrated over those shocks in previous periods when their preferences were not so linear.[45] This will be made clear in the sections below.

## 5.2 College

While in college, students have a major $m_{it}$ that was chosen in the last period,[46] and observe the current STEM-industry employment share $SE_t$. They also have some amount of STEM ability $A_{is13}$ and non-STEM ability $A_{in13}$ from high school.[47] We will not explicitly model human capital accumulation in college, but recall from the previous section that the $\tau_m$ terms will capture growth in human capital specific to a student's major $m$. For convenience, let $c$ represent whether major $m$ is in a STEM or non-STEM category. The state space is

$$\{A_{is13}, A_{in13}, m_{it}, SE_t\}$$

Which will be shortened below to

$$\{A_{i13}, m_{it}, SE_t\}$$

While in college, a student's flow utility depends upon a student's abilities and an idiosyncratic preference shock $\varepsilon_{imt}$:

$$\begin{aligned} U_{ict} &= \delta_{0m}^{col} + \delta_{1c}^{col} \log\left(A_{ic13}\right) + \delta_{2c}^{col} SE_t + transfer\ \ costs_{ct} + \varepsilon_{imt} \\ &= u_{imt} + \varepsilon_{imt} \end{aligned}$$

Here, $transfer\ \ costs_{ct}$ gives costs for transferring between subjects that can depend on a student's year in school. These depend upon the specific period, and so they will be explained in more detail below.

### 5.2.1 Period 4 (Junior and Senior Years of College)

If students decide to switch between a STEM and a Non-STEM major between period 3 and 4 in college, they will face a transfer cost in their utility function:

$$transfer\ \ costs_{c4} = \delta_{c4}^{transfer} \mathbb{1}\left\{c_{i3} \neq c_{i4}\right\} \quad (11)$$

[45] In the setup of Buchinsky and Leslie (2010), it is important that individuals take into account the fact that their forecasted estimates are uncertain. Students take this uncertainty into account in our context as well, but we are clarifying here that the only uncertainty in the $\omega_t^{SE}$ will actually affect student decisions.

[46] Since a student's college major defines their type in the labor market, we will use the same notation $m$ for both of these things. Here it is indexed by time $t$ because students can change their major while in college, but after graduating from college we will assume that their degree type $m$ is fixed.

[47] The "13" indicates that these abilities are taken from the end of twelfth grade, and the beginning of the period after.

At the end of period 4, students choose whether to enter the labor market as a college graduate with their major $m_{it}$ or to enter the labor market without graduating. For simplicity, we index these possible options with $m_{it+1}$. Their value function when making this decision is the value of entering the labor market, $v_{imt}^{labor}$. Students receive a type-1 extreme value shock $\varepsilon_{imt+1}$ between these choices, and use it to make their decision:

$$m_{it+1} = \text{argmax}\left\{v_{it+1}^{labor}\left(A_{is13}, A_{in13}, m_{it+1}, SE_t\right) + \varepsilon_{imt+1}\right\}$$

The value function in this period is given by

$$v_{it}\left(A_{i13}, m_{it}, SE_{t-1}\right) = \int_{\omega_{t-1}^{SE}} \left(u_{imt} + \beta \log\left[\sum_{m_{it+1}} \exp\left(v_{imt+1}^{labor}\left(A_{i13}, m_{it+1}, SE_t\right)\right)\right]\right) dF\left(\omega_{t-1}^{SE}\right) \tag{12}$$

where $dF\left(\omega_{t-1}^{SE}\right)$ is the normal distribution for $\omega_{t-1}^{SE}$.

#### 5.2.2 Period 3 (Freshman and Sophomore Years of College)

Upon entering college, students will pay a utility cost meant to capture the adjustment between a non-STEM intensive courseload in high school and a STEM intensive courseload in college, or vice versa.

$$transfer\ \ costs_{c3} = \delta_{c3}^{transfer} S_{it} \tag{13}$$

Here, $S_{it}$ is the share of a student's senior-year classes that were in STEM. This will be made more clear in the following section on high school.

At the end of period 3, students choose whether to stay in college or drop out. If they decide to stay in college, they also choose whether to switch their major. For simplicity, we will represent this as a single decision where entering the labor market is included as an option for the major.[48] Their value function when making this decision is

$$v_{it}\left(A_{i13}, m_{it}, SE_{t-1}\right) = \int_{\omega_{t-1}^{SE}} \left(u_{imt} + \beta \log\left[\sum_{m_{it+1}} \exp\left(v_{it+1}\left(A_{i13}, m_{it+1}, SE_t\right)\right)\right]\right) dF\left(\omega_{t-1}^{SE}\right) \tag{14}$$

Then, given an ideosyncratic preference shock $\varepsilon_{imt+1}$, students make their decision for next period as

$$m_{it+1} = \text{argmax}\left\{v_{it+1}\left(A_{i13}, m_{it+1}, SE_t\right) + \varepsilon_{imt+1}\right\}$$

### 5.3 High School

#### 5.3.1 Period 2 (Junior and Senior Years of High School)

In their junior and senior years of high school, students take a combination of STEM and Non-STEM classes. The classes that students take contribute to their STEM and non-STEM abilities, where STEM classes are more helpful for STEM ability and non-STEM classes are more helpful for non-STEM ability.

While in high school, a student's STEM or non-STEM human capital for the next grade level are produced as follows:

$$\log\left(A_{ict+1}\right) = \gamma_{0c} + \gamma_{1c} \log\left(A_{ist}\right) + \gamma_{2c} \log\left(A_{int}\right) + \gamma_{3c} P_{ist} + \gamma_{4c} P_{int} + \gamma_{5c} S_{it} \tag{15}$$

Here, the $P_{ict}$ terms are indicators for whether a student passes all of their classes of type $c$. The probability of passing all classes in a certain subject $c$ is

$$\mathbb{P}\left\{P_{ict} = 1\right\} = \Phi\left(\kappa_{0c} + \kappa_{1c} \log\left(A_{ist}\right) + \kappa_{2c} \log\left(A_{int}\right) + \kappa_{3c} S_{it}\right) \tag{16}$$

Equations (15) and (16) will together serve to update student human capital in the junior and senior years. We define them over an academic grade level rather than a model period for ease of estimating in the data, and so human capital will be updated twice in this period.

[48] In the future we can make this a nested logit, where "staying in school" and "going to the labor market" are the two nests, and after picking "staying in school" a student can also choose their major.

In the first update, representing a student's junior year, we give all modeled students the average STEM class share from the data for that grade level. For their senior year, however, we allow model students to choose either a "STEM intensive" or "non-STEM intensive" track. Index these tracks by $c$. Each track $c$ involves STEM classes and non-STEM classes, with the difference being that the share of classes taken in STEM $S$ is higher for the STEM track. Students commit to their senior-year track option at the end of the prior model period (period 1).

While in the second period, the student receives utility from their chosen track as follows:

$$\begin{aligned} U_{ict} &= \delta_c^{hs} + \delta_1^{hs} S_{it} SE_{it} + \varepsilon_{ict} \\ &= u_{ict} + \varepsilon_{ict} \end{aligned}$$

Here, $\varepsilon_{imt}$ is a type-1 extreme value preference shock for the course selection. We normalize the $\delta_c^{hs}$ constant to zero for the non-STEM track.

Then, students choose either to go to college or the labor market after high school. If they choose college, they also choose an initial major. Again, for simplicity, we index all of these choices with $m_{it+1}$. After receive an idiosyncratic preference shock between these options $\varepsilon_{imt}$, they then choose based on

$$m_{it+1} = \operatorname{argmax}\left\{v_{it+1}\left(A_{i13}, m_{it+1}, SE_t\right) + \varepsilon_{imt+1}\right\}$$

The value function in this period is

$$v_{it}\left(A_{ist}, A_{int}, c_{it}, SE_{t-1}\right) = \sum_{\mathbb{p}\in\mathscr{P}} \mathbb{P}\left(\mathbb{p}\right) \int_{\omega_{t-1}^{SE}} \left(u_{ict} + \beta \log\left[\sum_{m_{it+1}} \exp\left(v_{it+1}\left(A_{i13}, m_{it+1}, SE_t\right)\right)\right]\right) dF\left(\omega_{t-1}^{SE}\right) \tag{17}$$

where $\mathbb{p}$ represents a possible combination of passing STEM and non-STEM classes in the junior and senior year, and $\mathscr{P}$ is the set of all these possibilities.

### 5.3.2 Period 1 (Freshman and Sophomore Years of High School)

Students enter the model knowing their what their initial STEM and non-STEM abilities $A_{is11}$ and $A_{in11}$ will be at the beginning of grade 11. At the end of the period, students receive a preference shock $\varepsilon_{ict+1}$ and choose a high school track $c$. This decision is given by

$$c_{it+1} = \operatorname{argmax}\left\{v_{it+1}\left(A_{ist+1}, A_{int+1}, c_{it+1}, SE_t\right) + \varepsilon_{ict+1}\right\}$$

# 6 Identification and Moments

To ease the computational burden of solving this model, we will estimate a large number of the parameters outside the model. Then, we will estimate the remaining parameters using simulated method of moments estimation.

## 6.1 Parameters Identified Outside the Model

### 6.1.1 Terms for Human Capital Growth and Passing Classes in High School

The first terms that we will estimate outside of the model are the $\gamma$ terms related to human capital accumulation and the $\kappa$ terms related to the probability of passing classes outside of the model. This will involve estimating the equations (15) and (16).

There is no single measure of STEM or non-STEM ability within our data. Instead, we have a number of TAAS standardized test scores shown in table 8 that we will treat as measures of unobserved STEM and non-STEM ability. Specifically, we assume that the measure for ability type $c$ from test $r$ takes the form

$$Z_{icg}^r = \mu_c^r + \lambda_c^r \log\left(A_{icg}\right) + \omega_{icg}^r \tag{18}$$

where $Z_{icg}^r$ is a student's score on test $r$ in grade $g$, and the student's underlying ability is given by $\log\left(A_{icg}\right)$. The $\mu_c^r$ and $\lambda_c^r$ are terms that adjust for how test scores may be given on different scales,[49] and for how some tests are harder than others

[49] For example, one score may be given on a scale from 0 to 100, while another scale goes from 0 to 1000.

while measuring the same underlying ability. Finally, the $\omega^r_{icg}$ term reflects "measurement error," or differences between a student's score and what would be predicted based on their underlying ability. We assume that this measurement error has a mean of zero[50] and is independent across students and tests. However, we do not take a specific stance on where this measurement error comes from. It could be influenced by idiosyncratic shocks to students like distractions in the classroom, for example, and its importance may differ across tests if certain tests do a better job of measuring ability than others.[51]

Table 8: Standardized Test Scores in the TSP

| Grade | STEM | Non-STEM |
|---|---|---|
| 9[1] | TAAS Math, Science | TAAS Writing, Reading, Social Studies |
| 11[1] | TAAS Math | TAAS Writing, Reading |
| 12[2] | SAT Math | SAT Verbal |

[1] Each TAAS test listed here is typically taken toward the end of the previous year: The tests listed for ninth grade are generally taken in April at the end of the eighth grade year, and the tests for 11th grade are taken in February of the tenth grade year. We associate these tests with the following grade level to capture a student's ability when entering that grade level.

[2] We count SAT tests taken from May before the school year a student graduates high school through April of that school year as being taken in twelfth grade.

In this kind of measurement framework, it is straightforward to identify the $\mu^r_c$ and $\lambda^r_c$ terms in equation (18) for a given ability type $c$ when the average latent ability $E\left(\log\left(A_{icg}\right)\right)$ and the variance of latent ability $Var\left(\log\left(A_{icg}\right)\right)$ are both known. In the case that three tests $r$, $r'$, and $r''$ are available for ability type $c$, the equation for $\lambda^r_c$ is

$$\lambda^r_c = \frac{1}{\text{Var}\left(A_{icg}\right)}\sqrt{\frac{\text{Cov}\left(Z^r_{icg}, Z^{r'}_{icg}\right)\text{Cov}\left(Z^r_{icg}, Z^{r''}_{icg}\right)}{\text{Cov}\left(Z^{r'}_{icg}, Z^{r''}_{icg}\right)}}$$

and the equation for $\mu^r_c$ is

$$\mu^r_c = E\left(Z^r_{icg}\right) - \lambda^r_c E\left(A_{icg}\right)$$

Appendix section C.2 gives additional details on how we implement this estimation in the data, particularly in cases where three tests are not available in ability subject $c$, but each case follows the same general pattern presented here.

The problem comes when the the average latent ability $E\left(\log\left(A_{icg}\right)\right)$ and the variance of latent ability $Var\left(\log\left(A_{icg}\right)\right)$ are not known. If only one period were being used, this could be quickly resolved by normalizing the latent ability for the population so that $E\left(\log\left(A_{icg}\right)\right) = 0$ and $Var\left(\log\left(A_{icg}\right)\right) = 1$. However, our goal will be to identify human capital accumulation across subsequent grade levels. In that case normalization cannot fully solve the problem, because it is not reasonable to normalize that $E\left(\log\left(A_{icg}\right)\right) = 0$ both in a grade $g$ and the next grade $g+1$. Ideally, the average ability will have increased over that interval.

Agostinelli and Wiswall (2022) show that in contexts like these, in order to identify the development of human capital over time, we need to have either age-invariant measures or an assumption about the underlying distribution of human capital. The TAAS standardized tests are good to use in our application because all students are required to take them, and the same test is administered to all students. Unfortunately, since a different test is administered to each grade level, they do not provide age-invariant measures of ability over time. Instead of using age-invariant measures, then, we will take the second approach suggested by Agostinelli and Wiswall (2022), and make an assumption about the underlying distribution of ability. Specifically, we will assume that two features of the ability distribution in Texas match with data from the NLSY97 data:

1. We assume that the aggregate growth rate between grades in each average latent ability and in the variances of those abilities are the same between the TSP and a matched sample from the NLSY97.

[50] The assumption that $E\left(\omega^r_{icg}\right) = 0$ comes without loss of generality, because if it were the case that $E\left(\omega^r_{icg}\right)$ equaled some $x \neq 0$, this $x$ could be absorbed into the constant $\mu^r_c$. In other words, we would be estimating the coefficients of $Z^r_{icg} = \tilde{\mu}^r_c + \lambda^r_c \log\left(A_{ic}\right) + \tilde{\omega}^r_{icg}$, where $\tilde{\mu}^r_c = \mu^r_c + x$ and $\tilde{\omega}^r_{icg} = \omega^r_{icg} - x$.

[51] As a measure of how good each test is at capturing underlying abilities, we report the signal-to-noise ratios of each test that we consider in table 32.

2. We assume that the correlation between STEM and non-STEM ability in eleventh grade from the TSP matches this correlation from a matched sample in the NLSY97.

To make these assumptions as plausible as possible,[52] we take the following steps to make the NLSY97 match our data in the ERC: First, we limit the NLSY97 sample to students attending public high school who obtained a high school diploma. Second, we use data collected over a similar time frame from each sample. The ASVAB test was given to NLSY97 respondents from the summer of 1997 through the spring of 1998, and TAAS test was administered in Texas from 1991 until it was replaced in 2003. For our sample, we will use TAAS tests taken from 2000 to 2003. Third, when estimating things in the NLSY97, we use weights taken from the TSP for the sample of students that the NLSY97 estimates will be applied to. This is done by segmenting the student population into demographic groups $D$, which are defined by a combination of their race or ethnicity and sex. Then, the original sample weight $w_i$ provided by the NLSY97 is modified as follows:[53]

$$\text{new weight} = \left(\frac{\#\text{ in group } D_i \text{ in TSP}}{\#\text{ in TSP}}\right)\frac{w_i}{\sum_{j\in D_i} w_j} \tag{19}$$

Finally, we use the SAT Math and Verbal scores as "shared measures" of ability, which are present in both the NLSY97 and the TSP. This allows us to identify the relevant measurement parameters for the SAT tests in the NLSY97, and apply those same measurement parameters to the SAT data from the TSP.[54] If we did not have this shared measure, we would have to make the additional stronger assumption that the levels of aggregate expected ability and aggregate variance matched between the TSP and the NLSY97. As it is, through the SAT tests we are able to identify the differences in these levels between the two sources, and need only to assume that their growth rates from one grade level to the next are the same.

Table 9: Standardized Test Scores in the NLSY97

| Grade | STEM | Non-STEM |
|---|---|---|
| 9 | PIAT Math | |
| 12[1] | SAT Math | SAT Verbal |
| All Grades: | ASVAB STEM Tests: Mathematical Knowledge, Arithmetic Reasoning, Numerical Operations, General Science | ASVAB Non-STEM Tests: Word Knowledge, Paragraph Comprehension |

[1] To match our method in the ERC data, we count SAT tests taken from May before the school year a student graduates high school through April of that school year as being taken in twelfth grade.

The ASVAB test in the NLSY97 gives us four measures of STEM ability and two measures of non-STEM ability.[55] The PIAT Math Test was also given to ninth graders over the same time period, and serves as additional measure of STEM ability for that grade level.[56]

We make the normalization that $E\left(\log\left(A_{ict}\right)\right) = 0$ and $\text{Var}\left(\log\left(A_{ict}\right)\right) = 1$ for both ability types $c$ in grade 9 in the NLSY97. Then, we can identify the measurement parameters for all high school grade levels in the NLSY97, and the assumptions we have made above along with the presence of the SAT as a shared measure allows us to do the same for the TSP.[57] Table 32 in appendix section C.2 shows the estimated measurement parameters and their signal to noise ratios.

[52] Given the focus of our model on how students react to local employment conditions, one potential criticism of this approach is that it ignores how the growth rate in expected ability or the variance of ability depends upon student reactions to aggregate movements in employment. We address this criticism more directly in appendix section C.2.

[53] We calculate this "new weight" multiple times based on the TSP student group we are trying to match. For example, in calculating the $\mu$ and $\lambda$ terms for SAT tests, we calculate $\left(\frac{\#\text{ in group } D_i \text{ in TSP}}{\#\text{ in TSP}}\right)$ based on students who graduated at or before the 2003-2004 school year, had non-missing ethnic and sex data, and took their SAT with the highest recorded combined score sometime from the (prior) summer through the spring of the school year they graduated high school. For this restriction, we use May as the first month of summer.

[54] Students self-select into taking the SAT, but we are assuming that this self-selection is the same for students in the NLSY97 and in the TSP once we correct the weights as shown in equation (19).

[55] The specific measures of STEM and non-STEM ability that we selected are shown in Table 9.

[56] Specifically, the ASVAB was administered from the summer of 1997 to the spring of 1988, and the PIAT Math test was administered in round 1 to all ninth grades from either February through October of 1997 or March through May of 1998. After round 1 of the NLSY97 survey, there were age restrictions on who received the PIAT Math test in subsequent rounds, and so rather than finding a way to correct for that selection we will only use the PIAT Math Test scores from round 1.

[57] More details on this are included in Appendix C.2.

Then, we can use these measurement parameters to identify the gamma coefficients from equation (15). Let

$$\tilde{Z}^r_{icg} \equiv \frac{Z^r_{icg} - \mu^r_c}{\lambda^r_c} = \log\left(A_{icg}\right) + \frac{\omega^r_{icg}}{\lambda^r_c} \tag{20}$$

be the residualized version of measure $Z^r_{icg}$. If we plug this residualized version into equation (15), we get

$$\tilde{Z}^{r'}_{icg+1} = \gamma_{0c} + \gamma_{1c}\tilde{Z}^r_{isg} + \gamma_{2c}\tilde{Z}^r_{ing} + \gamma_{3c}P_{isg} + \gamma_{4c}P_{ing} + \gamma_{5c}S_{ig} \tag{21}$$

or

$$\tilde{Z}^{r'}_{icg+1} = \gamma_{0c} + \gamma_{1c}\log\left(A_{isg}\right) + \gamma_{2c}\log\left(A_{ing}\right) + \gamma_{3c}P_{isg} + \gamma_{4c}P_{ing} + \gamma_{5c}S_{ig} - \gamma_{1c}\frac{\omega^r_{isg}}{\lambda^r_m} - \gamma_{2c}\frac{\omega^r_{ing}}{\lambda^r_n} + \frac{\omega^{r'}_{icg+1}}{\lambda^{r'}_c} \tag{22}$$

where the last three terms are unobserved. This creates a problem for estimating $\gamma_{1m}$ and $\gamma_{2m}$ with OLS, because $\tilde{Z}^r_{ist}$ and $\tilde{Z}^r_{int}$ will be correlated with things in the error term. To deal with this, we follow Agostinelli and Wiswall (2022) in using our additional residualized measures for STEM and non-STEM ability at time $t$ as instruments for $\tilde{Z}^r_{ist}$ and $\tilde{Z}^r_{int}$, and estimate equation (21) with two stage least squares.[58] The estimates for equation (16) can be prepared in a similar way, using an IV-probit. This gives us estimates of the $\gamma$ and $\kappa$ coefficients, which are shown in tables 10 and 11.

Table 10: Estimates for $\gamma$ terms related to Human Capital Accumulation

| | (1) Underclassmen STEM Growth | (2) Underclassmen Non-STEM Growth | (3) Upperclassmen STEM Growth | (4) Upperclassmen Non-STEM Growth |
|---|---|---|---|---|
| STEM Ability at Beg. of Period | 1.2192*** | 0.4309*** | 0.4567*** | -0.4353*** |
| | (0.0336) | (0.0298) | (0.0899) | (0.1100) |
| Non-STEM Ability at Beg. of Period | -0.3577*** | 0.7035*** | 0.6089*** | 1.5728*** |
| | (0.0369) | (0.0321) | (0.1166) | (0.1455) |
| Received Credit for All STEM Classes | 0.0602*** | -0.0180** | 0.0777*** | 0.0694*** |
| | (0.0104) | (0.0063) | (0.0138) | (0.0145) |
| Received Credit for All Non-STEM Classes | 0.0627*** | 0.0582*** | -0.0250 | -0.0745* |
| | (0.0048) | (0.0044) | (0.0223) | (0.0302) |
| Percent of Classes that are STEM | 0.6293*** | -0.2028*** | 0.9088*** | 0.4453*** |
| | (0.0941) | (0.0511) | (0.0820) | (0.0759) |
| Constant | 0.8856*** | 1.0709*** | -0.2048** | -0.2622*** |
| | (0.0354) | (0.0242) | (0.0659) | (0.0515) |
| Observations | 310052 | 310067 | 71799 | 71799 |
| F-Stat | 214.003 | 211.994 | 202.661 | 202.661 |

Standard errors in parentheses
Observations are clustered at the level of High School Commuting Zones. The underclassmen columns show the growth in ability from the beginning of ninth grade through the beginning of eleventh grade. The upperclassmen columns show the growth in ability from the beginning of 11th grade through when the SAT is taken, which is typically around the beginning of the twelfth grade year.
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

### 6.1.2 Separating Majors and Industries into Groups

In our model, workers in the labor market work for either a STEM-intensive industry or a non-STEM intensive industry. To mimic this in the data, we divide industries into one of these two groups. We do this by ranking industries at the two-digit NAICS code level in order of the share of their average "relevance" $r_{skt}$ for students graduating with STEM degrees. Then,

[58] This will demand having multiple measures of STEM and non-STEM ability at each grade level. Referring back to table 8, this presents a problem for STEM in grade 11 and STEM and non-STEM in grade 12. For those grade levels, we are able to use ability measures from prior grade levels to construct new ability measures as the fitted values from equation (21).

Table 11: Estimates for the $\kappa$ Terms related to the Probability of Passing all Classes

| | (1) Passing all STEM Classes | (2) Passing all Non-STEM Classes |
|---|---|---|
| STEM Ability at Beg. of Period | -0.0329 | -0.2405$^{*}$ |
| | (0.1144) | (0.1054) |
| Non-STEM Ability at Beg. of Period | 0.6851$^{***}$ | 0.8863$^{***}$ |
| | (0.1423) | (0.1264) |
| Percent of Classes that are STEM | -0.2418 | 0.9441$^{***}$ |
| | (0.1451) | (0.1057) |
| Constant | 0.3114$^{**}$ | 0.0341 |
| | (0.1024) | (0.0923) |
| Observations | 71799 | 71799 |

Standard errors in parentheses
Observations are clustered at the level of High School Commuting Zones. These estimates are specifically for the likelihood of passing all classes as an upperclassmen, and were estimated using the junior year of high school.
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

the half with the highest relevance are grouped together as the “STEM-intensive” industry, and the rest are considered “Non-STEM-intensive.”[59] The remaining industries will be considered non-STEM intensive. The resulting groupings are shown in table 33 of appendix C.3.

