Machine Learning and Econometrics

Submitted by - Aryaman Garg(2015MT60556) and Aasavari Kakne(2015MT10582)

Q.1. Print mean and standard deviation of Y_i for i = 1, 2....15.

1. Y_1 = "If worked for pay" for all women sample

```
In [4]: Y1 = (Y.WEEKSM.dropna()>0).dropna();
        print 'The mean is: ',Y1.mean()
        print 'The STD is:', Y1.std()

The mean is:  0.5632344953153158
The STD is: 0.49598585876907014
```

2. Y_2 = "If worked for pay" for married women sample

```
In [5]: Y2 = Y[Y.TIMESMAR.dropna()>0].WEEKSM.dropna()>0;
        print 'The mean is: ',Y2.mean()
        print 'The STD is:', Y2.std()

The mean is:  0.5661555033011633
The STD is: 0.4956047214741414
```

3. Y_3 = "If worked for pay" for husbands of married women sample

```
In [6]: Y3 = Y.WEEKSD.dropna()>0;
        print 'The mean is: ',Y3.mean()
        print 'The STD is:', Y3.std()

The mean is:  0.9729730477766768
The STD is: 0.16216216839580402
```

4. Y_4 = "weeks worked" for all women sample

```
In [7]: Y4 = Y.WEEKSM.dropna();
        print 'The mean is: ',Y4.mean()
        print 'The STD is:', Y4.std()

The mean is:  20.666505032
The STD is: 22.2365411866
```

5. Y_5 = "weeks worked" for married women sample

```
In [8]: Y5 = Y[Y.TIMESMAR>0].WEEKSM.dropna();
        print 'The mean is: ',Y5.mean()
        print 'The STD is:', Y5.std()

The mean is:  20.7887976841
The STD is: 22.2449489973
```

6. Y_6 = "weeks worked" for husbands of married women sample

```
In [9]: Y6 = Y.WEEKSD.dropna();
        print 'The mean is: ',Y6.mean()
        print 'The STD is:', Y6.std()

The mean is:  47.39468207
The STD is: 11.2378905509
```

7. Y_7 = "hours worked" for all women sample

```
In [10]: Y7 = Y.HOURSM.dropna();
         print 'The mean is: ',Y7.mean()
         print 'The STD is:', Y7.std()

The mean is:  18.7538017163
The STD is: 18.9337119334
```

8. Y_8 = "hours worked" for married women sample

```
In [11]: Y8 = Y[Y.TIMESMAR>0].HOURSM.dropna();
         print 'The mean is: ',Y8.mean()
         print 'The STD is:', Y8.std()

The mean is:  18.8135552061
The STD is: 18.9077721081
```

9. Y_9 = "hours worked" for husbands of married women sample

```
In [12]: Y9 = Y.HOURSD.dropna();
         print 'The mean is: ',Y9.mean()
         print 'The STD is:', Y9.std()

The mean is:  43.1116366736
The STD is: 12.5962779687
```

10. Y_{10} = "annual labour income" for all women sample

```
In [42]: Y10 = ((pd.Series(x for x in Y.INCOME1M)  +
                 pd.Series(max(0,y) for y in Y.INCOME2M))*2.099173554).dropna();
         Y10.index = Y1.index
         print 'The mean is: ',Y10.mean()
         print 'The STD is:', Y10.std()

The mean is:  7113.17240341
The STD is: 10780.203499
```

11. Y_{11} = "annual labour income" for married women sample

```
In [48]: Y11 = ((pd.Series(x for x in Y[Y.TIMESMAR>0].INCOME1M)  +
                pd.Series(max(0,y) for y in Y[Y.TIMESMAR>0].INCOME2M))*2.099173554).dropna();
         Y11.index = Y2.index;
         print 'The mean is: ',Y11.mean()
         print 'The STD is:', Y11.std()

         The mean is:  7158.75754568
         The STD is: 10804.9067613
```

12. Y_{12} = "annual labour income" for husbands of married women sample

```
In [49]: Y12 = ((pd.Series(x for x in Y.INCOME1D)  +
                pd.Series(max(0,y) for y in Y.INCOME2D))*2.099173554).dropna();
         Y12.index = Y3.index;
         print 'The mean is: ',Y12.mean()
         print 'The STD is:', Y12.std()

         The mean is:  37178.4893897
         The STD is: 24535.7932891
```

