Assignment 1: Link to Collab file: Assignment 3_MBA753M

Q1. Use the Card & Krueger data to study the effect of an increase in minimum wage. Consider two covariates (or control or predictor variables) – whether the outlet is owned by a company and the number of hours open per day. [20 marks]

<u>Output of DID design:</u>

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Employment   R-squared:                       0.092
Model:                            OLS   Adj. R-squared:                  0.086
Method:                 Least Squares   F-statistic:                     15.19
Date:                Sat, 21 Sep 2024   Prob (F-statistic):           3.12e-14
Time:                        10:38:41   Log-Likelihood:                -2623.5
No. Observations:                 757   AIC:                             5259.
Df Residuals:                     751   BIC:                             5287.
Df Model:                           5
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
Intercept           3.2754      1.333      2.457      0.014       0.659       5.892
CompanyOwned       -1.0600      0.598     -1.774      0.077      -2.233       0.113
HoursOpen           0.5244      0.066      7.908      0.000       0.394       0.655
Treatment          -1.4339      1.293     -1.109      0.268      -3.972       1.104
After              -2.8678      0.995     -2.882      0.004      -4.821      -0.914
Treatment_After     4.1158      1.431      2.876      0.004       1.307       6.925
==============================================================================
Omnibus:                      178.665   Durbin-Watson:                   2.038
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              369.313
Skew:                           1.314   Prob(JB):                     6.38e-81
Kurtosis:                       5.191   Cond. No.                         105.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

<u>Falsification test using Soda</u>

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   Soda   R-squared:                       0.223
Model:                            OLS   Adj. R-squared:                  0.218
Method:                 Least Squares   F-statistic:                     43.18
Date:                Sat, 21 Sep 2024   Prob (F-statistic):           3.63e-39
Time:                        10:39:49   Log-Likelihood:                 837.42
No. Observations:                 757   AIC:                            -1663.
Df Residuals:                     751   BIC:                            -1635.
Df Model:                           5
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
Intercept           0.8857      0.014     64.280      0.000       0.859       0.913
CompanyOwned        0.0122      0.006      1.982      0.048       0.000       0.024
HoursOpen           0.0058      0.001      8.476      0.000       0.004       0.007
Treatment           0.0154      0.013      1.154      0.249      -0.011       0.042
After               0.0893      0.010      8.684      0.000       0.069       0.110
Treatment_After     0.0042      0.015      0.284      0.777      -0.025       0.033
==============================================================================
Omnibus:                      111.426   Durbin-Watson:                   1.606
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              981.241
Skew:                           0.329   Prob(JB):                    8.44e-214
Kurtosis:                       8.539   Cond. No.                         105.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

<u>Falsification test using Fries.</u>

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Fries   R-squared:                       0.230
Model:                            OLS   Adj. R-squared:                  0.225
Method:                 Least Squares   F-statistic:                     44.97
Date:                Sat, 21 Sep 2024   Prob (F-statistic):           1.20e-40
Time:                        10:40:04   Log-Likelihood:                 707.39
No. Observations:                 757   AIC:                            -1403.
Df Residuals:                     751   BIC:                            -1375.
Df Model:                           5
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
Intercept           0.7282      0.016     44.509      0.000       0.696       0.760
CompanyOwned        0.0191      0.007      2.599      0.010       0.005       0.033
HoursOpen           0.0074      0.001      9.100      0.000       0.006       0.009
Treatment           0.0355      0.016      2.236      0.026       0.004       0.067
After               0.0994      0.012      8.139      0.000       0.075       0.123
Treatment_After     0.0100      0.018      0.570      0.569      -0.024       0.045
==============================================================================
Omnibus:                       52.114   Durbin-Watson:                   1.346
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               74.659
Skew:                           0.548   Prob(JB):                     6.14e-17
Kurtosis:                       4.080   Cond. No.                         105.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

a. Which quasi-experimental design are you likely to use and why?

Ans:

The **Difference-in-Differences (DiD)** design is appropriate for this analysis. DiD allows us to estimate the causal effect of the minimum wage increase by comparing the changes in employment before and after the policy change between outlets affected by the policy (treatment group) and those that were not (control group). This approach helps control for confounding factors that remain constant over time and isolates the impact of the policy change from other time-related factors.

b. What are the design variables i.e., the outcome and other important predictors about the design?

