Senior Project

Department of Economics

A Decomposition of High School Effects on College STEM Major Choice

Kaleb McAbier

May 10, 2018

Abstract

STEM jobs are a fast-growing demanding field with plenty of opportunity. However, women are a minority in most STEM fields. This issue raises concern of why this is happening and what can be done to improve women participation in STEM. This paper aims to look at how highs school effects women graduating with a STEM major. Utilizing National Longitudinal Survey of Freshman data, this study provides some insight as to why women are underrepresented in STEM and if high school preparations has any effect on this. A linear probability model regression suggests that women seem to have an almost 3% lower chance of declaring and graduating with a STEM major. A decomposition was then used to see the effects social discrimination has on women in STEM, and it was found that out of a gender gap of 5.21%, 2.43% is representative of social discrimination in STEM major choice.

Table of Contents

I. Introduction

For years, it has been widely known that women in the STEM fields (Science, Technology, Engineering, Math) are few and far between (Bottia, 2015). This study will set out to see if underrepresentation of women in these fields is caused by social discrimination. More specifically, if women are discouraged or discriminated against in high school compared to men when choosing a STEM major for college. The exact question this paper sets out to answer is, "Why are women underrepresented in STEM fields prior to choosing a college major". To test this hypothesis, the Oaxaca decomposition will be utilized. First, a linear probability estimation will be used to analyze these effects. The linear probability model is used as a stepping stone toward computing the decomposition.

This study is aimed at providing new insight to why females are lacking in the STEM fields. The National Science Foundation (2009) states that universities and colleges award 40% of their female graduates with STEM major degrees compared to men. This statistic sounds high but most of that 40% is a biology type degree and not a PSEM major (Physical Science Engineering Math) such as physics or engineering. It is important to understand what discourages women from joining the STEM field so new policies can be created that encourage women to join STEM. Also, study of this topic can provide a stepping stone to further lessen the divide between men and women in STEM. Previous research on this topic has been done and the following articles detail many different findings and ways of going about determining the cause for low female participation in STEM, and what can be done to improve this.

II. Literature Review

The underrepresentation of women in STEM is problematic for the rising global importance for STEM occupations. Bottia's (2015) goal was to determine if female students are more inclined to major in STEM fields if there is a stronger presence of female teachers in math and science at the high school level. A sample size of around 350 schools in North Carolina consisting of 21,340 students were surveyed from 2005-2011 to gather data on a variety of variables. Using multilevel multinomial logistic models, this study concludes that graduating females with a biology or PSEM major, increased by 19% and chances of declaring a PSEM major increased by 14% when there was an increase in the female to male teacher ratio from 0.54 to 0.72.

The article by Yingyi Ma (2011) looks at the gender divide in STEM through three pathways of attainment; expected college major, first major declared, and bachelor degree obtained. This paper like that of Bottia (2015), uses longitudinal data from the National Education Longitudinal Study from 1988-2000 to study these effects. Ma (2011) found that women are as persistent as men to stick with STEM degree attainment but also came across the revolving door theory as an explanation as to why women tend to be underrepresented in STEM. The revolving door theory predicts that men in STEM are in power and choose to keep their male cohorts in power in turn "shunning" women from the field. This conclusion goes against most of the other literature is saying that women are as likely as men to graduate in a STEM major and the problem is when women are trying to gain positions in the professional STEM world.

Using longitudinal data from the NLSF and NELS Griffith (2010) looked into institutional effects on female and minority persistence in STEM fields. She found that higher grades during the freshman and sophomore years of a STEM field major increase the odds of that student holding true to their STEM major. However, it was also concluded that institutions with a higher undergraduate to graduate ratio tend to see higher persistence of women and minorities in STEM along with institutions that have higher graduate level minority and female “role models” (Griffith, 2010). These findings ring true with those of Bottia (2015) who found evidence of female “role models” in the high school setting having a positive influence on women in STEM.

A study was conducted at Binghamton University by Kokkelenberg (2010) which set out to test the success and persistence of STEM majors. The results showed that both men and women alike had similar reasons in choosing to drop out of the STEM major or pursue and eventually graduate in STEM. However, this study did not consider the effects of the background of the students tested. This could have skewed the results in a way to show that women at the university were similar to men, since Bottia (2015) shows past experiences and high school preparation has a large impact on women choosing and ultimately staying in the STEM fields.

