**EXERCISE 1**

**a)**

is the coefficient of the time dummies . It captures the aggregate shocks for a given year in the sample – shocks that affect all workers’ wages in that particular year.

The parameter , on the other hand, captures time-invariant characteristics of workers that affect wages and that possibly correlate with explanatory variables. It is important to include this term to try to avoid omitted variable bias problems.

**b)**

The sign of should be positive – this would mean that being unionized leads to an increase in the worker’s wage, on average, all else constant.

**c)**

We want to estimate the following equation:

Where the variables d81 through d87 are dummy variables that take the value 1 if the observation reports to that year (t=1981, t=1982, …, t=1987). Example: if the observation reports to 1984 (t=1984), d84 = 1, and all the other dummies will equal 0. The base category is t = 1980, occurring when d81 through d87 are zero.

And our coefficient of interest is .

Computing a t-test for individual significance of the coefficient:

The test statistic: ~

The critical value for t is +- 1.960 (this is a two-tailed test).

We get, from our regression output, t= 6.65 > |1.960|, so we reject the null hypothesis and conclude for statistical significance at a 5% level.

This could also be directly concluded from looking at p-value for this variable (since the p-value is 0.00 [smaller than 0.05] we reject the null).

Given this, there is statistically significant evidence for a positive union effect – being part of a union has a positive impact on the worker’s wage.

The assumptions required for the Pooled OLS model to be consistent are two. Given

1. Weak exogeneity:

* This assumption seems plausible. We are essentially saying that there is no correlation between the variables of the vector and the transitory (or random) part of the error term.

1. Uncorrelated unobserved effect:

* It is harder to find this assumption plausible. For instance, we can easily have endogeneity due to an unobserved effect of ability on labor market experience, which would simultaneously correlate with , thus violating this assumption. This can lead to an asymptotic underestimation of the union effect.

- We should also consider that observations for the same individual are likely correlated with each other – so it is not optimal to weigh them equally. This also points towards Pooled OLS inconsistency.

**d)**

We want to estimate the following regression:

With

This simplifies to:

The first thing we should observe, is that with our First Differences estimation, we lose one time period. We had 8 time periods initially (1980 to 1987), and we now have only 7 time periods (1981 to 1987). We excluded one time-dummy (d81) to avoid perfect collinearity.

We can also see by the model that all time-invariant variables disappear – here they correspond to all the variables from the vector . As such, we can conclude that it is not possible to estimate, using First Differences, the returns on education or race.

As for experience, even though it is a time-variant variable, we should be cautious as in this equation we do not have the variable in levels, but in differences. When the variable experience is first-differenced, it becomes a vector of 1s – and this is not very informative. It is more appropriate to get information from , and refrain from interpreting the estimates on the experience variable itself.

In fact, if we included a constant in the model, would be perfectly collinear and disappear. This is one more reason to refrain from interpreting estimates on it from our equation.

**e)**

We get an estimate result for the union effect of 0.0411497. However, as can be read from the p-value (0.061), the union effect is not statistically significant at a 5% significance level.

So we can see that the union effect is not statistically significant here, as opposed to the results found when estimating through Pooled OLS. What this suggests is that there probably exists a significant correlation between and , indicating the relevance of unobserved time-invariant heterogeneity that may be correlated with the explanatory variables.

**f)**

The time-varying variables are the ones associated to vector .

For the within estimator (fixed effects), we need to estimate the following equation:

The Random Effects estimation equation is the following:

(Note: for simplicity, from now on we are writing only the vector and not decomposing it in its different variables, as we have already done so before and repeating it is time-consuming)

With , T= 1, … , 8

And

(Note: The parameter here designed by is usually designed . However, since we already have a in our equation, we chose to avoid confusion).

Necessary assumptions for consistency of the Random Effects estimator:

1. Strict Exogeneity
2. Uncorrelated unobserved effects

**g)**

The Hausman test for this specific case will compare Fixed Effects and Random Effects estimators, and if the null hypothesis holds, both provide consistent estimates. Since Pooled OLS is also consistent under the Random Effects model assumptions, if the Random Effects is consistent, so is the Pooled OLS model.

This means that, under , the individual effects are uncorrelated and the Random Effects estimator as well as the Fixed Effects estimator give us consistent estimates.

Under , only the Fixed Effects estimator gives us consistent estimates.

Hausman Test Statistic:

The critical value for is 0.711, and our test statistic equals 77. The associated p-value is 0.00. Given this, we reject the null hypothesis at a 1% significance level.

This means that the Random Effects estimator is inconsistent. This means that the Pooled OLS estimator is also inconsistent. The most appropriate estimator is the Fixed Effects estimator.