Written problems:

Wooldridge: Chapter 7, Problem 9 (7.9); only parts (iii) and (iv)

(iii) We set the value of tocoll so the predicted log(wage) for women equal to the predicted log(wage) for men.

For men, where female = 0, the equation is 2.289 + 0.50 * totcoll

For women, the equation is 1.932 + 0.5 * totcoll + 0.030 * totcoll

Set the equation equal and we get totcoll approximately equal to 0.347

(iv) Various other factors contribute to the wage gap, such as discrimination, career choices, work experience, industry, and societal factors, among others. Therefore, achieving wage parity between men and women is a complex issue that cannot be solely addressed by educational attainment.

Chapter 6, 6.4

(i) If B2 is positive, it suggests that the return to an individual's education is enhanced by having more educated parents. In this context, it's possible that B1 might also be positively signed.

(ii) For individuals whose parents both have a high school education (pareduc=24), the return to an additional year of their own education (
educ) is enhanced by 0.00078. This means that each year of additional education is associated with an increase in wages by the sum of the coefficient on educ (
0.047) and the interaction term (0.00078), which is 0.04778.

In both cases, the interaction term educ−pareduc is positive, indicating that the return to an individual's education is positively influenced by the total education level of both parents. Essentially, it suggests that individuals with more educated parents tend to have a higher return on their own education, whereas individuals with less educated parents have a slightly lower return on their education. This implies an intergenerational impact on wages, where parental education positively influences the wage returns associated with an individual's education.

(iii) find t statistics for the coefficient edu * pareduc is estimated to be -0.016, and the standard error is 0.0012, so the t-statistics is about -13.33. This exceeds the critical value for a two-tailed test, which against the null hypothesis, and suggests that the return to education depends on parent education.

(iv)

not including pareduc as a separate variable in the regression constrains the analysis by limiting the direct interpretation of the effect of parental education on wages, particularly in cases where education (the variable used in the interaction term) is not present or zero.

2. Wooldridge: Chapter 6, Problem 4 (6.4); also you can answer the following:
(iv) Explain why the regression in part (ii) (which includes pareduc in the interaction but not as a variable by itself) is not desirable. Specifically, how does it restrict the effect of pareduc on log(wage)? Hint: First think about what the regression in (ii) says about the effect of pareduc on log(wage) when educ=0, then think about what the regression in (iii) says about this same effect.

Computer Problems

Wooldridge Computer Exercise C6

```
                    OLS Regression Results
==============================================================================
Dep. Variable:               voteA   R-squared:                       0.868
Model:                         OLS   Adj. R-squared:                  0.865
Method:              Least Squares   F-statistic:                     276.5
Date:             Tue, 07 Nov 2023   Prob (F-statistic):           9.03e-73
Time:                     17:25:09   Log-Likelihood:                -557.66
No. Observations:              173   AIC:                             1125.
Df Residuals:                  168   BIC:                             1141.
Df Model:                        4
Covariance Type:         nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         18.1954      2.568      7.086      0.000      13.126      23.265
prtystrA       0.1573      0.050      3.165      0.002       0.059       0.255
expendA       -0.0067      0.003     -2.354      0.020      -0.012      -0.001
expendB        0.0043      0.003      1.637      0.104      -0.001       0.009
shareA         0.4944      0.025     19.535      0.000       0.444       0.544
==============================================================================
Omnibus:                       36.148   Durbin-Watson:                   1.743
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              112.238
Skew:                           0.789   Prob(JB):                     4.24e-25
Kurtosis:                       6.616   Cond. No.                     3.05e+03
==============================================================================
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.05e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

(i)As the coefficient for 'expend' directly isn't present in the output, we can't directly interpret its isolated effect.
Regarding the expected sign for B4, typically, the sign for an interaction term like 'expendA-expend' (which is represented by B4 in the model) is not immediately obvious without further context or understanding of the variables involved.
The interaction term often signifies the joint effect of 'expendA' and 'expendB' and can't be independently interpreted without considering the individual variables involved in the interaction.

(ii)The variables 'prtystrA' and 'shareA' appear to be statistically significant in predicting 'voteA'.
However, 'expendA' and 'expendB' might need further scrutiny due to their lower significance levels or potential multicollinearity issues with other variables.

(iii)Whether this effect is considered large or not would depend on the context and the scale of the variable 'voteA'.
A coefficient of 0.0043 implies that for every $100,000 increase in Candidate B's expenditure, voteA would be expected to change by 0.43 votes, given other variables remain constant.
The significance of this effect would depend on the domain and the significance of a single vote in the context of the scenario.

Effect = $100 * coefficient for 'expendA'

This value represents the expected change in 'voteA' for every $100 increase in 'expendA', given 'expend' is fixed at $100.
This maybe reasonable depending on the cases or scenarios given.