13. Y_{13} = "log of family income" for all women sample

```
In [53]: Ytemp = Y.FAMINC*2.099173554;
         Y13 = np.log(pd.Series(max(1,x) for x in Ytemp))
         Y13.index = Y.index;
         print 'The mean is: ',Y13.mean()
         print 'The STD is:', Y13.std()

         The mean is:  10.3028235718
         The STD is: 1.36638171028
```

14. Y_{14} = "log of family income" for married women sample

```
In [55]: Ytemp = Y.FAMINC[Y.TIMESMAR>0]*2.099173554;
         Y14 = np.log(pd.Series(max(1,x) for x in Ytemp))
         Y14.index = Y2.index;
         print 'The mean is: ',Y14.mean()
         print 'The STD is:', Y14.std()

         The mean is:  10.3493584142
         The STD is: 1.31080402469
```

15. Y_{15} = "log of non-wife income" for married women sample

```
In [54]: Ytemp = ((Y.FAMINC[Y.TIMESMAR>0] - Y[Y.TIMESMAR>0].INCOME1M)*2.099173554);
         Ytemp2 = pd.Series(max(1,x) for x in Ytemp);
         Y15 = np.log(Ytemp2)
         Y15.index = Y2.index;
         print 'The mean is: ',Y15.mean()
         print 'The STD is:', Y15.std()

         The mean is:  9.78610058889
         The STD is: 2.32820877613
```

Q.2. Construct causal trees for every Y_i and compute R2 score for each of the trees.

1) Y_1 = "If worked for pay" for all women sample

```
In [23]: Evaluation_data_index1 = random.sample(Sample_indexM, np.int(np.ceil(len(XM)*0.1)))
         Evaluation_data_xi1 = XM.loc[Evaluation_data_index1]
         Evaluation_data_target1 = Y1.loc[Evaluation_data_index1]
         Training_data_index1 = list(set(Sample_indexM)-set(Evaluation_data_index1))
         Training_data_target1 = Y1.loc[Training_data_index1]
         Training_data_xi1 = XM.loc[Training_data_index1]
         clf1 = tree.DecisionTreeRegressor(criterion='mse')
         clf1 = clf1.fit(Training_data_xi1, Training_data_target1)

         tau_hat1 = clf1.predict(Evaluation_data_xi1)

         print(clf1.score(Evaluation_data_xi1, Evaluation_data_target1))

         0.02411821013733828
```

2) Y_2 = "If worked for pay" for married women sample

```
In [24]: Evaluation_data_index2 = random.sample(Sample_indexMAR, np.int(np.ceil(len(XMAR)*0.1)))
         Evaluation_data_xi2 = XMAR.loc[Evaluation_data_index2]
         Evaluation_data_target2 = Y2.loc[Evaluation_data_index2]
         Training_data_index2 = list(set(Sample_indexMAR)-set(Evaluation_data_index2))
         Training_data_target2 = Y2.loc[Training_data_index2]
         Training_data_xi2 = XM.loc[Training_data_index2]
         clf2 = tree.DecisionTreeRegressor(criterion='mse')
         clf2 = clf2.fit(Training_data_xi2, Training_data_target2)

         tau_hat2 = clf2.predict(Evaluation_data_xi2)

         print(clf2.score(Evaluation_data_xi2, Evaluation_data_target2))

         0.024006017879074637
```

3) Y_3 = "If worked for pay" for husbands of married women sample

In [25]:
```
Evaluation_data_index3 = random.sample(Sample_indexD, np.int(np.ceil(len(XD)*0.1)))
Evaluation_data_xi3 = XD.loc[Evaluation_data_index3]
Evaluation_data_target3 = Y3.loc[Evaluation_data_index3]
Training_data_index3 = list(set(Sample_indexD)-set(Evaluation_data_index3))
Training_data_target3 = Y3.loc[Training_data_index3]
Training_data_xi3 = XD.loc[Training_data_index3]
clf3 = tree.DecisionTreeRegressor(criterion='mse')
clf3 = clf3.fit(Training_data_xi3, Training_data_target3)

tau_hat3 = clf3.predict(Evaluation_data_xi3)

print(clf3.score(Evaluation_data_xi3, Evaluation_data_target3))
```

```
-0.1553990093948283
```