We likewise simplify the list of possible majors within our model, as maintaining a separate amount of labor for every major at the level of 2-digit CIP codes would be computationally burdensome. Drawing from our motivating evidence in section 4, many of the interesting dynamics we wish to capture are related to students’ choices either between STEM and non-STEM majors, or to their choices within STEM majors. Based on this, we elect to group all non-STEM majors into a single representative group for the model, and divide STEM majors into a smaller set of groups. In order to group majors which have similar implications for the labor market after college, we group them with a kmeans algorithm based upon the average relevance $\bar{r}_{mk}$ of major $m$ for industry $k$, where the average is taken across time. The four groups that we find are shown in table 34 of appendix C.3. Among these, we further restrict the STEM major groups to those in which at least one percent of students from a commuting zone major earn a degree in that area on average. This brings us to the STEM major groups “Computer and Biological Sciences" and "Engineering."

### 6.1.3 Terms for Calculating Effective Labor Amounts

The $\tau$ terms in our model determine how STEM and non-STEM human capital translate into wages for STEM and non-STEM workers. Specifically, the wages a worker of type $m$ receives in the labor market is $w_{imt} = l_{imt} w_{mt}$, where $l_{imt} = A_{is13}^{\tau_m^s} A_{in13}^{\tau_m^n} e_{it}^{\tau_m^e}$. This means that log wages are given by

$$\log(w_{imt}) = \tau_m^s \log(A_{is13}) + \tau_m^n \log(A_{in13}) + \tau_m^e \log(e_{it}) + \log(w_{mt}) \tag{23}$$

When estimating this equation in the Texas Data, there is more than one commuting zone in the dataset, and so we additionally include an $\ell$ location subscript on $\log(w_{mt})$. Then, we absorb the $\log(w_{m\ell t})$ terms with a worker type-by-location-by-time fixed effect $\nu_{m\ell t}$. These fixed effects will be saved and used in the section below.

For the ability terms, we can likely use almost the same method as was used in section 6.1.1, which was to replace the ability terms with estimates of ability based upon ability measures, and then to use other ability measures as instruments. First, we replace $\log(A_{is13})$ and $\log(A_{in13})$ with fitted values from (21) for what these abilities would be at the end of twelfth grade based on their course-taking. We complete these fitted values $\log\left(\hat{A}_{is13}\right)$ and $\log\left(\hat{A}_{in13}\right)$ using one set of TAAS standardized test scores taken at the beginning of eleventh grade. Then, equation (23) becomes

$$\log(w_{imt}) = \tau_m^s \log\left(\hat{A}_{is13}\right) + \tau_m^n \log\left(\hat{A}_{in13}\right) + \tau_m^e \log(e_{it}) + \log(w_{mt}) \tag{24}$$

[59] Another option would be to rank industries based on their the share of wages that are paid in each industry to workers with college STEM degrees.

or

$$\log\left(w_{imt}\right)=\tau_m^s\log\left(A_{is13}\right)+\tau_m^n\log\left(A_{in13}\right)+\tau_m^e\log\left(e_{it}\right)+\log\left(w_{mt}\right)-\tau_m^s\omega_{is}^f-\tau_m^n\omega_{in}^f \tag{25}$$

where $\omega_{ic}^f$ is meant to represent the measurement error in the fitted value $\log\left(\hat{A}_{ic13}\right)$.

To address this measurement error, we re-compute fitted values, which can be labeled them $\log\left(\tilde{A}_{is13}\right)$ and $\log\left(\tilde{A}_{in13}\right)$, using either SAT test scores or TAAS standardized test scores from grade 9.[60] This will allow us to instrument for $\log\left(\hat{A}_{is13}\right)$ and $\log\left(\hat{A}_{in13}\right)$ with $\log\left(\tilde{A}_{is13}\right)$ and $\log\left(\tilde{A}_{in13}\right)$, removing bias coming from measurement error. The estimates found by completing this estimation are shown in table 12.

### 6.1.4 Identifying the $\rho$ terms

Here, we use the saved these fixed effects from equation (23) above to identify the $\rho$ terms from equation (9) as follows:

$$\begin{aligned} v_{m\ell t+1} &= \rho_{m0}^w+\rho_{m1}^w SE_{\ell t}+\omega_{mt}^w \\ &= \delta_{m\ell}+\delta_{mt}+\rho_{m1}^w SE_{\ell t}+\omega_{mt}^w \end{aligned} \tag{26}$$

The final equation is one that is estimated, where $\delta_{m\ell}$ and $\delta_{mt}$ are fixed effects, and we instrument with $SE_{\ell t}$ using a shift-share instrument similar to the one described in section 3.2 for the regression results of section 4.[61] Including the fixed effects in this way helps to make sure that $\rho_{m1}^w$ is just being identified off of local variation in the data, but they are not necessary for the model. To map these estimates to the model, then, we set $\rho_{m0}^w=E\left[\delta_{m\ell}+\delta_{mt}\right]$.[62] The results of estimating equation (26) are shown in table 13.

Equation (8) is more simple to estimate in the data, because both $SE_{t+1}$ and $SE_t$ are directly observed. Here too, we use a shift share instrument to identify the coefficient on $SE_t$, and use the average of a location and a time fixed effect to pin down the constant $\rho_0^{SE}$. The results for the $\rho_1^{SE}$ term are shown in table 14.

[60] If both tests are available for a given student, we prefer the SAT test scores. Note that using the SAT scores to construct the instrument rather than the original $\log\left(\hat{A}_{ic13}\right)$ avoids a potential identification concern. Specifically, it could be worried that measurement error in the SAT score is part of what determines whether a student goes into college or certain majors within that college. This could occur, for example, because colleges use SAT scores as part of the application process. In that case, if the SAT scores were used to construct $\log\left(\hat{A}_{ic13}\right)$, there would be a concern of sample-selection bias generating a correlation between the measurement error in the SAT score and the underlying latent ability. By using SAT scores to construct the instrument instead, this concern about bias from SAT measurement error is removed.

[61] Specifically, the instrument that we use here is just like equation (6), but with $r_{mk0}$ replaced by an indicator that takes a value of 1 whenever industry $k$ is STEM-intensive, and 0 otherwise. The separation of industries into STEM and non-STEM intensive groups was described in section 6.1.2.

[62] We normalize so that the "no college degree" $m$ option satisfies $\rho_{m0}^w=0$.

Table 12: $\tau$ terms by College Degree

| | (1) Computer and Bio. Sciences STEM Degree | (2) Nat. Resources and Conservation STEM Degree | (3) Math, Physical Sci., and Sci. Tech. STEM Degree | (4) Engineering STEM Degree | (5) Non-STEM Degree | (6) No College Degree |
|---|---|---|---|---|---|---|
| Fitted STEM Ability after High School | 0.4372*** | 0.3452 | -0.2668* | 0.6942*** | 0.4415*** | 0.3568*** |
| | (0.0586) | (0.2547) | (0.1276) | (0.0682) | (0.0109) | (0.0050) |
| Fitted Non-STEM Ability after High School | -0.2993*** | -0.2376 | 0.1317 | -0.4346*** | -0.2899*** | -0.2382*** |
| | (0.0507) | (0.2284) | (0.1017) | (0.0568) | (0.0096) | (0.0042) |
| log experience | 1.4814*** | 2.3868* | 0.0365 | 1.4600*** | 1.1970*** | 0.4643*** |
| | (0.2710) | (1.0975) | (0.5076) | (0.2455) | (0.0889) | (0.0730) |
| Constant | 7.7320*** | 6.5913*** | 10.4900*** | 7.9710*** | 8.1905*** | 9.0435*** |
| | (0.3603) | (1.4388) | (0.6716) | (0.3288) | (0.1168) | (0.0951) |
| Observations | 26095 | 1921 | 8161 | 27428 | 319030 | 781321 |
| Number of Students | 4484.000 | 316.000 | 1334.000 | 4293.000 | 49532.000 | 128481.000 |
| F-Stat | 1604.856 | 51.021 | 455.527 | 1258.458 | 23052.566 | 76025.670 |

Standard errors in parentheses

All regressions include fixed effects for each combination of college degree type, high school commuting zone, and year.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 13: $\rho^{w}_{1m}$ terms by College Degree

| | (1) Computer and Bio. Sciences STEM Degree | (2) Nat. Resources and Conservation STEM Degree | (3) Math, Physical Sci., and Sci. Tech. STEM Degree | (4) Engineering STEM Degree | (5) Non-STEM Degree | (6) No College Degree |
|---|---|---|---|---|---|---|
| Share of Emp. in STEM, 2 Years Ago | 5.6988*** | 4.8858 | 4.5674 | 3.8258 | 2.2840*** | 1.2904*** |
| | (1.2916) | (3.6678) | (5.1879) | (2.5514) | (0.5929) | (0.3631) |
| Observations | 953 | 511 | 755 | 908 | 1103 | 1108 |
| Number of Students | 4484.000 | 316.000 | 1334.000 | 4293.000 | 49532.000 | 128481.000 |
| F-Stat | 118.333 | 46.078 | 16.496 | 25.790 | 89.883 | 109.274 |

Standard errors in parentheses

All regressions include separate fixed effects for the high school commuting zone and the year. 2016 is set as the base year for the year fixed effects.

Regressions are weighted by the number of student-and-year combinations that were used to estimate the dependent variable, which is a fixed effect from the corresponding regression shown in table 12.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 14: $\rho_1^{SE}$ estimation

| | (1)<br>SE share |
|---|---|
| Share of Emp. in STEM, 2 Years Ago | 0.8240*** |
| | (0.0710) |
| Observations | 1407 |
| F-Stat | 95.698 |
| St. Dev. of Residual | 0.014 |

Standard errors in parentheses
All regressions include separate fixed effects for the high school commuting zone and the year. 2016 is set as the base year for the year fixed effects.
Regressions are weighted by the number of students who graduated high school from a given commuting zone in 1998.
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

### 6.1.5 Other Parameters Identified Outside the Model

The share of classes that are "STEM" in the non-STEM track is set to be 23%. This is the approximate share of math and science credits required to graduate in Texas under both Foundation graduation standard with one endorsement and the prior MHSP standard.[63] Then, we will use the model to estimate the share of classes that are STEM in the STEM-intensive track. With the measurement terms for STEM and Non-STEM latent ability estimated from section 6.1.1 above, we are able to estimate the initial average latent STEM and Non-STEM ability in 11th grade, as well as the variances and covariance of those abilities. These are used for generating the simulated sample of students in the estimation below. The average share of classes that are "STEM" within 11th grade is also taken from the data.

We will use an estimate of the time preference parameter $\beta = .95$. The standard deviation of the $\omega_t^{SE}$ term is taken from the variance of the residual when estimating the $\rho^{SE}$ terms, and is shown in table 14.[64] For the $r_{mkt}$ terms within the model, we prepare a version of these terms within the data in which the industry $k$ is aggregated to the STEM and Non-STEM industry level, and then the average is taken across time. In this way, our model estimation abstracts away from changes in the relevance variable $r_{mkt}$ over time. These aggregated $r_{mkt}$ terms and the STEM employment share $SE_t$ together pin down the movement in relevant employment $RE_{mt}$ within the model.

## 6.2 Parameters Identified Inside the Model

The parameters left to be identified are the $\delta$ terms that shape a student's utility over time and the share of classes that are STEM in the STEM-intensive high school track. We estimate these using simulated method of moments, using moments like those described in section 4. Each period in the model represents two years, and so we only estimate the moments found by taking the two, four, six, eight and ten-year differences. For the high school moments, since both STEM and non-STEM classes are present in the model, we directly use results shown in tables 1 and 3.

For college moments like those presented in tables 4 through 6, we cannot match these in our model as directly because our model only has three degree category groupings, and those tables were estimated using the full set of degree categories at the two digit CIP code level. To resolve this, we aggregate to the college major groupings that we will be using within the model, and re-estimate the moments from college that we will be using. These estimates broadly show the same patterns as we presented in section 4, and are included for reference in appendix section C.4.

In addition to the high school and college moments described above, we also include as moments the shares of high school graduates who major in each of the three degree groupings. We include these shares taken both at the stage of declaring a first major and at the stage of graduating with final majors. The moments included in table 7 are not included as moments in the regression, because the influence of employment shares on wages through the lens of the model has already been pinned down by the $\rho_1^w$ estimates shown in table 13.

The central goal of the model is to understand how students react to the local labor market over time, and so the most important $\delta$ terms are those governing how students value $SE_t$ realizations while in school and wages once in the labor

[63] With one endorsement, the Foundation high school program requires 6 credits to be math or science out of 26 total credits (this is 23.1%). The MHSP requires 5 math or science credits out of a total 22 credits (this is 22.7%).

[64] Note that this regression included both a location fixed effect and a time fixed effect. Thus, Texas-wide changes would be absorbed in the time fixed effect, and so the residuals used here are more specific to commuting-zone-level variation across time.

market. In appendix section B.2, We argue that these terms are effectively pinned down by the regression moments described above .

To arrive at our weighting matrix for simulated method of moments, we first create a variance-covariance matrix for our moments by bootstrapping from the data.[65] This bootstrapping is done at the level of commuting zones, and we take 500 separate bootstrapped samples. The theoretically optimal weighting matrix is the inverse of the variance-covariance matrix, but we diverge from this matrix in two ways for our case: First, since the shares of students declaring certain majors as their first and final choices are unconditional averages, these moments are estimated far more precisely than the other moments that we are using. If we were to use the inverse variance-covariance matrix directly, then, a highly disproportionate weight would be placed on these moments, even though they are not as central to the patterns we are hoping to explain as the other moments. To fix this, before taking the matrix inverse we set the covariance between these shares and the other moments to 0, and raise the variances of these moments to a constant $1e{-4}$.[66]

Second, although we have 40 moments dedicated to the way college majors respond to employment shares and only 5 describing how high school class selections respond, we consider the response of high school course-taking to employment conditions equally if not more important to our research question. To address this imbalance in the moments, we multiply the rows and columns of the weighting matrix that correspond to the high school course-taking moments by 8.[67] In doing so, we multiply the differences between these simulated moments and their data counterparts by 8, and hopefully correct for the disproportionate number of moments included for each group.

The final estimation process is over-identified, with 61 moments and only 19 parameters to be estimated.

# 7 Structural Model Results

The parameter estimates and their standard errors are shown in table 15. The final estimation process was quite over-identified, with 61 moments and only 19 parameters to be estimated. Still, the model did a good job of matching these moments. For the "major share" moments, this is shown in table 16. The fit of the remaining moments is shown graphically in figures 4 and 5, where the confidence intervals are created based on the standard errors from the bootstrapped variance-covariance matrix. Tables with these simulated and data moments can be found in appendix section C.5.

For the high school class share moments presented in figure 4, the model does a good job of matching both the level and the trend. This is good, since the trend in coefficient estimates from table 1 was a key motivation for estimating a structural model. Within the model, this trend occurs for two reasons. First, students make their decisions between academic "tracks" for period two of the model while still in period one, before the STEM employment share in the second period has been realized. This means that the STEM employment share in period one has a non-contemporaneous effect on the course outcomes in period 2. Second, the STEM employment share in one period covaries positively with the STEM employment share in the next because they are both generated by the same AR-1 process. Equation (34) in appendix section B.2 shows how these two factors come together to produce the observed trend in coefficient estimates.

For the other moments, the model generally does a good job of matching the approximate levels of the coefficients, but replicates the trends in those coefficients with varying degrees of effectiveness. The STEM class passing moments in figure 4, for example, effectively match that the change in relevant employment has a negative coefficient, but the trend seems to be inverted. Figure 5 displays the moments related to college. One more noticeable failing of the model there is that it produces small positive coefficients in the first major regressions, but the data suggests that these should be negative. This will be a potential area for improvement in future drafts of this paper.

## 7.1 Counterfactual Estimation

With model estimates in hand, we turn to estimating a counterfactual. We are interested in testing how significant the responses that students make while in high school are for shaping their overall response to the labor market in college and beyond.Specifically, we will run a counterfactual in which we shut down students' abilities to choose a STEM-intensive track for their senior year STEM course share in high school. Instead, all students will be forced to take the non-STEM track. This will be informative about the extent to which the observed responsiveness of Texas students to the labor market is coming

[65]In section 12.24 of Hansen (2021), Bruce Hansen discourages against using bootstrapped standard error estimates for two-stage least squares. In the future, we can look into alternative methods of preparing this variance-covariance matrix.

[66]Before making this correction, the highest variance for these moments was $5.72e{-5}$ and the minimum was $3.69e{-7}$.

[67]This is done after the inverse of the variance-covariance matrix is taken, whereas the adjustment above to the shares was made beforehand. Since both the rows and columns are being multiplied by eight, the weights along the diagonal for the high school classes are effectively multiplied by 64.

Table 15: Parameter Estimates

**High School Parameters**

| Parameter | Estimate |
|---|---|
| $\delta^{hs}_{0,c=s}$ | -1.601 |
| | (0.144) |
| $\delta^{hs}_{1}$ | 2.184 |
| | (0.129) |
| $S_{c=s}$ | 0.900 |
| | (0.002) |

**Transferring Parameters**

| Parameter | Estimate |
|---|---|
| $\delta^{transfer}_{t=3,c=n}$ | 0.041 |
| | (0.042) |
| $\delta^{transfer}_{t=3,c=s}$ | 0.041 |
| | (0.733) |
| $\delta^{transfer}_{t=4,c=n}$ | 1.870 |
| | (1.292) |
| $\delta^{transfer}_{t=4,c=s}$ | 0.114 |
| | (4.014) |

**College Parameters**

| Parameter | Estimate |
|---|---|
| $\delta^{col}_{0,m=1}$ | -6.182 |
| | (0.418) |
| $\delta^{col}_{0,m=2}$ | -17.896 |
| | (0.896) |
| $\delta^{col}_{0,m=3}$ | -17.887 |
| | (2.716) |
| $\delta^{col}_{1,c=n}$ | 2.512 |
| | (0.245) |
| $\delta^{col}_{1,c=s}$ | 4.045 |
| | (0.308) |
| $\delta^{col}_{2,c=n}$ | 0.092 |
| | (2.577) |
| $\delta^{col}_{2,c=s}$ | 1.389 |
| | (3.888) |

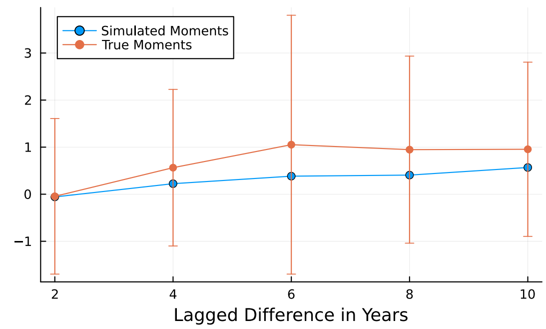
**Labor Market Parameters**

| Parameter | Estimate |
|---|---|
| $\delta^{lab}_{0,m=1}$ | -0.044 |
| | (0.143) |
| $\delta^{lab}_{0,m=2}$ | -0.038 |
| | (0.187) |
| $\delta^{lab}_{0,m=3}$ | -0.093 |
| | (0.592) |
| $\delta^{lab}_{1}$ | 0.097 |
| | (0.183) |
| $\delta^{dropout}$ | 0.247 |
| | (2.606) |

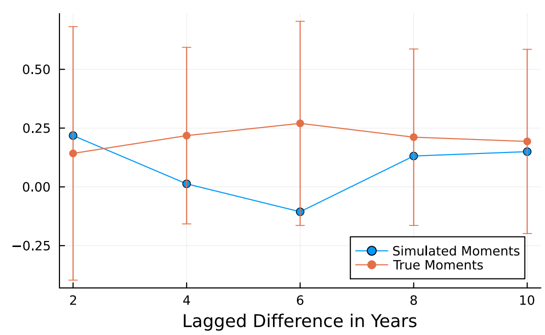
Table 16: Fit of "Major Share" Moments

| Major | % First Major | | % Graduation Major | |
|---|---|---|---|---|
| | Model | Data | Model | Data |
| Non-STEM Degree | 30.583 | 30.494 | 21.849 | 21.785 |
| STEM, Computer and Bio Sciences | 2.541 | 2.653 | 0.961 | 1.731 |
| STEM, Engineering | 2.558 | 2.225 | 0.961 | 1.689 |

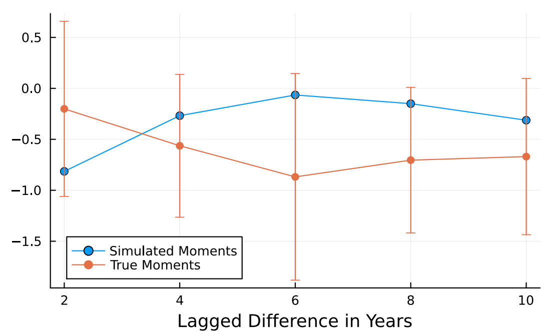
Figure 4: Fit of High School Class Moments




(a) HS Class Shares

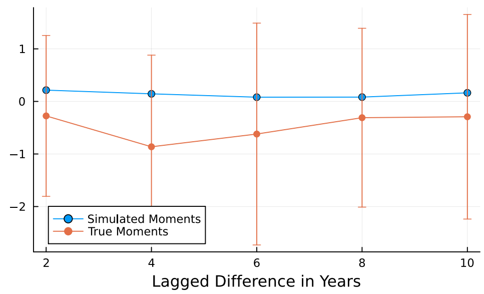



(b) Non-STEM HS Passing Rates

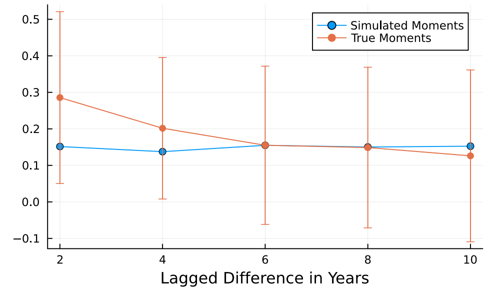



(c) STEM HS Passing Rates

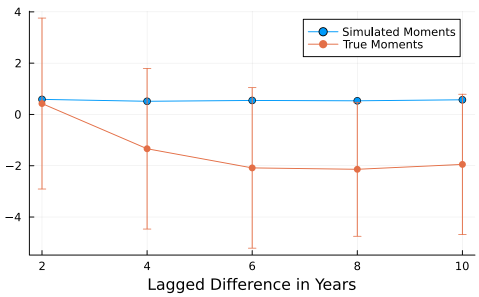
Figure 5: Fit Of College Moments



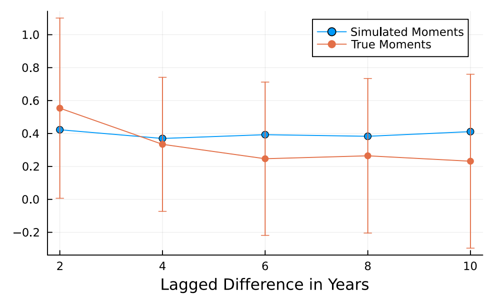

(a) Non-STEM First Major Share for All HS Graduates



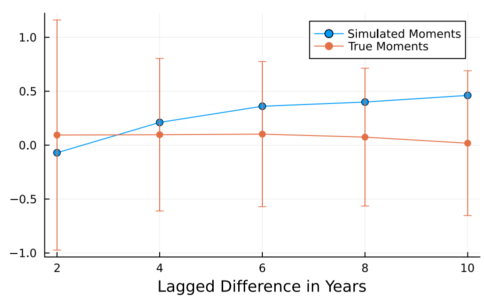

(b) STEM First Major Share for All HS Graduates



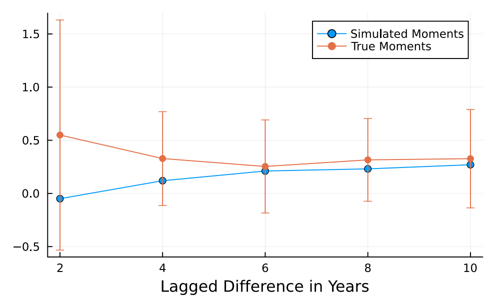

(c) Non-STEM First Major Share for College-Goers



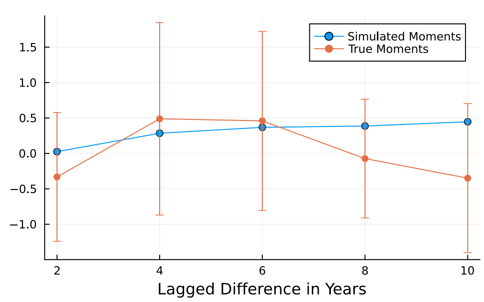

(d) STEM First Major Share for College-Goers



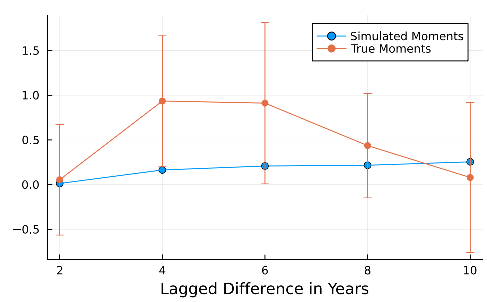

(e) Non-STEM Graduation Share, Effect of Employment Share at time of HS Graduation




(f) STEM Graduation Share, Effect of Employment Share at time of HS Graduation



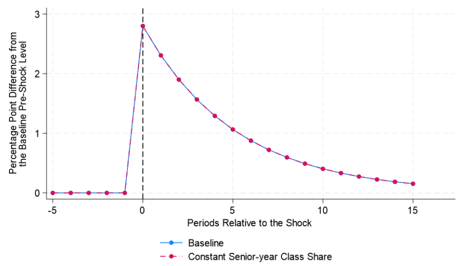

(g) Non-STEM Graduation Share, Effect of Employment Share at time of College Graduation



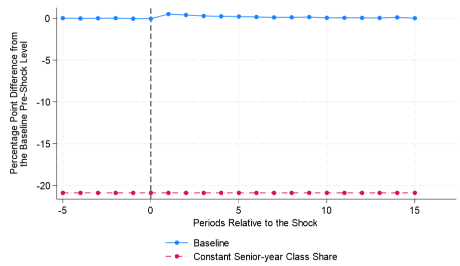

(h) STEM Graduation Share, Effect of Employment Share at time of College Graduation from the flexibility that they have to adjust their course-taking behavior while in their senior year of high school. If it is found that this flexibility has substantial positive effects, that will underline its importance in the Texas education system, and encourage greater flexibility to be adopted in any educational systems that currently have less flexibility than Texas.