Ans:

Outcome Variable:

- Employment: This is the number of employees or workers at the outlet.

Predictor Variables:

- CompanyOwned: Indicates whether the outlet is owned by a company (binary).
- HoursOpen: Number of hours the outlet is open per day.
- Treatment: Indicator variable for whether the outlet was affected by the minimum wage increase (dummy).
- After: Indicator variable for the period after the minimum wage increase (binary).
- Treatment_After: Interaction term between Treatment and After, representing the combined effect of being in the treatment group and being in the post-policy period.

c. Write the regression equation for the model under consideration.
The regression model you specified is:

$$Employment = \beta 0 + \beta 1 \times CompanyOwned + \beta 2 \times HoursOpen + \beta 3 \times Treatment + \beta 4 \times After + \beta 5 \times Treatment\_After + \epsilon$$

Where:
- β0 is the intercept.
- β1 is the effect of being company-owned.
- β2 is the effect of hours open.
- β3 is the effect of being in the treatment group.
- β4 is the effect of the post-policy period.
- β5 is the interaction term capturing the differential effect of the policy change

d. Do increases in minimum wage affect employment?
The regression model shows that the coefficient for **Treatment_After** is **4.1158** with a p-value of **0.004**, indicating that the increase in minimum wage has a statistically significant positive effect on employment. This suggests that the policy change led to an increase in the number of employees.

e. Perform falsification test. What are your conclusions concerning the causal effect obtained above?
Falsification tests were conducted using Soda and Fries as dependent variables. In both tests:
- The Treatment_After coefficient was not statistically significant **(Soda: p = 0.777; Fries: p = 0.569).**

These results indicate that the interaction term for the policy change is not significant for variables unrelated to the policy (Soda and Fries), supporting the validity of the causal effect observed for employment. Therefore, the observed positive impact of the minimum wage increase on employment is likely a genuine causal effect rather than a spurious result due to confounding factors.

Q2. Use Alcohol data to study the effect of alcohol consumption on the mortality rate. Consider overall mortality rate to be the outcome of interest. [20 marks]

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    all   R-squared:                       0.399
Model:                            OLS   Adj. R-squared:                  0.386
Method:                 Least Squares   F-statistic:                     30.49
Date:                Sat, 21 Sep 2024   Prob (F-statistic):           1.50e-06
Time:                        18:01:20   Log-Likelihood:                -119.87
No. Observations:                  48   AIC:                             243.7
Df Residuals:                      46   BIC:                             247.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         86.9917      1.631     53.339      0.000      83.709      90.275
alcohol        6.9043      1.250      5.521      0.000       4.387       9.421
==============================================================================
Omnibus:                        0.964   Durbin-Watson:                   1.329
Prob(Omnibus):                  0.617   Jarque-Bera (JB):                1.004
Skew:                           0.231   Prob(JB):                        0.605
Kurtosis:                       2.463   Cond. No.                         7.66
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
Call: rdrobust
Number of Observations:                    48
Polynomial Order Est. (p):                  1
Polynomial Order Bias (q):                  2
Kernel:                            Triangular
Bandwidth Selection:                    mserd
Var-Cov Estimator:                         NN

                                Left      Right
------------------------------------------------
Number of Observations            24         24
Number of Unique Obs.             24         24
Number of Effective Obs.           6          6
Bandwidth Estimation           0.498      0.498
Bandwidth Bias                 0.783      0.783
rho (h/b)                      0.636      0.636