4) Y_4 = "weeks worked" for all women sample

In [26]:
```
Evaluation_data_index4 = random.sample(Sample_indexM, np.int(np.ceil(len(XM)*0.1)))
Evaluation_data_xi4 = XM.loc[Evaluation_data_index4]
Evaluation_data_target4 = Y4.loc[Evaluation_data_index4]
Training_data_index4 = list(set(Sample_indexM)-set(Evaluation_data_index4))
Training_data_target4 = Y4.loc[Training_data_index4]
Training_data_xi4 = XM.loc[Training_data_index4]
clf4 = tree.DecisionTreeRegressor(criterion='mse')
clf4 = clf4.fit(Training_data_xi4, Training_data_target4)

tau_hat4 = clf4.predict(Evaluation_data_xi4)

print(clf4.score(Evaluation_data_xi4, Evaluation_data_target4))
```

```
0.04375511589686709
```

5) Y_5 = "weeks worked" for married women sample

In [28]:
```
Evaluation_data_index5 = random.sample(Sample_indexMAR, np.int(np.ceil(len(XMAR)*0.1)))
Evaluation_data_xi5 = XMAR.loc[Evaluation_data_index5]
Evaluation_data_target5 = Y5.loc[Evaluation_data_index5]
Training_data_index5= list(set(Sample_indexMAR)-set(Evaluation_data_index5))
Training_data_target5 = Y5.loc[Training_data_index5]
Training_data_xi5 = XM.loc[Training_data_index5]
clf5 = tree.DecisionTreeRegressor(criterion='mse')
clf5 = clf5.fit(Training_data_xi5, Training_data_target5)

tau_hat5 = clf5.predict(Evaluation_data_xi5)

print(clf5.score(Evaluation_data_xi5, Evaluation_data_target5))
```

```
0.045422377891715215
```

6) Y_6 = “weeks worked” for husbands of married women sample

In [29]:
```
Evaluation_data_index6 = random.sample(Sample_indexD, np.int(np.ceil(len(XD)*0.1)))
Evaluation_data_xi6 = XD.loc[Evaluation_data_index6]
Evaluation_data_target6 = Y6.loc[Evaluation_data_index6]
Training_data_index6 = list(set(Sample_indexD)-set(Evaluation_data_index6))
Training_data_target6 = Y6.loc[Training_data_index6]
Training_data_xi6 = XD.loc[Training_data_index6]
clf6 = tree.DecisionTreeRegressor(criterion='mse')
clf6 = clf6.fit(Training_data_xi6, Training_data_target6)

tau_hat6 = clf6.predict(Evaluation_data_xi6)

print(clf6.score(Evaluation_data_xi6, Evaluation_data_target6))
```

```
-0.07174901066898332
```

7) Y_7 = “hours worked” for all women sample

In [30]:
```
Evaluation_data_index7 = random.sample(Sample_indexM, np.int(np.ceil(len(XM)*0.1)))
Evaluation_data_xi7 = XM.loc[Evaluation_data_index4]
Evaluation_data_target7 = Y7.loc[Evaluation_data_index7]
Training_data_index7 = list(set(Sample_indexM)-set(Evaluation_data_index7))
Training_data_target7 = Y7.loc[Training_data_index7]
Training_data_xi7 = XM.loc[Training_data_index7]
clf7 = tree.DecisionTreeRegressor(criterion='mse')
clf7 = clf7.fit(Training_data_xi7, Training_data_target7)

tau_hat7 = clf7.predict(Evaluation_data_xi7)

print(clf7.score(Evaluation_data_xi7, Evaluation_data_target7))
```

```
-0.08751037708080944
```

8) Y_8 = “hours worked” for married women sample

In [31]:
```
Evaluation_data_index8 = random.sample(Sample_indexMAR, np.int(np.ceil(len(XMAR)*0.1)))
Evaluation_data_xi8 = XMAR.loc[Evaluation_data_index8]
Evaluation_data_target8 = Y8.loc[Evaluation_data_index8]
Training_data_index8 = list(set(Sample_indexMAR)-set(Evaluation_data_index8))
Training_data_target8 = Y8.loc[Training_data_index8]
Training_data_xi8 = XM.loc[Training_data_index8]
clf8 = tree.DecisionTreeRegressor(criterion='mse')
clf8 = clf8.fit(Training_data_xi8, Training_data_target8)

tau_hat8 = clf8.predict(Evaluation_data_xi8)

print(clf8.score(Evaluation_data_xi8, Evaluation_data_target8))
```

```
0.0302663401857578
```

