Stand still or Flee away? Evidence from the College-educated Female Occupational Choice

Yaling Xu

June 2019

Abstract: Over the last decade, the literature has increasingly focused on the importance of overwork in explain the female occupational choices. Using detailed survey data form US Census and 5 year ACS, this paper examines the prevalence of overwork in college-educated females occupational choices. Our results suggest that the prevalence of overwork significantly reduce the share of females given specific occupation. The effect is stronger for married females with children. Exploring detailed nature of data, we use the lead of overwork variable to exclude positivity of reverse causality. Our findings support the hypothesis that childcare is an important cost for long hours working commitments.

Keywords: occupational choices; college-educated females

1 Introduction

Numerous researches have been conducted on the gender pay gap. Traditional researches emphasized on factors such as human capital, the family division of labor and compensating wage differentials. Nonetheless, recently new researches have been shifting to study the impact of norms, psychological attributes and non-cognitive skills on the gender pay gap. Blau and Kahn (2017) documents that the role of occupation remains pronounced in explaining the gender wage gap. Specifically, occupational differences account for over 20 percent gender wage gap and are currently at the top of observed factors. However, it is found that male continue to choose male-dominated occupations while female prefer entering predominantly female occupations in the past three decades (Bettio Verashchagina 2009). At a high level, occupation choice by gender plays a significant role in understanding the gender wage gap.

Female's occupation displacement could be differed by males' because of different job attributes, i.e. workplace flexibility, corporate culture, family-career balance, and working hour. Compare to the male, less of the female choose to work on long hour occupations, even if they pay more. Fouad et al. (2012) document females' condition in management position: "Most of management is a male-dominated culture (male conversation topics, long hours, demanding lifestyle, career-focused expectations). ... Women usually choose to leave WITHOUT FIGHTING THE UPHILL BATTLE to make improvements. It is a self-sustaining cycle!" Current engineers work over twice hours than what in 2000. And extensive working hour has been one of the major reasons for female to leave engineering field. Meanwhile, there is a tendency in the U.S. where working hour will be further extended. There is also large increase in the share of male working overtime (which is defined as working over 48 hours per week) (Pencavel, 2016). Cortes and Pan (2016) calculate the prevalence of overwork for educated male in US and find that "the share of college educated males working 50+ hours per week in the US was 36%, larger than that in every Western European country, including France (31%), United Kingdom (28%), Germany (26%), Italy (22%), Spain (17%), and the Netherlands (10%)."

Working on longer hours could result to a positive change in wage. But such change is not considered by a large portion of the female when choosing occupation. Elke (2002) studies the "impact of working hours on gross hourly wage rates by using a simultaneous wage-hours model which fully accounts for labor supply effects" She finds that jobs working over 48 hours has lower marginal compensation. On the other hand, there is an undeniable fact that women tend to shoulder larger family responsibilities than men. Cortes and Pan (2017) use data from 2005 International Social Survey Programme (ISSP) and demonstrate that over 50 percent of the female prefer working fewer hours and earning less money than working more hours but earning more. But statistics shows only 20 percent of the male have the same preference. In addition, mothers are likely to exit occupations when the work hours over 50 hours. However, such effect isn't observed on male or childless female.

The switch towards less working-hours occupation and "flee" from male-dominated are even more prevalent among college-educated women since they are more "affordable" for potential wage loss. College-educated female tend to lean towards occupations with less working hours (in comparison with other jobs in the market). However, very few studies focus on this group of females and exanimate its effect to the overall gender wage gap. This paper will emphasis on the college-educated female group and compare their occupational choices with those male or female with partial similar characteristics.

I will focus on the educated-female group for several reasons: Firstly, it seems that educated-female are more "affordable" for potential wage loss when "fleeing" from long working hours occupations. Secondly, it has been reported that over 30 percent of college-educated female who leave STEM field express unsatisfied to the overworking environment. (Jaimovich the al., 2017). And work-family balance is usually taken into account by female when considering occupation choices. Thirdly, focusing on this narrow group helps me to compare their occupational choices with individuals who share partial similar characteristics. The following questions will be answered:

(1) Will college-educated females respond to overwork demands by switching to other occupations or exiting labor market?

(2) Will the response be different among different demographic groups? Specifically, will married college-educated females with children be more sensitive to overwork demands than other individuals who are in same age and occupation?

The key variable is the occupational distribution in a specific demographic group, which is measured by the share of a given demographic population working in a specific occupation. In a panel dataset, I believe that this variable helps us capture occupation switch and labor market exit. Hopefully, this paper will also complement previous papers by exploring how the long hours work commitment effect college-educated females' occupational choices, using data from 1990 to 2000 Census, and 2015 to 2017 five-year American Community Survey (ACS).