With the counterfactual limitation to senior-year course taking in place, we then simulate a large shock to the local STEM employment share. Specifically, we will hold the STEM employment value at the steady state value, then apply a two-standard-deviation shock. After the shock, we allow STEM employment share to slowly adjust back to the steady state level according to to the AR-1 process in equation (8).[68] Panel (a) of figure 6 shows this shock to the STEM employment share. As can be seen, the shock is very persistent—the STEM employment share remains over a percentage point higher five periods after the shock, which represents 10 years of real time.

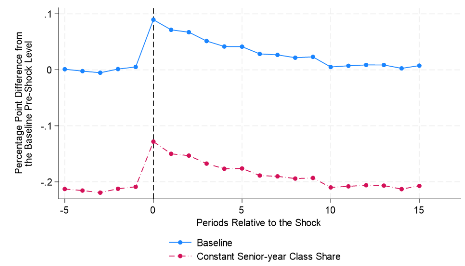
Figure 6: A Shock to the STEM Employment Share, Removing the Option of a STEM-Intensive Track in the Senior Year



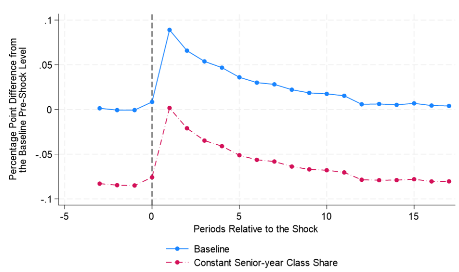
(a) STEM employment share



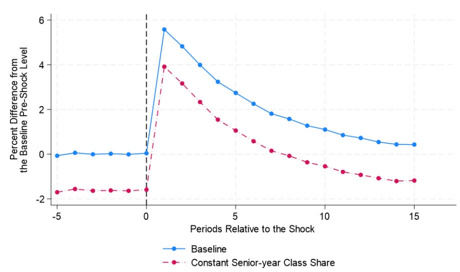
(b) Senior Year STEM Class Share



(c) Share of High School Graduates who Declare an Initial Major in STEM



(d) Share of High School Graduates who Graduate College in STEM



(e) Average Period 5 Earnings

(f) Average Present Value of Lifetime Earnings

Panel (b) of figure 6 shows what has been enforced for this counterfactual, that the average share of senior year classes taken in STEM is lowered by about 20 percentage points prior to the shock, and is unable to react even once the shock has taken place. The next two panels are more interesting because they show the result of that change on student's future outcomes. Panel (c) shows that the share of students declaring a first major in STEM drops by about .21 percentage points prior to the shock, which is equivalent to a 4.20 percent reduction in the number of students declaring first majors in STEM. This loss in initial STEM majors translates into a loss in STEM college graduates, as shown in panel (d). There, share of all high school graduates who major in STEM declines by about .08 percentage points prior to the shock, signaling a 4.42 percent decline in the number of STEM graduates.

Panel (e) plots the response of student's earnings in period 5, which is the first period of the model in which all students have entered the labor force. The observed loss in these "initial" labor market earnings, down about 1.63 percent prior to the shock, reflects two channels in the model: students changing their final degrees, and students having reduced STEM human capital from missing out on STEM classes in their senior year. Panel (f) presents the corresponding changes in the present value of lifetime earnings. Each point in that graph represents the present value of lifetime earnings for a student who enters the model as a freshman in that period. As such, lifetime wages are elevated prior to the shock because the students born in those time periods will live through the STEM employment share shock. In the counterfactual, this life time present value of wages drops by just over two percent prior to the shock.

Note that these graphs are describing movements in average wages across all students, the majority of whom did not take the STEM-intensive senior course load even in the baseline case. Their wages remained fixed in the counterfactual, so

[68] More specifically, we are apply a shock equal to two standard deviations of $\omega_t^{SE}$. After the shock, we hold future $\omega_t^{SE}$ values at 0 to allow the STEM employment share to adjust back down to the steady state level.

the observed differences are being driven by the students who took the STEM intensive course share. Prior to the shock, the share of students taking this STEM-intensive course load in their senior year was 31.19%. If we condition the earnings results on being a student who would have taken the STEM-intensive course load in their senior year, the differences are more dramatic. This is shown in figure 11 of appendix section C.6. Prior to the shock, we find that students who would have taken the STEM-intensive senior year course share would lose about 4.96 percent of their period 5 income, and that the present value of the lifetime income is reduced by around 6.72 percent. That figure also shows that these results are not being driven by students of high or low ability—both the highest and lowest terciles of initial STEM ability have remarkably similar losses in the counterfactual.[69]

All of the disparities in student responses that figure 6 shows prior to the shock persist afterword as well. However, there are subtle differences between the baseline and the counterfactual in the degree to which students respond to the employment share shock. To make this more clear, figure 7 plots the difference between the "baseline" and "counterfactual" lines from figure 6. The drop following the shock in each panel reflects the decreased responsiveness of these variables to the employment share shock when high school seniors cannot choose a STEM-intensive course load.

Figure 7: Change in Student Responsiveness when the Option of the STEM-Intensive Track is Removed.


(a) Share of High School Graduates who Declare an Initial Major in STEM


(b) Share of High School Graduates who Graduate College in STEM


(c) Average Period 5 Earnings


(d) Average Present Value of Lifetime Earnings

For the share of high school graduates declaring an initial major in STEM, the responsiveness declines by up to around .008 percentage points. In other words, during employment shocks, the decline in student responsiveness when they cannot alter their senior class shares can further reduce initial STEM major shares by up to another 3.5 percent of the pre-shock decline. The effect is similar for graduation majors. There, the response in the share of students graduating in STEM drops by up to around .004 percentage points, or 5.1 percent of the pre-shock decline. For income, the response of period 5 earnings drops by up to around .06 percent, which is about $1/26$ of the original pre-shock effect. Finally, the response of the present value of lifetime earnings drops by .05 percent, or about $1/47$ of the original effect. In all of these ways, removing student's ability to choose STEM-Intensive coursework reduces the degree to which they are able to adjust to a shock in the labor market. These reductions in responsiveness are small in absolute terms, but non-negligible when compared to the size of the

[69] High and low ability students show fairly similar responses regardless of whether we condition on them having taken STEM-intensive track in their senior year. This is shown in figure 12 of appendix C.6.

pre-shock effects that were measured.

With all of the results from this counterfactual, it is important to remember that the only thing being changed is a student's course load during a single year of high school, and that effectively a change is only being made for the students that would have taken the STEM-intensive course load if they had the option—just 31.19 percent of students prior to the shock. Even among those students, the counterfactual did not entail a total removal of STEM classes, but rather a reduction down to 23 percent of their classes being in STEM. As shown above, this relatively small change has lasting implications for the share of students pursuing STEM degrees, labor market earnings, and the responsiveness of these two things to STEM employment shocks.

# 8 Conclusion

In this paper, we looked at the effects of local labor markets on the academic subjects students pursue in school. We found that a student's high school labor market has a large influence on their course taking while in high school, as well as likelihood that students will pass their STEM classes. Importantly, we found evidence from these regressions that lagged relevant employment shares can affect a student's present decisions.

This motivated a structural model in which students were able to react to STEM employment shares each period, and the decisions that they made in one period affected their decisions in the next. Our results from that model underlined the importance of allowing students flexibility in choosing their senior-year coursework. Specifically, we find that removing the option of a STEM-intensive course load lowers the number of students who graduate college with STEM degrees by about 4.4 percent, and decreases a measure of early career earnings by about 1.6 percent. As would be expected, the declines in earnings are much more pronounced when we look only at students who would have pursued the STEM-Intensive senior-year track if they had the chance. Further, the growth of student STEM degree shares and eventual wages after a STEM employment share shock is diminished in the counterfactual, suggesting that the ability to choose senior year course loads has important effects on the overall responsiveness of students to the labor market.

From a policy perspective, our work underlines the importance of providing students with meaningful choices between career-relevant courses while in high school. Doing so can help them to better respond to changes in their local labor markets, increasing their abilities to be successful in college and earn higher wages upon entry into the labor force. In light of these results, it is interesting that we found no evidence of school systems changing their course offerings in response to local economic conditions. If they were to begin doing so, these efforts may compliment the existing responses of students, so that more STEM courses are available to students precisely when local labor market conditions make taking those courses more attractive. We view this as a promising avenue for future policy work and research.

# References


Table 203.20. enrollment in public elementary and secondary schools, by region, state, and jurisdiction: Selected years, fall 1990 through fall 2031, 2023. URL https://nces.ed.gov/programs/digest/d23/tables/dt23_203.20.asp.

Francesco Agostinelli and Matthew Wiswall. Estimating the technology of children's skill formation. Working Paper, mar 2022.

Tom Ahn, Peter Arcidiacono, Amy Hopson, and James Thomas. Equilibrium grade inflation with implications for female interest in stem majors. Working Paper, feb 2022.

Joseph G. Altonji. The effects of high school curriculum on education and labor market outcomes. *The Journal of Human Resources*, 30(3):409–438, 1995.

Joseph G. Altonji, Erica Blom, and Costas Meghir. Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics*, 4:185–223, sep 2012.

Peter Arcidiacono, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom. College attrition and the dynamics of information revelation. Working Paper, apr 2023.

Adrien Auclert, Matthew Rognlie, and Ludwig Straub. Micro jumps, macro humps: Monetary policy and business cycles in an estimated hank model. Working Paper, mar 2020.

David H. Autor, David Dorn, and Gordon H. Hanson. The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, 103(6):2121–2168, 2013.

Ashok Bardhan, Daniel L. Hicks, and Dwight Jaffee. How responsive is higher education? the linkages between higher education and the labour market. *Applied Economics*, 45:1239–1256, 2013.

TEXAS HIGHER EDUCATION COORDINATING BOARD. 60x30tx progress report. Technical report, TEXAS HIGHER EDUCATION COORDINATING BOARD, jul 2022.

Elizabeth S. Bradley. The effect of the business cycle on freshman major choice. Technical Report 42412, MPRA, nov 2012.

Moshe Buchinsky and Phillip Leslie. Educational attainment and the changing u.s. wage structure: Dynamic implications on young individuals' choices. *Journal of Labor Economics*, 28(3):541–594, 2010.

A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427, aug 2008.

Raj Chetty, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. Income segregation and intergenerational mobility across colleges in the united states. *Quarterly Journal of Economics*, 135(3):1567–1633, feb 2020.

Johnathan G. Conzelmann, Steven W. Hemelt, Brad Hershbein, Shawn M. Martin, Andrew Simon, and Kevin M. Stange. Skills, majors, and jobs: Does higher education respond? Working Paper 31572, National Bureau of Economic Research, aug 2023.

David J. Deming and Kadeem Noray. Earnings dynamics, changing job skills, and stem careers. *The Quarterly Journal of Economics*, 135(4):1965–2005, nov 2020.

James A. Freeman and Barry T. Hirsch. College majors and the knowledge content of jobs. *Economic of Education Review*, 27:517–535, 2008.

Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift. Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624, 2020.

Michel Grosz. Do postsecondary training programs respond to changes in the labor market? *Journal of Human Capital*, 16 (4):461–487, 2022.

Luyi Han and John V Winters. Industry fluctuations and college major choices: Evidence from an energy boom and bust. *Economics of Education Review*, 77:101996, 2020.

Bruce E. Hansen. Econometrics. Textbook, mar 2021.

James J. Heckman, Lance Lochner, and Christopher Taber. Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics*, 1:1–58, 1998.

Steven W. Hemelt, Brad Hershbein, Shawn M. Martin, and Kevin M. Stange. College majors and skills: Evidence from the universe of online job ads. Technical Report 29605, National Bureau of Economic Research, dec 2021.

Daniel Hubbard. The impact of local labor market shocks on college choice: Evidence from plant closings in michigan. Working Paper, sep 2018.

Daniel Kreisman and Kevin Stange. Vocational and career tech education in american high schools: The value of depth over. Working Paper 23851, NBER, sep 2017.

Per Krusell and Anthony A. Smith. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896, 1998.

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

James G. MacKinnon and Matthew D. Webb. Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2):233–254, mar 2017.

Jonathan Meer. Evidence on the returns to secondary vocational education. *Economics of Education Review*, 26:559–573, 2007.

NBER. Census core-based statistical area (cbsa) to federal information processing series (fips) county crosswalk, 2022. URL https://www.nber.org/research/data/census-core-based-statistical-area-cbsa-federal-information-processing-series-fips-county-crosswalk.

NCES/BLS. Guidelines for using the cip-soc crosswalk. Technical report, National Center for Education Statistics and Bureau of Labor Statistics, mar 2011.

Heather Rose and Julian R. Betts. The effect of high school courses on earnings. *The Review of Economics and Statistics*, 86(2):497–513, may 2004.

Jaewoo Ryoo and Sherwin Rosen. The engineering labor market. *Journal of Political Economy*, 112(S1), feb 2004.

Ralph Stinebrickner and Todd R. Stinebrickner. A major in science? initial beliefs and final outcomes for college major and dropout. *The Review of Economic Studies*, 81(1 (286)):426–472, 2014.

TEXAS HIGHER EDUCATION COORDINATING BOARD. Texas higher education strategic plan: 2015–2030. Technical report, TEXAS HIGHER EDUCATION COORDINATING BOARD, jul 2015.

Russell Weinstein. Local labor markets and human capital investments. *Journal of Human Resources*, 57(5):1498–1525, 2022.

# Appendix

## A Preparing the Data

Each college in the TSP data is identified using a FICE code. To link these colleges to specific commuting zones, we first used the Master College Code list from WebAdMIT by Liaison to match each FICE code to an IPEDs Code. Then, we used IPEDS directory data to get the county FIPS code for each college.[70] Some remaining FIPS County codes were also filled in manually by looking at the addresses of the colleges.[71] To link FIPS county codes to specific commuting zones and MSAs, we used a crosswalk provided by the NBER (NBER, 2022).

When grouping majors at the two-digit CIP code level, we combine certain two-digit CIP categories that are highly similar. Specifically, we combine “History” (54) and “Social Sciences” (45) under the header “Social Sciences.” Additionally, we combine “Theology and Religious Vocations” (39) and “Philosophy and Religious Studies” (38) under the header “Philosophy and Religious Studies.” There are other two digit CIP codes that we exclude from the analysis because they are not vocational or occupational in the private sector: “Undecided” (99), “Military Science, Leadership and Operational Art” (28), “Military Technologies and Applied Sciences” (29), “Basic Skills and Developmental/Remedial Education” (32), “Citizenship Activities” (33), “Health-Related Knowledge and Skills” (34), “Interpersonal and Social Skills” (35), “Leisure and Recreational Activities” (36), and “Personal Awareness and Self-Improvement” (37).

We also take additional steps to prepare the CBP data. First, we exclude the two-digit NAICS codes “00”, “95”, and “99.” Then, we combine certain two digit CIP codes into sectors as the Census does: Codes 31-33 are combined as “Manufacturing,” 44-45 are combined as “Retail Trade,” and 48-49 are combined as “Transportation and Warehousing.”

The CBP suppresses employment counts or annual payrolls for certain industry, county, and year combinations in order to protect respondent confidentiality. Up until 2017, however, the approximate value of employment counts can be deduced from other variables.[72] One option for how to do this uses the EMPFLAG variable, which gives the range that total employment falls into. An EMPFLAG value of “B,” for example, tells us that total employment is somewhere from the lower bound of 20 to the upper bound of 99. A second option is to use the number of establishments that fall within certain employment ranges. If there is one establishment in the 10-19 employee range (variable N10_19) and two in the 20-49 employee range (variable N20_49), then we can construct a lower bound for total employment of 1*10+2*20=50 and an upper bound for employment of 1*19+2*49=117. To estimate the approximate value of total employment, we begin by finding the maximum lower bound and minimum upper bound from these two methods. Then, we take the average of those bounds and use it as the approximate employment level.

Unfortunately, these methods are not available for finding approximate annual payroll levels. For our measure of average relevant wages $RW_{m\ell t}$, then, we must exclude these suppressed industry, county, and year combinations. More details on this are provided in section A.5 below.

### A.1 Additional Motivational Figures

In section 2, we showed the relationship between movement in a STEM Industry (Health Care/Social Assistance) and the share of people declaring their first major in STEM for different commuting zones in Texas. Here, we will provide a similar graph showing this relationship for Non-STEM degrees using “Retail Trade” as a Non-STEM industry. Health Care/Social Assistance and Retail Trade were the two-digit NAICS codes with the highest employment shares in Texas as of 2005. Figure 8 shows the relationship between Retail Trade employment shares in Dallas and non-STEM first major shares, and figure 9 shows this relationship for a number of other commuting zones in Texas.

As can be seen, while there is some suggestive evidence of a relationship here, it is certainly not as strong as the relationships for STEM shown in figures 1 and 2. This same pattern will be present in our other main results from section 4

[70] The IPEDS directory data started including counties in 2009, but before that they recorded zip codes. If a school had the same zip code before 2009 as they had in 2009, we assume that they did not move, and that their county matches the county fips code given in 2009. For schools that were still missing FIPS county codes before 2009, we used a crosswalk file between zip codes and county FIPS codes from the Housing and Urban Development (HUD) and the United States Postal Service (USPS).

[71] One college that we did not match to a county was Texas State Technical College-West Texas, because they have campuses in Nolan County, Taylor Counter, Stephens County, and Brown County. This means that we cannot identify the FIPS county code for students going to that college, and so we exclude them from regressions that use the college location.

[72] Starting in 2017, a industry, county, and year combination is entirely removed from the publication if it contains two or fewer establishments. For these years we cannot differentiate these suppressed cases from industry, county, and year combinations where there is no employment. The EMPFLAG variable was removed in 2018.

Figure 8: Retail Trade Employment Shares and non-STEM College Major Shares in Dallas

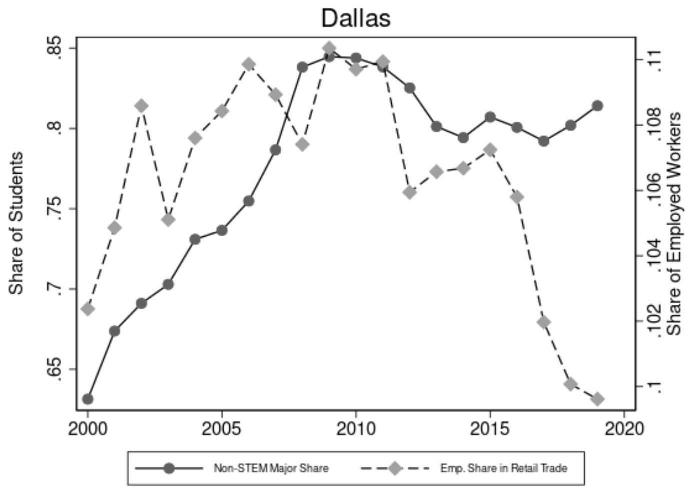
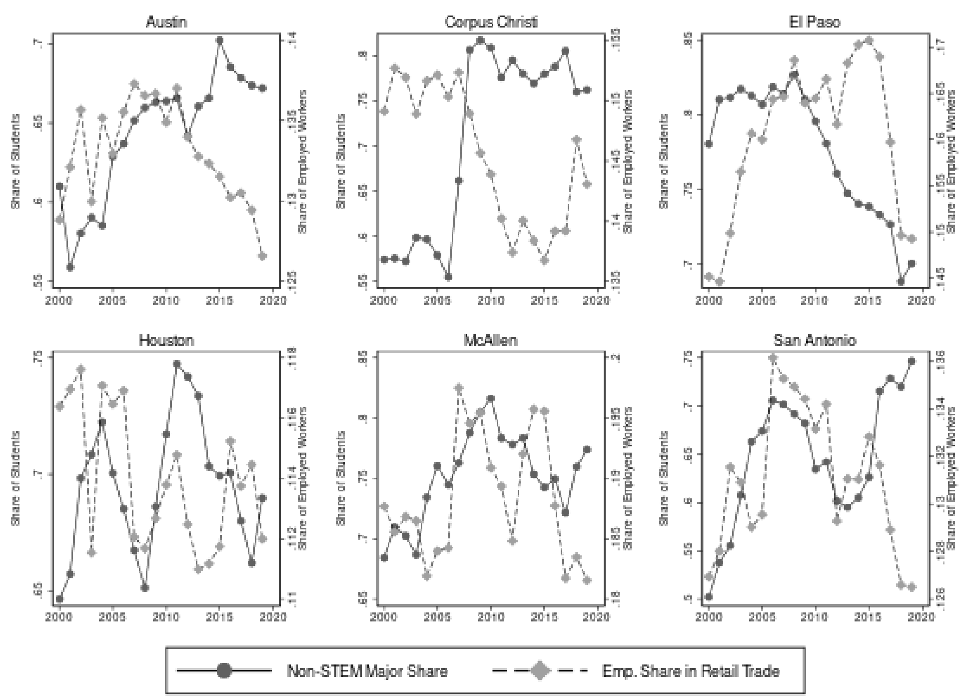



Figure 9: Retail Trade and non-STEM Major Shares Across Texas



as well—the response of STEM majors to employment conditions will consistently be more significant than the response of non-STEM majors.

### A.2 Descriptive Statistics

Table 17 displays the mean values of various characteristics for the students in our sample. For example, among students who enrolled in college, 46.3% of them were male. Here, "no college" means that the student did not enroll in any public university or community college in Texas, and "Enrolled in college" means that they did enroll in one of those options. The standard errors are given in parenthesis.