9) Y_9 = "hours worked" for husbands of married women sample

In [32]:
```
Evaluation_data_index9 = random.sample(Sample_indexD, np.int(np.ceil(len(XD)*0.1)))
Evaluation_data_xi9 = XD.loc[Evaluation_data_index9]
Evaluation_data_target9 = Y9.loc[Evaluation_data_index9]
Training_data_index9 = list(set(Sample_indexD)-set(Evaluation_data_index9))
Training_data_target9 = Y9.loc[Training_data_index9]
Training_data_xi9 = XD.loc[Training_data_index9]
clf9 = tree.DecisionTreeRegressor(criterion='mse')
clf9 = clf9.fit(Training_data_xi9, Training_data_target9)

tau_hat9 = clf9.predict(Evaluation_data_xi9)

print(clf9.score(Evaluation_data_xi9, Evaluation_data_target9))
```

```
-0.059486756571655475
```

10) Y_{10} = "annual labour income" for all women sample

In [46]:
```
Evaluation_data_index10 = random.sample(Sample_indexM, np.int(np.ceil(len(XM)*0.1)))
Evaluation_data_xi10 = XM.loc[Evaluation_data_index4]
Evaluation_data_target10 = Y10.loc[Evaluation_data_index10]
Training_data_index10 = list(set(Sample_indexM)-set(Evaluation_data_index10))
Training_data_target10 = Y10.loc[Training_data_index10]
Training_data_xi10 = XM.loc[Training_data_index10]
clf10 = tree.DecisionTreeRegressor(criterion='mse')
clf10 = clf10.fit(Training_data_xi10, Training_data_target10)

tau_hat10 = clf10.predict(Evaluation_data_xi10)

print(clf10.score(Evaluation_data_xi10, Evaluation_data_target10))
```

```
-0.10444901742330615
```

11) Y_{11} = "annual labour income" for married women sample

In [50]:
```
Evaluation_data_index11 = random.sample(Sample_indexMAR, np.int(np.ceil(len(XMAR)*0.1)))
Evaluation_data_xi11 = XMAR.loc[Evaluation_data_index11]
Evaluation_data_target11 = Y11.loc[Evaluation_data_index11]
Training_data_index11 = list(set(Sample_indexMAR)-set(Evaluation_data_index11))
Training_data_target11 = Y11.loc[Training_data_index11]
Training_data_xi11 = XM.loc[Training_data_index11]
clf11 = tree.DecisionTreeRegressor(criterion='mse')
clf11 = clf11.fit(Training_data_xi11, Training_data_target11)

tau_hat11 = clf11.predict(Evaluation_data_xi11)

print(clf11.score(Evaluation_data_xi11, Evaluation_data_target11))
```

```
0.053297816569354506
```

12) $Y_{12} =$ "annual labour income" for husbands of married women sample

```
In [51]: Evaluation_data_index12 = random.sample(Sample_indexD, np.int(np.ceil(len(XD)*0.1)))
Evaluation_data_xi12 = XD.loc[Evaluation_data_index12]
Evaluation_data_target12 = Y12.loc[Evaluation_data_index12]
Training_data_index12 = list(set(Sample_indexD)-set(Evaluation_data_index12))
Training_data_target12 = Y12.loc[Training_data_index12]
Training_data_xi12 = XD.loc[Training_data_index12]
clf12 = tree.DecisionTreeRegressor(criterion='mse')
clf12 = clf12.fit(Training_data_xi12, Training_data_target12)

tau_hat12 = clf12.predict(Evaluation_data_xi12)

print(clf12.score(Evaluation_data_xi12, Evaluation_data_target12))
```

```
0.1268138134378861
```

13) $Y_{13} =$ "log of family income" for all women sample

```
In [56]: Evaluation_data_index13 = random.sample(Sample_indexM, np.int(np.ceil(len(XM)*0.1)))
Evaluation_data_xi13 = XM.loc[Evaluation_data_index4]
Evaluation_data_target13 = Y13.loc[Evaluation_data_index13]
Training_data_index13 = list(set(Sample_indexM)-set(Evaluation_data_index13))
Training_data_target13 = Y13.loc[Training_data_index13]
Training_data_xi13 = XM.loc[Training_data_index13]
clf13 = tree.DecisionTreeRegressor(criterion='mse')
clf13 = clf13.fit(Training_data_xi13, Training_data_target13)

tau_hat13 = clf13.predict(Evaluation_data_xi13)

print(clf13.score(Evaluation_data_xi13, Evaluation_data_target13))
```

```
-0.12568998131289
```