The main independent variable in my analysis is the prevalence of overwork in an occupation. However, there was no existing metric in measuring the degree of overwork quantitatively among different occupations. Intuitively, most overwork commitments impose on males, so I use the share of overwork among males as a quantitive method to measure overwork prevalence given occupation.

In this paper, I focus on the married college-educated females with children, and analyze how the overwork prevalence affect occupational distribution in this demographic group using fixed effects model. Intuitively, this population group should be more sensitive to long hours commitments since they are likely to shoulder most responsibilities to their family.Since there may exist time-invariant factors and year-specific shocks, I include year and occupation fixed effects.

Then I expand my analysis by controlling for other demographic groups, for example, married females without children, single females, married males who are at the same age and in the same occupation. Before the formal estimation, I apply the LASSO feature selection to extract two most significant control variables from 74 candidates, which are education and earnings. Thus in some specifications, I include these two control variables to address the concern that there may exist occupation-specific factors which are related to overwork and thus affect occupational choices of college-educated female. I find that overwork prevalence negatively affect the share of married college-educated females with children working in this occupation, and results are robust controlling other demographic groups. I assume that family responsibility is the major cost for working in a overworking occupation. I assume that childcare is the most significant factor hindering female career development. In order to exclude other time-varying occupational factors or unobservable characteristics in married female group, I include a "placebo test" to test the degree of overwork effect on female occupational choices. I test the casual effect on other subgroups: female without children, males and single females, controlling for education and log wages. Thus I confirm that long hours working commitment has the largest and most significant effect only on young married female with children.

The remainder of this paper is organized as follows. Section 2 introduces the data. Section 3 describes the estimation model and computational approach. Section 4 presents the main empirical results while Section 5 explores potential mechanisms. Finally, Section 6 discusses and concludes.

2 Data Description

I use data from 1990 to 2010 Census, and five-year American Community Survey (ACS) from 2015 to 2017. I choose OCC1990 (OCC1990 is a modified version of the 1990 Census Bureau occupational classification scheme) as my occupation basis to keep a consistent classification of occupations with the 1990 coding scheme.

In all specifications, occupation and year fixed effects are included. I use the way that Dorn (2009) developed to construct 301 consistently defined occupations. I drop the occupations that are not consistent over time, and then drop the Not in LF or in Military occupation in the following analysis since the sample is too small and limited.

I limit the data to native-born Americans who have at least a Bachelor's degree and age 25 to 57. I drop the self-employed individuals and only keep individuals whose usual work time is at least 30 hours or 3 weeks last year. I also discount the wage to the 2018 level using CPI index.

Overwork is defined as working over 50+ hours. The dependent variable is the occupational distribution in a specific demographic group of interest. It is measured by the share of a given demographic population working in a specific occupation. The main independent variable is the prevalence of overwork in an occupation, which is measured as the share of overwork among all the males in an occupation. Control variables candidates will be discussed in detail in Section 3.

3 Estimation Model and Computational Approach

3.1 Estimation Model

For main analysis, I plan to examine the relationship between the female occupational distribution and overworking prevalence based on following baseline regression:

$$\frac{female\_gr_{it}}{female\_gr_t} = \alpha + \beta * share\_male\_overworkk_{it} + \gamma * \frac{control\_gr_{it}}{control\_gr_t} + \delta * X_{it} + \pi_i + \pi_t + \epsilon_{ict} \tag{1}$$

where i refers to an specific occupation and t refers to year. $female_{gr}$ is the female demographic group I interest; $control_{gr}$ means another demographic group at the same age. We assume the control group has relatively low cost for working long hours. For example, female age 25-44 without children may have lower cost than female who is at the same age but have children controlling other factors. π_i and π_r refer to occupation i and year t control variables. By including year fixed effect, we could catch some year-specific effect such as economic status. And we also could capture some occupational time-invariant characteristics that may relate to working hours. X_{it} are other control variables that are related to overwork and may have unobservable effect on occupational distribution.

3.2 Feature selection by LASSO

"The feature selection is the process that choose a reduced number of explanatory variable to describe a response variable" (Fonti, 2017). By using feature selection, we could remove redundant variables and make the model easier to interpret. This helps us to deal with high-dimensional data and reduce overfitting. LASSO (Least Absolute Shrinkage and Selection Operator) is a powerful method to perform feature selection. In this process, LASSO apply a shrinking regularization process, penalizing coefficients that shrinks to zero. Thus we could still have some non-zero coefficients variables in regression model by minimizing the prediction error.