Table 17: Descriptive Statistics

| | Total | No College | Enrolled in College |
|---|---|---|---|
| Male | 0.494 | 0.543 | 0.463 |
| | (0.500) | (0.498) | (0.499) |
| Native American | 0.00324 | 0.00356 | 0.00303 |
| | (0.0568) | (0.0596) | (0.0550) |
| Asian | 0.0389 | 0.0312 | 0.0438 |
| | (0.193) | (0.174) | (0.205) |
| Black | 0.128 | 0.134 | 0.124 |
| | (0.334) | (0.340) | (0.330) |
| Hispanic | 0.416 | 0.474 | 0.379 |
| | (0.493) | (0.499) | (0.485) |
| White | 0.408 | 0.352 | 0.443 |
| | (0.491) | (0.478) | (0.497) |
| Two or More Races | 0.00599 | 0.00626 | 0.00582 |
| | (0.0771) | (0.0789) | (0.0760) |
| Bilingual | 0.000670 | 0.000559 | 0.000740 |
| | (0.0259) | (0.0236) | (0.0272) |
| Economic Disadvantage | 0.437 | 0.534 | 0.375 |
| | (0.496) | (0.499) | (0.484) |
| ESL | 0.0480 | 0.0865 | 0.0238 |
| | (0.214) | (0.281) | (0.152) |
| LEP | 0.0706 | 0.121 | 0.0392 |
| | (0.256) | (0.326) | (0.194) |
| At-Risk | 0.362 | 0.479 | 0.289 |
| | (0.481) | (0.500) | (0.453) |
| Gifted | 0.126 | 0.0904 | 0.149 |
| | (0.332) | (0.287) | (0.356) |
| First Major is in STEM | 0.0671 | 0 | 0.109 |
| | (0.250) | (0) | (0.312) |
| Observations | 4,744,298 | 1,830,741 | 2,913,557 |

Table 17 shows several of the patterns that might be expected for college attendance. Around half of the students who do not attend one of our college options were "at-risk" of dropping out of high school, whereas only 28.9% of students who enrolled in college were "at-risk." Having an economic disadvantage, English as a second language (ESL), or limited English proficiency (LEP) are associated with similar gaps.

### A.3 Measuring Relevant Employment

$RE_{m\ell t}$, defined in (1), is our measure of the share of employment in location $\ell$ and time $t$ that is relevant for major $m$. To construct an instrument for this variable, we make use of the fact that the share of local employment that is in industry $k$, $s_{k\ell t}$ can be expressed in terms of the share of employment in that industry in the base year $s_{k\ell 0}$ and the growth rates of employment in each industry $g_{k\ell t}$. Specifically,

$$
\begin{aligned}
s_{k\ell t} &= \frac{(\text{\# workers employed in } k)_{\ell t}}{(\text{\# workers employed})_{\ell t}} \\
&= \frac{(\text{\# workers employed in } k)_{\ell 0}}{(\text{\# workers employed})_{\ell 0}} \frac{(\text{\# workers employed in } k)_{\ell t}}{(\text{\# workers employed})_{\ell t}} \frac{(\text{\# workers employed})_{\ell 0}}{(\text{\# workers employed in } k)_{\ell 0}} \\
&= \frac{s_{k\ell 0} g_{k\ell t}}{\frac{(\text{\# workers employed})_{\ell t}}{(\text{\# workers employed})_{\ell 0}}} \\
&= \frac{s_{k\ell 0} g_{k\ell t}}{\frac{(\text{\# workers employed})_{\ell t}}{(\text{\# workers employed})_{\ell 0}}} \\
&= \frac{s_{k\ell 0} g_{k\ell t}}{\sum\limits_{k'} \frac{(\text{\# workers employed})_{k'\ell t}}{(\text{\# workers employed})_{\ell 0}}} \\
&= \frac{s_{k\ell 0} g_{k\ell t}}{\sum\limits_{k'} s_{k'\ell 0} g_{k'\ell t}}
\end{aligned}
$$

This is how we reached equation (2).

Literally, this measure is the weighted average share of jobs for new graduates going to those with major $m$, where the weights are the size of each industry $k$. When all local employment is in industries where 100% of the jobs go to graduates with major $m$, $RE_{m\ell t}$ is equal to 1. To give a slightly more complicated example, if half of local employment is in an industry where 50% of the jobs go to graduates with major $m$, and the other industries do not hire any new graduates, then $RE_{m\ell t} = (.5)(.5) = .25$.

If the number of jobs going to new graduates in each industry is proportional to the size of that industry, then $RE_{m\ell t}$ is exactly equal to the share of jobs for new graduates that are going to students of major $m$. To see this, say that the fraction of workers that are new graduates is $\alpha$ in all industries. Then,

$$
\begin{aligned}
RE_{m\ell t} &= \sum_{k=1}^{K} r_{mkt} s_{k\ell t} \\
&= \sum_{k=1}^{K} \left( \frac{(\text{\# new graduates in major } m \text{ and industry } k)_t}{(\text{\# new graduates in industry } k)_t} \right) \left( \frac{(\text{\# workers employed in } k)_{\ell t}}{(\text{\# workers employed})_{\ell t}} \right) \\
&= \sum_{k=1}^{K} \left( \frac{(\text{\# new graduates in major } m \text{ and industry } k)_t}{(\text{\# new graduates in industry } k)_t} \right) \left( \frac{\alpha\, (\text{\# workers employed in } k)_{\ell t}}{\sum_{k'=1}^{K} \alpha \left(\text{\# workers employed in } k'\right)_{\ell t}} \right) \\
&= \sum_{k=1}^{K} \left( \frac{(\text{\# new graduates in major } m \text{ and industry } k)_t}{(\text{\# new graduates in industry } k)_t} \right) \left( \frac{(\text{\# new graduates employed in } k)_{\ell t}}{\sum_{k'=1}^{K} \left(\text{\# new graduates employed in } k'\right)_{\ell t}} \right).
\end{aligned}
$$

Next, using the assumption that the share of new graduates in industry $k$ that have major $m$ does not vary across locations, we can say that

$$
\begin{aligned}
RE_{m\ell t} &= \sum_{k=1}^{K} \left( \frac{(\text{\# new graduates in major } m \text{ and industry } k)_{\ell t}}{(\text{\# new graduates in industry } k)_{\ell t}} \right) \left( \frac{(\text{\# new graduates employed in } k)_{\ell t}}{\sum_{k'=1}^{K} \left(\text{\# new graduates employed in } k'\right)_{\ell t}} \right) \\
&= \sum_{k=1}^{K} \left( \frac{(\text{\# new graduates in major } m \text{ and industry } k)_{\ell t}}{\sum_{k'=1}^{K} \left(\text{\# new graduates employed in } k'\right)_{\ell t}} \right) \\
&= \sum_{k=1}^{K} \left( \frac{(\text{\# new graduates in major } m \text{ and industry } k)_{\ell t}}{(\text{\# new graduates that are employed})_{\ell t}} \right) \\
&= \frac{(\text{\# new graduates in major } m \text{ that are employed})_{\ell t}}{(\text{\# new graduates that are employed})_{\ell t}}.
\end{aligned}
$$

This is the share of jobs for new graduates that are going to students of major $m$.

### A.4 Other Measures for Relevant Employment

To our knowledge, we are the first paper to measure the relevance of industries to certain degrees using data that is specific to new college graduates. As described in the main text above, this is important because the types of jobs people with a given major take may change as they gain experience in the labor force.

A second contribution is that we allow this relevance to change over time. The only other paper we know of that does this is Conzelmann et al. (2023). Their measure of labor demand is the number of new online job ads that are relevant for a given major and posted in the geographic areas where alumni of a student's college live. Since only about half of job ads list a major, the authors are required to impute the majors that are likely relevant for job ads based on information from the ads which do list a major. It is a strong assumption that job ads which do not list majors are actually associated with certain majors in the same way as the jobs which do list majors. As the authors suggest, it could be possible that the absence of listed majors indicates that any college major could be potentially suitable. In support of concerns like these, four of the authors find that the job ads which offer majors are different than those which do not along observable characteristics, and note that the observable characteristics can only explain three quarters of the variation in which job ads list a major (Hemelt et al., 2021).

In our paper, we use the more direct measure of the share of new graduates in an industry that have each degree. Even if the job ad data could perfectly record the majors that an industry desires to hire for a certain job, looking directly at who is hired has the advantage of reflecting true matches. For example, even if a particular business listed "English" as a possible major, if they only higher finance majors than the true "relevance" of that job for English majors would be different than the job ad suggested.

Other papers have used different methods to identify the jobs in the labor market that are relevant for specific degrees. These methods generally require that this relevance is constant over time and the number of years a worker has spent in the labor force. One common method is to use a crosswalk between CIP college major codes and occupation codes that was prepared by the National Center for Education Statistics and the Bureau of Labor Statistics (NCES/BLS, 2011).[73] Their crosswalk links a single CIP code to multiple possible occupations, and so this also requires some way of deciding how important each of those possible occupations should be for a specific major.

Our method of finding the "relevance" of an industry to a degree looks at the share of new graduates in an industry that have each degree. Long et al. (2015) have a similar approach, which is use the ACS to find share of people with a degree that work in a particular occupation. Since the ACS only recorded majors for a limited time frame, they are required to assume that the mapping between occupations and degrees is constant over time. Freeman and Hirsch (2008) used a different method, which was to link the "knowledge areas" of occupations described in O*Net to specific degrees. Then, they linked O*Net to specific occupations, and asked if people major in a degree more often when the occupations that require the related knowledge areas account for a large share of employment.

### A.5 Measuring Relevant Wages

In this section, we describe how we measure relevant wages $RW_{m\ell t}$. Our goal is to calculate the expected average industry wage for a worker with major $m$, conditional on that worker being employed in one of the two digit NAICS code industries $k$ in our sample. This can be calculated as the sum across industries $k$ of the average wage in industry $k$ times the probability that an employed worker with major $m$ would be working in industry k:

$$RW_{m\ell t} = \sum_{k=1}^{K} P\left\{\text{Worker gets a job in k} \mid \text{Worker Gets a Job \& has Major } m\right\}_{m\ell t} \bar{w}_{k\ell t}$$

Here, $\bar{w}_{k\ell t}$ is the average wage in industry $k$, location $\ell$, and year $t$. We calculate it by dividing the annual payroll from this industry by the number of workers employed in that industry in the CBP data. The CBP does sometimes suppress wage information for a given industry, location, and year combination. We will deal with this below.

[73] This method was used by Grosz (2022) and Bardhan et al. (2013), and also by Long et al. (2015) in a robustness check.

In order to measure $RW_{m\ell t}$, we perform the following steps:

$$\begin{aligned}
RW_{m\ell t} &= \sum_{k=1}^{K} \left( \frac{(\text{\# with major } m \text{ who get a job in } k)_{m\ell t}}{(\text{\# with major } m \text{ who get a job})_{m\ell t}} \right) \bar{w}_{k\ell t} \\
&= \sum_{k=1}^{K} \left( \frac{\frac{(\text{\# with major } m \text{ who get a job in } k)_{m\ell t}}{(\text{total \# of jobs})_{\ell t}}}{\frac{(\text{\# with major } m \text{ who get a job})_{m\ell t}}{(\text{total \# of jobs})_{\ell t}}} \right) \bar{w}_{k\ell t} \\
&= \sum_{k=1}^{K} \left( \frac{\frac{(\text{\# with major } m \text{ who get a job in } k)_{m\ell t}}{(\text{total \# of jobs})_{\ell t}}}{\sum\limits_{k'=1}^{K} \frac{\left(\text{\# with major } m \text{ who get a job in } k'\right)_{m\ell t}}{(\text{total \# of jobs})_{\ell t}}} \right) \bar{w}_{k\ell t} \\
&= \sum_{k=1}^{K} \left( \frac{\frac{(\text{\# with major } m \text{ who get a job in } k)_{m\ell t}}{(\text{total \# of jobs in } k)_{\ell t}} \frac{(\text{total \# of jobs in } k)_{\ell t}}{(\text{total \# of jobs})_{\ell t}}}{\sum\limits_{k'=1}^{K} \frac{\left(\text{\# with major } m \text{ who get a job in } k'\right)_{m\ell t}}{\left(\text{total \# of jobs in } k'\right)_{\ell t}} \frac{\left(\text{total \# of jobs in } k'\right)_{\ell t}}{(\text{total \# of jobs})_{\ell t}}} \right) \bar{w}_{k\ell t}
\end{aligned}$$

This expression makes use of two shares. The first one, $\frac{(\text{\# with major } m \text{ who get a job in } k)_{m\ell t}}{(\text{total \# of jobs in } k)_{\ell t}}$, is the way that we measure the relevance of industry $k$ to graduates with major $m$. As with the calculation for relevant employment in section 3.1, we will make the assumption that this relevance does not vary geographically, and measure it using the share for all students in Texas $r_{mkt}$.

The second share, $\frac{(\text{total \# of jobs in } k)_{\ell t}}{(\text{total \# of jobs})_{\ell t}}$, we will measure with using the share $s_{k\ell t}$ of employment in location $\ell$ and year $t$ that is industry $k$. This is not an exact measurement, because there is a difference between $(\text{total \# of jobs in } k)_{\ell t}$ as we are using it in the relevance share $r_{mkt}$ and in the employment share $s_{k\ell t}$. The number of jobs in the denominator of $r_{mkt}$ is the number of jobs in industry $k$ going to new college graduates, and the number of jobs in the numerator of $s_{k\ell t}$ is the number of workers employed in industry $k$. Similar to the case of relevant employment described in appendix section A.3 above, this approximation is exactly correct under the simplification that the number of jobs going to new graduates in an industry is proportional to the total number of jobs in that industry. Assume that this is true, and that a constant share $\alpha$ of jobs in each industry $k$ go to new graduates. Then,

$$\frac{(\text{total \# of jobs in } k)_{\ell t}}{(\text{total \# of jobs})_{\ell t}} = \frac{\alpha\,(\text{total \# employed in } k)_{\ell t}}{\sum\limits_{k=1}^{K} \alpha\,(\text{total \# of jobs in } k)_{\ell t}} = \frac{(\text{total \# employed in } k)_{\ell t}}{(\text{total \# employed})_{\ell t}} = s_{k\ell t}$$

With these two shares replaced, relevant employment can be written more simply as

$$RW_{m\ell t} = \sum_{k=1}^{K} \left( \frac{r_{mkt} s_{k\ell t}}{\sum\limits_{k'=1}^{K} r_{mk't} s_{k'\ell t}} \right) \bar{w}_{k\ell t}$$

We make one adjustment to this measure to account for the fact that the CBP suppresses payroll information in some cases to protect respondent confidentiality. In those cases, the average wage $\bar{w}_{k\ell t}$ is missing for the associated industry, location, and year combination. To get around this issue, we exclude those industries from the calculation of $RW_{m\ell t}$ as follows:

$$RW_{m\ell t} = \sum_{k=1}^{K} \left( \frac{r_{mkt} s_{k\ell t} \mathbb{1}\left\{\text{wage data is present}\right\}_{k\ell t}}{\sum\limits_{k'=1}^{K} r_{mk't} s_{k'\ell t} \mathbb{1}\left\{\text{wage data is present}\right\}_{k'\ell t}} \right) \bar{w}_{k\ell t}$$

### A.6 Separating Two-Digit CIP Codes into "STEM" and "Non-STEM"

In our regressions related to the majors that students select in college, we group majors into categories based on their two-digit CIP codes. However, the DHS's definition of STEM degrees does not define degrees to be STEM at the two-digit

CIP level, and so when we are presenting results separately by "STEM" and "Non-STEM" degrees we need to make some decisions on which two-digit CIP codes to mark as STEM. Fortunately, the DHS's STEM definitions can be captured quite accurately at the two-digit CIP level.

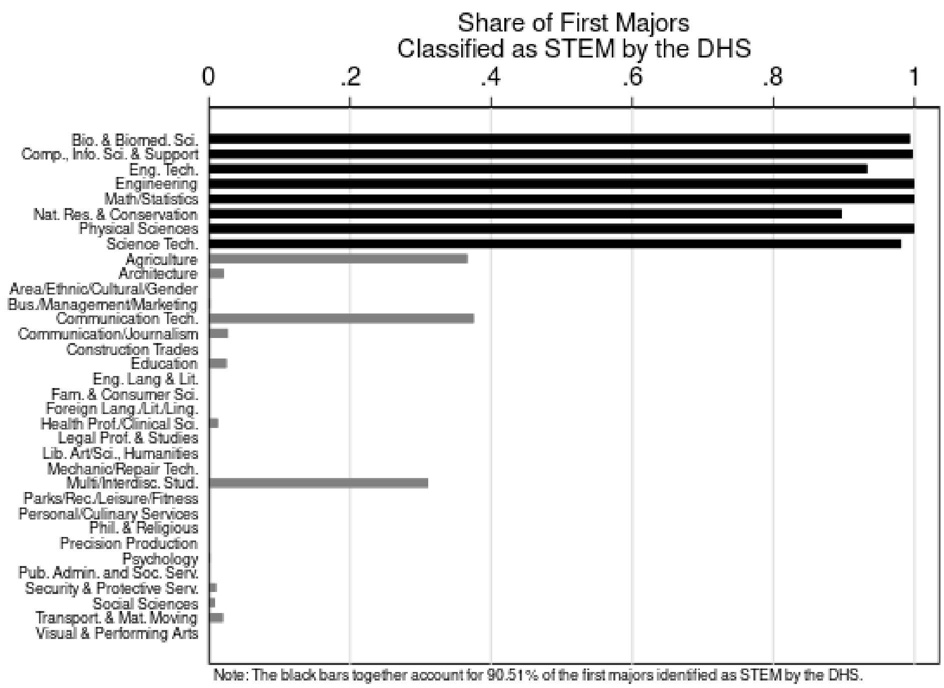



Figure 10: Comparing the DHS definition of STEM with a Definition Based on Two-Digit CIP Codes

In figure 10, each of the bars is associated with different grouping of majors at the two-digit CIP code level. The height of these bars shows the percentage of first majors declared in this two-digit category which were classified as STEM by the DHS.[74] We will define the black bars as representing "STEM" two-digit CIP codes. As can be seen, there is a clear difference between these bars and the bars from any other major category. Where over 80% of first majors declared in each of the black bar categories are considered STEM by the DHS, none of the other first major categories had even 40% of their declared majors designated as STEM. Further, the black bars together account for 90.51% of the first majors identified as STEM by the DHS.

## B Additional Information on the Regressions

### B.1 Evidence that Employment shares have Lagged Effects

The upward trend in the coefficients from table 1 gives evidence against a model in which the relevant employment share only has effects on contemporaneous student decisions. Specifically, consider the model

$$Y_{m\ell t} = \gamma_{m\ell} + \beta_1 RE_{m\ell t} + \varepsilon_{m\ell t} \tag{27}$$

where $\gamma_{m\ell}$ is a major-location fixed effect. Since we estimate our equation in differences, this would become

$$\Delta_d Y_{m\ell t} = \beta_1 \Delta_d RE_{m\ell t} + \Delta_d \varepsilon_{m\ell t} \tag{28}$$

One implication of this equation that our estimate for $\beta_1$ should not depend on the number of years $d$ that we use to take the difference. Thus, our result of different coefficients for different values of $d$ would be unexpected if this were the true data generating process. In the future, we intend to use a statistical test to more formally check whether we can reject the null hypothesis that the coefficients shown in the first seven lags of table 1 are all equal.

[74]If a student declared two separate first majors in "Agriculture," and only one of them was categorized by the DHS as STEM, we counted this as a single first major in the "Agriculture" category that was categorized as STEM.

By contrast, the trend shown in table 1 and the fact that our coefficient estimates eventually level off as $d$ increases are both more consistent with a general distributed lag model. Consider the model

$$Y_{m\ell t} = \gamma_{m\ell} + \sum_{j=0}^{J} \beta_j RE_{m\ell t-j} + \varepsilon_{m\ell_t} \tag{29}$$

in which all of the lags of the relevant employment share up to $J$ are allowed to have separate effects on the outcomes $Y_{m\ell t}$. Differencing this model results in the expression

$$\begin{aligned} \Delta_d Y_{m\ell t} &= \sum_{j=0}^{J} \beta_j \Delta_d RE_{m\ell t-j} + \Delta_d \varepsilon_{m\ell t} \\ &= \beta_0 \Delta_d RE_{m\ell t} + \sum_{j=1}^{J} \beta_j \Delta_d RE_{m\ell t-j} + \Delta_d \varepsilon_{m\ell t} \end{aligned} \tag{30}$$

In our estimation results from section 4, we omit the $\sum_{j=1}^{J} \beta_j \Delta_d RE_{m\ell t-j}$ terms above. If this distributed lag model is true, that creates omitted variable bias as follows:

$$\begin{aligned} \frac{\text{Cov}\left(\Delta_d Y_{m\ell t}, \Delta_d RE_{m\ell t}\right)}{\text{Var}\left(\Delta_d RE_{m\ell t}\right)} &= \frac{\text{Cov}\left(\beta_0 \Delta_d RE_{m\ell t} + \sum_{j=1}^{J} \beta_j \Delta_d RE_{m\ell t-j} + \Delta_d \varepsilon_{m\ell t}, \Delta_d RE_{m\ell t}\right)}{\text{Var}\left(\Delta_d RE_{m\ell t}\right)} \\ &= \beta_0 + \frac{\text{Cov}\left(\sum_{j=1}^{J} \beta_j \Delta_d RE_{m\ell t-j}, \Delta_d RE_{m\ell t}\right)}{\text{Var}\left(\Delta_d RE_{m\ell t}\right)} \\ &= \beta_0 + \frac{\sum_{j=1}^{J} \beta_j \left(\text{Cov}\left(RE_{m\ell t-j}, RE_{m\ell t}\right) + \text{Cov}\left(RE_{m\ell t-j-d}, RE_{m\ell t-d}\right)\right)}{\text{Var}\left(\Delta_d RE_{m\ell t}\right)} - \\ &\qquad \frac{\sum_{j=1}^{J} \beta_j \left(\text{Cov}\left(RE_{m\ell t-j-d}, RE_{m\ell t}\right) + \text{Cov}\left(RE_{m\ell t-j}, RE_{m\ell t-d}\right)\right)}{\text{Var}\left(\Delta_d RE_{m\ell t}\right)} \end{aligned} \tag{31}$$

To simplify this expression, we can make a stationarity assumption, that $\text{Cov}\left(RE_{m\ell t-j}, RE_{m\ell t}\right) = \text{Cov}\left(RE_{m\ell t'-j}, RE_{m\ell t'}\right)$ for all years $t$ and $t'$. Then, $\text{Cov}\left(RE_{m\ell t-j}, RE_{m\ell t}\right) = \text{Cov}\left(RE_{m\ell t-j-d}, RE_{m\ell t-d}\right)$, and we can rewrite equation (31) as

$$\begin{aligned} \frac{\text{Cov}\left(\Delta_d Y_{m\ell t}, \Delta_d RE_{m\ell t}\right)}{\text{Var}\left(\Delta_d RE_{m\ell t}\right)} &= \beta_0 + \frac{2\sum_{j=1}^{J} \beta_j \text{Cov}\left(RE_{m\ell t-j}, RE_{m\ell t}\right)}{\text{Var}\left(\Delta_d RE_{m\ell t}\right)} - \frac{\sum_{j=1}^{J} \beta_j \text{Cov}\left(RE_{m\ell t-j-d}, RE_{m\ell t}\right)}{\text{Var}\left(\Delta_d RE_{m\ell t}\right)} - \\ &\qquad \frac{\sum_{j=1}^{J} \beta_j \text{Cov}\left(RE_{m\ell t-j}, RE_{m\ell t-d}\right)}{\text{Var}\left(\Delta_d RE_{m\ell t}\right)} \end{aligned} \tag{32}$$

The first two terms in the above expression do not depend on $d$, and so they should not be contributing to the trend in coefficients seen in table 1.

To learn about the last two terms, we make an additional assumption that for all $t$, the absolute value of $\text{Cov}\left(RE_{m\ell t-d}, RE_{m\ell t}\right)$ decreasing monotonically toward a limit of zero as $d$ increases. In other words, we assume the relevant employment share from this year has more to do with the relevant employment share from last year than it does with any prior year, and that the relationship between relevant employment share and historical values of the share becomes negligible as we look far enough into the past. Under this assumption, the absolute value of the third term in equation (32) decreases monotonically toward 0.

