“The Wage Differential among Males and Females"

Estimations for HW2

By

SHUKHRAT KHUSEYNOV

Department of Economics
Bilkent University

In Partial Fulfilment of the Requirements for Econ 426

December 2018

I. Introduction

The difference in the wages of males and females has always been an interesting question in economics, especially for empirical studies. It is widely believed that women get paid less than men on average, given the same conditions. In the estimations of Cornwell and Rupert (1988), which were also reproduced by Baltagi and Khanti-Akkom (1990), negative impact of female binary variable in all their models can be observed. Furthermore, Kim and Polachek (1994) studied this issue in detail in one of their research papers, showing that the perceptible wage inequality favoring men is a result of distinct human capital characteristics, work discontinuity and external factors. Here comes the importance of work experience. Therefore, it is critical to analyze the differences in wages with regards to experience and gender of an individual to clarify the case. Econometrically, many factors may cause distortions to the original OLS model, such as heterogeneity and endogeneity, which violate the Gauss-Markov assumptions. Hence, many different models and tests should be applied to find both unbiased and consisted results. Besides, it is also necessary to conduct the Chow tests for different samples of males and females to observe the difference.

II. Data

The data used in the paper is Panel Study of Income Dynamics (PSID), which is cross-section data gathering of approximately 5000 families in USA starting from 1968 through 1987. The group of people was asked similar questions continuously throughout the period. Since some people were dying and due to missing data, the panel data was not perfectly balanced. Thus, although more observations were collected, only a sample of 7 year period of 595 individuals, making up 4165 observations in total, was employed in the paper. Moreover, several smaller samples were also used to divide the population to males and females in order to differentiate them and complete several tests.

In the data, ID and T variables stand for individual and time period, respectively, LWAGE represent the logarithm of wage of an individual at a particular time period, EXP stands for work experience, EXP2 is experience taken to the power of 2, WKS is a number of weeks worked, OCC is a dummy variable for occupation with blue collar being equal to 1, IND is a dummy being equal to 1 for manufacturing industry, SOUTH is another binary variable for region with south being equal to 1, SMSA is also a dummy with residing in a city being equal to 1, MS stands for a dummy being equal to 1 for married individuals, and UNION is a binary variable being equal to 1 for those with a wage set by union contract. In addition, several time invariant dummy variables were also deployed, where FEM is a dummy representing gender with female being equal to 1, BLK is also a dummy being equal to 1 if the individual is black, and ED stands for the years of education.

Furthermore, there are 528 men and 67 women observed in the sample within the given time period. As seen in the Table 1, the averages of work experience and weeks worked of males is noticeably higher than those of female individuals with females having more variability in the number of weeks worked. Additionally, compared to females, male individuals seem to be considerably more working in manufacturing industry, as a blue collar, and with a union contract wage. On the other hand, female respondents are observed to be more inclined to reside in the south region and in the cities. Most of black respondents are females. Remarkably, 91% of male individuals are

Table 1: Descriptive statistics

	Whole Sample				Male				Female			
Variables	Mean	Standard Dev.	Min	Max	Mean	Standard Dev.	Min	Max	Mean	Standard Dev.	Min	Max
EXP =Work experience,	19.85	10.97	1	51	20.21	10.99	1	51	17.014	10.33	1	50
WKS =Weeks worked,	46.81	5.13	5	52	46.96	4.98	5	52	45.61	6.03	11	52
OCC =Occupation, 1 if blue collar,	0.51	0.50	0	1	0.53	0.50	0	1	0.39	0.49	0	1
IND =1 if manufacturing industry,	0.40	0.49	0	1	0.43	0.49	0	0	0.15	0.36	0	1
SOUTH =1 if resides in south,	0.29	0.45	0	1	0.28	0.45	0	1	0.36	0.48	0	1
SMSA =1 if resides in a city (SMSA),	0.65	0.48	0	1	0.64	0.48	0	1	0.79	0.41	0	1
MS =1 if married,	0.81	0.39	0	1	0.91	0.28	0	1	0.026	0.16	0	1
FEM =1 if female,	0.11	0.32	0	1	0	0	0	0	1	0	1	1
UNION =1 if wage set by union contract,	0.36	0.48	0	1	0.38	0.49	0	1	0.21	0.41	0	1
ED =Years of education,	12.85	2.79	4	17	12.85	2.82	4	17	12.84	2.50	7	17
BLK =1 if individual is black,	0.072	0.26	0	1	0.053	0.22	0	1	0.22	0.42	0	1
LWAGE=Log of wage.	6.68	0.46	4.61	8.54	6.73	0.44	5.02	8.54	6.26	0.42	4.61	7.28
ID= Cross section identifier,	298	171.78	1	595	297.73	171.75	1	594	300.13	172.24	4	595
T=time dimension	4	2.00024	1	7	4	2.00027	1	7	4	2.0021	1	7