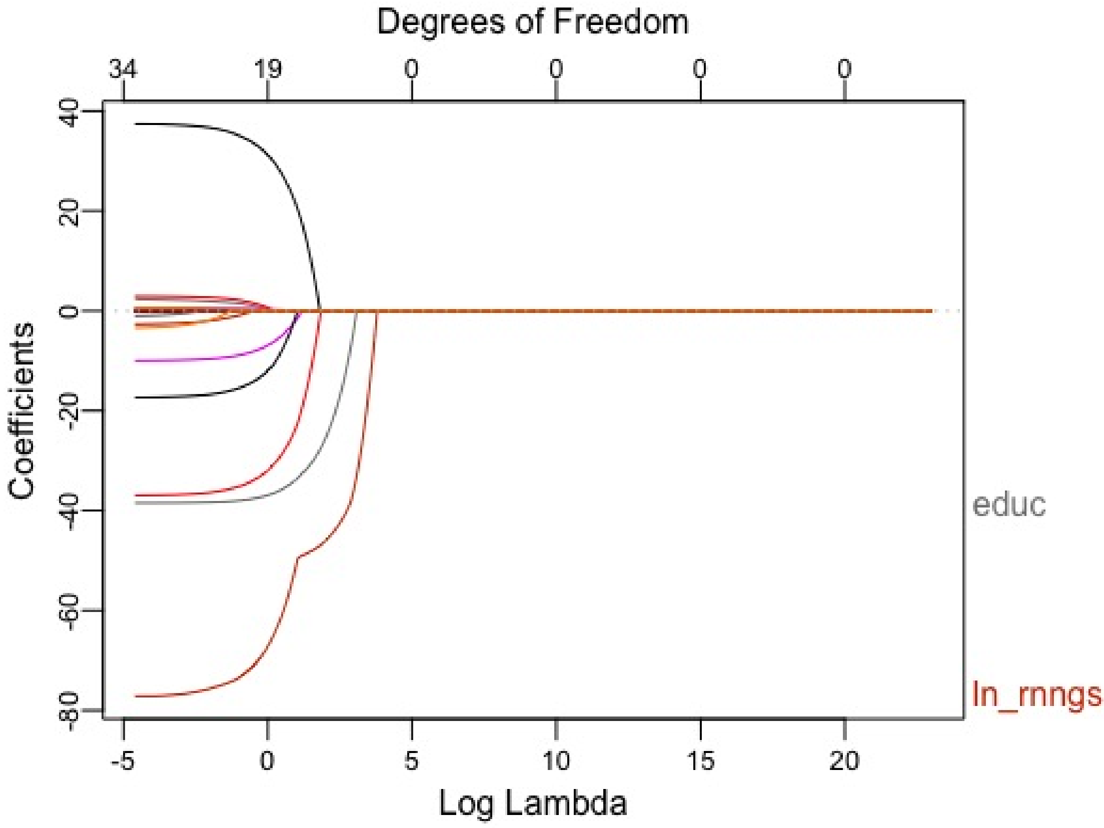
Here, I use LASSO to select control variables that are related to overwork and may have unobservable effect on occupational distribution. I construct a statistical model:

- Dependent variable: occupation.
- Independent variables: 79 variables such as education and wage, see more details in Appendix.

During the modeling process of feature selection, I choose the right control variables that have the most significant effect on dependent variable.

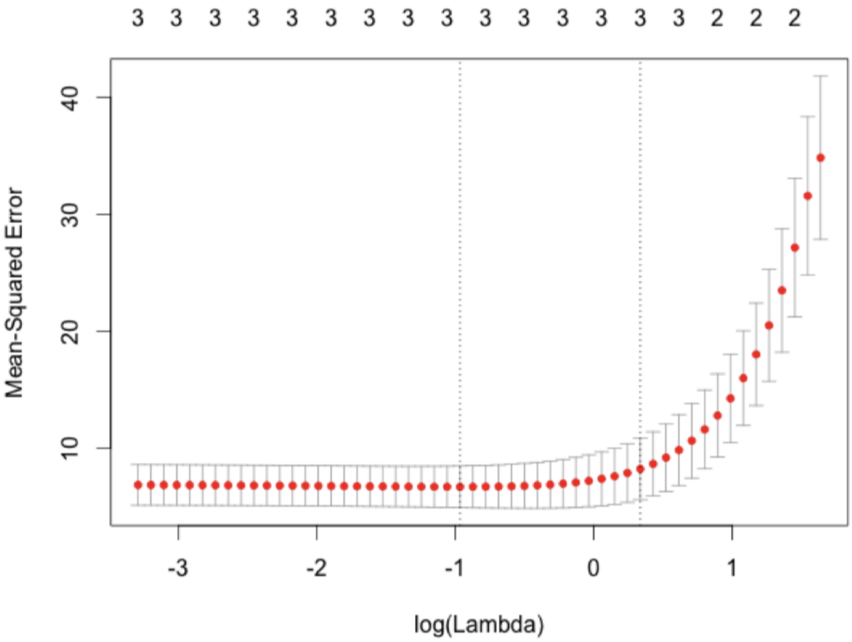
Figure 1 shows variables that coefficient is larger than 0.01. We could see that education and wage are the top two significant explanatory variables, thus we can expect that these variables will be included as control variables in the final model. In the dataset, we build a model to predict occupation response variable. To choose probable variables, we use glmnet package in R to do LASSO feature selection. The results is shown in figure 4. Each line represents one of independent variables and its performance in the model. From the plot above, we could see that education and wage are