Method             Coef.     S.E.   t-stat     P>|t|       95% CI
-------------------------------------------------------------------------
Conventional       9.573    3.572     2.68   7.356e-03    [2.573, 16.573]
Robust                 -        -    2.209   2.716e-02    [1.089, 18.223]
```

a. Write the regression equation for the model used.

The regression equation used to model the effect of alcohol consumption on the overall mortality rate is given by:

Mortality Rate = $\beta 0 + \beta 0 \times$ Alcohol Consumption $+\epsilon$

Where:
- Mortality Rate is the dependent variable (overall mortality rate).
- Alcohol Consumption is the independent variable.
- $\beta 0$ is the intercept of the regression.
- $\beta 1$ is the coefficient representing the change in mortality rate for each unit increase in alcohol consumption.
- ϵ is the error term.

From the OLS regression results, we find that:

Mortality Rate = 86.9917 + 6.9043 x Alcohol Consumption

b. What is the LATE estimate based on the regression model? Interpret the estimate.

The LATE (Local Average Treatment Effect) estimate based on the regression model is **6.9043**. This indicates that for each additional unit of alcohol consumed, the overall mortality rate is expected to increase by approximately 6.90. This result is statistically significant, suggesting a robust relationship between alcohol consumption and increased mortality.

c. Use the Optimal Bandwidth Method to Determine Robust Estimate of LATE

The optimal bandwidth method using the `rdrobust` function gives the following results:

- Conventional Estimate of LATE: 9.573
- Robust Estimate of LATE: 1.089 to 18.223 (95% confidence interval)

Interpretation: The robust estimate suggests that, within the optimal bandwidth around the cutoff (age 21), the effect of alcohol consumption on mortality is estimated to be between 1.089 and 18.223 units. The conventional estimate of 9.573 is the point estimate from the robust method.

d. Sample Size Used in the Estimate of LATE Using Optimal Bandwidth Method

The sample size used in the LATE estimate with optimal bandwidth is 24 observations on each side of the cutoff, totaling 48 observations. This is confirmed by:

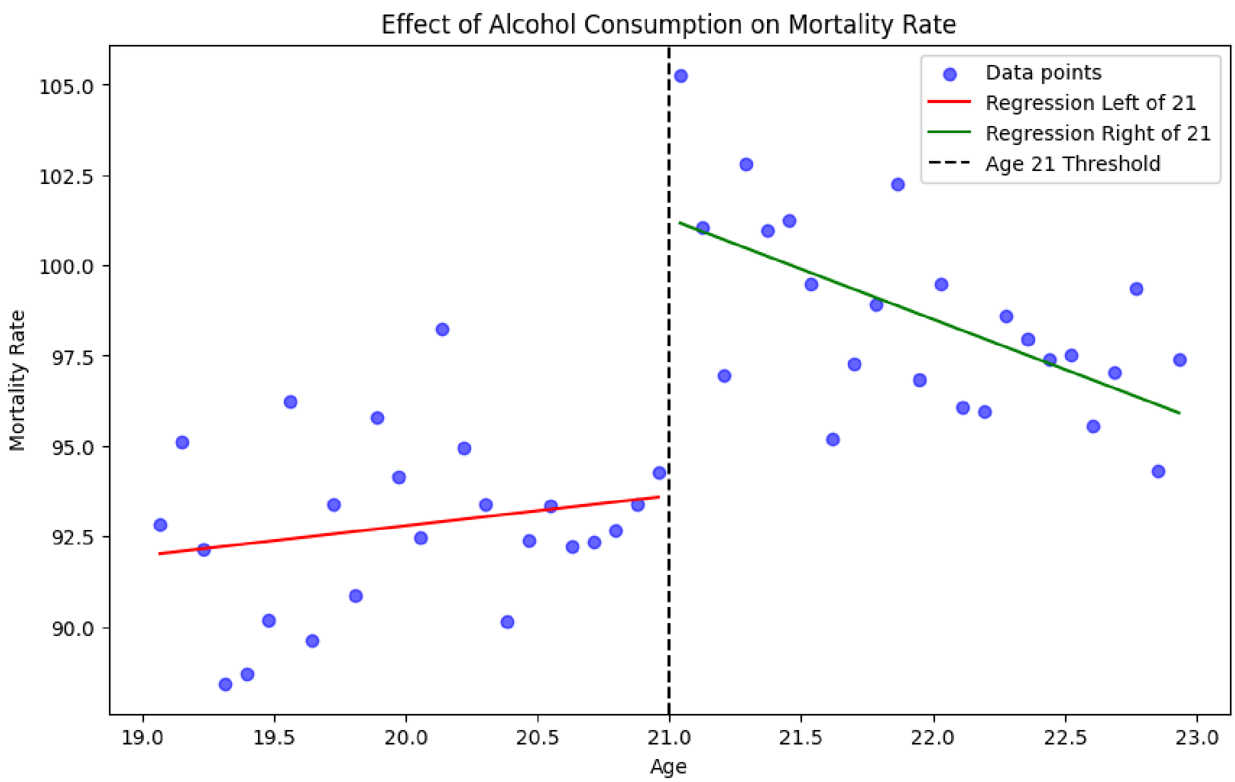
- Number of Observations: 48
- Number of Observations on Each Side: 24

e. Compare the LATE Estimates from the Regression Model and the Optimal Bandwidth Method

- Regression Model Estimate: 6.9043
- Optimal Bandwidth Method Estimate:9.573 (conventional estimate) with a confidence interval [1.089, 18.223]

