Two Small Empirical Exercises

Hazel Chui

November 2020

To demonstrate some of the empirical skills that I have acquired through both undergraduate education and self-learning, I conducted two small empirical exercises drawing on data from some publicly available datasets. It is divided into two parts, namely fixed effect regressions and difference in differences estimation.

I performed the following analysis in the exact same manner using two different statistical packages, Stata and R. The codes can be found on my GitHub: https://github.com/hazelchui/Empirical_Exercises/blob/main/R%20Code.R for R and https://github.com/hazelchui/Empirical_Exercises/blob/main/Stata%20Code.do for Stata.

1. Fixed Effects Regressions – Relationship between Marriage and Wages for Females

This section addresses one of the heated debates in Sociology – how marital status affects wages. A large body of literature has been examining whether married men earn more than their unmarried counterparts. Many earlier studies have discovered the male marital wage premium while more recent studies have argued that there is no causal effect of marriage on wages (Ludwig & Bruderl, 2018; Killewald & Lundberg, 2017). For example, Ludwig and Bruderl (2018) explain that married men earn more since those who have higher wage levels and wage growth self-select into marriage. Despite abundant findings on the effects of marriage on wages among men, there is a lack of similar studies focusing on women.

In this article, using panel data of National Longitudinal Surveys data on young women aged 14-26 years in 1968, I analyse the marriage effects on wages for women. It traces 5159 women from 1968 to 1988 with occasional gaps.

1.1.Graphical Analysis

Figure 1 plots the trend of (natural) log real wages by marital status, separated by race. As we can observe from the figure, it is ambiguous whether females enjoy the wage premium after marriage. While the unmarried White women seem to earn more income in the later part of the sample period, the results are generally mixed when we consider both racial groups and the whole sample period.

Figure 1: Wage Trends – by Race and Marital Status

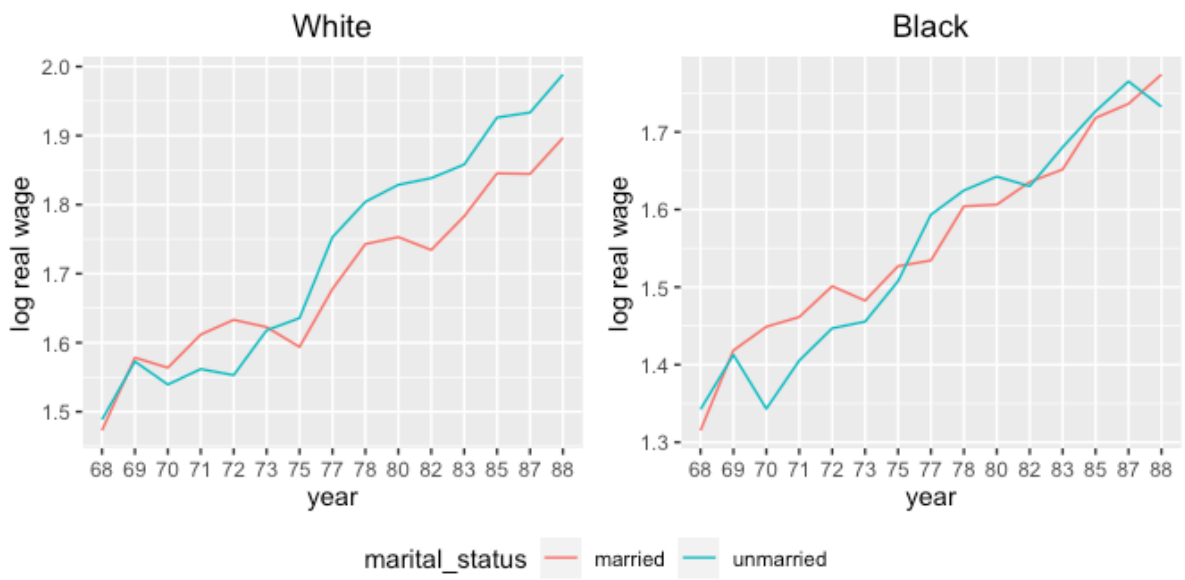
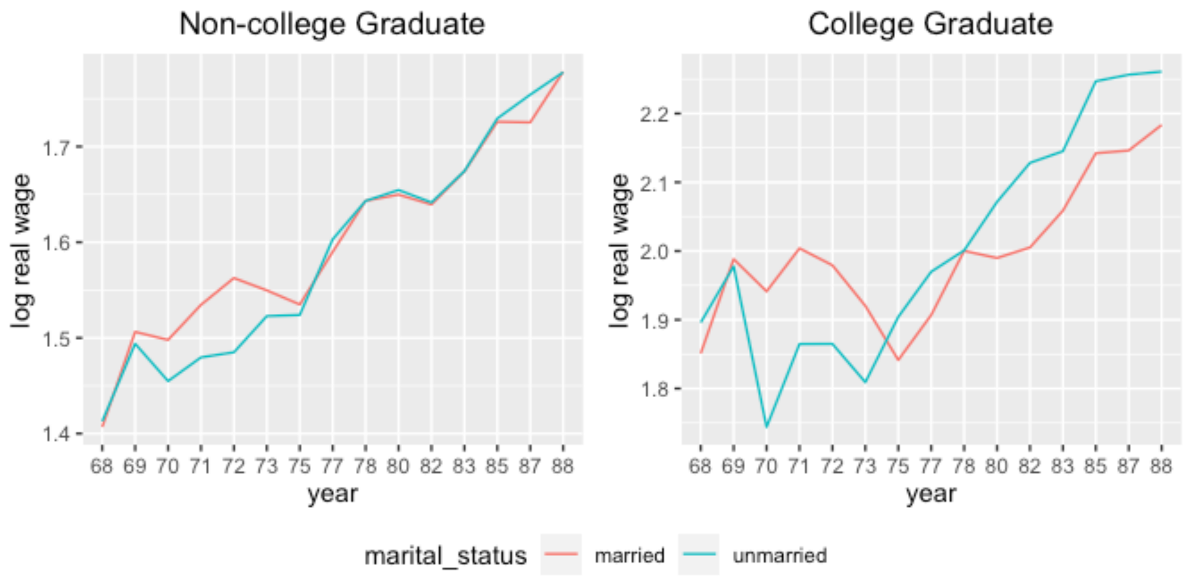