The fourth term is more complicated, because the magnitude of each $\beta_j \text{Cov}\left(RE_{m\ell t-j}, RE_{m\ell t-d}\right)$ will increase monotonically as $d$ approaches $j$. Nevertheless, once $d$ becomes greater than $J$, the amount of time between $t-j$ and $t-d$ will grow monotonically with $d$ for all $j$, and so the full fourth term will begin decreasing monotonically toward 0. In the limit as $d$ increases, then, the last two terms converge to zero, and equation (32) becomes the constant

$$\frac{\text{Cov}\left(\Delta_d Y_{m\ell t}, \Delta_d RE_{m\ell t}\right)}{\text{Var}\left(\Delta_d RE_{m\ell t}\right)} = \beta_0 + \frac{2\sum_{j=1}^{J} \beta_j \text{Cov}\left(RE_{m\ell t-j}, RE_{m\ell t}\right)}{\text{Var}\left(\Delta_d RE_{m\ell t}\right)} \tag{33}$$

This provides intuition for why the coefficient effects shown in table 1 eventually level off as $d$ increases.

As can be clearly seen from equation (32), however, the path that the coefficient estimates take to this limit will depend on both on the values of the $\beta_j$ terms and the covariances between relevant employment shares over different lengths of time. For this reason, the trend of the coefficients in table 1 contains valuable information about these values. We should also point out that just like our results in table 1 provide evidence that the relevant employment shares have lagged effects, they also provide evidence against a model in which relevant employment shares have no effect. If $\beta_j$ had been 0 for $j = 0$ through $J$, equation (32) suggests that we should have found a null effect in our estimates of $\beta_0$ as well. The fact that our estimated coefficients were statistically significant therefore suggests at least one $\beta_j$ term was nonzero, and that the local labor market has effects on student course taking. We intend on using our estimates from table 1 and other tables in the regression section as moments when estimating our full model.

An alternative method would be to try and estimate equation (29) directly from the data. We do not take this approach for two reasons: First, it would demand the use of $J$ separate instrumental variables to address the possible connections between each $RE_{m\ell t}$ term and unobserved variables. This is infeasible given our sample size. Second, even if we were to use $J$ separate instrumental variables, the form of (29) is quite restrictive. For example, it requires that each lagged value of relevant employment would have linear effects on the outcome, and that each of these effects would not depend at all on the values of relevant employment in other years.

In section 4.1, we mention that lagged conditions in the local labor market could affect a student's present course-taking choices if students begin to prepare for those courses in earlier years, and so students do not have sufficient time to react to contemporaneous changes in the labor market. This is the explanation that we will largely focus on in our structural model. There are at least two other possible explanations for the upward trend in table 1. The first one is that short-term movements in industries are volatile, and students may be less likely to react to them because they do not think that they will persist. This may be especially relevant for high school students, who likely will not enter the labor market and earn the rewards of a college degree there until several years later. Longer term changes, by contrast, may be more indicative of long term conditions in industries, and students may react to them more for that reason. Second, it could be that not all students learn about changes in the industry market instantly, and it may take several years before all students are aware of the changes. This explanation is in line with Auclert et al. (2020), who explain the seemingly delayed response of macroeconomic variables to changing conditions by arguing that individual agents may learn about these changing conditions with some delay. In the future, we hope to more directly consider these stories by including additional features to the model. For example, we hope to add a mechanism by which students only update their beliefs on $SE_t$ with a certain probability each period, and otherwise continue making decisions as though the expected realization of $SE_t = \rho_0^{SE} + \rho_1^{SE} SE_{t-1}$ had occurred in time $t$.

### B.2 Using Regression Results to Identify Parameters in the Structural Model

The above section showed how coefficient estimates like those prepared in section 4 encode information about the effects of relevant employment shares in prior time periods. To more easily see how these moments will be useful in identifying our model, we can rewrite equation (32) through the lens of the model. Within the model, the $r_{mkt}$ relevance terms are held fixed over time, the share of employment in STEM $SE_t$ follows an AR-1 process, and there are only two industries—one STEM-intensive and the other non-STEM intensive. In this case, the relevant employment share for a major $m$ can be re-written

$$RE_{m\ell t} = r_{ms} SE_t + r_{mn}\left(1 - SE_t\right)$$

Then, keeping in mind the AR-1 process of $SE_{t+1} = \rho_0^{SE} + \rho_1^{SE} SE_t + \omega_t$, equation (32) simplifies down to

$$\frac{\text{Cov}\left(\Delta_d Y_{m\ell t}, \Delta_d RE_{m\ell t}\right)}{\text{Var}\left(\Delta_d RE_{m\ell t}\right)} = \beta_0 + \frac{\sum_{j=1}^{J} \beta_j \left(2\left(\rho_1^{SE}\right)^{j} - \left(\rho_1^{SE}\right)^{(j+d)} - \left(\rho_1^{SE}\right)^{|j-d|}\right)}{2\left(1 - \left(\rho_1^{SE}\right)^{d}\right)} \tag{34}$$

This expression is linear in the $\beta_j$ terms. Thus, equation (34) shows that preparing some number $J$ of coefficient estimates at different lags $d$ will be sufficient to pin down $J$ separate $\beta_j$ terms,[75] each representing the influence of a lagged value of relevant employment conditions on students' present decisions in $Y_{m\ell t}$. We include 5 separate lags for each of the regression moments used in our simulated method of moments procedure, suggesting that the effects of relevant employment are pinned down in the current period and in each of the previous 4 periods. Since there are only 4 periods in which students are able to make decisions within the model, we argue that this should be more than sufficient to identify the $\delta$ parameters governing how students value SEt realizations while in school and wages once in the labor market.

### B.3 Additional Regression Results

In this section, we examine the robustness of the results presented in section 4, and also add other regressions for completeness. To begin, we will explain table 18 in detail, and then the remaining tables will follow similar formats. The top row shows the overall OLS estimates. These are generally lower than the 2SLS estimates, suggesting that the instrument is correcting for downward bias. The 2SLS row shows the results from table 1, and the "2SLS STEM" and "2SLS Non-STEM" rows show the results of estimating equation (4) separately for $m$ =STEM and $m$ =Non-STEM. In this case, the effects are going in the same direction for both rows, suggesting that the overall effect presented in the "2SLS" row is not driven exclusively by student's responses in either class group.

The next three rows of table 18 break these results down for different types of classes. Of the classes listed here, it appears that the results in the "2SLS" row are most being driven most prominently by student enrollment in Career and Technical Education (CTE) courses. This makes sense, given their close connection to the labor market. There are positive but statistically insignificant effects for "advanced" courses, and statistically insignificant negative effects for courses in which students can earn college credit.

Finally, the last two rows present estimates where each of the industry-specific terms summed in equation 5 and each year are included as separate instruments. In other words, we include

$$r_{mk0}s_{k\ell 0}\left(\frac{g_{kt}^{TX}}{\sum_{k'} s_{k'\ell 0}g_{k't}^{TX}} - \frac{g_{kt-d}^{TX}}{\sum_{k'} s_{k'\ell 0}g_{k't-d}^{TX}}\right)$$

as a separate instrument for every combination of $k$ and $t$. This kind of robustness check was suggested by Goldsmith-Pinkham et al. (2020), and serves to check whether we are properly specifying equation (4) when we choose to sum the relevant employment across industries. We show this done in two ways: first, by including all of these instruments in a single 2SLS estimation, and second by doing a LIML estimation. The fact that the signs and magnitudes of these estimates are similar to those presented in the 2SLS row provides encouraging evidence that our equations are properly specified. However, as described in section B.1, the variation in the coefficient estimates across the columns still suggests that there are non-contemporaneous effects that this regression method will not be able to capture directly. To do that, we will use a structural model.

[75] Barring, of course, the unlikely scenario in which the resulting versions of equation (34) were linearly dependent.

Table 18: Change in the Share of Courses High School Seniors take in subject *m*

| | (1) 1-year | (2) 2-year | (3) 3-year | (4) 4-year | (5) 5-year | (6) 6-year | (7) 7-year | (8) 8-year | (9) 9-year | (10) 10-year |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | -0.0384 | 0.0069 | 0.1471 | 0.1911 | 0.2783* | 0.3451* | 0.4031** | 0.4610** | 0.5500** | 0.6591** |
| | (0.0683) | (0.0820) | (0.0845) | (0.0976) | (0.1234) | (0.1366) | (0.1340) | (0.1450) | (0.1780) | (0.2050) |
| 2SLS | -0.2817 | -0.0450 | 0.1283 | 0.5640 | 0.9458* | 1.0525** | 1.0373* | 0.9471* | 0.9465* | 0.9549* |
| | (0.3323) | (0.3986) | (0.4030) | (0.4110) | (0.4120) | (0.4041) | (0.4059) | (0.3914) | (0.4390) | (0.4540) |
| 2SLS STEM | -0.2890 | -0.1884 | -0.0512 | 0.2616 | 0.5283 | 0.6450* | 0.6728* | 0.5661 | 0.5897 | 0.6164 |
| | (0.3265) | (0.3925) | (0.3627) | (0.3332) | (0.3052) | (0.3110) | (0.3029) | (0.3004) | (0.3161) | (0.3309) |
| 2SLS Non-STEM | -0.5650 | -0.2217 | -0.1282 | 0.2802 | 0.6059 | 0.7296 | 0.6929 | 0.5912 | 0.5905 | 0.5912 |
| | (0.3950) | (0.4423) | (0.3969) | (0.3699) | (0.3480) | (0.3750) | (0.3648) | (0.3486) | (0.3601) | (0.3802) |
| Advanced | 0.1124 | 0.1305 | 0.1725 | 0.0901 | 0.2071 | 0.1406 | 0.1069 | 0.0915 | 0.0780 | 0.0473 |
| | (0.1497) | (0.1360) | (0.1682) | (0.2497) | (0.2766) | (0.2932) | (0.2965) | (0.2832) | (0.3239) | (0.3553) |
| College Credit | -0.0860 | -0.0647 | -0.0114 | -0.0528 | -0.0134 | -0.0208 | -0.0117 | -0.0048 | 0.0058 | 0.0151 |
| | (0.1276) | (0.0925) | (0.0804) | (0.0779) | (0.0838) | (0.0894) | (0.0929) | (0.0947) | (0.0994) | (0.1019) |
| CTE | 0.0228 | 0.1961 | 0.3067* | 0.3735* | 0.4455* | 0.4265* | 0.4052 | 0.3809 | 0.4087 | 0.4338* |
| | (0.1993) | (0.1468) | (0.1402) | (0.1746) | (0.2010) | (0.2087) | (0.2075) | (0.1964) | (0.2114) | (0.2091) |
| All Instruments | -0.0872 | -0.0836 | 0.0790 | 0.3499** | 0.5985** | 0.7738*** | 0.7513*** | 0.7722*** | 0.7765** | 0.8908** |
| | (0.1002) | (0.1374) | (0.1339) | (0.1326) | (0.1868) | (0.2128) | (0.2070) | (0.2116) | (0.2402) | (0.2789) |
| LIML | -0.1921 | -0.2317 | -0.0578 | 0.6579* | 1.2676** | 1.3727*** | 1.2056*** | 1.1019*** | 0.9902** | 1.1114** |
| | (0.2718) | (0.3176) | (0.3575) | (0.2680) | (0.4266) | (0.3772) | (0.3488) | (0.3225) | (0.3363) | (0.3997) |
| Observations | 2,546 | 2,412 | 2,278 | 2,144 | 2,010 | 1,876 | 1,742 | 1,608 | 1,474 | 1,340 |
| Boot. P Val. for 2SLS | 0.428 | 0.864 | 0.012 | 0.003 | 0.116 | 0.011 | 0.001 | 0.002 | 0.004 | 0.010 |
| Boot. P Val. for STEM | 0.876 | 0.001 | 0.966 | 0.363 | 0.216 | 0.122 | 0.118 | 0.153 | 0.160 | 0.880 |
| Boot. P Val. for Non-STEM | 0.221 | 0.006 | 0.883 | 0.448 | 0.226 | 0.131 | 0.187 | 0.226 | 0.259 | 0.741 |
| F-Stat | 9.689 | 22.654 | 27.969 | 21.810 | 19.815 | 20.166 | 20.274 | 20.389 | 18.545 | 18.323 |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |

Standard errors in parentheses

The STEM and Non-STEM rows only used classes from those subjects, so they have half as many observations.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Change in the Share of Course Offerings that are in subject $m$

| | (1)<br>1-year | (2)<br>2-year | (3)<br>3-year | (4)<br>4-year | (5)<br>5-year | (6)<br>6-year | (7)<br>7-year | (8)<br>8-year | (9)<br>9-year | (10)<br>10-year |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | -0.0236 | 0.0101 | 0.0045 | -0.0030 | -0.0218 | -0.0104 | -0.0122 | -0.0159 | -0.0181 | -0.0147 |
| | (0.0581) | (0.0587) | (0.0505) | (0.0478) | (0.0433) | (0.0506) | (0.0574) | (0.0649) | (0.0677) | (0.0805) |
| 2SLS | 0.0951 | -0.0915 | -0.0122 | 0.0045 | 0.0689 | 0.0989 | 0.1182 | 0.1115 | 0.1289 | 0.1592 |
| | (0.3114) | (0.2204) | (0.2290) | (0.1963) | (0.1986) | (0.1754) | (0.1725) | (0.1691) | (0.1866) | (0.1850) |
| 2SLS STEM | 0.0627 | -0.1986 | -0.1921 | -0.2238 | -0.2693 | -0.1979 | -0.1602 | -0.1777 | -0.1794 | -0.1725 |
| | (0.2794) | (0.2025) | (0.2055) | (0.1804) | (0.1704) | (0.1373) | (0.1435) | (0.1308) | (0.1417) | (0.1511) |
| 2SLS Non-STEM | -0.0239 | -0.1878 | -0.1312 | -0.1159 | -0.0347 | 0.0227 | 0.0186 | -0.0236 | -0.0278 | -0.0167 |
| | (0.3426) | (0.2380) | (0.2145) | (0.1932) | (0.1840) | (0.1633) | (0.1579) | (0.1468) | (0.1543) | (0.1647) |
| Advanced | 0.0428 | -0.0130 | 0.0831 | -0.0278 | -0.0204 | -0.1298 | -0.1386 | -0.1565 | -0.2059 | -0.2468 |
| | (0.1530) | (0.1124) | (0.1352) | (0.1727) | (0.2009) | (0.2128) | (0.2193) | (0.2154) | (0.2450) | (0.2644) |
| College Credit | -0.1420 | -0.1002 | -0.0630 | -0.1550 | -0.1656 | -0.2053 | -0.2074 | -0.2068 | -0.2159 | -0.2082 |
| | (0.1155) | (0.1055) | (0.0950) | (0.1014) | (0.1128) | (0.1225) | (0.1186) | (0.1232) | (0.1301) | (0.1324) |
| CTE | 0.1452 | 0.1145 | 0.1940 | 0.2227 | 0.2846 | 0.2858 | 0.2506 | 0.2069 | 0.2301 | 0.2721 |
| | (0.1657) | (0.1235) | (0.1480) | (0.1695) | (0.1978) | (0.1865) | (0.1847) | (0.1709) | (0.1783) | (0.1775) |
| All Instruments | -0.0574 | -0.1288 | -0.0657 | -0.0807 | -0.0830 | -0.0332 | -0.0421 | -0.0510 | -0.0555 | -0.0326 |
| | (0.0953) | (0.0878) | (0.0799) | (0.0709) | (0.0662) | (0.0793) | (0.0862) | (0.0879) | (0.0877) | (0.1036) |
| LIML | -0.1103 | -0.3063 | -0.1392 | -0.1576 | -0.1436 | -0.0605 | -0.0746 | -0.0868 | -0.0915 | -0.0545 |
| | (0.2017) | (0.1728) | (0.1345) | (0.1193) | (0.1142) | (0.1412) | (0.1443) | (0.1397) | (0.1360) | (0.1741) |
| Observations | 2,546 | 2,412 | 2,278 | 2,144 | 2,010 | 1,876 | 1,742 | 1,608 | 1,474 | 1,340 |
| Boot. P Val. for 2SLS | 0.792 | 0.428 | 0.097 | 0.007 | 0.191 | 0.105 | 0.053 | 0.102 | 0.122 | 0.158 |
| Boot. P Val. for STEM | 0.512 | 0.028 | 0.513 | 0.370 | 0.142 | 0.177 | 0.256 | 0.221 | 0.258 | 0.059 |
| Boot. P Val. for Non-STEM | 0.784 | 0.039 | 0.690 | 0.709 | 0.911 | 0.917 | 0.976 | 0.922 | 0.912 | 0.292 |
| F-Stat | 9.689 | 22.654 | 27.969 | 21.810 | 19.815 | 20.166 | 20.274 | 20.389 | 18.545 | 18.323 |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |

Standard errors in parentheses

The STEM and Non-STEM rows only used classes from those subjects, so they have half as many observations.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 20: Change in the Share of Senior-year Classes in subject *m* where the student earns credit

| | (1)<br>1-year | (2)<br>2-year | (3)<br>3-year | (4)<br>4-year | (5)<br>5-year | (6)<br>6-year | (7)<br>7-year | (8)<br>8-year | (9)<br>9-year | (10)<br>10-year |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | -0.0472 | -0.0541$^{*}$ | -0.0606 | -0.0856$^{*}$ | -0.1098 | -0.1127 | -0.1272 | -0.1355 | -0.1568 | -0.1775 |
| | (0.0268) | (0.0274) | (0.0340) | (0.0410) | (0.0561) | (0.0638) | (0.0684) | (0.0799) | (0.0888) | (0.0961) |
| 2SLS | 0.0153 | 0.0313 | 0.0468 | -0.0718 | -0.1152 | -0.1665 | -0.1266 | -0.1072 | -0.0753 | -0.0802 |
| | (0.1756) | (0.0810) | (0.0751) | (0.0758) | (0.0902) | (0.0934) | (0.0927) | (0.0822) | (0.0949) | (0.0966) |
| 2SLS STEM | 0.0188 | -0.2011 | -0.2025 | -0.5634$^{***}$ | -0.7212$^{***}$ | -0.8684$^{***}$ | -0.7652$^{***}$ | -0.7042$^{***}$ | -0.6690$^{***}$ | -0.6694$^{***}$ |
| | (0.3349) | (0.2022) | (0.1753) | (0.1684) | (0.1938) | (0.2045) | (0.1932) | (0.1677) | (0.1833) | (0.1937) |
| 2SLS Non-STEM | -0.0806 | 0.1424 | 0.1457 | 0.2181$^{*}$ | 0.2382$^{*}$ | 0.2702$^{*}$ | 0.2447$^{*}$ | 0.2114 | 0.2023 | 0.1933 |
| | (0.1619) | (0.1128) | (0.1155) | (0.0935) | (0.1117) | (0.1137) | (0.1098) | (0.1152) | (0.1189) | (0.1138) |
| Advanced | -0.1554 | -0.1353 | -0.1993 | -0.2189 | -0.2409 | -0.2370 | -0.2173 | -0.2020 | -0.1504 | -0.1431 |
| | (0.2067) | (0.1132) | (0.1067) | (0.1164) | (0.1352) | (0.1351) | (0.1373) | (0.1290) | (0.1466) | (0.1452) |
| College Credit | -0.5842$^{*}$ | -0.3146$^{*}$ | -0.3159$^{**}$ | -0.2538$^{*}$ | -0.2175 | -0.2415$^{*}$ | -0.2864$^{*}$ | -0.2961$^{**}$ | -0.2207 | -0.2165 |
| | (0.2876) | (0.1379) | (0.1172) | (0.1191) | (0.1265) | (0.1196) | (0.1160) | (0.1060) | (0.1141) | (0.1144) |
| CTE | -0.1602 | 0.0501 | -0.0687 | -0.0219 | 0.0824 | 0.1928 | 0.2583$^{*}$ | 0.1838 | 0.2005 | 0.1846 |
| | (0.2212) | (0.1188) | (0.1451) | (0.1512) | (0.1498) | (0.1440) | (0.1208) | (0.1280) | (0.1313) | (0.1230) |
| All Instruments | -0.0143 | 0.0262 | -0.0114 | -0.0999 | -0.1067 | -0.1376 | -0.1476 | -0.1644 | -0.1673 | -0.1787 |
| | (0.0609) | (0.0646) | (0.0505) | (0.0632) | (0.0883) | (0.0908) | (0.0909) | (0.0893) | (0.0966) | (0.1002) |
| LIML | 0.0735 | 0.1649 | 0.0679 | -0.1168 | -0.1034 | -0.1641 | -0.1677 | -0.1880 | -0.1765 | -0.1798 |
| | (0.2010) | (0.1544) | (0.1113) | (0.1116) | (0.1501) | (0.1481) | (0.1427) | (0.1288) | (0.1427) | (0.1489) |
| Observations | 2,546 | 2,412 | 2,278 | 2,144 | 2,010 | 1,876 | 1,742 | 1,608 | 1,474 | 1,340 |
| Boot. P Val. for 2SLS | 0.943 | 0.931 | 0.168 | 0.623 | 0.277 | 0.443 | 0.817 | 0.738 | 0.382 | 0.204 |
| Boot. P Val. for STEM | 0.691 | 0.410 | 0.304 | 0.045 | 0.005 | 0.001 | 0.001 | 0.001 | 0.015 | 0.002 |
| Boot. P Val. for Non-STEM | 0.452 | 0.441 | 0.221 | 0.042 | 0.120 | 0.066 | 0.073 | 0.141 | 0.303 | 0.106 |
| F-Stat | 10 | 23 | 28 | 22 | 20 | 20 | 20 | 20 | 19 | 18 |
| Obs for Advanced | 2529.000 | 2399.000 | 2265.000 | 2131.000 | 1997.000 | 1863.000 | 1729.000 | 1596.000 | 1461.000 | 1329.000 |
| Obs for College-Credit | 2490.000 | 2358.000 | 2224.000 | 2094.000 | 1959.000 | 1826.000 | 1694.000 | 1563.000 | 1431.000 | 1306.000 |
| Obs for CTE | 1745.000 | 1619.000 | 1489.000 | 1362.000 | 1237.000 | 1113.000 | 990.000 | 866.000 | 741.000 | 673.000 |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |

Standard errors in parentheses

The STEM and Non-STEM rows only used classes from those subjects, so they have half as many observations.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 21: Change in the Share of College-Bound High School Graduates Choosing *m* as their First Major in College

| | (1)<br>1-Yr. Change | (2)<br>2-Yr. Change | (3)<br>3-Yr. Change | (4)<br>4-Yr. Change | (5)<br>5-Yr. Change | (6)<br>6-Yr. Change | (7)<br>7-Yr. Change | (8)<br>8-Yr. Change | (9)<br>9-Yr. Change | (10)<br>10-Yr. Change |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | -0.1687* | -0.0642 | 0.0370 | 0.1410 | 0.1098 | 0.1061 | 0.1042 | 0.0956 | 0.0509 | 0.1224 |
| | (0.0817) | (0.0952) | (0.1357) | (0.1992) | (0.2456) | (0.2903) | (0.3104) | (0.3297) | (0.3696) | (0.3821) |
| 2SLS | 2.0352 | 1.5376 | 1.3721 | 1.3149 | 1.7473 | 1.5956 | 1.9044 | 1.8343 | 2.3994 | 2.4063 |
| | (1.4322) | (0.8706) | (0.9841) | (1.3821) | (2.4997) | (2.9268) | (3.4250) | (3.1406) | (4.0287) | (4.1204) |
| 2SLS STEM | 0.6409 | 0.7077** | 0.6921** | 0.5621** | 0.6067** | 0.4929** | 0.4944** | 0.4770** | 0.4635* | 0.4790* |
| | (0.3392) | (0.2699) | (0.2225) | (0.2131) | (0.1943) | (0.1711) | (0.1694) | (0.1740) | (0.1869) | (0.1949) |
| 2SLS NonSTEM | 3.1831 | 2.0733 | 1.8239 | 2.2386 | 8.6893 | -9.7656 | -7.5778 | -15.9694 | -12.8769 | -13.1304 |
| | (2.7656) | (1.4753) | (1.8105) | (3.7166) | (55.1529) | (95.7088) | (46.7693) | (181.2749) | (95.2053) | (103.7415) |
| All Instruments | -0.0120 | 0.2880 | 0.2098 | 0.3195 | 0.4120 | 0.3159 | 0.0061 | -0.0201 | -0.0726 | -0.0412 |
| | (0.1595) | (0.1471) | (0.1826) | (0.2045) | (0.2467) | (0.2966) | (0.2859) | (0.2742) | (0.2584) | (0.2757) |
| LIML | 0.0095 | 0.3280* | 0.2287 | 0.3379 | 0.4565 | 0.3492 | -0.0072 | -0.0338 | -0.0852 | -0.0579 |
| | (0.1786) | (0.1605) | (0.1985) | (0.2217) | (0.2767) | (0.3376) | (0.3207) | (0.3029) | (0.2817) | (0.3009) |
| Observations | 43,282 | 41,004 | 38,726 | 36,448 | 34,170 | 31,892 | 29,614 | 27,336 | 25,058 | 22,780 |
| Boot. P Value for 2SLS | 0.059 | 0.023 | 0.089 | 0.238 | 0.309 | 0.403 | 0.389 | 0.376 | 0.353 | 0.360 |
| Boot. P Value for 2SLS STEM | 0.051 | 0.012 | 0.005 | 0.019 | 0.003 | 0.009 | 0.014 | 0.016 | 0.023 | 0.027 |
| Boot. P Value for 2SLS NonSTEM | 0.099 | 0.060 | 0.168 | 0.334 | 0.780 | 0.858 | 0.778 | 0.879 | 0.814 | 0.837 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 3.407 | 6.034 | 4.321 | 2.396 | 0.976 | 0.700 | 0.629 | 0.732 | 0.588 | 0.542 |
| First-Stage F, STEM | 10.140 | 21.642 | 30.831 | 29.739 | 26.818 | 26.055 | 26.798 | 25.450 | 22.919 | 21.734 |
| First-Stage F, Non-STEM | 1.942 | 3.581 | 1.998 | 0.627 | 0.025 | 0.011 | 0.031 | 0.008 | 0.020 | 0.018 |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the NonSTEM row used the other 76.471%.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 22: Change in the Share of Community College-Bound High School Graduates Choosing *m* as their First Major in College

| | (1)<br>1-Yr. Change | (2)<br>2-Yr. Change | (3)<br>3-Yr. Change | (4)<br>4-Yr. Change | (5)<br>5-Yr. Change | (6)<br>6-Yr. Change | (7)<br>7-Yr. Change | (8)<br>8-Yr. Change | (9)<br>9-Yr. Change | (10)<br>10-Yr. Change |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | -0.2298* | -0.0947 | 0.0031 | 0.1793 | 0.1512 | 0.1717 | 0.1789 | 0.1554 | 0.1738 | 0.2685 |
| | (0.0978) | (0.1242) | (0.1863) | (0.2779) | (0.3449) | (0.4040) | (0.4177) | (0.4334) | (0.4780) | (0.4908) |
| 2SLS | 2.8703 | 2.0970 | 1.9682 | 2.1485 | 3.2663 | 3.1925 | 3.6434 | 3.3716 | 4.3276 | 4.4572 |
| | (2.0499) | (1.2303) | (1.3811) | (2.0366) | (3.9798) | (4.7126) | (5.5157) | (4.9352) | (6.4729) | (6.7506) |
| 2SLS STEM | 0.5936 | 0.5775 | 0.6156 | 0.5999 | 0.6381* | 0.5164 | 0.5211* | 0.4574* | 0.4998* | 0.4642* |
| | (0.4596) | (0.3608) | (0.3323) | (0.3348) | (0.3105) | (0.2686) | (0.2445) | (0.2312) | (0.2366) | (0.2317) |
| 2SLS NonSTEM | 4.7364 | 3.1246 | 2.9912 | 4.2290 | 20.0884 | -25.9447 | -17.7332 | -37.0580 | -26.7230 | -28.7616 |
| | (4.0265) | (2.0747) | (2.5911) | (6.0052) | (123.2762) | (256.2020) | (108.2091) | (442.6217) | (197.8784) | (229.0172) |
| All Instruments | 0.0943 | 0.4178* | 0.1919 | 0.3019 | 0.3986 | 0.2992 | -0.1102 | -0.1451 | -0.1576 | -0.1582 |
| | (0.1880) | (0.2012) | (0.2353) | (0.2768) | (0.3439) | (0.4414) | (0.3823) | (0.3588) | (0.3222) | (0.3358) |
| LIML | 0.1443 | 0.4879* | 0.2155 | 0.3169 | 0.4415 | 0.3235 | -0.1574 | -0.1906 | -0.2018 | -0.2176 |
| | (0.2142) | (0.2228) | (0.2591) | (0.3055) | (0.3952) | (0.5163) | (0.4392) | (0.4074) | (0.3620) | (0.3798) |
| Observations | 43,078 | 40,800 | 38,590 | 36,312 | 34,034 | 31,756 | 29,478 | 27,200 | 24,922 | 22,678 |
| Boot. P Value for 2SLS | 0.078 | 0.029 | 0.079 | 0.193 | 0.248 | 0.317 | 0.327 | 0.319 | 0.315 | 0.317 |
| Boot. P Value for 2SLS STEM | 0.193 | 0.115 | 0.070 | 0.079 | 0.048 | 0.068 | 0.049 | 0.070 | 0.053 | 0.066 |
| Boot. P Value for 2SLS NonSTEM | 0.103 | 0.040 | 0.111 | 0.273 | 0.775 | 0.859 | 0.776 | 0.882 | 0.813 | 0.839 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 3.406 | 6.019 | 4.309 | 2.385 | 0.972 | 0.698 | 0.629 | 0.732 | 0.586 | 0.542 |
| First-Stage F, STEM | 10.085 | 21.998 | 31.915 | 31.411 | 28.282 | 27.285 | 27.833 | 26.171 | 23.441 | 22.256 |
| First-Stage F, Non-STEM | 1.940 | 3.574 | 1.997 | 0.627 | 0.025 | 0.011 | 0.029 | 0.007 | 0.019 | 0.016 |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the NonSTEM row used the other 76.471%.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 23: Change in the Share of 4-year public university-Bound High School Graduates Choosing *m* as their First Major in College

| | (1)<br>1-Yr. Change | (2)<br>2-Yr. Change | (3)<br>3-Yr. Change | (4)<br>4-Yr. Change | (5)<br>5-Yr. Change | (6)<br>6-Yr. Change | (7)<br>7-Yr. Change | (8)<br>8-Yr. Change | (9)<br>9-Yr. Change | (10)<br>10-Yr. Change |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | 0.0199 | 0.0548 | 0.0888 | 0.0414 | 0.0511 | 0.0045 | -0.0421 | -0.0277 | -0.1599 | -0.1707 |
| | (0.1551) | (0.1295) | (0.1117) | (0.1198) | (0.1267) | (0.1432) | (0.1631) | (0.1789) | (0.2066) | (0.2272) |
| 2SLS | 0.8770 | 0.2978 | 0.0715 | -0.2788 | -1.3935 | -1.2935 | -1.7060 | -1.2792 | -1.2202 | -1.4412 |
| | (1.4344) | (0.6711) | (0.6602) | (0.7924) | (1.8301) | (2.3364) | (2.8302) | (2.2172) | (2.3547) | (2.6290) |
| 2SLS STEM | 0.9857 | 0.9446 | 0.8477 | 0.4050 | 0.5642 | 0.2008 | 0.0874 | 0.1664 | 0.1111 | 0.2530 |
| | (0.8156) | (0.5758) | (0.5673) | (0.5185) | (0.5694) | (0.5220) | (0.5424) | (0.5226) | (0.4920) | (0.5183) |
| 2SLS NonSTEM | 0.7703 | -0.2199 | -0.6910 | -1.4110 | -14.9844 | 18.6419 | 11.4995 | 23.7250 | 12.0759 | 15.1267 |
| | (2.6095) | (1.2261) | (1.2331) | (2.2413) | (86.5275) | (209.0948) | (72.7483) | (340.7319) | (104.8641) | (138.5299) |
| All Instruments | -0.3718 | -0.3571 | -0.2412 | -0.1083 | -0.0435 | 0.3060 | 0.1054 | -0.0013 | -0.1879 | -0.1843 |
| | (0.2835) | (0.1929) | (0.1993) | (0.2453) | (0.2849) | (0.2948) | (0.2910) | (0.2810) | (0.2757) | (0.3025) |
| LIML | -0.4059 | -0.3920 | -0.2706 | -0.1225 | -0.0573 | 0.3578 | 0.1307 | 0.0025 | -0.1920 | -0.1861 |
| | (0.3066) | (0.2061) | (0.2150) | (0.2661) | (0.3225) | (0.3375) | (0.3339) | (0.3161) | (0.3103) | (0.3369) |
| Observations | 41,854 | 39,678 | 37,468 | 35,258 | 32,980 | 30,838 | 28,628 | 26,418 | 24,140 | 21,998 |
| Boot. P Value for 2SLS | 0.472 | 0.647 | 0.903 | 0.673 | 0.253 | 0.380 | 0.325 | 0.360 | 0.397 | 0.365 |
| Boot. P Value for 2SLS STEM | 0.264 | 0.159 | 0.206 | 0.505 | 0.366 | 0.716 | 0.880 | 0.766 | 0.831 | 0.650 |
| Boot. P Value for 2SLS NonSTEM | 0.700 | 0.838 | 0.448 | 0.316 | 0.757 | 0.872 | 0.780 | 0.902 | 0.845 | 0.858 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 3.418 | 6.077 | 4.322 | 2.385 | 0.979 | 0.711 | 0.634 | 0.745 | 0.603 | 0.555 |
| First-Stage F, STEM | 10.320 | 21.753 | 30.408 | 28.710 | 26.026 | 25.466 | 26.645 | 25.445 | 23.231 | 21.908 |
| First-Stage F, Non-STEM | 1.966 | 3.648 | 2.019 | 0.631 | 0.027 | 0.008 | 0.026 | 0.005 | 0.014 | 0.012 |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the NonSTEM row used the other 76.471%.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 24: Change in the Share of College-Bound High School Graduates who Choose *m* as their First Major, using their College Location for the Relevant Employment Variable

| | (1) 1-Yr. Change | (2) 2-Yr. Change | (3) 3-Yr. Change | (4) 4-Yr. Change | (5) 5-Yr. Change | (6) 6-Yr. Change | (7) 7-Yr. Change | (8) 8-Yr. Change | (9) 9-Yr. Change | (10) 10-Yr. Change |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | 0.0166 | 0.0850 | 0.2260 | 0.2067 | 0.3125 | 0.3596 | 0.4104* | 0.5064* | 0.5929* | 0.7534* |
| | (0.1797) | (0.1604) | (0.2305) | (0.1865) | (0.1864) | (0.1908) | (0.2007) | (0.2511) | (0.2982) | (0.3698) |
| 2SLS | 0.9349 | 0.3390 | 0.3468 | 0.6979 | 0.7580 | 1.4553 | 1.4484 | 1.4217 | 1.6687 | 1.5656 |
| | (0.8007) | (0.7372) | (0.6438) | (0.6783) | (0.8887) | (1.1344) | (1.1532) | (1.2609) | (1.4515) | (1.4621) |
| 2SLS STEM | -0.1009 | -0.1364 | 0.1330 | 0.3320 | 0.4834 | 0.8242* | 0.8252* | 0.7894 | 0.7172 | 0.8096 |
| | (0.1990) | (0.2864) | (0.2443) | (0.2596) | (0.3141) | (0.4116) | (0.3955) | (0.4435) | (0.4471) | (0.4620) |
| 2SLS NonSTEM | 1.5327 | 0.6141 | 0.4417 | 0.9157 | 0.9891 | 2.0119 | 2.0936 | 2.1247 | 2.6138 | 2.5034 |
| | (1.2495) | (1.0727) | (0.9203) | (1.0317) | (1.3672) | (1.9164) | (2.0248) | (2.2465) | (2.7196) | (2.7724) |
| All Instruments | -0.0023 | 0.1067 | 0.0639 | 0.3049 | 0.3504 | 0.4696 | 0.4964 | 0.1550 | 0.3113 | 0.5973* |
| | (0.2820) | (0.2851) | (0.3026) | (0.2677) | (0.3201) | (0.3205) | (0.2559) | (0.2233) | (0.2382) | (0.2678) |
| LIML | -0.0065 | 0.1103 | 0.0425 | 0.3186 | 0.3580 | 0.5055 | 0.5178 | 0.0898 | 0.2637 | 0.5676 |
| | (0.3353) | (0.3247) | (0.3347) | (0.2985) | (0.3766) | (0.4167) | (0.3153) | (0.2545) | (0.2676) | (0.3072) |
| Observations | 23,222 | 21,930 | 20,638 | 19,346 | 18,054 | 16,762 | 15,470 | 14,178 | 12,886 | 11,594 |
| Boot. P Value for 2SLS | 0.273 | 0.696 | 0.622 | 0.321 | 0.453 | 0.146 | 0.164 | 0.253 | 0.221 | 0.254 |
| Boot. P Value for 2SLS STEM | 0.650 | 0.728 | 0.597 | 0.299 | 0.265 | 0.235 | 0.243 | 0.295 | 0.329 | 0.252 |
| Boot. P Value for 2SLS NonSTEM | 0.239 | 0.646 | 0.668 | 0.392 | 0.523 | 0.281 | 0.301 | 0.363 | 0.309 | 0.329 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 22.888 | 27.281 | 20.726 | 15.975 | 13.181 | 12.072 | 11.891 | 9.189 | 7.379 | 6.775 |
| First-Stage F, STEM | 16.898 | 27.530 | 50.466 | 41.313 | 27.452 | 21.441 | 22.734 | 21.527 | 19.396 | 15.925 |
| First-Stage F, Non-STEM | 18.997 | 19.515 | 12.503 | 9.912 | 8.675 | 7.754 | 7.361 | 5.438 | 4.012 | 3.727 |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the NonSTEM row used the other 76.471%.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 25: Change in the Share of All High School Graduates who Attend College and Choose *m* as their First Major

| | (1)<br>1-Yr. Change | (2)<br>2-Yr. Change | (3)<br>3-Yr. Change | (4)<br>4-Yr. Change | (5)<br>5-Yr. Change | (6)<br>6-Yr. Change | (7)<br>7-Yr. Change | (8)<br>8-Yr. Change | (9)<br>9-Yr. Change | (10)<br>10-Yr. Change |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | -0.0507 | -0.0172 | 0.0371 | 0.0815 | 0.0788 | 0.0895 | 0.0926 | 0.0973 | 0.0799 | 0.1119 |
| | (0.0324) | (0.0331) | (0.0480) | (0.0715) | (0.0934) | (0.1146) | (0.1251) | (0.1350) | (0.1497) | (0.1546) |
| 2SLS | 0.5801 | 0.3256 | 0.2275 | 0.1059 | 0.1618 | 0.0603 | 0.1749 | 0.1542 | 0.2499 | 0.1506 |
| | (0.4475) | (0.2451) | (0.2841) | (0.4064) | (0.7112) | (0.9222) | (1.0268) | (0.9513) | (1.1092) | (1.1548) |
| 2SLS STEM | 0.3195$^{*}$ | 0.3005$^{**}$ | 0.2974$^{**}$ | 0.2438$^{**}$ | 0.2500$^{**}$ | 0.2156$^{**}$ | 0.2150$^{**}$ | 0.2033$^{**}$ | 0.1973$^{**}$ | 0.2021$^{*}$ |
| | (0.1530) | (0.1060) | (0.0942) | (0.0899) | (0.0863) | (0.0772) | (0.0749) | (0.0736) | (0.0761) | (0.0802) |
| 2SLS NonSTEM | 0.8184 | 0.3622 | 0.1903 | -0.0610 | -0.2555 | 1.3111 | 0.0136 | -0.1143 | -0.7051 | -0.0486 |
| | (0.8138) | (0.4130) | (0.5090) | (1.0038) | (6.0592) | (13.2155) | (5.8091) | (11.8791) | (10.0951) | (8.5038) |
| All Instruments | -0.0387 | 0.0870 | 0.0531 | 0.1285 | 0.1874 | 0.1632 | 0.0764 | 0.0882 | 0.0717 | 0.0971 |
| | (0.0724) | (0.0635) | (0.0815) | (0.0915) | (0.1182) | (0.1428) | (0.1278) | (0.1203) | (0.1098) | (0.1143) |
| LIML | -0.0362 | 0.1055 | 0.0556 | 0.1354 | 0.2104 | 0.1802 | 0.0730 | 0.0865 | 0.0705 | 0.0948 |
| | (0.0852) | (0.0721) | (0.0920) | (0.1028) | (0.1395) | (0.1721) | (0.1513) | (0.1391) | (0.1241) | (0.1292) |
| Observations | 43,282 | 41,004 | 38,726 | 36,448 | 34,170 | 31,892 | 29,614 | 27,336 | 25,058 | 22,780 |
| Boot. P Value for 2SLS | 0.091 | 0.124 | 0.385 | 0.759 | 0.739 | 0.920 | 0.787 | 0.799 | 0.717 | 0.833 |
| Boot. P Value for 2SLS STEM | 0.046 | 0.008 | 0.005 | 0.018 | 0.008 | 0.015 | 0.018 | 0.021 | 0.025 | 0.026 |
| Boot. P Value for 2SLS NonSTEM | 0.158 | 0.289 | 0.623 | 0.925 | 0.943 | 0.869 | 0.997 | 0.987 | 0.907 | 0.993 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 3.407 | 6.034 | 4.321 | 2.396 | 0.976 | 0.700 | 0.629 | 0.732 | 0.588 | 0.542 |
| First-Stage F, STEM | 10.187 | 21.937 | 32.072 | 31.302 | 28.532 | 26.535 | 27.103 | 25.816 | 23.356 | 22.299 |
| First-Stage F, Non-STEM | 1.944 | 3.587 | 2.000 | 0.626 | 0.024 | 0.012 | 0.032 | 0.008 | 0.020 | 0.017 |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the NonSTEM row used the other 76.471%.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 26: Change in the Share of All High School Graduates who go to Community College and Choose *m* as their First Major

| | (1)<br>1-Yr. Change | (2)<br>2-Yr. Change | (3)<br>3-Yr. Change | (4)<br>4-Yr. Change | (5)<br>5-Yr. Change | (6)<br>6-Yr. Change | (7)<br>7-Yr. Change | (8)<br>8-Yr. Change | (9)<br>9-Yr. Change | (10)<br>10-Yr. Change |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | -0.0582 | -0.0178 | 0.0332 | 0.0815 | 0.0829 | 0.1032 | 0.1131 | 0.1176 | 0.1145 | 0.1495 |
| | (0.0300) | (0.0314) | (0.0462) | (0.0655) | (0.0843) | (0.1017) | (0.1061) | (0.1122) | (0.1239) | (0.1281) |
| 2SLS | 0.5818 | 0.3804 | 0.3567 | 0.3505 | 0.6161 | 0.5555 | 0.6728 | 0.5808 | 0.7602 | 0.7248 |
| | (0.4264) | (0.2470) | (0.2856) | (0.4045) | (0.8024) | (0.9371) | (1.1020) | (0.9670) | (1.2396) | (1.2419) |
| 2SLS STEM | 0.2666 | 0.2393* | 0.2431* | 0.2160* | 0.2081* | 0.1733* | 0.1700* | 0.1491* | 0.1590* | 0.1496* |
| | (0.1460) | (0.1123) | (0.1001) | (0.0958) | (0.0836) | (0.0744) | (0.0681) | (0.0647) | (0.0673) | (0.0692) |
| 2SLS NonSTEM | 0.8641 | 0.5016 | 0.4824 | 0.5951 | 3.7284 | -3.9248 | -3.0563 | -5.9110 | -4.4962 | -4.5404 |
| | (0.7919) | (0.4088) | (0.5152) | (1.0406) | (23.5567) | (37.1018) | (18.6183) | (67.3707) | (33.2901) | (36.7918) |
| All Instruments | 0.0054 | 0.1112* | 0.0575 | 0.1112 | 0.1555 | 0.1515 | 0.0719 | 0.0909 | 0.0839 | 0.0913 |
| | (0.0570) | (0.0540) | (0.0682) | (0.0727) | (0.1023) | (0.1405) | (0.1234) | (0.1174) | (0.0974) | (0.0990) |
| LIML | 0.0171 | 0.1335* | 0.0611 | 0.1152 | 0.1694 | 0.1620 | 0.0642 | 0.0864 | 0.0791 | 0.0822 |
| | (0.0661) | (0.0610) | (0.0764) | (0.0811) | (0.1192) | (0.1678) | (0.1441) | (0.1347) | (0.1106) | (0.1123) |
| Observations | 43,078 | 40,800 | 38,590 | 36,312 | 34,034 | 31,756 | 29,478 | 27,200 | 24,922 | 22,678 |
| Boot. P Value for 2SLS | 0.073 | 0.059 | 0.136 | 0.287 | 0.267 | 0.368 | 0.349 | 0.367 | 0.344 | 0.365 |
| Boot. P Value for 2SLS STEM | 0.053 | 0.029 | 0.017 | 0.027 | 0.021 | 0.033 | 0.025 | 0.043 | 0.037 | 0.054 |
| Boot. P Value for 2SLS NonSTEM | 0.116 | 0.109 | 0.199 | 0.360 | 0.774 | 0.852 | 0.777 | 0.875 | 0.817 | 0.841 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 3.406 | 6.019 | 4.309 | 2.385 | 0.972 | 0.698 | 0.629 | 0.732 | 0.586 | 0.542 |
| First-Stage F, STEM | 10.186 | 21.916 | 32.059 | 31.276 | 28.441 | 26.523 | 27.076 | 25.791 | 23.324 | 22.286 |
| First-Stage F, Non-STEM | 1.943 | 3.573 | 1.989 | 0.619 | 0.023 | 0.012 | 0.032 | 0.008 | 0.020 | 0.017 |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the NonSTEM row used the other 76.471%.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 27: Change in the Share all High School Graduates who Attend a 4-year Public University and Choose *m* as their First Major in College

| | (1)<br>1-Yr. Change | (2)<br>2-Yr. Change | (3)<br>3-Yr. Change | (4)<br>4-Yr. Change | (5)<br>5-Yr. Change | (6)<br>6-Yr. Change | (7)<br>7-Yr. Change | (8)<br>8-Yr. Change | (9)<br>9-Yr. Change | (10)<br>10-Yr. Change |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | 0.0086 | 0.0050 | 0.0044 | 0.0003 | -0.0023 | -0.0122 | -0.0209 | -0.0211 | -0.0356 | -0.0383 |
| | (0.0131) | (0.0112) | (0.0140) | (0.0180) | (0.0216) | (0.0271) | (0.0321) | (0.0351) | (0.0387) | (0.0411) |
| 2SLS | 0.0121 | -0.0607 | -0.1285 | -0.2459 | -0.4608 | -0.4914 | -0.4960 | -0.4250 | -0.4986 | -0.5675 |
| | (0.1225) | (0.0670) | (0.0899) | (0.1919) | (0.5132) | (0.6654) | (0.7104) | (0.5786) | (0.7165) | (0.8323) |
| 2SLS STEM | 0.0540 | 0.0616 | 0.0553 | 0.0282 | 0.0423 | 0.0426 | 0.0460 | 0.0553 | 0.0389 | 0.0539 |
| | (0.0728) | (0.0676) | (0.0650) | (0.0606) | (0.0568) | (0.0542) | (0.0522) | (0.0493) | (0.0462) | (0.0519) |
| 2SLS NonSTEM | -0.0205 | -0.1483 | -0.2890 | -0.6556 | -3.8925 | 5.8565 | 3.1824 | 6.5751 | 4.2172 | 4.9880 |
| | (0.2236) | (0.1408) | (0.2261) | (0.8451) | (24.0671) | (58.0062) | (18.1066) | (80.4135) | (32.4767) | (41.4866) |
| All Instruments | -0.0442 | -0.0242 | -0.0059 | 0.0170 | 0.0305 | 0.0091 | 0.0035 | -0.0026 | -0.0122 | 0.0043 |
| | (0.0533) | (0.0225) | (0.0290) | (0.0354) | (0.0372) | (0.0401) | (0.0429) | (0.0430) | (0.0419) | (0.0433) |
| LIML | -0.0633 | -0.0358 | -0.0101 | 0.0241 | 0.0526 | 0.0288 | 0.0211 | 0.0079 | -0.0009 | 0.0213 |
| | (0.0707) | (0.0311) | (0.0392) | (0.0477) | (0.0578) | (0.0710) | (0.0685) | (0.0631) | (0.0587) | (0.0574) |
| Observations | 41,854 | 39,678 | 37,468 | 35,258 | 32,980 | 30,838 | 28,628 | 26,418 | 24,140 | 21,998 |
| Boot. P Value for 2SLS | 0.919 | 0.335 | 0.103 | 0.120 | 0.211 | 0.273 | 0.278 | 0.268 | 0.283 | 0.288 |
| Boot. P Value for 2SLS STEM | 0.501 | 0.461 | 0.504 | 0.735 | 0.571 | 0.514 | 0.452 | 0.314 | 0.422 | 0.332 |
| Boot. P Value for 2SLS NonSTEM | 0.911 | 0.242 | 0.094 | 0.251 | 0.793 | 0.871 | 0.773 | 0.892 | 0.836 | 0.854 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 3.418 | 6.077 | 4.322 | 2.385 | 0.979 | 0.711 | 0.634 | 0.745 | 0.603 | 0.555 |
| First-Stage F, STEM | 10.159 | 21.877 | 31.767 | 31.061 | 28.417 | 26.530 | 27.083 | 25.757 | 23.323 | 22.340 |
| First-Stage F, Non-STEM | 1.960 | 3.639 | 2.014 | 0.627 | 0.025 | 0.010 | 0.030 | 0.006 | 0.016 | 0.014 |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the NonSTEM row used the other 76.471%.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 28: Change in the Share of College Graduates Choosing $m$ as their Final Major in College, with labor market conditions measured in the Senior Year of High School in the Labor Market of the High School