14) $Y_{14} =$ "log of family income" for married women sample

```
In [57]: Evaluation_data_index14 = random.sample(Sample_indexMAR, np.int(np.ceil(len(XMAR)*0.1)))
Evaluation_data_xi14 = XMAR.loc[Evaluation_data_index14]
Evaluation_data_target14 = Y14.loc[Evaluation_data_index14]
Training_data_index14 = list(set(Sample_indexMAR)-set(Evaluation_data_index14))
Training_data_target14 = Y14.loc[Training_data_index14]
Training_data_xi14 = XM.loc[Training_data_index14]
clf14 = tree.DecisionTreeRegressor(criterion='mse')
clf14 = clf14.fit(Training_data_xi14, Training_data_target14)

tau_hat14 = clf14.predict(Evaluation_data_xi14)

print(clf14.score(Evaluation_data_xi14, Evaluation_data_target14))
```

```
0.009589090503039022
```

15) Y_{15} = “log of non-wife income” for married women sample

In [58]:
```
Evaluation_data_index15 = random.sample(Sample_indexMAR, np.int(np.ceil(len(XMAR)*0.1)))
Evaluation_data_xi15 = XMAR.loc[Evaluation_data_index15]
Evaluation_data_target15 = Y15.loc[Evaluation_data_index15]
Training_data_index15 = list(set(Sample_indexMAR)-set(Evaluation_data_index15))
Training_data_target15 = Y15.loc[Training_data_index15]
Training_data_xi15 = XM.loc[Training_data_index15]
clf15 = tree.DecisionTreeRegressor(criterion='mse')
clf15 = clf15.fit(Training_data_xi15, Training_data_target15)

tau_hat15 = clf15.predict(Evaluation_data_xi15)

print(clf15.score(Evaluation_data_xi15, Evaluation_data_target15))
```

```
0.007875120602259034
```

Judging by the R2 score of the decision trees,

1) Clf 12 has positive R2 score implying that it is a good model for prediction of Y12.
2) Clf 1,2,3,4,5,6,7,8,9,10,11,14,15 have R2 score very close to 0. Therefore, these trees don’t do a lot better than just taking the average on Y_i’s.
3) Clf 3 and 13 have negative R2 score, therefore the models do worse than taking average value of Y_i.

Q.3. OLS estimation of Y_i on X_i's and W:

1. Y_i = "If worked for pay" for all women sample

```
In [35]: est1.summary()
```

Out[35]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.043
Model:	OLS	**Adj. R-squared:**	0.043
Method:	Least Squares	**F-statistic:**	3237.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	02:33:50	**Log-Likelihood:**	-2.9837e+05
No. Observations:	428951	**AIC:**	5.968e+05
Df Residuals:	428944	**BIC:**	5.968e+05
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.2346	0.006	41.288	0.000	0.223	0.246
x1	0.0184	0.000	80.381	0.000	0.018	0.019
x2	-0.0301	0.000	-102.668	0.000	-0.031	-0.030
x3	0.0308	0.000	89.501	0.000	0.030	0.031
x4	0.0240	0.002	11.115	0.000	0.020	0.028
x5	0.1068	0.002	43.775	0.000	0.102	0.112
x6	0.0387	0.004	10.357	0.000	0.031	0.046
x7	0.0652	0.004	16.469	0.000	0.057	0.073

Omnibus:	3877.888	**Durbin-Watson:**	1.918
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	60043.959
Skew:	-0.236	**Prob(JB):**	0.00
Kurtosis:	1.229	**Cond. No.**	2.47e+16