Table 2: Estimation Results

Explanatory Variables	OLS-Pooled		With-in group Fixed Effect (Mean Diff)		Between-Group		Random Effects (RE) (GLS)	
EXP	0.039*** (17.49)	0.0401*** (18.56)	0.11*** (27.95)	0.11*** (27.95)	0.030*** (6.13)	0.032*** (6.68)	0.083*** (20.71)	0.082*** (20.43)
EXP2	-0.00065*** (-13.10)	0.000048*** (-14.04)	-0.00042*** (-5.07)	-0.00042*** (-5.07)	-0.00053*** (-4.93)	-0.00057*** (-5.39)	-0.0008*** (-8.92)	-0.00081*** (-9.03)
WKS	0.0052*** (4.34)	0.0042*** (3.68)	0.00084 (0.97)	0.00084 (0.97)	0.011*** (3.02)	0.0092** (2.55)	0.0011 (1.17)	0.0010 (1.10)
OCC	-0.13*** (-8.54)	-0.14*** (-9.36)	-0.021 (-1.13)	-0.021 (-1.13)	-0.16*** (-4.57)	-0.17*** (-4.96)	-0.045** (-2.21)	-0.050** (-2.41)
IND	0.060*** (4.92)	0.047*** (3.89)	0.019 (0.85)	0.019 (0.85)	0.067*** (2.57)	0.058** (2.27)	0.012 (0.50)	0.0037 (0.16)
SOUTH	-0.058*** (-4.43)	-0.056*** (-4.36)	-0.0019 (-0.02)	-0.0019 (-0.02)	-0.060** (-2.23)	-0.057** (-2.20)	-0.016 (-0.34)	-0.017 (-0.36)
SMSA	0.146*** (11.77)	0.152*** (12.54)	-0.042 (-1.44)	-0.042 (-1.44)	0.17*** (6.53)	0.18*** (6.82)	-0.025 (-0.81)	-0.014 (-0.46)
MS	0.26*** (16.12)	0.048** (2.36)	-0.030 (-1.11)	-0.030 (-1.11)	0.32*** (9.43)	0.11** (2.41)	-0.018 (-0.68)	-0.075*** (-2.72)
UNION	0.101*** (7.95)	0.093*** (7.50)	0.033 (1.31)	0.033 (1.31)	0.12*** (3.89)	0.11*** (3.73)	0.068*** (2.69)	0.063** (2.54)
ED	0.059*** (21.28)	0.057*** (20.77)	0.14*** (10.42)	0.14*** (10.42)	0.053*** (9.33)	0.051*** (9.26)	0.102*** (12.39)	0.100*** (12.42)
FEM		-0.37*** (-15.98)		-0.13 (-1.12)		-0.32*** (-5.79)		-0.34*** (-5.38)
BLK	-0.19*** (-9.02)	-0.17*** (-8.03)	-0.31* (-1.78)	-0.28 (-1.62)	-0.17*** (-3.68)	-0.16*** (-3.51)	-0.28*** (-3.03)	-0.21** (-2.54)
const	4.98*** (66.83)	5.25*** (70.52)	4.65*** (59.60)	4.65*** (59.60)	4.82*** (23.75)	5.12*** (25.07)	4.13*** (30.02)	4.26*** (31.36)
N	4165	4165	4165	4165	4165	4165	4165	4165
Cross Section units	595	595	595	595	595	595	595	595
Time series units	7	7	7	7	7	7	7	7
Cross Section Fixed Effects	No	No	Yes	Yes	No	No	No	No
Time Series Fixed Effects	No	No	No	No	No	No	No	No
R-square (with-in)			0.6581	0.6581	0.0425	0.0763	0.6188	0.6124
R-square (between)			0.0261	0.0261	0.5180	0.5443	0.2068	0.2539
R-square (overall)	0.3991	0.4286	0.0461	0.0461	0.3854	0.4173	0.2131	0.2512
F-stat / Wald test (overall significance)	238.02	256.02	377.62	377.62	56.96	57.93	1540.96	1528.82
MODEL COMPARISONS								
F-stat (OLS vs FE)			38.25	38.25				
Hausman c^2								
$b_{RE} = b_{MD}$							5160.84	5075.25
$b_{RE} = b_{BE}$							24889.6	121.77
Breusch Pagan LM test (OLS ve RE)							3585.50	3497.02
Tests of GM assumptions								
Heteroscedasticity Correction	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Type of correction het. correction	Robust	Robust	Cluster id	Cluster id	none	none	Cluster id	Cluster id