```
             marr         wage        exper          age         coll  \
count  269.000000   269.000000  269.000000  269.000000  269.000000
mean     0.442379  1423.827515    5.118959   27.394052    3.717472
std      0.497595   999.774048    3.400062    3.391292    0.754410
min      0.000000   150.000000    1.000000   21.000000    0.000000
25%      0.000000   650.000000    2.000000   25.000000    4.000000
50%      0.000000  1186.000000    4.000000   27.000000    4.000000
75%      1.000000  2014.500000    7.000000   30.000000    4.000000
max      1.000000  5740.000000   18.000000   41.000000    4.000000

            games      minutes       guard     forward      center  \
count  269.000000   269.000000  269.000000  269.000000  269.000000
mean    65.724907  1682.193309    0.420074    0.408922    0.171004
```

```
std     18.851110   893.327771    0.494491    0.492551    0.377214
min      3.000000    33.000000    0.000000    0.000000    0.000000
25%     57.000000   983.000000    0.000000    0.000000    0.000000
50%     74.000000  1690.000000    0.000000    0.000000    0.000000
75%     79.000000  2438.000000    1.000000    1.000000    0.000000
max     82.000000  3533.000000    1.000000    1.000000    1.000000

           points    rebounds     assists      draft     allstar      avgmin  \
count  269.000000  269.000000  269.000000  240.00000  269.000000  269.000000
mean    10.210409    4.401115    2.408922   20.20000    0.115242   23.979254
std      5.900667    2.892572    2.092986   18.73582    0.319909    9.731176
min      1.200000    0.500000    0.000000    1.00000    0.000000    2.888889
25%      5.400000    2.300000    0.900000    7.00000    0.000000   16.731340
50%      9.300000    3.800000    1.900000   14.50000    0.000000   24.816900
75%     14.200000    5.500000    3.400000   28.25000    0.000000   33.256100
max     29.799999   17.299999   12.600000  139.00000    1.000000   43.085369

            lwage       black    children     expersq        agesq   marrblck
count  269.000000  269.000000  269.000000  269.000000   269.000000  269.00000
mean     6.952296    0.806691    0.345725   37.721190   761.892193    0.33829
std      0.881376    0.395629    0.476491   46.537021   195.149406    0.47401
min      5.010635    0.000000    0.000000    1.000000   441.000000    0.00000
25%      6.476973    1.000000    0.000000    4.000000   625.000000    0.00000
50%      7.078341    1.000000    0.000000   16.000000   729.000000    0.00000
75%      7.608126    1.000000    1.000000   49.000000   900.000000    1.00000
max      8.655214    1.000000    1.000000  324.000000  1681.000000    1.00000
                            OLS Regression Results
==============================================================================
Dep. Variable:                 points   R-squared:                       0.051
Model:                            OLS   Adj. R-squared:                  0.044
Method:                 Least Squares   F-statistic:                     7.164
Date:                Tue, 07 Nov 2023   Prob (F-statistic):           0.000932
Time:                        17:25:09   Log-Likelihood:                -851.63
No. Observations:                 269   AIC:                             1709.
Df Residuals:                     266   BIC:                             1720.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          8.5392      0.637     13.413      0.000       7.286       9.793
exper          0.2193      0.116      1.885      0.060      -0.010       0.448
expersq        0.0201      0.010      2.068      0.040       0.001       0.039
==============================================================================
```

```
Omnibus:                 19.340   Durbin-Watson:              2.277
Prob(Omnibus):              0.000   Jarque-Bera (JB):            21.495
Skew:                    0.681   Prob(JB):                2.15e-05
Kurtosis:                 3.251   Cond. No.                    89.6
==============================================================================

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

(i) he overall model with experience and its quadratic term appears to have some explanatory power for the points variable.
The individual effects of exper and expersq are statistically suggestive but might not be highly significant,
given the higher p-values.It seems that the quadratic relationship between experience and points scored
might provide some added explanation beyond a linear relationship, but the evidence is not overwhelmingly strong

(ii)The absence of including all three position dummy variables might be due to a choice made in the analysis to use only two of the three position dummies as reference groups.
we want to avoid multicollinearity.

(iii)To compare guards and centers, a statistical test could be performed on the coefficient of the guard dummy variable to ascertain whether guards
score significantly more points than centers, holding experience fixed.

(iv)Adding marital status to the equation allows for an examination of the productivity difference between married and unmarried players while controlling for position and experience.
A statistical test can determine if marital status significantly affects points per game.

(v)Adding interactions between marital status and experience variables provides insight into whether being married influences how experience relates to points per game.
Statistical tests can ascertain if these interactions have a significant effect on points per game.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             assists   R-squared:                   0.338
Model:                       OLS   Adj. R-squared:              0.325
```

```
Method:              Least Squares   F-statistic:                26.83
Date:             Tue, 07 Nov 2023   Prob (F-statistic):          6.68e-22
Time:                     17:25:09   Log-Likelihood:             -524.44
No. Observations:              269   AIC:                         1061.
Df Residuals:                  263   BIC:                         1082.
Df Model:                        5
Covariance Type:         nonrobust
==============================================================================
             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        0.4277      0.310      1.380      0.169      -0.182      1.038
exper        0.0486      0.036      1.340      0.181      -0.023      0.120
expersq      0.0085      0.003      2.918      0.004       0.003      0.014
guard        2.5956      0.302      8.598      0.000       2.001      3.190
forward      0.6114      0.303      2.019      0.044       0.015      1.208
marr         0.3616      0.224      1.615      0.107      -0.079      0.802
==============================================================================
Omnibus:                    76.699   Durbin-Watson:                   2.099
Prob(Omnibus):               0.000   Jarque-Bera (JB):              179.601
Skew:                        1.357   Prob(JB):                     1.00e-39
Kurtosis:                    5.943   Cond. No.                         222.
==============================================================================

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

(vi)With the given data, he guard position seems to significantly influence the number of assists, while forward players have a less influential but still notable impact. Marital status and experience, including its quadratic form,
do not seem to significantly affect assists per game in this model.