Figure 2 depicts the trend of log real wage by completion of college education and by marital status. Similarly, in the later part of the sample period, the married women seem to suffer from a negative wage gap relative to the unmarried, particularly for college graduates. Figure 2 also conveys an unclear conclusion on the relationship between marriage and wages.

Figure 2: Wage Trends – by College Education and Marital Status



The above graphical analysis only controls for race and college education respectively. To control for a set of different effects simultaneously, I proceed to the following regression analysis.

1.2. Regression Analysis

To examine the relationship between marriage and wages more precisely, I propose the following model:

$$ln(wage_{it}) = \theta \cdot marriage_{it} + x_{it}'\beta + \alpha_i + \gamma_t + \epsilon_{it} \text{-------} \; (1)$$

where $ln(wage_{it})$ is the log real wage of the individual i in year t, and $marriage_{it} = 1$ if i is married in year t, and 0 otherwise. The vector x_{it} is a collection of control variables which will be explained further below. α_i and γ_t refer to the individual and year fixed effects respectively. Table 1 summarises the regression results for this section:

<u>Table 1: Estimation Results of Equation 1</u>

Table 1 summarises the estimation results of equation 1. Standard errors are reported inside the parentheses: robust standard errors in the curly brackets {}, individual-clustered standard errors in the round brackets (), and bootstrapped standard errors in the square brackets. Statistical significance is indicated by the stars, with ***:$p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

	(1)	(2)	(3)	(4)	(5)
$marriage$	-0.022 {0.006***} [0.006***]	-0.015 (0.008*) [0.07**]	-0.014 (0.008*) [0.006**]	-0.015 (0.008**) [0.007**]	-0.014 (0.008*) [0.008*]
age	0.003 {0.001***} [0.005***]	0.010 (0.001***) [0.001***]	0.029 (0.011***) [0.012**]	0.010 (0.001***) [0.001***]	0.027 (0.011**) [0.012**]
$tenure$	0.032 {0.001***} [0.008***]	0.018 (0.001***) [0.001***]	0.018 (0.001***) [0.001***]	0.017 (0.001***) [0.174***]	0.017 (0.001***) [0.017***]
$hours$	0.001 {0.000***} [0.000***]	-0.002 (0.001***) [0.001***]	-0.002 (0.001***) [0.001***]	-0.003 (0.001***) [0.001***]	-0.003 (0.001***) [0.001***]
$union$	0.163 {0.007***} [0.007***]	0.101 (0.010***) [0.010***]	0.103 (0.010***) [0.010***]	0.088 (0.009***) [0.101***]	0.089 (0.009***) [0.089***]
Industry code				✓	✓
$black$	-0.148 {0.006***} [0.006***]				
$other\ race$	0.069 {0.031**} [0.029**]				
$collegegrad$	0.360 {0.008***} [0.007***]				
Intercept	1.445 {0.025***} [0.027***]				
Individual FE		✓	✓	✓	✓
Year FE			✓		✓

A natural approach to the question is to regress the variables of interest, $ln(wage_{it})$, $marriage_{it}$ together with various potential control variables that may affect wages using the pooled ordinary least square (POLS). Control variables include age, job tenure in years, number of usual working hours, a binary variable $union$ indicating whether the individual has joined a labour union, race (white compared to black and other races) and college education.

Column (1) shows the results of this POLS. Unlike men enjoying a wage premium after being married, married women actually earn less than the unmarried. The negative coefficient associated with $marriage$ is highly significant under both robust and bootstrapped standard errors. Furthermore, the coefficients of controls also confirm results previously found in literature: with other factors being constant, a black person earns less than a white person; there is a huge and significant return to college education; and joining a labour union can also help improve wages.