| | (1) 1-Yr. Change | (2) 2-Yr. Change | (3) 3-Yr. Change | (4) 4-Yr. Change | (5) 5-Yr. Change | (6) 6-Yr. Change | (7) 7-Yr. Change | (8) 8-Yr. Change | (9) 9-Yr. Change | (10) 10-Yr. Change |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | 0.1223 | 0.1317 | 0.1384* | 0.1768** | 0.1918** | 0.2570** | 0.2771** | 0.3024** | 0.2996** | 0.2872* |
| | (0.0911) | (0.0693) | (0.0685) | (0.0671) | (0.0723) | (0.0878) | (0.0969) | (0.1023) | (0.1129) | (0.1291) |
| 2SLS | -0.2541 | -0.1949 | -0.2615 | -0.7927 | -1.3232 | -0.9348 | -0.1569 | 0.0561 | 0.0410 | 0.0332 |
| | (1.8075) | (0.6499) | (0.6318) | (1.1593) | (2.6607) | (1.8148) | (0.9487) | (0.5678) | (0.5995) | (0.6927) |
| 2SLS STEM | 1.7964 | 0.6342 | 0.2523 | 0.3523* | 0.3362 | 0.2962* | 0.3049* | 0.2683 | 0.2129 | 0.3337* |
| | (1.4170) | (0.4082) | (0.2292) | (0.1689) | (0.1769) | (0.1485) | (0.1460) | (0.1377) | (0.1443) | (0.1661) |
| 2SLS NonSTEM | -2.9250 | -0.8122 | -0.7735 | -3.2027 | 11.3298 | 12.3443 | 98.3013 | -0.8434 | -0.7358 | -1.3981 |
| | (6.5815) | (1.2271) | (1.4327) | (7.1581) | (81.0752) | (113.3491) | (16679.7711) | (4.2624) | (4.2830) | (8.0831) |
| All Instruments | 0.2728 | 0.1051 | 0.1012 | 0.1049 | 0.2069 | 0.4502** | 0.3641* | 0.1894 | 0.1433 | 0.2166 |
| | (0.1660) | (0.1177) | (0.1209) | (0.1061) | (0.1496) | (0.1714) | (0.1764) | (0.1555) | (0.1574) | (0.1521) |
| LIML | 0.2978 | 0.1011 | 0.0967 | 0.0944 | 0.2102 | 0.4871* | 0.3764 | 0.1753 | 0.1204 | 0.2076 |
| | (0.1921) | (0.1330) | (0.1334) | (0.1195) | (0.1785) | (0.2013) | (0.1981) | (0.1726) | (0.1776) | (0.1690) |
| Observations | 34,102 | 31,824 | 29,546 | 27,268 | 24,990 | 22,746 | 20,468 | 18,190 | 15,912 | 13,634 |
| Boot. P Value for 2SLS | 0.840 | 0.736 | 0.664 | 0.408 | 0.473 | 0.425 | 0.792 | 0.884 | 0.923 | 0.936 |
| Boot. P Value for 2SLS STEM | 0.111 | 0.117 | 0.297 | 0.068 | 0.102 | 0.085 | 0.072 | 0.081 | 0.180 | 0.096 |
| Boot. P Value for 2SLS NonSTEM | 0.421 | 0.384 | 0.468 | 0.503 | 0.826 | 0.842 | 0.991 | 0.721 | 0.759 | 0.762 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 0.772 | 3.497 | 3.074 | 1.513 | 0.460 | 0.593 | 0.787 | 1.523 | 1.259 | 0.977 |
| First-Stage F, STEM | 3.267 | 13.712 | 23.599 | 22.327 | 21.490 | 23.479 | 26.945 | 28.478 | 24.893 | 22.191 |
| First-Stage F, Non-STEM | 0.260 | 1.897 | 1.278 | 0.262 | 0.018 | 0.011 | 0.000 | 0.113 | 0.096 | 0.056 |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the Non-STEM row used the other 76.471%.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 29: Change in the Share of College Graduates Choosing $m$ as their Final Major in College, with Labor Market Conditions Measured at the Year of College Graduation in the Labor Market of the High School

| | (1)<br>1-Yr. Change | (2)<br>2-Yr. Change | (3)<br>3-Yr. Change | (4)<br>4-Yr. Change | (5)<br>5-Yr. Change | (6)<br>6-Yr. Change | (7)<br>7-Yr. Change | (8)<br>8-Yr. Change | (9)<br>9-Yr. Change | (10)<br>10-Yr. Change |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | 0.2145 | 0.1663 | 0.3605** | 0.3543** | 0.3948* | 0.3702 | 0.4208 | 0.3276 | 0.2740 | 0.2543 |
| | (0.1233) | (0.1067) | (0.1280) | (0.1325) | (0.1711) | (0.2148) | (0.2877) | (0.3794) | (0.4102) | (0.4315) |
| 2SLS | 0.2078 | 0.2176 | 0.4830 | 0.7759 | 1.1781 | 1.4049 | 0.6893 | -0.6419 | -3.0709 | 5.0134 |
| | (0.4252) | (0.2515) | (0.3422) | (0.5142) | (1.1305) | (2.0767) | (1.6715) | (2.7116) | (8.2941) | (9.4967) |
| 2SLS STEM | -0.0831 | -0.0772 | 0.4943* | 1.0337*** | 1.1297** | 1.0643** | 0.8252** | 0.4844 | 0.2682 | 0.2694 |
| | (0.3106) | (0.2314) | (0.2372) | (0.3060) | (0.4120) | (0.3994) | (0.2833) | (0.2477) | (0.2445) | (0.3470) |
| 2SLS NonSTEM | 0.3949 | 0.3344 | 0.4015 | 0.4698 | 0.9372 | 40.2345 | 1.8805 | 2.6526 | 3.7308 | 2.2475* |
| | (0.6756) | (0.3849) | (0.4947) | (0.8293) | (2.3578) | (8927.6760) | (10.8068) | (5.7570) | (5.2224) | (1.1193) |
| All Instruments | 0.4730* | 0.4529** | 0.4113** | 0.4135** | 0.2733 | 0.3865 | 0.6698* | 0.7273** | 0.7492** | 0.9097*** |
| | (0.1894) | (0.1381) | (0.1585) | (0.1367) | (0.2362) | (0.3469) | (0.2836) | (0.2678) | (0.2381) | (0.2499) |
| LIML | 0.5074* | 0.4822** | 0.4171* | 0.4211** | 0.2535 | 0.3895 | 0.7020* | 0.7699** | 0.7952** | 0.9714*** |
| | (0.2156) | (0.1523) | (0.1742) | (0.1519) | (0.2708) | (0.4064) | (0.3167) | (0.2948) | (0.2585) | (0.2692) |
| Observations | 31,892 | 29,614 | 27,336 | 25,058 | 22,780 | 20,502 | 18,224 | 15,946 | 13,668 | 11,390 |
| Boot. P Value for 2SLS | 0.630 | 0.370 | 0.133 | 0.072 | 0.169 | 0.310 | 0.510 | 0.676 | 0.545 | 0.427 |
| Boot. P Value for 2SLS STEM | 0.806 | 0.753 | 0.040 | 0.004 | 0.016 | 0.016 | 0.010 | 0.094 | 0.326 | 0.529 |
| Boot. P Value for 2SLS NonSTEM | 0.549 | 0.366 | 0.399 | 0.528 | 0.534 | 0.994 | 0.750 | 0.446 | 0.319 | 0.059 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 10.030 | 17.565 | 11.372 | 4.971 | 1.623 | 0.517 | 0.494 | 0.316 | 0.190 | 0.220 |
| First-Stage F, STEM | 17.130 | 27.479 | 26.843 | 22.915 | 18.943 | 22.156 | 28.782 | 23.576 | 17.721 | 11.847 |
| First-Stage F, Non-STEM | 7.870 | 15.099 | 8.840 | 2.884 | 0.492 | 0.000 | 0.033 | 0.174 | 0.371 | 3.011 |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the Non-STEM row used the other 76.471%.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 30: Change in the Share of College Graduates Choosing *m* as their Final Major in College, with Labor Market Conditions Measured at the Year of College Graduation in the Labor Market of the College

| | (1) 1-Yr. Change | (2) 2-Yr. Change | (3) 3-Yr. Change | (4) 4-Yr. Change | (5) 5-Yr. Change | (6) 6-Yr. Change | (7) 7-Yr. Change | (8) 8-Yr. Change | (9) 9-Yr. Change | (10) 10-Yr. Change |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | 0.2145 | 0.1663 | 0.3605$^{**}$ | 0.3543$^{**}$ | 0.3948$^{*}$ | 0.3702 | 0.4208 | 0.3276 | 0.2740 | 0.2543 |
| | (0.1233) | (0.1067) | (0.1280) | (0.1325) | (0.1711) | (0.2148) | (0.2877) | (0.3794) | (0.4102) | (0.4315) |
| 2SLS | 0.2078 | 0.2176 | 0.4830 | 0.7759 | 1.1781 | 1.4049 | 0.6893 | -0.6419 | -3.0709 | 5.0134 |
| | (0.4252) | (0.2515) | (0.3422) | (0.5142) | (1.1305) | (2.0767) | (1.6715) | (2.7116) | (8.2941) | (9.4967) |
| 2SLS STEM | -0.0831 | -0.0772 | 0.4943$^{*}$ | 1.0337$^{***}$ | 1.1297$^{**}$ | 1.0643$^{**}$ | 0.8252$^{**}$ | 0.4844 | 0.2682 | 0.2694 |
| | (0.3106) | (0.2314) | (0.2372) | (0.3060) | (0.4120) | (0.3994) | (0.2833) | (0.2477) | (0.2445) | (0.3470) |
| 2SLS NonSTEM | 0.3949 | 0.3344 | 0.4015 | 0.4698 | 0.9372 | 40.2345 | 1.8805 | 2.6526 | 3.7308 | 2.2475$^{*}$ |
| | (0.6756) | (0.3849) | (0.4947) | (0.8293) | (2.3578) | (8927.6760) | (10.8068) | (5.7570) | (5.2224) | (1.1193) |
| All Instruments | 0.4730$^{*}$ | 0.4529$^{**}$ | 0.4113$^{**}$ | 0.4135$^{**}$ | 0.2733 | 0.3865 | 0.6698$^{*}$ | 0.7273$^{**}$ | 0.7492$^{**}$ | 0.9097$^{***}$ |
| | (0.1894) | (0.1381) | (0.1585) | (0.1367) | (0.2362) | (0.3469) | (0.2836) | (0.2678) | (0.2381) | (0.2499) |
| LIML | 0.5074$^{*}$ | 0.4822$^{**}$ | 0.4171$^{*}$ | 0.4211$^{**}$ | 0.2535 | 0.3895 | 0.7020$^{*}$ | 0.7699$^{**}$ | 0.7952$^{**}$ | 0.9714$^{***}$ |
| | (0.2156) | (0.1523) | (0.1742) | (0.1519) | (0.2708) | (0.4064) | (0.3167) | (0.2948) | (0.2585) | (0.2692) |
| Observations | 31,892 | 29,614 | 27,336 | 25,058 | 22,780 | 20,502 | 18,224 | 15,946 | 13,668 | 11,390 |
| Boot. P Value for 2SLS | 0.630 | 0.370 | 0.133 | 0.072 | 0.169 | 0.310 | 0.510 | 0.676 | 0.545 | 0.427 |
| Boot. P Value for 2SLS STEM | 0.806 | 0.753 | 0.040 | 0.004 | 0.016 | 0.016 | 0.010 | 0.094 | 0.326 | 0.529 |
| Boot. P Value for 2SLS NonSTEM | 0.549 | 0.366 | 0.399 | 0.528 | 0.534 | 0.994 | 0.750 | 0.446 | 0.319 | 0.059 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 10.030 | 17.565 | 11.372 | 4.971 | 1.623 | 0.517 | 0.494 | 0.316 | 0.190 | 0.220 |
| First-Stage F, STEM | 17.130 | 27.479 | 26.843 | 22.915 | 18.943 | 22.156 | 28.782 | 23.576 | 17.721 | 11.847 |
| First-Stage F, Non-STEM | 7.870 | 15.099 | 8.840 | 2.884 | 0.492 | 0.000 | 0.033 | 0.174 | 0.371 | 3.011 |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the Non-STEM row used the other 76.471%.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 31: Percentage change in Relevant Wages

| | (1)<br>1-Yr. Change | (2)<br>2-Yr. Change | (3)<br>3-Yr. Change | (4)<br>4-Yr. Change | (5)<br>5-Yr. Change | (6)<br>6-Yr. Change | (7)<br>7-Yr. Change | (8)<br>8-Yr. Change | (9)<br>9-Yr. Change | (10)<br>10-Yr. Change |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | 2.5849* | 2.1112** | 1.8039** | 1.6620** | 1.6809** | 1.5914* | 1.5526* | 1.4681* | 1.6074** | 1.7173** |
| | (1.0109) | (0.7802) | (0.6459) | (0.6130) | (0.6075) | (0.6262) | (0.6171) | (0.5845) | (0.5984) | (0.6393) |
| 2SLS | 1.4798 | 4.4736 | 3.9071 | 6.1818 | 13.5725 | 20.8868 | 20.6469 | 17.5381 | 15.1137 | 20.3676 |
| | (4.5737) | (3.1330) | (3.1099) | (5.0250) | (14.0753) | (33.6415) | (31.7914) | (22.2037) | (18.0723) | (31.1140) |
| 2SLS STEM | -6.8341 | 5.2375* | 6.3739** | 7.2805*** | 13.4750*** | 14.5189*** | 15.5789*** | 16.3991*** | 14.9479*** | 15.9682*** |
| | (3.7781) | (2.2587) | (2.0275) | (1.9690) | (2.1222) | (2.1652) | (2.1320) | (2.0423) | (1.9451) | (2.1182) |
| 2SLS NonSTEM | 8.3187 | 3.4740 | 1.0981 | 2.6421 | 5.6149 | 3.8840 | 7.0717 | 14.6670 | 19.0834 | 8.7188 |
| | (8.9667) | (4.4242) | (4.3794) | (9.8271) | (81.5334) | (23.8056) | (29.5119) | (76.8673) | (134.7628) | (39.5854) |
| All Instruments | -0.2995 | 0.5759 | -0.2570 | 0.3760 | 1.4377 | 3.0182* | 2.6194* | 2.5600* | 2.3002* | 2.7180** |
| | (0.8982) | (0.9949) | (1.0158) | (0.9783) | (1.2071) | (1.3052) | (1.2032) | (1.1079) | (1.0672) | (0.9843) |
| LIML | -2.1699 | -1.0251 | -2.4140 | -1.6175 | 0.6021 | 11.3513 | 9.2342 | 8.5881 | 6.0731 | 9.4969 |
| | (1.5337) | (2.0050) | (2.0908) | (2.5527) | (4.9766) | (13.8391) | (11.6797) | (9.5832) | (7.3273) | (12.1031) |
| Observations | 45,492 | 45,492 | 43,214 | 40,936 | 38,658 | 36,414 | 34,136 | 31,858 | 29,580 | 27,302 |
| Boot. P Value for 2SLS | 0.735 | 0.149 | 0.153 | 0.120 | 0.147 | 0.300 | 0.284 | 0.197 | 0.181 | 0.277 |
| Boot. P Value for 2SLS STEM | 0.096 | 0.052 | 0.009 | 0.006 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Boot. P Value for 2SLS NonSTEM | 0.259 | 0.336 | 0.760 | 0.661 | 0.902 | 0.779 | 0.681 | 0.746 | 0.810 | 0.712 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 3.284 | 6.675 | 4.749 | 2.709 | 1.035 | 0.392 | 0.422 | 0.624 | 0.723 | 0.419 |
| First-Stage F, STEM | 10.268 | 24.152 | 34.067 | 44.935 | 34.153 | 32.501 | 31.271 | 28.072 | 27.950 | 24.248 |
| First-Stage F, Non-STEM | 1.831 | 4.042 | 2.205 | 0.623 | 0.016 | 0.145 | 0.161 | 0.062 | 0.030 | 0.114 |
| Clustering | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 23.529% of the observations from the 2SLS row, and the NonSTEM row used the other 76.471%.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

## C Additional Details on the Model and Model Estimation

### C.1 Comparing Student Beliefs to Rational Expectations

In our model, equations (8) and (9) constitute limitations on the ability of students to predict future wages and STEM employment shares. They are constrained to be linear predictions, and only make use of the current STEM employment share. An extreme alternative to these limitations would be to expand the student's information set to include all the local labor market characteristics a student could potentially access as predictors in their forecast, or even to assume that students could somehow predict the future perfectly.

The main problem with each of these alternatives is that they seem highly unrealistic, particularly in our context when the initial and most interesting decisions of the model occur while a student is still in high school. Were it not for the evidence presented above in section 4 that these students react to local employment shares, it might have been doubted whether even local employment shares would factor into high school decision making. As it is, we allow students access to the key components that were shown to be relevant in section 4, namely, the local employment share and the implications of that employment share for wages.

A second problem with including additional variables in the student's set of information is that it quickly makes the model intractable. For example, allowing past realized $w_{mt}$ terms to inform predictions about future ones would add four continuous variables to the state space, one for each of the possible degree types $m$ within the model. In practice, some restriction must be made, and our restriction is to assume that students observe only $SE_t$.

The two problems mentioned here often lead researchers to apply similar restrictions even in models where the model specifies the path of these future values to be something different than what the students think. Buchinsky and Leslie (2010) is an example of researcher's responding to the first problem, that it seems unlikely that students would have perfectly rational expectations. Instead, they provide a setup in which agents estimate a VAR model for predicting future wage distributions. Krusell and Smith (1998) propose a solution to the second problem, the "curse of dimensionality," by giving modeled agents reduced form expectations and arguing that reduced form expressions of this kind can be made sufficiently accurate to capture essentially perfect rational expectations. Our justification is closer to that of Buchinsky and Leslie (2010). However, we intend to estimate a general equilibrium version of this model, which would allow us to check the accuracy of student predictions and see how well the justification of Krusell and Smith (1998) could fit our case as well.

### C.2 Finding the Measurement Parameters

In this section we describe how we identify the $\mu$ and $\lambda$ parameters from equation (18). We start with the measures from the NLSY97, and make the normalization that $E\left(\log\left(A_{img}\right)\right) = 0$ and $\text{Var}\left(\log\left(A_{img}\right)\right) = 1$ for both ability types $m$ in grade 9. Then, we can identify the $\mu^r$ terms for all measures present in grade 9 like this:

$$\mu_m^r = E\left(Z_{img}^r\right) \tag{35}$$

For the $\lambda^r$ terms, we can use the three measures for STEM ability to identify each of the STEM lambda terms for the measures present in ninth grade:

$$\lambda_s^r = \sqrt{\frac{\hat{\text{Cov}}\left(Z_{ist}^r, Z_{isg}^{r'}\right)\hat{\text{Cov}}\left(Z_{isg}^r, Z_{isg}^{r''}\right)}{\hat{\text{Cov}}\left(Z_{isg}^{r'}, Z_{isg}^{r''}\right)}} \tag{36}$$

Since we have more than three measures of STEM ability, this is over-identified. To deal with this, we calculate each $\lambda_s^r$ parameter separately with all of the possibilities for the other two measures $r'$ and $r''$ to include, and use the average as our final estimate of $\lambda_s^r$.

For the non-STEM ability, pick a STEM ability measure from the ninth grade $Z_{isg}^r$. Then,

$$\lambda_n^r = \sqrt{\frac{\hat{\text{Cov}}\left(Z_{isg}^r, Z_{ing}^{r'}\right)\hat{\text{Cov}}\left(Z_{ing}^{r'}, Z_{ing}^{r''}\right)}{\hat{\text{Cov}}\left(Z_{isg}^r, Z_{ing}^{r''}\right)}} \tag{37}$$

Here, the STEM ability measure $Z_{isg}^r$ was arbitrarily chosen, and so we will estimate $\lambda_n^r$ with each of the three possible $Z_{is}^r$ measures and take the average. Not that this equation assumes that the covariance between STEM and non-STEM ability is non-zero.

Once the measurement parameters for the NLSY97 have been found, we use the SAT measurement parameters to identify the expected latent ability and the variance of that latent ability for students in our TSP sample that took the SAT. Then, in each case matching the relevant students from our TSP sample to a sample from the NLSY97 using weights from (19), we back-solve the average latent abilities and the variance of those in the TSP by assuming that these two terms grow at the same rate between grade levels for the TSP sample as they do for the NLSY97 sample. For 11th grade students, we also need to make the additional assumption that the correlation between STEM and Non-STEM ability is the same in the TSP as it is in the NLSY97.

With the movement of aggregate abilities described in this way, we have everything that we need to identify the measurement parameters for the TAAS standardized tests even though no age-invariant measurement is present in the TSP data. For the STEM measure in 11th grade, take the TAAS Math test as a measure $r$, and then the two Non-STEM TAAS measures for that year $r'$ and $r''$.

$$\lambda_s^r = \frac{1}{\operatorname{Corr}\left(A_{isg}, A_{ing}\right)\operatorname{Var}\left(A_{isg}\right)}\sqrt{\frac{\operatorname{Cov}\left(Z^r, Z^{r'}\right)\operatorname{Cov}\left(Z^r, Z^{r''}\right)}{\operatorname{Cov}\left(Z^{r'}, Z^{r''}\right)}} \tag{38}$$

For all other cases, there are at least two TAAS measures present by grade level for each kind of ability. We follow this procedure to identify their lambda terms: Take a TAAS measure $r$ for ability type $c$, which here could be either STEM or Non-STEM. In order to identify $\lambda_c^r$, we need two other measures from the same grade level as $r$. If there exist two other measures with the same type $c$, then we will use those and label them $r'$ and $r''$. Otherwise, we will use the other measure of type $c$ as $r'$, and take the last measure $r''$ from the other type.[76] Then, use this formula:

$$\lambda_c^r = \frac{1}{\operatorname{Var}\left(A_{ict}\right)}\sqrt{\frac{\operatorname{Cov}\left(Z^r, Z^{r'}\right)\operatorname{Cov}\left(Z^r, Z^{r''}\right)}{\operatorname{Cov}\left(Z^{r'}, Z^{r''}\right)}} \tag{39}$$

Once we have found the lambda terms, the process for finding the $\mu_m^r$ terms is the same for all cases:

$$\mu_c^r = E\left(Z_t^r\right) - \lambda_c^r E\left(A_{ict}\right) \tag{40}$$

Table 32 provides all of the resulting estimates for the measurement parameters.

