Project Proposal: Modeling opinions and the Gender Labor Force Participation gap in the United States

Eric Karsten [*] Ezra Max [†] Deniz Turkcapar [‡]

January 24, 2020

Contents

[*]ekarsten@uchicago.edu
[†]emax@uchicago.edu
[‡]dturkcapar@uchicago.edu

1 Introduction

1.1 Our Question

The labor market looks different for men and women, both in labor activity and in remuneration. There remains a sizeable gender pay gap in the United States, although it has narrowed substantially over the last 40 years and continues to diminish. On the other hand, the gap between men and women's participation in the labor market has not diminished since the financial crisis: women's labor force participation rate (LFPR) was 79.5% of men's LFPR in 2000, 82.3% of men's LFPR in 2010, and 82.1% of men's LFPR in 2019 (WBG). Moreover, there is substantial geographical heterogeneity in the gender labor force gap across the United States (figure 1). This raises the question of which factors drive men and women's labor force participation, and which factors are responsible for women's persistently lower rates of labor force participation.

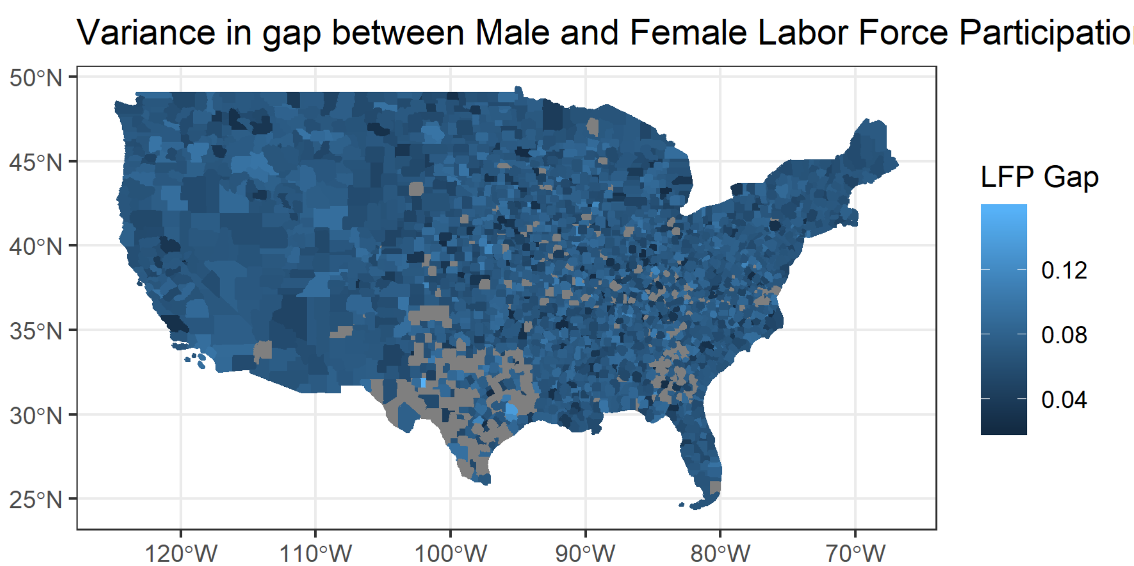

Figure 1: Female labor force participation gap by county

Our project will link temporal and geographic data for women's aggregate labor participation to a wide variety of gendered and non-gendered features–including many not previously studied in predictive models for labor participation–and then fit multiple regressions on these data. By leveraging variation in the data over time across different localities, we hope to find an improved predictive model for labor participation and to gain insight into which factors may explain the gender gap in labor participation.

1.2 Prior Work

We can broadly divide the relevant literature to date into two camps: predictive models for aggregate labor force participation built on regression analysis of population-level data, and analyses of how social and political structures affect women's labor participation that are carried out in small-scale experimental settings. We use this literature to inform features we might like to include in our analysis. We also use this as a starting point to specify a "state of the art" regression as an outside option to compare with our machine learning techniques.

1.2.1 Population-Level Predictive Models for Gender LFP Gap

In general, this literature builds causal linear models and then estimates their parameters. Grigoli, Koczan, and Topalova (2018) fit a linear regression for labor participation of different demographic groups across 23 countries since 1980. They find that secondary and tertiary education rates, public spending on childhood education and care, and the share of part-time employment in the job market have positive and statistically significant effects on female labor force participation. The paper also finds that the shape of women's age participation profile is consistent with a large share of women leaving the labor force to start their own family and, in the process, become the main caretakers of the family. We will use this paper to inform how we use the "age" feature we will include in our analysis.

Toossi (2009) fits a regression on both aggregate and individual features using time-series census data in the United States from individual data from 1970-2006 to predict labor force participation, finding statistically significant impact of marital status and education on women working. We will be using a similar dataset, but we will build on this work by using different modeling techniques and by merging the census data with additional opinion data at the geographic level.