Figure 1: Lasso Feature Selection(Glmnet): most significant variables



the top two significant explanatory variable. They enter the model first, and steadily negatively affect our dependent variable. All the other variables seem less significant than these two. The next step is to choose the most appropriate value for λ. I use the cross-validation to choose the number of nfold validation. The "min" lambda is minimum average out of sample deviance, represented by the left most vertical line, while the "1se" lambda value/oos deviance is at least 1 standard deviation away from the minimum value and still penalizes the large number of coefficeints.$R^2_{1se} = 0.3924649$ and $R^2_{min} = 0.3991116$.Detailed results show in Table 4-6.

Figure 2: Cross-Validation (nfold=5)



4 Descriptive Analysis

To capture the bigger picture of the occupational distribution, I compressed fine-grained 301 occupations to 16 general occupations **only for descriptive analysis**. These 16 occupations are: Managers; Natural, Math and Computing Science Professionals; Architects and Engineers; Health Professionals; Educators; Business Professionals (Accountants, HR, etc); Legal Professionals; Social Scientists; Writers and Artists; Technicians; Religious/social workers; Office clerks; Market salespersons; Precision Production, Operators, craft, repair Professionals; Agricultural workers.

Figure 3: Prevalence of Overwork vs Gender Difference in Employment share per Occupation