HW8/HW8.py

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm

# Include similar code as HW1 for the basic steps
file_location = "/Users/amyliang/Eco441K/HW8/VOTE1.DTA"
df= pd.read_stata(file_location)
f2 = "/Users/amyliang/Eco441K/HW8/nbasal.dta"
df2 = pd.read_stata(f2)

pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

# Wooldridge Computer Exercise C6
print("Wooldridge Computer Exercise C6")

# Define the independent variables (explanatory variables)
X = df[['prtystrA', 'expendA', 'expendB', 'shareA']]

# Add a constant term (intercept) to the model
X = sm.add_constant(X)

# Define the dependent variable
Y = df['voteA']

# Fit the linear regression model
model = sm.OLS(Y, X).fit()

# Print the regression results summary
print(model.summary())
ans1 = """
(i)As the coefficient for 'expend' directly isn't present in the output, we can't directly interpret its isolated effect.
Regarding the expected sign for B4, typically, the sign for an interaction term
like 'expendA-expend' (which is represented by B4 in the model) is not immediately obvious without further context or understanding of the variables involved.
The interaction term often signifies the joint effect of 'expendA' and 'expendB' and can't be independently interpreted without considering the individual
variables involved in the interaction.
"""
ans = """
(ii)The variables 'prtystrA' and 'shareA' appear to be statistically significant in predicting 'voteA'.
However, 'expendA' and 'expendB' might need further scrutiny due to their lower significance levels or potential multicollinearity issues
with other variables.
"""
```

```python
print(ans1)
print(ans)

ans2 = """
(iii)Whether this effect is considered large or not would depend on the context and the scale of the variable 'voteA'.
A coefficient of 0.0043 implies that for every $100,000 increase in Candidate B's expenditure,
voteA would be expected to change by 0.43 votes, given other variables remain constant.
The significance of this effect would depend on the domain and the significance of a single vote in the context of the scenario.
"""
print(ans2)

ans3 = """
Effect = $100 * coefficient for 'expendA'

This value represents the expected change in 'voteA' for every $100 increase in 'expendA', given 'expend' is fixed at $100.
This maybe reasonable depending on the cases or scenarios given.
"""

print(ans3)

print(df2.describe())
# Create the quadratic term for 'Experience'
df2['expersq'] = df2['exper'] ** 2

# Define the dependent variable (Points)
Y = df2['points']

# Define the independent variables including Experience and Experience_Squared
X = df2[['exper', 'expersq']]

# Add a constant term (intercept) to the model
X = sm.add_constant(X)

# Fit the linear regression model
model = sm.OLS(Y, X).fit()

# Print the regression results summary
print(model.summary())

ans4 = """
(i) he overall model with experience and its quadratic term appears to have some explanatory power for the points variable.
The individual effects of exper and expersq are statistically suggestive but might not be highly significant,
given the higher p-values.It seems that the quadratic relationship between experience and points scored
might provide some added explanation beyond a linear relationship, but the
```

```python
evidence is not overwhelmingly strong
"""
print(ans4)

ans5 = """
(ii)The absence of including all three position dummy variables might be due to a choice made in the analysis to use only two of the three position dummies as reference groups.
we want to avoid multicollinearity.
"""
print(ans5)

ans6 = """
(iii)To compare guards and centers, a statistical test could be performed on the coefficient of the guard dummy variable to ascertain whether guards
score significantly more points than centers, holding experience fixed.
"""
print(ans6)

ans7 = """
(iv)Adding marital status to the equation allows for an examination of the productivity difference between married and unmarried players while controlling for position and experience.
A statistical test can determine if marital status significantly affects points per game.
"""
print(ans7)

ans8= """
(v)Adding interactions between marital status and experience variables provides insight into whether being married influences how experience relates to points per game.
Statistical tests can ascertain if these interactions have a significant effect on points per game.
"""
print(ans8)
# Define the dependent variable (Assists per Game)
Y = df2['assists']

# Define the independent variables (including marital status, position, experience, and their interactions)
X = df2[['exper', 'expersq', 'guard', 'forward', 'marr']]

# Add a constant term (intercept) to the model
X = sm.add_constant(X)

# Fit the linear regression model
model = sm.OLS(Y, X).fit()

# Print the regression results summary
print(model.summary())

```

```python
ans9 = """
(vi)With the given data, he guard position seems to significantly influence the number of assists,
while forward players have a less influential but still notable impact. Marital status and experience, including its quadratic form,
do not seem to significantly affect assists per game in this model.
"""

print(ans9)
```