Özerkek, Özbal, et al. (2017) fit a probit-regression model on individual data from Turkey and find that being married has a negative correlation with women participating in the labor force, while secondary education has a positive correlation. Taşseven, Altaş, and Turgut (2016) study aggregate FLFP rates across the OECD from 1980-2013 using a panel logit model and find a small and not statistically significant effect from the ratio of female to male enrollment in tertiary education, and a large and highly statistically significant effect from fertility rates on female labor force participation.

These findings square with the broad economic consensus established in Bertrand, Goldin, and Katz (2010) that maternity decisions are responsible for the vast majority of the gender wage gap that remains in the modern workplace. Furthermore, Albanesi and Olivetti (2016) use historical data to suggest that improved maternal health from 1930-1960 may have been responsible for gains in FLFP in the United States over the same period. We therefore will make sure to include maternity decisions in our analysis. Associated with maternal decisions is of course access to birth control and family planning. Goldin and Katz (2002) suggest a positive impact on women's labor participation due to the advent of oral birth control using a study of panel data.

1.2.2 Setting-Specific Evidence of Female LFP Shifters

In addition to population-level studies, economists has made valuable contributions to explaining the female labor force participation gap by looking at particular institutional dimensions of labor supply. In a seminal paper, Bertrand, Kamenica, and Pan (2015) find that the institution of marriage generally constrains a wife's earnings to below those of her husband and that even in marriages where the wife's earnings exceed those of her husband, the wife tends to do more of the household work.[1] This suggests the importance of the gendered institution of marriage in influencing female labor force decisions as a function of her marital partner's labor force decisions.

In the field of experimental economics, we see from Bursztyn, Fujiwara, and Pallais (2017) that unmarried female MBA students tend to be less ambitious compared to married MBA students in settings where their ambition might be visible to their male peers. This suggests a trade-off between success in marriage markets and success in labor markets (at least in this setting). In light of this, our analysis could be improved by including proxies for a husband's income and labor force decisions in our models of female labor force participation.

Additionally, Bursztyn, González, and Yanagizawa-Drott (2018) find that within the marriage norms in Saudi Arabia, many men privately do not mind if their wife works, but publicly oppose women's entry into the workforce. This suggests that institutional norms, and perceptions of public beliefs, can have a meaningful impact on whether women participate in the labor force. Our analysis is constrained to the United States, but we will incorporate similar public-opinion data in our analysis.

1.3 Our Contribution

Building on the existing literature, our project will study both conventional drivers of (female) labor force participation and proposed non-conventional drivers of female labor force participation in an effort to build more effective predictive models for the aggregate female LFPR. By including a wide set of features in our data and applying a variety of machine learning techniques, we hope to capture more of the complexity behind women's choice to participate in the labor market. In particular, our group aims to: (1) evaluate which data features are relevant in modeling the labor force participation gap (using techniques such as LASSO); (2) evaluate which classes of machine learning techniques have strong predictive power in this setting; (3) make predictions at a granular geographic level about the evolution of the gender LFPR gap.

1. Potentially, this is because wives want to offset the psychological frictions that arise when a female partner out earns her male partner.

2 Data

Our project will rely primarily on three public data sources. First, we will use public data from the American Community Survey (ACS), which allows us to link husbands and wives and study households' demographic and geographic characteristics. Secondly, we will use the American National Election Study (ANES) to join opinion data geographically (probably at the state level because of privacy concerns) and in time (for three election cycles: 2008, 2012, 2016) to see how labor force decisions vary with opinion data. The substantial geographic variation in public opinion data–for instance, in how respondents feel about women's participating in the labor market (figure 2)–may allow us to gain insight into relevant the social-psychological factors driving female labor force participation. Finally, we will join aggregate CDC Health Data that reports metrics such as access to and use of contraceptives at the county level, allowing us to study whether access to family planning and other healthcare services affects women's decision to work.