Earnings function coefficients, estimated with PSID data. All regressions are corrected for cross section heteroscedasticity. Pooled OLS data is estimated without panel data properties. Fixed Effect models have two versions: (i) variables defined as Mean Difference (ME), (ii) variables defined as First Difference (FD). Between- Group estimator has no cross section variation (BE). GLS estimator is the Random Effects (RE) estimator. In all regressions robust heterocedasticity consistent error term are used in t-stat computations. The test for model comparisons are conducted in models without any heteroscedasiticity corrections. Different R^2 are reported when available. ***, **, * shows statistical significance at 1%, 5% and 10% levels respectively.

married, whereas among female individuals only 2.6% are married, demonstrating a likely negative correlation between work and marital status of women. The education level of working males and females is not significantly different in the sample (see Table 1). The percentage change in wages (LWAGE) depicts a difference between those of men and women, which has to be clarified in the estimation part.

III. Estimation

The estimation part of the paper consists of numerous regression methods, several hypotheses testing, and Chow tests. In the general model, the logarithm of average labor income (LWAGE) was taken as a dependent variable. Among the independent variables work experience and its square, weeks worked, and several individual variables, such as region, residence, and industry indicators together with time invariant variables, which are education level, race and gender dummy variables, were placed. Additionally, parameter alpha that represents the effect of unobserved characteristics being constant over time and the error term are on the right-hand side of the equation. In order to observe the difference between genders, each regression model was estimated twice: once with a female dummy and once without it. In general, among the main econometric problems of panel models are endogeneity and heterogeneity. The former, endogeneity, can be cured by using instrumental variables but is neglected in this paper for the sake of simplicity. To remedy the heterogeneity, various models were applied and tested to choose the most unbiased and consistent results.

First of all, Pooled Ordinary Least Squares (OLS) method with robust standard errors, to adjust for heteroskedasticity, was implemented to compare other models with its output (see Table 2). The observed coefficients are mostly significant, showing positive impact of work experience on individuals' wages. After adding a female binary variable in the second model it can be observed that the females are expected to get lower wages by 0.37% according to the Pooled OLS. After that, to adjust for heterogeneity issue, different models, such as Fixed Effects (Mean deviated model), the Between-Group Estimator model and also the Random Effects model (GLS) were analyzed. For most of them, except the Between-Group model, robust errors of clustering adjustment were utilized. In the Mean Deviated Fixed Effects model, as depicted in Table 2, although three time invariant variables were dropped to avoid the multicollinearity, they were estimated as a second stage by regressing the residuals on the dropped variables. Before the second stage, only work experience and its square was significant with acceptable error amount of 5%, where the coefficient of experience increased from around 0.04 in OLS to 0.11, which is a huge difference. Moreover, among the time invariant variables estimated later, only the education level is significant. Apparently, all dummy variables in this model are estimated with insignificant coefficients, as seen in the table above. Albeit insignificant, the female dummy demonstrates a decline compared to OLS model. In the Between Group Estimator model, almost all coefficients are significant (see Table 2). The experience coefficients of this model are a bit lower than those of OLS, also demonstrating a positive impact of work experience on wage. Similarly, the female binary variable is near to its OLS counterpart, being slightly more and implying that females get paid less than males. Lastly, the Random Effects model has both many significant and insignificant coefficients in the estimated regression. The work experience coefficient is both positive and significant at 5% significance level with higher magnitude than in OLS and Between-Group models, but considerably less than in Mean Deviated model. The female dummy is also significant and follows a common trend with -0.34 being close to Fixed Effects and OLS models. The implications of work experience coefficient and female dummy are same throughout all models. To choose the best model, several hypotheses tests have to be conducted.