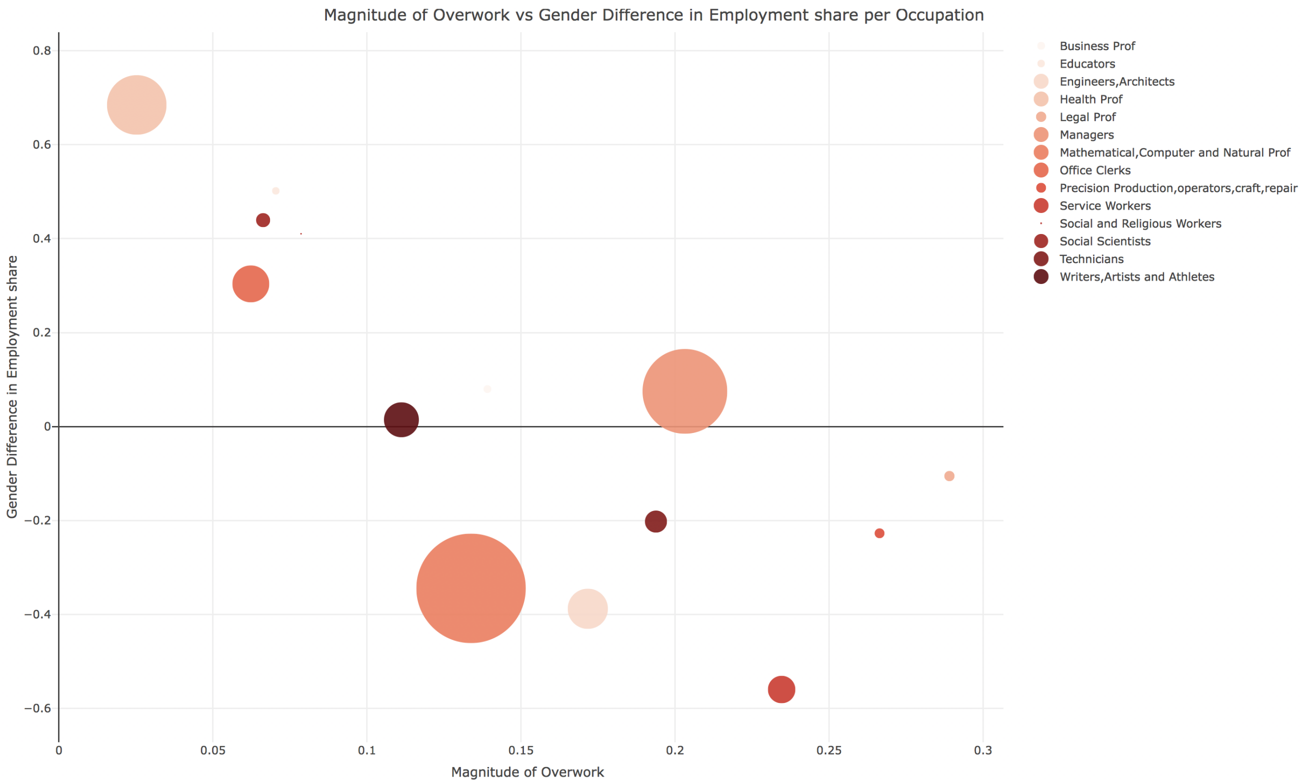


Figure 4: Prevalence of Overwork per Occupation

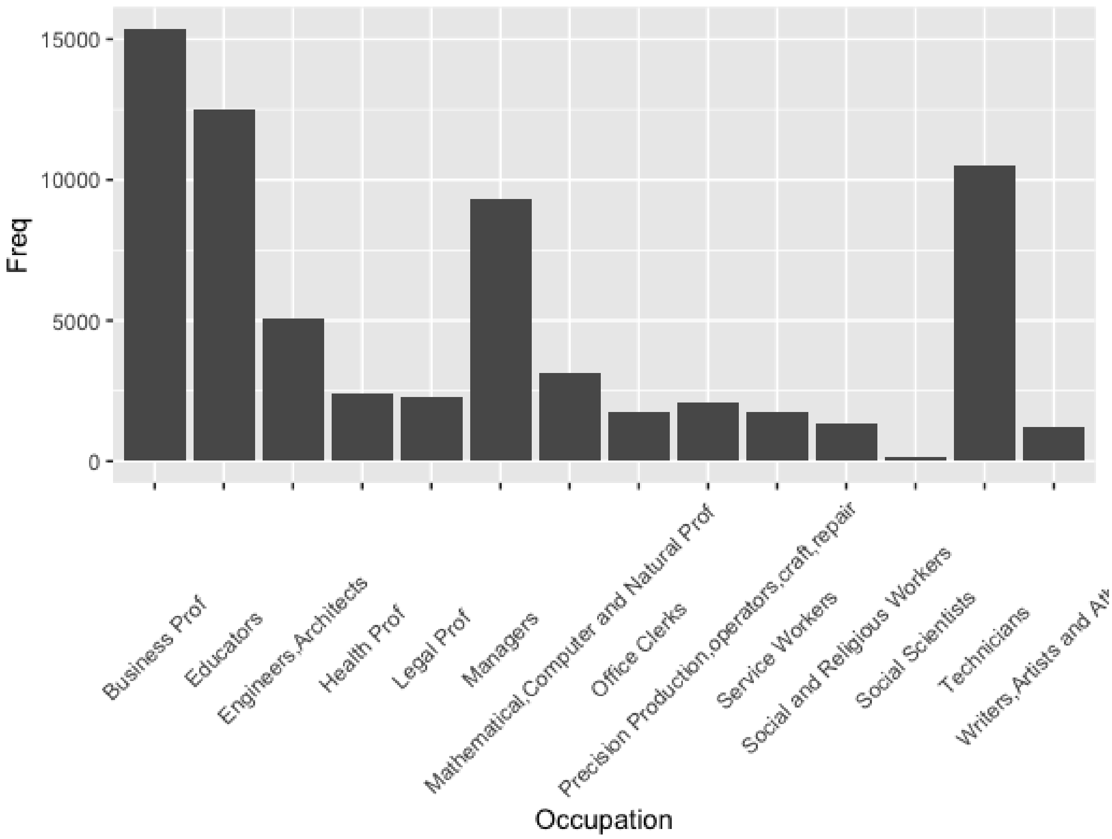
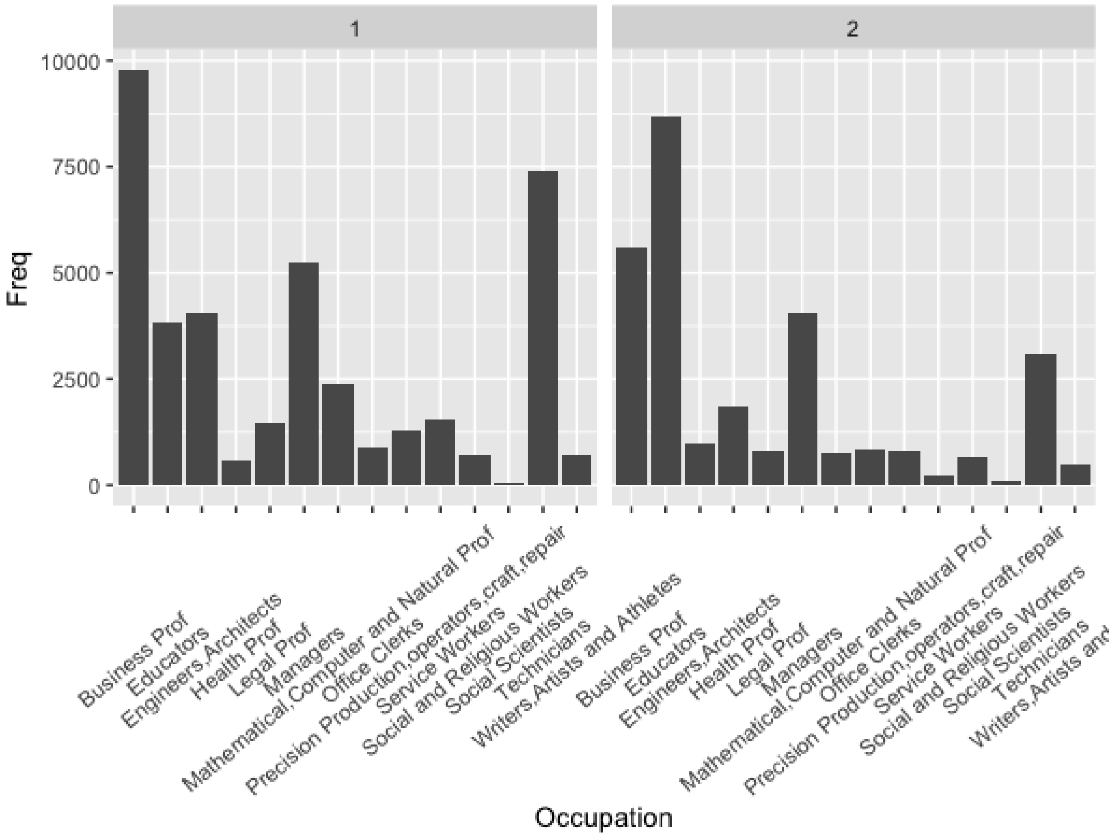


