Gender and Racial Disparities in STEM Workforce: A Machine Learning Approach

Luxin Tian

April 26, 2020

1 Literature Review

1.1 Gender and Racial Gaps in STEM Workforce

As Science, Technology, Engineering, and Mathematics, collectively being referred to as STEM, is creating significant economic impacts on human society, policymakers and employers have placed increasing emphasis on fostering the growth of STEM workforce as well as promoting equal opportunities. While the market preference for STEM workers is generic, it has been a historical and persisting pattern that there exist significant gaps in the population distribution and the labor earnings across different groups by gender, race, and Hispanic origin in the STEM workforce.

Empirical evidence shows that, in the United States, women, African Americans, and Hispanic have been underrepresented in the STEM workforce in recent decades. As Landivar (2013) suggest, according to the micro-level data from the decennial US Census from 1970 to 2000 and the American Community Survey in 2011, although women had increased their presence in STEM workforce since the 1970s, they remained underrepresented in engineering and computer science occupations, which comprised over 80% of all STEM employment. According to Beede et al. (2011)'s calculation using the 2000 and 2009 American Community Survey (ACS) micro-level data, across the 2000s, women's representation in STEM jobs remains 25% lower than that of men, although they comprised nearly half of all the workers in the US, and the share of women with college-level education had increased in the workforce. US

Census data also suggests that African Americans and Hispanics have been underrepresented in the STEM workforce as well(Landivar, 2013).

While women and racial minorities are underrepresented in STEM fields, their engagement in STEM education and workforce turns out to be an important force driving the alleviation of earning gaps across genders and races. Beede et al. (2011) and Baird et al. (2017) find that education backgrounds in STEM fields and working in a STEM occupation can render a large increase in monetary earnings, and women's wage premium is significantly larger than that of men on average for working in a STEM occupation. Similar mechanisms are also found between racial groups. Carnevale et al. (2011) finds that, between 2005 and 2009, earning gaps across racial minority groups are much narrower in the STEM workforce than that in the non-STEM workforce, after controlling for the age variable.

Existing research has elaborated on the direct evidence from demographical data of gender and racial gaps that persisted in the US labor market. Given potentially complicated self-selection processes underlying people's decisionmaking on education and employment, there is very little empirical research making efforts on identifying the causal relationship between gender and racial identity and engagement in STEM occupations and employment outcomes. On the contrary, many researchers tried to understand the gender and gaps related to the STEM workforce qualitatively, from the perspectives of economics, psychology, sociology, as well as cognitive science, and empirical evidence has been presented to verify the hypothesis. The gender and racial gaps can form from an early stage, even from as early as high school and college when students choose their subjects and field of studies, and the selection can compound and reinforce over years in their career path (Fryer Jr and Levitt, 2010; Hill et al., 2010; Jaeger et al., 2017). Furthermore, Kanny et al. (2014) review the explanations of gender gaps in STEM fields in the past forty years and identify five dominant narratives that emerged in previous research: "Based on a systematic review of 324 full texts spanning the past 40 years of scholarly literature, five dominant meta-narrative explanations emerged: individual background characteristics; structural barriers in K-12 education; psychological factors, values, and preferences; family influences and expectations; and perceptions of STEM fields". Wang and Degol (2017) further summarize six explanations for the minor presence of women in STEM fields from biological and sociocultural perspectives: "(a) cognitive ability, (b) relative cognitive strengths, (c) occupational interests or preferences, (d) lifestyle values or work-family balance preferences, (e) field-specific ability beliefs, and (f) gender-related stereotypes and biases". These explanations can provide evidence on model specifications for research that examine the role of gender and racial identity in employment status and outcomes.

1.2 Classfications of STEM Occupations

While plenty of research has tried to estimate the inequalities related to the STEM workforce, there has been a lack of consensus on the classification standards of STEM occupations (Landivar, 2013; Anderson et al., 2018). Many efforts have been made to clarify the boundary of STEM fields by both the government and researchers.

As is summarized by Anderson et al. (2018), there are two dominant existing methods for defining STEM occupations at different levels. Most of the previous academic and policy research employs the official standards published by the Standard Occupational Classification Policy Committee (SOCPC), which collaborated with nine federal agencies for the 2010 revision of the Standard Occupational Classification manual by the Bureau of Labor Statistics. As is introduced by Landivar (2013), the SOC manual, which provides 539 specific occupational categories associated with 23 occupational groups, classifies occupations based on the type of work performed, and the SOCPC creates the classifications of STEM occupations based on the SOC codes. This method is characterized by its top-down approach of identifying work features, as the classification standard is suggested by experts and researchers from general to discrete occupation categories. However, there have been criticisms of the SOCPC approach considering that it does not reflect on the potential variations in the extent to which an occupation actually requires STEM skills and knowledge but rather hinges on top-down subjective perceptions by the public sector representatives and experts(Anderson et al., 2018).