The final piece of literature to look over covers why women in the STEM pipeline seem to drop out of that pipeline more so than men. Fischer’s (2017) study aimed to see how first experiences in STEM affected students’ persistence in the field. Data from 1997-2007 was collected at a public university for first year college students taking general chemistry. Fischer (2017) narrowed in on seeing how fellow classmate grades or “achievement” in this class affected the students who did not do as well in the class. The results showed no effect for men in persisting in STEM but did show women being effected. These results showed that an increase in the number of on-track students by 15% reduced the probability of women graduating in the

STEM pipeline by 3.1% (Fischer, 2017). Women may be more influenced by their peers' grades or achievement than men, so much so that they choose to switch majors to non-STEM.

III. Theoretical Model

As mentioned in the motivation section a linear probability model regression will be taken out to test whether women are socially discriminated against or not in high school toward choosing a STEM major in college. This is the first step toward utilizing the Oaxaca decomposition which will come after the linear probability model is estimated. The testable hypothesis in this case will be, a decrease in the social discrimination towards women in STEM will increase the STEM participation rate of women. The model I will use is as follows;

$$stem = ß_0 + ß_1 X + ß_2 Female + ß_3 MAbetter + ß_4 SAbetter + ß_5 HSyrscalc + ß_6 HSyrsbio + ß_7 HSyrschem + ß_8 HSyrsphy + ß_9 HSqual + ß_{10} HSrep + ß_{11} Public + ß_{12} HSprepColl + \varepsilon$$

The variable *stem* is the dependent variable and it takes the value of 1 if the student graduated with a STEM major, 0 otherwise. X is the set of exogenous variables such as gender, ethnicity, family background and so on. *Female* is a dummy variable for the gender of the student, 1 being female 0 being male. The expected sign on this variable is negative since previous literature has shown that females have a lower chance of declaring a STEM major in college. *MAbetter* and *SAbetter* are dummy variables representing getting and A or better in math and science classes, respectively. The expected sign for these variables should be positive as a high grade in these subjects should better prepare a student for a STEM major in college. The variables *HSyrscalc*, *HSyrsbio*, *HSyrschem*, and *HSyrsphy* represent the number of years of calculus, biology, chemistry, and physics that were taken in high school. The expected signs for all these variables would be positive since these high school classes should positively impact students' choice in a STEM major. The *HSqual*, *HSrep*, and Public variables are a way to measure the quality of education. The *HSqual* and *HSrep* variables are expected to be positive as a high quality and reputable school should have a positive impact on a better education and thus higher chance of going into STEM. The *Public* variable is expected to be negative since going to a public school opposed to a private school should have a worse impact on going into a STEM major. The *HSprepColl* variable is a measure of how well the student thinks high school has prepared them for college. This variable is expected to be positive since being prepared for college should increase the likelihood of graduating with a degree. The next section explains the data utilized for testing my hypothesis.

The Oaxaca Decomposition formula is below which will determine the effects of social discrimination of high school preparation on STEM major choice for women.

$$\alpha^m_1(\bar{X}^m - \bar{X}^w) + (\alpha^m_0 - \alpha^w_0) + \bar{X}^w(\alpha^m_0 - \alpha^w_0)$$

Table 3 details how this formula was used in determining the gap between men and women and how much of this gap is due to social discrimination.

IV. Data

The data utilized for this study was gathered by Princeton University and is their National Longitudinal Survey of Freshman from 1999-2004. This data set has a multitude of different variables but only the previously mentioned ones will be used for the linear probability model and decomposition. The survey gathers data over 5 years from the same students. The data set records information about the students' background before, during, and after college graduation.

I took the data and merged the five years together to obtain one continuous data set to run the regressions on. In the appendix, the descriptive statistics and definitions/sources of the variables are shown. The next section will explain the results I gathered from both the linear probability model and also the decomposition.

V. Results

The outcome from this initial regression show some promising results. Females have an almost 3% lower chance of choosing a STEM major compared to men. This result lines up with previous research on the topic. Another surprising variable is the years of calculus, it seems to be very insignificant as to describing STEM major achievement. Also, two of the three variables to control for quality of education seem to be insignificant. The rest of the variables such as getting an A or better in both math and science have a large impact on STEM major choice. For an A in math, a STEM major choice increases by 6% and for science, increases almost 10%, which aligns with my theoretic prediction. The years that a person takes biology, chemistry, and physics all have a positive effect on STEM major choice which aligns again with my theoretic prediction. The overall conclusion for these results are nor good nor bad. The results provided crucial insight in starting to explain this issue in the STEM field. The results also provided a path to the decomposition results which I will discuss next.