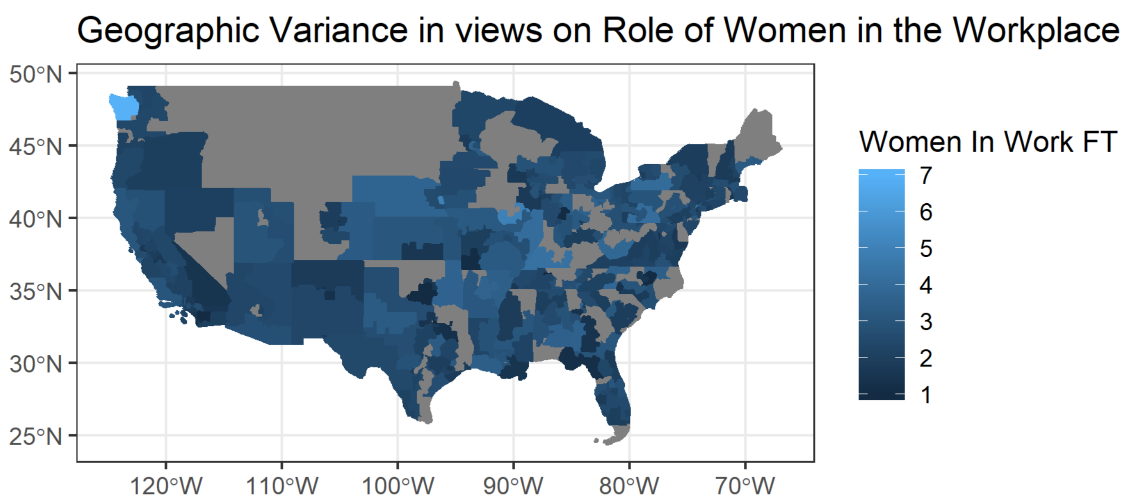

Figure 2: Example of the sort of rich geographical variation that can come from opinions about Women's role in the workplace vs the home. This connects closely with the experimental work by Bursztyn et. al. (2017)

3 Methods

We will follow a methodology that is tailored to the particular dimensions we want to describe in our models, but generally, we can split our methodology into two parts: individual level analysis and spatio-temporal level analysis.

3.1 Individual Prediction

Here, we will not be able to merge the rich features that exist outside the ACS, but we will be able to perform a much more robust analysis at the individual level since we can link individuals with their spouses and explore some of the predictive power that we might get from including these features, as suggested by our literature review.

We will use a variety of ML methods with the aim of predicting individual labor force participation for men and women separately. After we calibrate our models, we will evaluate their performance on the test data, and then we will see if we can get more accurate results than traditional regression models. We will then use the predictions from these models to examine the expected difference in the population between male and female labor force participation as we change different parameters in the calibrated model (i.e., how much of an effect does the first or second child have on the gap) by feeding simulated data into the calibrated models.

Finally, time-permitting, we will aim to train models based on prior years' data in order to make predictions about future female labor force participation. This plays to the strengths of the machine learning toolbox.

3.2 Geographic by Time level Prediction

At the county-year or state-year level, we will be able to merge many more datasets onto the ACS to create a richer set of features to perform our prediction on. Here, we will be predicting the gender labor force participation gap using similar techniques to those described above, this time accounting only for aggregate features of communities.

3.3 The Toolbox

For both of the data settings described above, we hope to use the following techniques: Linear Regression as a baseline, LASSO Regression to evaluate important features, K nearest neighbors, Random Forest, and a Neural Network as three competing ML models whose performance we will evaluate in each setting. We hope that by leveraging these different techniques we can identify both stronger predictive models and the key drivers of female labor force participation across models.

References

Albanesi, Stefania, and Claudia Olivetti. 2016. "Gender roles and medical progress." *Journal of Political Economy* 124 (3): 650–695.

Bertrand, Marianne, Claudia Goldin, and Lawrence F Katz. 2010. "Dynamics of the gender gap for young professionals in the financial and corporate sectors." *American economic journal: applied economics* 2 (3): 228–55.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender identity and relative income within households." *The Quarterly Journal of Economics* 130 (2): 571–614.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. 2017. "'Acting Wife': Marriage Market Incentives and Labor Market Investments." *American Economic Review* 107 (11): 3288–3319.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott. 2018. *Misperceived social norms: Female labor force participation in Saudi Arabia.* Technical report. National Bureau of Economic Research.

Goldin, Claudia, and Lawrence F Katz. 2002. "The power of the pill: Oral contraceptives and women's career and marriage decisions." *Journal of political Economy* 110 (4): 730–770.

Grigoli, Francesco, Zsoka Koczan, and Petia Topalova. 2018. *A Cohort-Based Analysis of Labor Force Participation for Advanced Economies.* International Monetary Fund.

Özerkek, Yasemin, Yasemin Özbal, et al. 2017. "The Effects of Education and Marital Status on Women's Labor Force Participation: A Regional Analysis of Turkey." *Ekonomi-tek-International Economics Journal* 6 (3): 15–38.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, and Erin Meyer. 2018. "Jose Pacas, and Matthew Sobek." *Ipums usa: Version* 8.

Taşseven, Özlem, Dilek Altaş, and ÜN Turgut. 2016. "The determinants of female labor force participation for OECD countries." *Uluslararası Ekonomik Araştırmalar Dergisi* 2 (2): 27–38.

Toossi, Mitra. 2009. "Employment outlook: 2008-18-labor force projections to 2018: older workers staying more active." *Monthly Lab. Rev.* 132:30.

WBG. *World Bank Data: Ratio of female to male labor force participation rate (%) (modeled ILO estimate).* `https://data.worldbank.org/indicator/SL.TLF.CACT.FM.ZS`. Accessed: 2020-01-23.