Figure 3 shows that there are significant variations in overwork across different occupations. The average overwork percentage is 23.9 percent. Figure 4 and Figure 5 show the prevalence of overwork per occupation. The top three occupations that require longest working hours are Legal Professionals, Precision Production, Operators, craft, repair Professionals and Managers. They all have overwork percentage that over 30 percent. On the other hand, the percentage of overwork among Religious/social workers, Office Clerks and Health Professionals are lower than 15.5 percent. Legal Professionals, precision production, operators, craft, repair professionals and managers still appear to make up the largest proportion of overworking occupation among females. In other words, for females working in these three occupations, they have over 29 percent probability to work for over 50 hours per week. The results are same for males. Religious/social workers, office clerks and health professionals still have lowest probability to overwork for females, while natural, math and computing science Professionals seem easier for males than religious/social workers.

Figure 5: Prevalence of Overwork per Occupation(1=male, 2=female)



5 Empirical Analysis

For initial analysis, I control for occupation and year fixed effects. Note that for empirical analysis, we continue to use 301 occupations basis. Then I add the log earnings and education as the control variable manually. In this analysis, standard errors are clustered at the occupation level. Table 2 reports the results for initial analysis. Column (1) reports the coefficient of β only controlling for year and occupation fixed effects. In column (2), log earnings is added as another control variable. These two coefficients are both significant in 10 percent level, which means that as the share of male overworking in an occupation increases, the employment share of females with bachelor's degree and at least one child in the same occupation decreases.

Occupation-specific demand shocks might introduce noise to the overwork pattern and female occupational distribution in our result. In order to address this concern, I include controls for the share of individuals of the same age working in occupation i in other control groups. Here, I control for males, single females and married females without children in column (3), (4) and (5). The coefficients are essentially unchanged. Since the control group should face the similar demand shocks with our interest group, our result after controlling control group should reflect the lower willingness of married women with children remain in one occupation that requires long hours commitment to some extent. In sum, this result in accordance with our previous assumption: once overwork prevail in an occupation, college-educated females who have children are likely to "flee" from the occupation.

I include a "placebo test" to illustrate that this effect is likely to be driven by cost of working long hours, instead of other time-varying occupational factors or unobservable characteristics in married female group. Table 3 shows the results of placebo tests. Here, I consider 4 demographic groups (all are college-educated) married females age 25-40 without children, single females age 25-40, married females age 40-57 with children and males age 25-40. In all specifications, I include log wages and education year as control variables, controlling for year and occupation fixed effects. I also include the share of males at the same age and working in the same occupation as control group except the last demographic group. Overall, the coefficients of the overwork prevalence on occupational choices are smaller and not significant enough. This result proves our assumptions that only married college-educated females age 25-40 with children are significantly sensitive to the prevalence of overwork. In this test, four control have relatively less family responsibilities, especially in childcare. Single or married females without children have no commitment in childcare. The children of older married females are mostly in older age group, thus eliminating childcare pressure on their mothers. Males usually contribute less on the raising of. Thus, it indicates the overwork mainly reduce workers' desirability for individuals who have relatively large childcare responsibilities, because married females struggle the demands of parenthood with the intense working hours required to some occupations. Comparing Table 2 with Table 3 eliminates our concern about occupational demand shocks since they should have same effect on all population who are in the same occupation.