Now that I had a base to build from for my estimation, I embarked on the decomposition. The overall gap between men and women in the STEM field major choice was 5.21%. This means that 5% fewer females choose this field as a major in college. The next step in the decomposition was to determine how much of this effect was explained and unexplained. The explained portion of the decomposition amounted to 2.78%, that left 2.43% unexplained which can be seen in table 3. This unexplained portion is caused by unobserved characteristics not included in the model which includes social discrimination. The results of the decomposition show that before adjustments for each variable there is a gap between men and women in STEM of almost 16% which aligns with previous literature on the topic. A few notable coefficients to look at are those of the school reputation variable, A or better in math variable, and high school years of biology. To start, the school reputation variable shows that school reputation closes the gap of unexplained STEM differences by 10% which is the largest percent looking at the other variables. This shows that the quality of school may be an important factor when determining the differences between men and women. For the A or better in math variable the gap closed by l% which may show that women are more likely to choose STEM as a major if their grades in math are higher. The most interesting and theory aligning variables is the high school years of biology. Previous literature and the National Science Foundation (2009) have found that most female STEM degrees are awarded to biology majors. The decomposition shows a 7% close in the gap between men and women in STEM which aligns with this fact. In the conclusion I will discuss what these results determined and what the next steps would be in this research.

VI. Conclusion

The results of my analysis show interesting insight pertaining to my question of “Why are women underrepresented in STEM fields prior to choosing a college major”. The 5.21% gap between men and women was narrowed down to 2.43% which is the unexplained portion of the gap that has many different factors. When looking further into this percent, it seems that women are affected largely when it comes to the years of biology class they take in high school and also the schools reputation in the community when choosing a college major. At this time I cannot say that 2.43% of this gap is in fact social discrimination because more variables need to be added into the model. If I had more time to work on this I would look more thoroughly through the data to find better explanatory variables and only then feel confident in stating that X amount of underrepresentation of women in STEM is cause by social discriminatory factors in high school.

VII. References

Bottia, M. C., Stearns, E., Mickelson, R. A., Moller, S., & Valentino, L. (2015). Growing the roots of STEM majors: Female math and science high school faculty and the participation of students in STEM. *Economics of Education Review, 45*, 14-27. Retrieved from http://ezproxy.uakron.edu:2048/login?url=https://search-proquest-com.ezproxy.uakron.edu:2443/docview/1775617352?accountid=14471

Borjas, J. G. (1996). *Labor Economics.* United States: McGraw-Hill Higher Education. (Pages 363-367).

Fischer, S. (2017). The downside of good peers: How classroom composition differentially affects men's and women's STEM persistence. *Labour Economics, 46*, 211-226. doi:http://dx.doi.org.ezproxy.uakron.edu:2048/10.1016/j.labeco.2017.02.003

Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters? *Economics of Education Review, 29*(6), 911-922.doi:http://dx.doi.org. ezproxy.uakron.edu:2048/10.1016/j.econedurev.2010.06.010

Kokkelenberg, E. C., & Sinha, E. (2010). Who succeeds in STEM studies? an analysis of binghamton university undergraduate students. *Economics of Education Review, 29*(6), 935-946. doi:http://dx.doi.org.ezproxy.uakron.edu:2048/10.1016/j.econedurev.2010.06.016

Ma, Y. (2011). Gender differences in the paths leading to a STEM baccalaureate.*Social Science Quarterly, 92*, 1169-1190. Retrieved from http://ezproxy.uakron.edu:2048 /login?url= https://search-proquest com.ezproxy.uakron.edu:2443/docview/919222197? accounted=14471

This research is based on data from the Nation Longitudinal Survey of Freshmen, a project designed by Douglas S. Massey and Camille Z. Charles and funded by the Mellon Foundation and the Atlantic Philanthropies.