2. Y_i = "If worked for pay" for married women sample

In [36]:
```
est2.summary()
```

Out[36]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.045
Model:	OLS	**Adj. R-squared:**	0.045
Method:	Least Squares	**F-statistic:**	3250.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	02:49:45	**Log-Likelihood:**	-2.8866e+05
No. Observations:	415914	**AIC:**	5.773e+05
Df Residuals:	415907	**BIC:**	5.774e+05
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.2729	0.006	46.820	0.000	0.262	0.284
x1	0.0177	0.000	75.690	0.000	0.017	0.018
x2	-0.0305	0.000	-102.868	0.000	-0.031	-0.030
x3	0.0302	0.000	86.247	0.000	0.029	0.031
x4	0.0237	0.002	10.847	0.000	0.019	0.028
x5	0.1360	0.003	51.798	0.000	0.131	0.141
x6	0.0468	0.004	12.214	0.000	0.039	0.054
x7	0.0664	0.004	16.580	0.000	0.059	0.074

Omnibus:	4055.219	**Durbin-Watson:**	1.925
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	57948.259
Skew:	-0.245	**Prob(JB):**	0.00
Kurtosis:	1.238	**Cond. No.**	3.86e+16

3. Y_i = "If worked for pay" for husbands of married women sample

```
In [37]: est3.summary()
```

Out[37]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.021
Model:	OLS	**Adj. R-squared:**	0.021
Method:	Least Squares	**F-statistic:**	1267.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	02:51:04	**Log-Likelihood:**	1.4836e+05
No. Observations:	361306	**AIC:**	-2.967e+05
Df Residuals:	361299	**BIC:**	-2.966e+05
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.7606	0.002	452.732	0.000	0.757	0.764
x1	-0.0012	7.74e-05	-15.806	0.000	-0.001	-0.001
x2	-0.0011	8.74e-05	-12.393	0.000	-0.001	-0.001
x3	0.0053	8.92e-05	59.668	0.000	0.005	0.005
x4	0.2154	0.001	306.216	0.000	0.214	0.217
x5	0.1756	0.001	182.928	0.000	0.174	0.178
x6	0.1925	0.001	143.045	0.000	0.190	0.195
x7	0.1771	0.001	122.404	0.000	0.174	0.180

Omnibus:	381665.219	**Durbin-Watson:**	1.975
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	16235259.185
Skew:	-5.660	**Prob(JB):**	0.00
Kurtosis:	33.827	**Cond. No.**	1.33e+17

4. Y_i = “weeks worked” for all women sample

In [38]:
```
est4.summary()
```

Out[38]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.059
Model:	OLS	**Adj. R-squared:**	0.059
Method:	Least Squares	**F-statistic:**	4447.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	02:53:10	**Log-Likelihood:**	-1.9262e+06
No. Observations:	428951	**AIC:**	3.852e+06
Df Residuals:	428944	**BIC:**	3.852e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.2634	0.253	1.042	0.297	-0.232	0.759
x1	1.1915	0.010	117.307	0.000	1.172	1.211
x2	-1.4013	0.013	-107.559	0.000	-1.427	-1.376
x3	1.2758	0.015	83.395	0.000	1.246	1.306
x4	-2.1591	0.096	-22.518	0.000	-2.347	-1.971
x5	3.1550	0.108	29.084	0.000	2.942	3.368
x6	-1.2747	0.166	-7.679	0.000	-1.600	-0.949
x7	0.5422	0.176	3.080	0.002	0.197	0.887

Omnibus:	9218.103	**Durbin-Watson:**	1.916
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	45945.441
Skew:	0.371	**Prob(JB):**	0.00
Kurtosis:	1.578	**Cond. No.**	2.47e+16

5. Y_i = "weeks worked" for married women sample

In [39]:
```
est5.summary()
```

Out[39]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.060
Model:	OLS	**Adj. R-squared:**	0.060
Method:	Least Squares	**F-statistic:**	4435.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	02:54:34	**Log-Likelihood:**	-1.8675e+06
No. Observations:	415914	**AIC:**	3.735e+06
Df Residuals:	415907	**BIC:**	3.735e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.0354	0.260	7.842	0.000	1.527	2.544
x1	1.1583	0.010	111.564	0.000	1.138	1.179
x2	-1.4207	0.013	-107.611	0.000	-1.447	-1.395
x3	1.2463	0.016	79.983	0.000	1.216	1.277
x4	-2.1506	0.097	-22.106	0.000	-2.341	-1.960
x5	4.5730	0.117	39.121	0.000	4.344	4.802
x6	-1.0189	0.171	-5.970	0.000	-1.353	-0.684
x7	0.6320	0.178	3.544	0.000	0.282	0.982