Another classification approach that alleviates the concerns about the SOCPC approach is created by the Brookings Institution (Rothwell, 2013). This approach uses occupation characteristics data provided by the Occupational Information Network (O*Net) of the US Department of Labor. Each occupation in the SOC occupation codes is evaluated and scored by its level of knowledge required from one of the four STEM fields, which gives each occupation four scores from 0 to 7 corresponding to the extent in which this occupation requires knowledge in each STEM field. Then, occupations with a score at least 1.5 standard deviations above the mean in at least one STEM field would be classified as a high STEM occupation, and, those whose average score of the four scores for the four fields is at least 1.5 standard deviations above the mean would be classified as super STEM occupations(Anderson et al., 2018; Rothwell, 2013). This approach is based on the characteristics of occupations and thus is expected to be more accurate to reflect on the extent to which occupation is STEM-like. Nevertheless, the evaluation process is still somewhat subjective and only based on a limited number of characteristics, and the threshold, 1.5 standard deviations, is also somewhat an arbitrary cut-off.

In light of such caveats of using these two classification standards, Anderson et al. (2018) suggests a new bottom-up approach based on a micro-level survey that interviewed workers in the real workforce whether their job requires STEM knowledge and skills. This survey also includes the categorical label of interviewee's occupation in terms of the SOC standard codes, and thus this self-reported classification can be compared with the existing "occupation-based classifications". Anderson et al. (2018) identified that 10% to 15% occupations that actually require STEM knowledge and skills are not counted by the traditional STEM classifications. With classifications more sensitive to practical variations across workers, they find that women are more likely to engage in jobs that are not consistently classified as STEM with their self-reported outcome. Furthermore, they find that there is no significant gender gap in STEM workforce engagement, suggesting that the results of previous research had overestimated the inequality. They also verified that the wage premium for women working in STEM fields that are consistently classified by different methods is larger than other women working in periphery STEM fields that are inconsistently classified. This paper contributes to the literature, although being rich already, of the STEM workforce and provides strong evidence that a more accurate classification method that takes into accounts the variations across specific jobs.

Therefore, by integrating multiple data sources and employing a machine learning approach, this paper is expected to construct a more accurate classification of STEM occupations, which, although remains at the occupation level, can quantitatively reflect on the probability that one occupation should be counted as STEM workforce. As is suggested by Anderson et al. (2018), such enhancement in the reliability of classification can potentially challenge the existing literature by providing a more unbiased estimation of gender and racial inequalities in the STEM workforce.

References

Anderson, Drew, Matthew Baird, and Robert Bozick, "Who Gets Counted as Part of America's STEM Workforce? The Implications of Different Classification Approaches for Understanding the Gender Gap in STEM," 2018.

Baird, Matthew D, Robert Bozick, and Mark Harris, "Postsecondary Education and STEM Employment in the United States: An Analysis of National Trends with a Focus on the Natural Gas and Oil Industry," 2017.

Beede, David N, Tiffany A Julian, David Langdon, George McKittrick, Beethika Khan, and Mark E Doms, "Women in STEM: A gender gap to innovation," *Economics and Statistics Administration Issue Brief*, 2011, (04-11).

Carnevale, Anthony P, Nicole Smith, and Michelle Melton, "STEM: Science Technology Engineering Mathematics.," *Georgetown University Center on Education and the Workforce*, 2011.

Hill, Catherine, Christianne Corbett, and Andresse St Rose, *Why so few? Women in science, technology, engineering, and mathematics.*, ERIC, 2010.

Jaeger, Audrey J, Tara D Hudson, Penny A Pasque, and Frim D Ampaw, "Understanding how lifelong learning shapes the career trajectories of women with STEM doctorates: The life experiences and role negotiations (LEARN) model," *The Review of Higher Education*, 2017, *40* (4), 477–507.

Jr, Roland G Fryer and Steven D Levitt, "An empirical analysis of the gender gap in mathematics," *American Economic Journal: Applied Economics*, 2010, *2* (2), 210–40.

Kanny, Mary Allison, Linda J Sax, and Tiffani A Riggers-Piehl, "Investigating forty years of STEM research: How explanations for the gender gap have evolved over time," *Journal of Women and Minorities in Science and Engineering*, 2014, *20* (2).

Landivar, Liana Christin, "Disparities in STEM employment by sex, race, and Hispanic origin," *Education Review*, 2013, *29* (6), 911–922.

Rothwell, Jonathan, *The hidden STEM economy*, Metropolitan Policy Program at Brookings, 2013.

Wang, Ming-Te and Jessica L Degol, "Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions," *Educational psychology review*, 2017, *29* (1), 119–140.