VIII. Appendix

Table 1 (Linear Probability) High School preparation on College STEM major choice Dependent Variable: College STEM Major		
Variable	**Male**	**Female**
A or Better in math [MAbetter]	0.04902 (1.82)*	0.06755 (3.28)***
A or Better in science [SAbetter]	0.14602 (5.15)***	0.06297 (2.91)***
High school years of calculus [HSyrscalc]	-0.00552 (-0.53)	0.00443 (0.53)
High school years of biology [HSyrsbio]	-0.00510 (-0.44)	0.02427 (2.60)***
High school years of chemistry [HSyrschem]	0.01406 (1.14)	0.02179 (1.99)**
High school years of physics [HSyrsphy]	0.03401 (3.18)***	0.03007 (3.30)***
High school overall quality [HSqual]	0.02090 (1.13)	-0.00330 (-0.21)
High school reputation in community [HSrep]	-0.03972 (-2.43)**	-0.00877 (-0.66)
Attended a Public School [Public]	-0.02529 (-0.96)	-0.00205 (-0.10)
Student is 'Prepared' for College [HSprepColl]	0.03792 (1.57)	0.04177 (2.19)**
Student Ethnicity is Black [DBlack]	-0.05457 (-1.66)*	0.04319 (1.65)*
Student Ethnicity is Asian [DAsian]	0.03671 (1.19)	0.01335 (0.50)
Student Ethnicity is Spanish/Latino [DLatino]	-0.02503 (-0.79)	0.00421 (0.15)
R-Squared	0.0614	0.0359
Adjusted R-Squared	0.0537	0.0303
Number of Observations	1608	2228
Note: The figures in parentheses are t-statistics; ***, ** and *, respectively, denote statistical significance at the 1%, 5%, and 10% levels.		

Table 2 Variable Definitions, Summary Statistics, and Data Sources

Variable	Male		Female	
	Mean	**Standard Deviation**	**Mean**	**Standard Deviation**
stem	0.3115	0.4632	0.2594	0.4384
MAbetter	0.6853	0.4645	0.6036	0.4892
SAbetter	0.7207	0.4487	0.6831	0.4653
HSyrscalc	1.3843	1.0977	1.4088	1.0979
HSyrsbio	2.5478	0.9991	2.6306	1.0107
HSyrschem	2.4458	0.9678	2.3366	0.8751
HSyrsphy	2.0702	1.1125	1.7149	1.0407
HSqual	3.2997	0.7338	3.2728	0.6957
HSrep	3.4446	0.8285	3.4676	0.8257
Public	0.7095	0.4540	0.7131	0.4523
HSprepColl	0.3202	0.4667	0.3702	0.4829
DBlack	0.2238	0.4169	0.2971	0.4570
DAsian	0.2518	0.4342	0.2396	0.4269
DLatino	0.2338	0.4233	0.2342	0.4236

This research is based on data from the Nation Longitudinal Survey of Freshmen, a project designed by Douglas S. Massey and Camille Z. Charles and funded by the Mellon Foundation and the Atlantic Philanthropies.

Table 3 Oaxaca Decomposition						
Variable	α Male	α Female	X male	X female	Explained	Unexplained
intercept	0.1711	0.0133	1.0000	1.0000	0.0000	0.1578
HSprepColl	0.0379	0.0418	0.3203	0.3703	-0.0019	-0.0014
DBlack	-0.0546	0.0432	0.2239	0.2971	0.0040	-0.0290
DAsian	0.0367	0.0134	0.2519	0.2397	0.0004	0.0056
DLatino	-0.0250	0.0042	0.2338	0.2343	0.0000	-0.0069
MAbetter	0.0490	0.0676	0.6853	0.6037	0.0040	-0.0112
SAbetter	0.1460	0.0630	0.7208	0.6831	0.0055	0.0567
HSyrscalc	-0.0055	0.0044	1.3843	1.4089	0.0001	-0.0140
HSyrsbio	-0.0051	0.0243	2.5479	2.6306	0.0004	-0.0773
HSyrschem	0.0141	0.0218	2.4459	2.3366	0.0015	-0.0181
HSyrsphy	0.0340	0.0301	2.0703	1.7150	0.0121	0.0068
SchoolQual	0.0209	-0.0033	3.2998	3.2729	0.0006	0.0792
SchoolRep	-0.0397	-0.0088	3.4447	3.4677	0.0009	-0.1073
Public	-0.0253	-0.0021	0.7096	0.7132	0.0001	-0.0166
Total					2.78%	2.43%