However, we still concern that possibility of reverse causality. Specifically, female-concentrated occupations (which has large proportion of females) may accommodate working hours in order to retain women workers. In this condition, it is hard to distinguish whether overwork leads to occupational choices or occupational distribution affect working hours demands. I include a ten-year lead of overwork variable in order to examine the reverse causality. If the overwork prevalence leads females to leave this occupation, the lead of the overwork variable at time t+1 should not be related to occupational distribution at time t.

Table 4 shows reverse causality tests results. Column (1) indicates a coefficient of -0.172, which is similar with the result -0.142 in Table 2 Column (3). In Column (2), a lead of overwork variable for share of males age 25 to 54 who works over 50 hours is included. The coefficient of long hours working at time t is -0.147, while the coefficient of lead at time t+1 is -0.082, which is much smaller to contemporaneous effect. However, these coefficients are significant at 0.1 significance level. Then we control for single females and married females without children in occupation distribution. The coefficient of lead at time t+1 is -0.044 in Column (3) and -0.040 in Column (4), both of them are much smaller comparing with baseline estimations. Neither of them is significant. Thus I could define the exist of reverse causality to some extent.

6 Conclusions

In this paper, we use fixed effects model to analyze how the prevalence of overwork influence the occupational choice of college-educated females.

We found that the share of college-educated mothers in an occupation will decrease significantly as the magnitudes of overwork increases in an occupation. This result is robust when controlling for occupational distribution of partial similar demographic groups. And the possibility of reserve causality has been excluded.

Individuals with larger family responsibilities face larger cost for long hours working commitments. Married females with children are the most sensitive to such cost. Thus, they are more likely to switch occupations or exit labor force market in response to the prevalence of overwork in their original occupations.

However, one limitation of this analysis is that we are still not clear why some occupations have loner working hours demands and how these demands change over time and across occupation. For example, technological professionals tend to overwork in recent years, but it may correlate with technology development and globalization. One interesting direction would be to learn what are the significant attractions or deficiencies of long-working-hour occupations for females. And how to encourage more females in those occupations. For example, if the government and company provide more subsidies for married females, will they have strong desirability and motivation to pursue the career in this occupation? We plan to explore more directions in the future.

References

Aguiar, Mark and Erik Hurst. 2007. ìMeasuring Trends in Leisure: The Allocation of Time Over Five Decades." Quarterly Journal of Economics. Vol. 122(3): 969-1006

Pan, Jessica, 2015. "Gender segregation in occupations: the role of tipping and social interactions", J Labor Econ 33(2):365–408

Goldin C. A Pollution Theory of Discrimination: Male and Female Occupations and Earnings. 2002.

Cortes, Patricia Pan, Jessica, 2017. "Occupation and Gender," IZA Discussion Papers 10672, Institute for the Study of Labor (IZA).

Francine D. Blau Lawrence M. Kahn, 2017. "The Gender Wage Gap: Extent, Trends, and Explanations," Journal of Economic Literature, vol 55(3), pages 789-865.

Nir Jaimovich Henry Siu Guido Matias Cortes, 2017. "The End of Men and Rise of Women in the High-Skilled Labor Market," 2017 Meeting Papers 809, Society for Economic Dynamics.

John Pencavel, 2016. "Whose Preferences Are Revealed In Hours Of Work?," Economic Inquiry, Western Economic Association International, vol. 54(1), pages 9-24, January.

Wolf, Elke. (2002). Lower wage rates for fewer hours? A simultaneous wage-hours model for Germany. Labour Economics. 9. 643-663. 10.1016/S0927-5371(02)00055-6. Jessica Pan Patricia Cortes, 2016. "Prevalence of Long Hours and Skilled Womenâ™s Occupational Choices," Working Papers id:11408, eSocialSciences.

Cortes, Patricia Pan, Jessica, 2017. "Occupation and Gender," IZA Discussion Papers 10672, Institute for the Study of Labor (IZA).

Occupation	share of male	share of female	share of overwork	share of male in overwork	share of female in overwork
Business Prof	0.460099	0.539901	0.2185882	0.6363518	0.3636482
Educators	0.2491277	0.7508723	0.229625	0.3066219	0.6933781
Engineers,Architects	0.6939343	0.3060657	0.2127211	0.8071287	0.1928713
Health Prof	0.1576793	0.8423207	0.1077594	0.2345169	0.7654831
Legal Prof	0.5526839	0.4473161	0.4500994	0.6422261	0.3577739
Managers	0.4625329	0.5374671	0.3597306	0.5648806	0.4351194
Mathematical,Computer and Natural Prof	0.6722128	0.3277872	0.1763441	0.758665	0.241335
Office Clerks	0.3481325	0.6518675	0.121494	0.512761	0.487239
Precision Production, operators,craft,repair	0.6136082	0.3863918	0.4313402	0.6175908	0.3824092
Service Workers	0.7798913	0.2201087	0.2672101	0.8779661	0.1220339
Social and Religious Workers	0.2948041	0.7051959	0.151255	0.5196507	0.4803493
Social Scientists	0.280303	0.719697	0.1590909	0.4166667	0.5833333
Technicians	0.60114	0.39886	0.2748666	0.7050989	0.2949011
Writers,Artists and Athletes	0.4927764	0.5072236	0.1909548	0.5822368	0.4177632