The approach above relies upon an assumption that the growth rates in expected ability and the variance of ability are the same between a matched sample of the NLSY97 and the TSP. Given the focus of our model on how students react to local employment conditions, one potential criticism of this approach is that it ignores how these growth rates depend upon student reactions to aggregate movements in employment. For example, if there were a substantial Texas-wide growth in the STEM employment share between the time that we take the NLSY97 data and the time of our TSP sample, a student reaction to take more STEM classes may alter the ability growth rate in the TSP sample compared to the NLSY97 sample. Note this concern is specifically about aggregate movement in employment shares across our full Texas sample, and the potential for this to result aggregate changes to ability growth rates between grade levels. Idiosyncratic movements in local employment shares do not carry this same level of concern, because different localities could have rising and falling ability growth while leaving the aggregate Texas wide average unchanged.[77]

We hope to take further steps to address this concern, but for now have two primary responses. First, this underlines the importance of using NLSY97 data and TSP data that are quite close to each other in time. Since we are able to do this, it mitigates the concern that there could develop substantial differences in aggregate employment shares and then aggregate ability growth rates in this time between the samples.

Second, this concern is only valid for changes to high school course taking before a student's senior year. This is because the SAT is typically taken toward the beginning of the senior year, and so our estimates for the upperclassmen $\gamma$ and $\kappa$ terms are formed based on growth that occurs in the junior year.[78] In order to bias our estimates, then, any changes in Texas-wide employment shares would have to affect course-taking in the junior year or earlier. Our estimates on student course-taking

[76] For example, to identify the $\lambda_c^r$ for the TAAS Math test in ninth grade, we would use the TAAS Science test as $r'$ and take the third measure $r''$ from the Non-STEM measures, which include the TAAS Writing, Reading, and Social Studies exams. It is an arbitrary choice which of these Non-STEM measures to use as $r''$, and so in cases like this we calculate $\lambda_m^r$ for all possible choices of $r''$ and take the average. By taking using measures from both STEM and Non-STEM to identify these lambda terms, we are assuming that there is a non-zero covariance between STEM and Non-STEM ability.

[77] Specifically, this intuition rests upon idiosyncratic movements in local employment conditions having linear effects on the growth rate in ability. Say, for example, that $g_\ell = \alpha\varepsilon_\ell$, where $g_\ell$ was the ability growth rate in location $\ell$ and $\varepsilon_\ell$ represented idiosyncratic changes in local employment shares. If $E\left(\varepsilon_\ell\right) = 0$, then the aggregate growth rate $E\left(g_\ell\right)$ will be zero as well, suggesting that the local $\varepsilon_\ell$ terms do not affect the aggregate growth rate.

[78] We then use these estimated terms to predict human capital growth in the senior year as well, assuming that these coefficients across the student's "upperclassman" years in high school.

responses in table 1 specifically referred to classes taken in the senior year, and within our model high school students are only able to alter their course-taking in response to local employment shares in their senior year. Thus, at least through the simplified lens of the model, there is no potential for this sort of concern to bias results. In the future, we can report more direct tests of these restrictions from regressions like those of table 1 to test whether students react to employment conditions in their junior year of high school or earlier. If it is the case that student course-taking reacts to employment shares in those years as well, we can likely then take further action to directly control for bias coming from course-taking by using estimates of how aggregate course-taking would have changed.

Table 32: Measurement Parameters

| Measure | Type | Estimated from | $\lambda_m^r$ | $\lambda_m^r$ Obs. | $\mu_m^r$ | $\mu_m^r$ Obs. | Signal To Noise Ratio | STN Ratio Obs. |
|---|---|---|---|---|---|---|---|---|
| ASVAB Paragraph Comprehension | Non-STEM | NLSY97 | 722.304 | 957 | -197.892 | 957 | .765 | 957 |
| ASVAB Word Knowledge | Non-STEM | NLSY97 | 675.398 | 957 | -508.309 | 957 | .701 | 957 |
| SAT Verbal Score | Non-STEM | NLSY97 | 131.616 | 122 | 328.257 | 140 | .839 | 140 |
| ASVAB Arithmetic Reasoning | STEM | NLSY97 | 760.803 | 958 | -281.826 | 957 | .812 | 957 |
| ASVAB General Science | STEM | NLSY97 | 501.76 | 958 | -336.72 | 958 | .436 | 958 |
| ASVAB Math. Knowledge | STEM | NLSY97 | 756.861 | 958 | 1.643 | 956 | .851 | 956 |
| ASVAB Numerical Operations | STEM | NLSY97 | 3488.942 | 958 | 16746.15 | 950 | .376 | 950 |
| PIAT Math Test | STEM | NLSY97 | 10.122 | 958 | 97.227 | 1,065 | .561 | 1,065 |
| SAT Math Score | STEM | NLSY97 | 127.152 | 122 | 298.096 | 140 | .828 | 140 |
| TAAS Reading, Grade 11 | Non-STEM | TSP | 8.011 | 82,902 | 80.513 | 82,739 | .614 | 82,739 |
| TAAS Reading, Grade 9 | Non-STEM | TSP | 13.486 | 332,595 | 87.906 | 331,051 | .64 | 331,051 |
| TAAS Soc. Studies, Grade 9 | Non-STEM | TSP | 137.328 | 332,595 | 1584.894 | 331,572 | .575 | 331,572 |
| TAAS Writing, Grade 11 | Non-STEM | TSP | 175.502 | 82,902 | 1564.504 | 82,626 | .302 | 82,626 |
| TAAS Writing, Grade 9 | Non-STEM | TSP | 203.622 | 332,595 | 1694.212 | 327,642 | .407 | 327,642 |
| TAAS Math, Grade 11 | STEM | TSP | 7.177 | 82,902 | 75.477 | 82,708 | .608 | 82,708 |
| TAAS Math, Grade 9 | STEM | TSP | 7.681 | 334,503 | 83.44 | 331,473 | .564 | 331,473 |
| TAAS Science, Grade 9 | STEM | TSP | 105.924 | 334,503 | 1643.108 | 330,165 | .582 | 330,165 |

The NLSY97 measurement parameters were found using weights from the TSP Student data. Each of these weights used observations from at least 83,147 students.

Table 33: Groups of Industries for the Model

| Two Digit NAICS Code | Industry Group | Number of Student Observations |
|---|---|---|
| 71 | Non-Stem Intensive | 27164 |
| 62 | Non-Stem Intensive | 396861 |
| 52 | Non-Stem Intensive | 117313 |
| 81 | Non-Stem Intensive | 44947 |
| 61 | Non-Stem Intensive | 393901 |
| 44–45 | Non-Stem Intensive | 212173 |
| 53 | Non-Stem Intensive | 33634 |
| 72 | Non-Stem Intensive | 118011 |
| 56 | Stem-Intensive | 151732 |
| 31–33 | Stem-Intensive | 108878 |
| 51 | Stem-Intensive | 48671 |
| 48–49 | Stem-Intensive | 39592 |
| 55 | Stem-Intensive | 7640 |
| 22 | Stem-Intensive | 10766 |
| 23 | Stem-Intensive | 66540 |
| 21 | Stem-Intensive | 30093 |
| 11 | Stem-Intensive | 5238 |
| 42 | Stem-Intensive | 74668 |
| 54 | Stem-Intensive | 188458 |

Table 34: Groups of Majors for the Model

| Major Type | Label for Major Group | Included Two-Digit CIP codes |
|---|---|---|
| Non-Stem | Non-Stem Degrees | 01,04,05,09,10,12,13,16,19,22,23,24,30,31,38,42,43,44,45,46,47,48,49,50,51,52 |
| Stem | Computer and Bio. Sciences | 11,26 |
| Stem | Engineering | 14,15 |
| Stem | Math, Physical Sci., and Sci. Tech. | 27,40,41 |
| Stem | Nat. Resources and Conservation | 03 |

The variables used to sort these two-digit CIP codes into groups were constructed using information from at least 1,905,531 students.

## C.3 Estimated groups of Majors and Industries

Table 33 displays the industries by two-digit NAICS code that we sort into the STEM and Non-STEM intensive groups. These are largely as would be expected. For example, "Mining, Quarrying, and Oil and Gas Extraction" (21), Utilities (22), Construction (23), Manufacturing (31-33), and "Professional, Scientific, and Technical Services" (54) are all in the "STEM-Intensive" category. On the other hand, "Retail Trade" (44-45), "Educational Services" (61), "Arts, Entertainment, and Recreation" (71), and "Accommodation and Food Services" (72) are all in the Non-STEM category. One surprising categorization is that "Health Case and Social Assistance" was found to be in the Non-STEM category. We had expected this to be in the STEM category, and actually used it as a motivational example for the movement of STEM majors in section 2, which goes to show the importance of doing the full analysis to see what results come up.

Table 34 shows the groups of majors that we will form from the two-digit CIP codes. The STEM groups were chosen by a kmeans algorithm, but largely seem to be sensible groupings of majors. We give each a label to help clarify the sort of majors that belong to each group.[79] Of these, we will use the STEM major groups "Computer and Biological Sciences" and "Engineering" when estimating the full model.

## C.4 Regression Results for Estimating the Structural Model

[79] A fuller description of these two digit CIP codes and the college majors that they encompass is avaliable at www.txhighereddata.org/Interactive/CIP/.

Table 35: Change in the Share of College-Bound High School Graduates Choosing $m$ as their First Major in College, using the Three Groups of College Majors for the Structural Model

| | (2) 2-Yr. Change | (4) 4-Yr. Change | (6) 6-Yr. Change | (8) 8-Yr. Change | (10) 10-Yr. Change |
|---|---|---|---|---|---|
| 2SLS | 0.5365 | -0.6216 | -1.0731 | -1.1737 | -1.2536 |
| | (0.9540) | (1.1294) | (1.2299) | (1.0670) | (1.0887) |
| 2SLS STEM | 0.5539* | 0.3344* | 0.2468 | 0.2646 | 0.2318 |
| | (0.2288) | (0.1660) | (0.1971) | (0.1949) | (0.2213) |
| 2SLS NonSTEM | 0.4249 | -1.3374 | -2.0843 | -2.1386* | -1.9477 |
| | (1.1911) | (1.2559) | (1.2567) | (1.0774) | (1.1653) |
| Observations | 3,618 | 3,216 | 2,814 | 2,412 | 2,010 |
| Boot. P Value for 2SLS | 0.498 | 0.626 | 0.428 | 0.330 | 0.317 |
| Boot. P Value for 2SLS STEM | 0.031 | 0.067 | 0.254 | 0.229 | 0.352 |
| Boot. P Value for 2SLS NonSTEM | 0.580 | 0.441 | 0.159 | 0.091 | 0.177 |
| Covariates | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 23.918 | 24.743 | 20.005 | 19.011 | 16.743 |
| First-Stage F, STEM | 22.211 | 31.146 | 27.857 | 25.512 | 21.967 |
| First-Stage F, Non-STEM | 24.373 | 38.836 | 38.611 | 33.398 | 31.215 |
| Clustering | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 66.667% of the observations from the 2SLS row, and the NonSTEM row used the other 33.333%.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 36: Change in the Share of All High School Graduates Who attend College and choose $m$ as their First Major, using the Three Groups of College Majors for the Structural Model

| | (2) 2-Yr. Change | (4) 4-Yr. Change | (6) 6-Yr. Change | (8) 8-Yr. Change | (10) 10-Yr. Change |
|---|---|---|---|---|---|
| 2SLS | -0.2574 | -0.6527 | -0.5198 | -0.4583 | -0.5995 |
| | (0.4084) | (0.5385) | (0.6467) | (0.5400) | (0.5735) |
| 2SLS STEM | 0.2857** | 0.2017** | 0.1551 | 0.1488 | 0.1261 |
| | (0.0991) | (0.0781) | (0.0925) | (0.0882) | (0.0961) |
| 2SLS NonSTEM | -0.2755 | -0.8627 | -0.6203 | -0.3096 | -0.2917 |
| | (0.6112) | (0.6637) | (0.7722) | (0.6059) | (0.7340) |
| Observations | 3,618 | 3,216 | 2,814 | 2,412 | 2,010 |
| Boot. P Value for 2SLS | 0.569 | 0.304 | 0.459 | 0.438 | 0.273 |
| Boot. P Value for 2SLS STEM | 0.013 | 0.031 | 0.160 | 0.148 | 0.244 |
| Boot. P Value for 2SLS NonSTEM | 0.703 | 0.334 | 0.522 | 0.659 | 0.623 |
| Covariates | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 23.921 | 24.767 | 20.011 | 19.012 | 16.747 |
| First-Stage F, STEM | 22.744 | 33.380 | 28.367 | 25.986 | 22.801 |
| First-Stage F, Non-STEM | 25.042 | 44.723 | 40.787 | 34.494 | 34.113 |
| Clustering | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 66.667% of the observations from the 2SLS row, and the NonSTEM row used the other 33.333%.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 37: Change in the Share of College Graduates Choosing *m* as their Final Major in College, with labor market conditions measured in the senior year of high school, using the Three Groups of College Majors for the Structural Model

| | (2)<br>2-Yr. Change | (4)<br>4-Yr. Change | (6)<br>6-Yr. Change | (8)<br>8-Yr. Change | (10)<br>10-Yr. Change |
|---|---|---|---|---|---|
| 2SLS | 0.2831 | 0.2545 | 0.2502 | 0.2277 | 0.2397 |
| | (0.3328) | (0.2565) | (0.2850) | (0.2566) | (0.2821) |
| 2SLS STEM | 0.5494 | 0.3282$^{*}$ | 0.2538 | 0.3157$^{*}$ | 0.3270 |
| | (0.3819) | (0.1609) | (0.1796) | (0.1591) | (0.1925) |
| 2SLS NonSTEM | 0.0940 | 0.0970 | 0.1023 | 0.0745 | 0.0183 |
| | (0.3532) | (0.2365) | (0.2468) | (0.2384) | (0.2450) |
| Observations | 2,808 | 2,406 | 2,007 | 1,605 | 1,203 |
| Boot. P Value for 2SLS | 0.364 | 0.353 | 0.377 | 0.406 | 0.548 |
| Boot. P Value for 2SLS STEM | 0.165 | 0.069 | 0.196 | 0.069 | 0.136 |
| Boot. P Value for 2SLS NonSTEM | 0.457 | 0.718 | 0.665 | 0.843 | 0.765 |
| Covariates | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 14.725 | 17.688 | 17.297 | 20.599 | 16.100 |
| First-Stage F, STEM | 14.801 | 23.506 | 25.592 | 29.519 | 22.599 |
| First-Stage F, Non-STEM | 16.791 | 33.742 | 41.547 | 37.438 | 30.900 |
| Clustering | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 66.667% of the observations from the 2SLS row, and the Non-STEM row used the other 33.333%.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 38: Change in the Share of College Graduates Choosing *m* as their Final Major in College, with labor market conditions measured at the year of college graduation, using the Three Groups of College Majors for the Structural Model

| | (2)<br>2-Yr. Change | (4)<br>4-Yr. Change | (6)<br>6-Yr. Change | (8)<br>8-Yr. Change | (10)<br>10-Yr. Change |
|---|---|---|---|---|---|
| 2SLS | -0.0393 | 1.0759$^{*}$ | 0.9246$^{*}$ | 0.2695 | 0.2950 |
| | (0.3183) | (0.4742) | (0.4339) | (0.3424) | (0.5784) |
| 2SLS STEM | 0.0547 | 0.9355$^{**}$ | 0.9117$^{**}$ | 0.4365 | 0.0796 |
| | (0.2781) | (0.2987) | (0.3488) | (0.2479) | (0.3464) |
| 2SLS NonSTEM | -0.3321 | 0.4888 | 0.4597 | -0.0727 | -0.3491 |
| | (0.3789) | (0.5436) | (0.4184) | (0.3145) | (0.3825) |
| | (0.2151) | (0.2315) | (0.1918) | (0.1957) | (0.2017) |
| Observations | 2,613 | 2,211 | 1,809 | 1,407 | 1,005 |
| Boot. P Value for 2SLS | 0.952 | 0.046 | 0.050 | 0.492 | 0.985 |
| Boot. P Value for 2SLS STEM | 0.830 | 0.004 | 0.021 | 0.101 | 0.850 |
| Boot. P Value for 2SLS NonSTEM | 0.430 | 0.968 | 0.284 | 0.758 | 0.123 |
| Covariates | Yes | Yes | Yes | Yes | Yes |
| Major-Time FE | Yes | Yes | Yes | Yes | Yes |
| First-Stage F | 29.208 | 25.901 | 21.851 | 21.345 | 11.287 |
| First-Stage F, STEM | 30.111 | 25.887 | 25.208 | 25.343 | 13.246 |
| First-Stage F, Non-STEM | 30.986 | 35.493 | 42.072 | 31.955 | 17.967 |
| Clustering | CZ | CZ | CZ | CZ | CZ |
| Weight | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads | 1998 HS Grads |

Standard errors in parentheses

The STEM row used 66.667% of the observations from the 2SLS row, and the Non-STEM row used the other 33.333%.

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

## C.5 Additional Tables Displaying the Fit of the Model

Table 39: Fit of High School Moments

| Lag: | 2-Year | | 4-Year | | 6-Year | | 8-Year | | 10-Year | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Model | Data | Model | Data | Model | Data | Model | Data | Model | Data |
| HS Classes | -0.058 | -0.045 | 0.224 | 0.564 | 0.383 | 1.053 | 0.406 | 0.947 | 0.568 | 0.955 |
| Credit, STEM Classes | -0.815 | -0.201 | -0.268 | -0.563 | -0.065 | -0.868 | -0.150 | -0.704 | -0.313 | -0.669 |
| Credit, Non-STEM Classes | 0.218 | 0.142 | 0.013 | 0.218 | -0.106 | 0.270 | 0.131 | 0.211 | 0.150 | 0.193 |

Table 40: Fit of Non-STEM College Moments

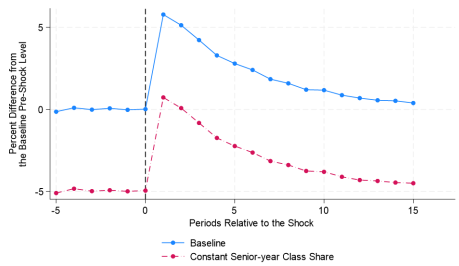
| Lag: | 2-Year | | 4-Year | | 6-Year | | 8-Year | | 10-Year | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Model | Data | Model | Data | Model | Data | Model | Data | Model | Data |
| First Major Choice, Intensive | 0.587 | 0.425 | 0.514 | -1.337 | 0.546 | -2.084 | 0.532 | -2.139 | 0.571 | -1.948 |
| First Major Choice, Extensive | 0.215 | -0.275 | 0.145 | -0.863 | 0.080 | -0.620 | 0.082 | -0.310 | 0.164 | -0.292 |
| Graduation Major, any year | -0.071 | 0.094 | 0.211 | 0.097 | 0.361 | 0.102 | 0.400 | 0.075 | 0.462 | 0.018 |
| Graduation Major, by year | 0.025 | -0.332 | 0.285 | 0.489 | 0.368 | 0.460 | 0.387 | -0.073 | 0.446 | -0.349 |

Table 41: Fit of STEM College Moments

| Lag: | 2-Year | | 4-Year | | 6-Year | | 8-Year | | 10-Year | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Model | Data | Model | Data | Model | Data | Model | Data | Model | Data |
| First Major Choice, Intensive | 0.423 | 0.554 | 0.370 | 0.334 | 0.393 | 0.247 | 0.383 | 0.265 | 0.411 | 0.232 |
| First Major Choice, Extensive | 0.152 | 0.286 | 0.138 | 0.202 | 0.155 | 0.155 | 0.150 | 0.149 | 0.153 | 0.126 |
| Graduation Major, any year | -0.049 | 0.549 | 0.120 | 0.328 | 0.211 | 0.254 | 0.231 | 0.316 | 0.270 | 0.327 |
| Graduation Major, by year | 0.013 | 0.055 | 0.164 | 0.935 | 0.208 | 0.912 | 0.217 | 0.436 | 0.255 | 0.080 |

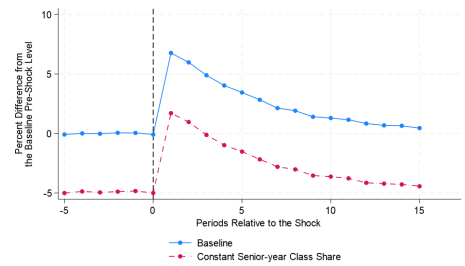
## C.6 Additional Figures for the Counterfactuals

Figure 11: Additional Figures Showing the Results of Removing the Option of a Senior-year STEM-Intensive track, just for Students who Chose the STEM-Intensive Senior Year track in the baseline case.

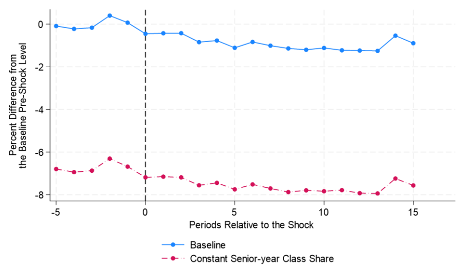



(a) Average Period 5 Earnings


(b) Average Period 5 Earnings, Lowest Tercile of Initial STEM Ability

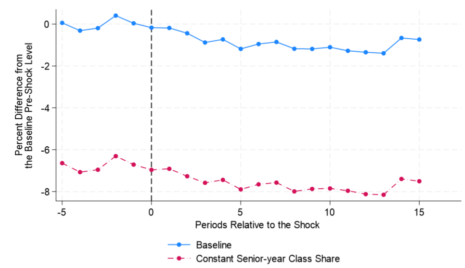



(c) Average Period 5 Earnings, Highest Tercile of Initial STEM Ability



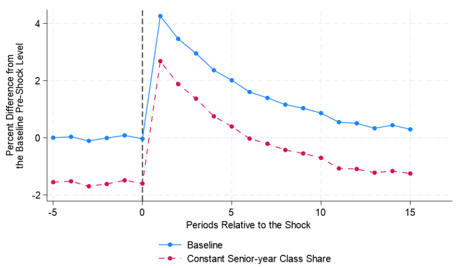

(d) Average Present Value of Lifetime Earnings


(e) Average Present Value of Lifetime Earnings, Lowest Tercile of Initial STEM Ability



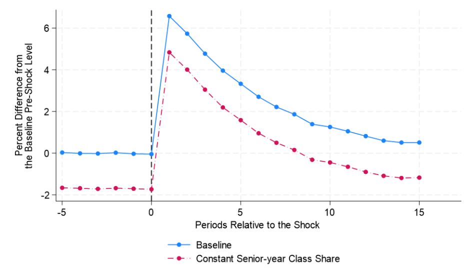

(f) Average Present Value of Lifetime Earnings, Highest Tercile of Initial STEM Ability

Figure 12: Additional Figures Showing the Results of Removing the Option of a Senior-year STEM-Intensive track for All Students, broken down by Tercile of Initial STEM Ability

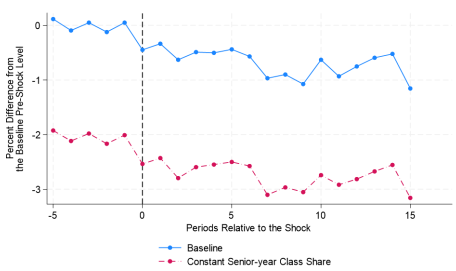



(a) Average Period 5 Earnings, Lowest Tercile of Initial STEM Ability

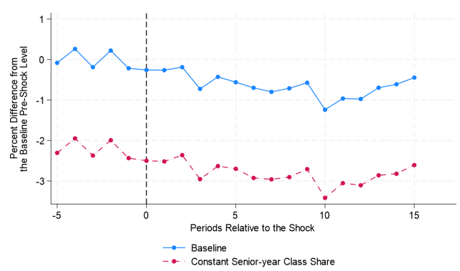

(b) Average Period 5 Earnings, Second Tercile of Initial STEM Ability




(c) Average Period 5 Earnings, Highest Tercile of Initial STEM Ability




(d) Average Present Value of Lifetime Earnings, Lowest Tercile of Initial STEM Ability




(e) Average Present Value of Lifetime Earnings, Second Tercile of Initial STEM Ability


(f) Average Present Value of Lifetime Earnings, Highest Tercile of Initial STEM Ability