Since the Between-Group model uses averages of each individual in regression analysis making it different, it was not compared with Pooled OLS. All other models were compared by using several testing methodologies. As reported in Table 2, when Pooled OLS is compared with the Fixed Effects model by using F test, the null hypothesis of having no observable difference between the two is rejected with 5% acceptable error amount. In other words, there is a fixed effect in this framework and Pooled OLS is inferior to that model. Moreover, the Random Effects model was also tested against OLS by using Breusch-Pagan Lagrangian Multiplier test (see Table 2). The test concluded that Random Effects model is far better than Pooled OLS. To compare the Random Effects model with others, Hausman specification test was conducted. While comparing with Between-Group Estimator model, the null of having same coefficient was rejected, meaning that the results of both are different (see Table 2). When compared with Fixed Effects model, the null hypothesis was similarly rejected. Knowing that the Fixed Effects model is consistent, albeit inefficient, with different results the Random Effects model is accepted as inconsistent. With Between-Group model being dubious, the implications of Mean Deviated Fixed Effects model should be deployed.

Finally, to observe the difference between the genders using separate samples, the Chow tests were conducted. In most of the tests, except for Random Effects model where χ^2 test was used, F test for comparing restricted and unrestricted models was employed.

Table 3: Estimation Results for Male and Female and Whole samples and Chow Tests

Dependent Variable: Log(wage)															
Explanatory Variables	OLS-Pooled			With-in group Fixed Effect						Between-Group			Random Effects (GLS)		
				Mean Diff			First Diff (optional)								
	M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
$H_o:b_{Mj}=b_{Fj}=b_j$															
Sum of Square Residuals	466.225	35.826	532.977	not reported	not reported	not reported	-	-	-	not reported	not reported	not reported	not reported	not reported	not reported
Chow Test	n=3696	n=469	n=4165 F test			F test			-			F test			chi_sq test
F (df1, df2) or χ^2, critical value or p-value			23.20, 1.83 (α=5%)			6.20, 0.0000 (p-value)			-			1.10, 0.3630 (p-value)			30.67, 0.0003 (p-value)
Test Decision			Reject Ho			Reject Ho			-			Do not Reject Ho			Reject Ho

Earnings function coefficients, estimated with PSID data. All regressions are corrected for cross section heteroscedasticity. Regresssion are conducted for Male (M), Female (F), and Whole (W) sub groups. Pooled OLS data is estimated without panel data properties. Ficed Effect models have two versions: (i) variables defined as Mean Difference, (ii) variables defined as First Difference. Between- Group estimator has no cross section variation. GLS estimator is the Random Effects estimator. The SSR used in the Chow tests are models **no**t corrected for heteroscedasticity. *SSR m is the regression for the male group, SSR_f is the regression for the female group, SSR_tot is the regression for both groups at the same time. The Unrestricted model is SSR_m+SSR_f and the Restricted model is SSR_tot.*
In Random effects model all regressions are conducted with robustness correction, hence you may obtain SSR as e(RSS) *and conduct the test but the test might be biased.*

As demonstrated in Table 3, in most Chow tests, except for Between-Group model, the null hypothesis of having equivalent samples in terms of coefficients was rejected. It means that both males and female individuals are assumed to have different wages and several other parameters.

IV. Conclusion

In a nutshell, the data and tests conclude that there is a significant positive effect of work experience on wages, which was justified by all implemented models. The gender issue, on the other hand, is challenging. After using robust standard errors and accounting for heterogeneity, the hypotheses testing indicated the Fixed Effects as a consistent model, where female dummy variable was not significant at 5% acceptable error amount. However, the Chow tests reported that the samples of males and females have distinguishable results. Consequently, the track of negative effect of gender on women's wages is observed but has to be proven by gathering more data.

V. References

Baltagi, B., & Khanti-Akom, S. (1990). On Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators. *Journal of Applied Econometrics, 5*(4), 401-406. Retrieved from http://www.jstor.org/stable/2096482

Cornwell, C., & Rupert, P. (1988). Efficient Estimation With Panel Data: An Empirical Comparison of Instrumental Variables Estimators. *Journal of Applied Econometrics, 3*(2), 149-155. Retrieved from http://www.jstor.org/stable/2096586

Kim, M., & Polachek, S. (1994). Panel Estimates of Male-Female Earnings Functions. *The Journal of Human Resources, 29*(2), 406-428. doi:10.2307/146104