In the POLS, I am only able to control individual characteristics of race and college education. Yet, there are a lot of unobserved traits which may affect wages. Therefore, I estimate an individual fixed-effects regression which is reported in column (2). Race, college education and the intercepts are dropped to avoid perfect collinearity. The (individual-) clustered standard errors are reported to take into account the possible correlated observations within each individual. I also report the bootstrapped standard errors to provide inference without assuming any statistical structure. As revealed in column (2), the negative wage gap for married females is still statistically significant, confirming the results in POLS.

Although the time effects on wage are less of a concern than individual effects, I have also included the two-way fixed-effects model using the same set of variables and the results are reported in column (3). The interpretation is essentially the same as that in column (2).

A potential challenge to the above analysis is that there may be industry effect on wages – people in a particular industry may earn higher wages systematically than those in other industries[1]. In view of this, I redo the time and two-way fixed-effects estimation with additional controls on industry codes, which are reported in column (4) and column (5) respectively. It is

[1] This specification is valid since people can switch industry during the sample period.

demonstrated that the negative wage gap for married women is still significant in 5% and 10% respectively.

Due to limited availability of data, I am not able to include all possible control variables in the model. Also, I would expect more accurate estimates and powerful tests if a more sophisticated dataset with abundant data points were available.

2. Difference in Differences – Effects of Training Subsidies on Job Trainings

Quasi-experimental designs are widely used in social sciences and are gaining popularity in sociology (Gangl, 2010). In this section, I try to use one of the major quasi-experimental designs, difference in differences (DiD), to estimate the treatment effects of training subsidies on hours of job trainings firms provide to their employees.

The dataset contains a panel of 157 Michigan firms in 1987, 1988 and 1989. Some firms received training subsidies in 1988 or 1989, while other firms received no treatment (i.e., no training subsidies). Observing the average hours of training per employee of each firm in each of the sample years, I estimate the DiD equation:

$$training_{igt} = \gamma_i + \gamma_t + \theta \cdot grant_{gt} + \sum_s \beta_s (firm_{si} \times t) + \epsilon_{igt} \text{ ------- } ②,$$

where i refers to firm i, g refers to the treatment status (i.e., receive training subsidies or not), and t stands for time. Therefore, $training_{igt}$ is the average training hours per employee in firm i in year t; $grant_{gt} = 1$ if the firm belongs to the treated group (i.e., receive training subsidies) in year t, and 0 otherwise. Since I have panel data at the firm level, I am able to do the firm level fixed effects γ_i, absorbing the treatment status fixed-effects standard in DiD and controlling for firm level unobserved traits at the same time. γ_t is the time fixed-effects standard in DiD. The DiD coefficient θ captures the desired treatment effects.

As in every DiD estimation, the parallel trend assumption is crucial. With observations of 157 firms and a 3-year panel, I am able to model a firm-specific linear trend $\sum_s \beta_s (firm_{si} \times t)$

where $firm_{si} = 1$ if $s = i$ and 0 otherwise. Therefore, β_i is the trend for firm i if it is not treated. Table 2 summarises the results of equation 2:

Table 2: Estimation Results of Equation 2

Table 2 summarises the estimation results of equation 2. Firm-clustered standard errors are reported inside the parentheses. Statistical significance is indicated by the stars, with ***:$p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

	(1) Without firm-specific trend	(2) With firm-specific trend
$grant$	30.10	25.86
	(4.18***)	(6.12***)

Similar to Section 1, I report the (firm-) clustered standard errors to consider the possible correlated observations within each firm. The results here show that subsidies increase hours of job trainings significantly: employees in the treated firms on average receive 30.1 or 25.86 hours more job trainings than those in the untreated firms, without or with firm-specific trend being modelled respectively.

3. References

Angrist, J. D., & Pischke, J. (2015). *Mastering 'metrics: The path from cause to effect*. Princeton, NJ: Princeton University Press.

Hanck, C., Arnold, M., Gerber, A., & Schmelzer, M. (2020). Introduction to Econometrics with R. Retrieved from https://www.econometrics-with-r.org/

Gangl, M. (2010). Causal Inference in Sociological Research. *Annual Review of Sociology, 36*(1), 21-47.

Holzer, H. J., Block, R. N., Cheatham, M., & Knott, J. H. (1993). Are Training Subsidies for Firms Effective? The Michigan Experience. *ILR Review, 46*(4), 625-636.

Killewald, A., & Lundberg, I. (2017). New evidence against a causal marriage wage premium. *Demography*, *54*(3), 1007-1028.

Ludwig, V., & Brüderl, J. (2018). Is There a Male Marital Wage Premium? New Evidence from the United States. *American Sociological Review, 83*(4), 744-770.

Rodríguez, G. (2019, September). Stata Tutorial. Retrieved November 08, 2020, from https://data.princeton.edu/stata