Table 1: Descriptive statistics of overwork and occupation distribution

	Married Female sage 25-57 With Children				
Share of males age 25 to 57 working 50+ hours	-0.182*	-0.148**	0.142**	-0.122**	-0.149**
Share of Males of the same age working in occupation i			0.839***		
Share of Single Females of the same age working in occupation i				0.732***	
Share of Married Females without children of the same age working in occupation i					0.577***
Standard deviation of log male and female wages and education years		X	X	X	X
Occupation FE	X	X	X	X	X
Year FE	X	X	X	X	X

Table 2: Relationship between the prevalence of overwork and college-educated female occupational choices

	Share of College-educated Married Females age 25-40 with Children working in occupation i			
Share of males age 25 to 54 working 50+ hours	-0.172* (0.082)	-0.147* (0.077)	-0.144* (0.087)	-0.123* (0.081)
Share of males age 25 to 54 working 50+ hours, t+1		-0.082* (0.057)	-0.044 (0.031)	-0.040 (0.027)
Controlling for occupation distribution of:	Males	Males	Single Females	Married Females without Children
Average and standard deviation of log male and female wages and education years	X	X	X	X
Occupation FE	X	X	X	X
Year FE	X	X	X	X

Table 3: Reverse Causality Tests

Variable	Coefficient
Marital status	
Married, spouse present	0.0082
Married, spouse absent	-0.043
Separated	-0.1235
Divorced	-0.0458
Widowed	-0.1635
English language ability	
Does not speak English	0.0021
Yes, speaks very well	-0.0442
Yes, speaks well	0.0554
Yes, but not well	0.0065
Race, ethnicity	
Black/African American/Negro	0.0793
Chinese	0.0581
Other Asian or Pacific Islander	0.0950
Hispanic	0.0057
Veteran status	
Veteran	0.0092
U.S. census region	
Middle Atlantic Div.	-0.0035
Pacific Div.	-0.0112
East North Central Div.	-0.0156
West North Central Div.	-0.0210
South Atlantic Division	-0.0186
East South Central Div.	-0.0322
West South Central Div.	-0.0314
Mountain Division	-0.0055

Table 4: LASSO Feature Selection Result 1

Variable	Coefficient
Child	
Age 18 or younger	-0.0762
Age 4 or younger	0.0945
Usual hours worked per week	
50 to 59	-0.0094
60 to 69	-0.0024
>70	-0.0452
Years of education	-0.5249
log earnings	-0.7820
Experience	
exp	-0.0024
exp2	-0.0019
Field	
bachelor degree	-0.1719
master degree	-0.1301
Ph.D.	-0.1009
Poverty	0.0812
Household earnings	
household investment	-0.2981
househould insurance	0.1342

Table 5: LASSO Feature Selection Result 2

Variable	Coefficient
College major	
Agri	-0.0092
Envir/Nat Res	-0.0439
Archit	-0.0331
Area/Ethnic/Civiliz Stud	-0.0356
Comm	-0.0492
Comm Tech	-0.0492
Comp/Inform Sci	-0.0727
Cosmet Serv/Culin Arts	0.1671
Engin	-0.0535
Engin Techn	0.0190
Ling/Foreign Lang	-0.0234
Fam/Consum Sci	-0.0398
Law	-0.0952
English/Lit/Compos	-0.0112
Lib Arts/Hum	-0.0332
Lib Sci	-0.0553
Bio/Life Sci	-0.0827
Math/Stats	-0.0564
Milit Techn	-0.0534
Inter-/Multi-Disc Stud (gen)	-0.0832
Phys Fit/Parks/Recr/Leis	0.0166
Philos/Rel Stud	0.0234
Theol/Rel Voc	0.0267
Phys Sci	-0.0613
Nucl/Ind Rad/Bio Techn	0.0871
Psych	-0.0711
Crim Just/Fire Prot	-0.071
Publ Aff/Policy/Soc Wo	-0.0701
Soc Sci	-0.0324
Constr Serv	-0.0867
Electr/Mech Rep/Techn	-0.1051
Transp	0.1074
Fine Arts	-0.0378
Med/Hlth Sci Serv	-0.0121

Table 6: LASSO Feature Selection Result 3