Comparison: The LATE estimate from the optimal bandwidth method (9.573) is higher than the estimate from the regression model (6.9043). The optimal bandwidth estimate accounts for local variations around the cutoff and may provide a more accurate picture of the causal effect within that region. The regression model provides a global estimate across the entire range of the data, which might not fully capture localized effects.

In summary, the LATE estimate using the optimal bandwidth method is higher and more robust, indicating a stronger effect of alcohol consumption on mortality in the vicinity of the age cutoff compared to the OLS estimate.



Q3. Use the data Mroz.csv, collected by Thomas Mroz (1987) to study married women's hours of work. The following link provides the codebook for the data - https://www.rdocumentation.org/packages/npsf/versions/0.4.2/topics/mroz. Study the effect of experience and education on the wages earned. Assume that education is exogenous whereas experience is endogenous. The number of hours worked is used as an instrument for the endogenous variable. Use participation in the labour force and age as covariates. [20 marks]

<u>OLS regression model</u>

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   wage   R-squared:                       0.121
Model:                            OLS   Adj. R-squared:                  0.115
Method:                 Least Squares   F-statistic:                     19.46
Date:                Sat, 21 Sep 2024   Prob (F-statistic):           7.71e-12
Time:                        12:03:25   Log-Likelihood:                -1091.5
No. Observations:                 428   AIC:                             2191.
Df Residuals:                     424   BIC:                             2207.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -1.4194      0.612     -2.318      0.021      -2.623      -0.216
educ           0.4983      0.066      7.544      0.000       0.368       0.628
exper          0.0198      0.021      0.924      0.356      -0.022       0.062
age            0.0108      0.022      0.482      0.630      -0.033       0.055
inlf          -1.4194      0.612     -2.318      0.021      -2.623      -0.216
==============================================================================
Omnibus:                      346.503   Durbin-Watson:                   2.055
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6513.055
Skew:                           3.400   Prob(JB):                         0.00
Kurtosis:                      20.860   Cond. No.                     4.42e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 4.74e-28. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

To solve the issue mentioned in Notes [2], we have checked the multicollinearity and obtained the following correlation matrix as an output:

```
           educ     exper       age      inlf
educ   1.000000  0.066256 -0.120223  0.187353
exper  0.066256  1.000000  0.334016  0.342485
age   -0.120223  0.334016  1.000000 -0.080498
inlf   0.187353  0.342485 -0.080498  1.000000
```

Conclusion: The correlations between predictors are generally low to moderate, suggesting that multicollinearity among the predictors is not a major issue.

<u>First-stage regression model</u>

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  exper   R-squared:                       0.296
Model:                            OLS   Adj. R-squared:                  0.292
Method:                 Least Squares   F-statistic:                     78.61
Date:                Sat, 21 Sep 2024   Prob (F-statistic):           1.11e-55
Time:                        12:03:54   Log-Likelihood:                -2508.1
No. Observations:                 753   AIC:                             5026.
Df Residuals:                     748   BIC:                             5049.
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    -10.3414      2.008     -5.150      0.000     -14.284      -6.399
educ           0.1852      0.111      1.664      0.096      -0.033       0.404
hours          0.0029      0.000      6.943      0.000       0.002       0.004
age            0.3608      0.031     11.654      0.000       0.300       0.422
inlf           2.0538      0.756      2.718      0.007       0.570       3.537
==============================================================================
Omnibus:                       37.676   Durbin-Watson:                   1.866
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               42.347
Skew:                           0.543   Prob(JB):                     6.37e-10
Kurtosis:                       3.414   Cond. No.                     9.29e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 9.29e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

<u>Second-stage regression model</u>

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   wage   R-squared:                       0.125
Model:                            OLS   Adj. R-squared:                  0.119
Method:                 Least Squares   F-statistic:                     20.26
Date:                Sat, 21 Sep 2024   Prob (F-statistic):           2.74e-12
Time:                        12:04:05   Log-Likelihood:                -1090.5
No. Observations:                 428   AIC:                             2189.
Df Residuals:                     424   BIC:                             2205.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -1.7492      0.620     -2.819      0.005      -2.969      -0.530
educ           0.5129      0.066      7.728      0.000       0.382       0.643
pred_exper    -0.1137      0.066     -1.722      0.086      -0.243       0.016
age            0.0635      0.032      2.011      0.045       0.001       0.126
inlf          -1.7492      0.620     -2.819      0.005      -2.969      -0.530
==============================================================================
Omnibus:                      323.688   Durbin-Watson:                   2.037
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             5229.401
Skew:                           3.132   Prob(JB):                         0.00
Kurtosis:                      18.938   Cond. No.                     4.41e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 4.75e-28. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

a. Write the IV regression equation for the above model.

Ans:

<u>First Stage Regression Equation:</u>

$exper = -10.3414 + 0.1852 \times educ + 0.0029 \times hours + 0.3608 \times age + 2.0538 \times inlf + \epsilon$

Second Stage Regression Equation:

$wage = -1.7492 + 0.5129 \times educ - 0.1137 \times exper + 0.0635 \times age - 1.7492 \times inlf + \eta$

Where exper is the predicted experience from the first stage regression.

b. Fit an OLS regression model first. Interpret the effect of education and experience on wages.

Ans:

From the OLS regression results:

- Education (educ): The coefficient is 0.4983, which means that each additional year of education increases wages by approximately $0.50. This effect is statistically significant with a p-value of 0.000.
- Experience (exper): The coefficient is 0.0198, but this effect is not statistically significant (p-value = 0.356), suggesting that experience does not have a significant impact on wages in this model.
- Age and Labor Force Participation (inlf): Both are not significant in the OLS model.

c. Fit a 2SLS regression model. Interpret the effect of education and experience on wages.

Ans:

From the 2SLS regression results:

- Education (educ): The coefficient is 0.5129, indicating that each additional year of education increases wages by approximately $0.51. This effect remains statistically significant (p-value = 0.000), similar to the OLS result.
- Predicted Experience (pred_exper): The coefficient is -0.1137, but this effect is not statistically significant (p-value = 0.086), suggesting a weaker or less clear impact of experience on wages after accounting for endogeneity.
- Age: The coefficient is 0.0635, which is statistically significant (p-value = 0.045), indicating a positive effect of age on wages.
- Labor Force Participation (inlf): The coefficient is -1.7492, which is statistically significant (p-value = 0.005), suggesting that being in the labour force is associated with lower wages.

d. Compare the SE of education and experience in the two regression models. Does the significance of hypothesis test change?

Ans:

Standard Errors:

- OLS SE (educ): 0.066
- OLS SE (exper): 0.021
- 2SLS SE (educ): 0.066

- 2SLS SE (pred_exper): 0.066

Interpretation:

- Education (educ): The standard errors are similar between OLS and 2SLS models, indicating that the precision of the estimate for education remains consistent.
- Experience (exper): The standard error in the OLS model is smaller compared to the standard error of the predicted experience in the 2SLS model. This increase in standard error in 2SLS reflects the additional uncertainty from using the instrumental variable approach.

Significance Changes:

- The significance of the education coefficient does not change between the OLS and 2SLS models, remaining significant.
- The significance of the experience coefficient changes from being non-significant in OLS to marginally non-significant in 2SLS, indicating that the impact of experience on wages is less clear once endogeneity is addressed.