Omnibus:	8596.948	**Durbin-Watson:**	1.922
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	44108.436
Skew:	0.363	**Prob(JB):**	0.00
Kurtosis:	1.580	**Cond. No.**	3.86e+16

6. Y_i = “weeks worked” for husbands of married women sample

```
In [40]: est6.summary()
```

Out[40]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.042
Model:	OLS	**Adj. R-squared:**	0.042
Method:	Least Squares	**F-statistic:**	2625.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	03:01:19	**Log-Likelihood:**	-1.3791e+06
No. Observations:	361306	**AIC:**	2.758e+06
Df Residuals:	361299	**BIC:**	2.758e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	30.9877	0.115	269.071	0.000	30.762	31.213
x1	0.0278	0.005	5.245	0.000	0.017	0.038
x2	-0.0814	0.006	-13.597	0.000	-0.093	-0.070
x3	0.6016	0.006	98.374	0.000	0.590	0.614
x4	10.2354	0.048	212.250	0.000	10.141	10.330
x5	6.5047	0.066	98.840	0.000	6.376	6.634
x6	7.9678	0.092	86.381	0.000	7.787	8.149
x7	6.2799	0.099	63.323	0.000	6.085	6.474

Omnibus:	212908.486	**Durbin-Watson:**	1.937
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1540081.208
Skew:	-2.888	**Prob(JB):**	0.00
Kurtosis:	11.302	**Cond. No.**	1.33e+17

7. Y_i = "hours worked" for all women sample

In [41]:
```
est7.summary()
```

Out[41]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.047
Model:	OLS	**Adj. R-squared:**	0.047
Method:	Least Squares	**F-statistic:**	3539.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	03:02:19	**Log-Likelihood:**	-1.8598e+06
No. Observations:	428951	**AIC:**	3.720e+06
Df Residuals:	428944	**BIC:**	3.720e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	13.2595	0.216	61.260	0.000	12.835	13.684
x1	0.6848	0.009	78.704	0.000	0.668	0.702
x2	-1.2928	0.011	-115.838	0.000	-1.315	-1.271
x3	0.8691	0.013	66.322	0.000	0.843	0.895
x4	0.7843	0.082	9.549	0.000	0.623	0.945
x5	5.5509	0.093	59.736	0.000	5.369	5.733
x6	2.4413	0.142	17.169	0.000	2.163	2.720
x7	4.4829	0.151	29.732	0.000	4.187	4.778

Omnibus:	66922.114	**Durbin-Watson:**	1.902
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	23249.027
Skew:	0.351	**Prob(JB):**	0.00
Kurtosis:	2.102	**Cond. No.**	2.47e+16

8. Y_i = "hours worked" for married women sample

In [42]:
```
est8.summary()
```

Out[42]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.051
Model:	OLS	**Adj. R-squared:**	0.051
Method:	Least Squares	**F-statistic:**	3691.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	03:03:13	**Log-Likelihood:**	-1.8020e+06
No. Observations:	415914	**AIC:**	3.604e+06
Df Residuals:	415907	**BIC:**	3.604e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	14.8960	0.222	67.180	0.000	14.461	15.331
x1	0.6560	0.009	73.963	0.000	0.639	0.673
x2	-1.3129	0.011	-116.404	0.000	-1.335	-1.291
x3	0.8411	0.013	63.185	0.000	0.815	0.867
x4	0.7885	0.083	9.487	0.000	0.626	0.951
x5	6.7925	0.100	68.018	0.000	6.597	6.988
x6	2.7532	0.146	18.882	0.000	2.467	3.039
x7	4.5619	0.152	29.939	0.000	4.263	4.861

Omnibus:	61611.144	**Durbin-Watson:**	1.910
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	21928.183
Skew:	0.346	**Prob(JB):**	0.00
Kurtosis:	2.113	**Cond. No.**	3.86e+16

9. Y_i = "hours worked" for husbands of married women sample

In [43]:
```
est9.summary()
```

Out[43]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.030
Model:	OLS	**Adj. R-squared:**	0.030
Method:	Least Squares	**F-statistic:**	1883.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	03:04:25	**Log-Likelihood:**	-1.4224e+06
No. Observations:	361306	**AIC:**	2.845e+06
Df Residuals:	361299	**BIC:**	2.845e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	29.9670	0.130	230.771	0.000	29.712	30.221
x1	-0.0172	0.006	-2.882	0.004	-0.029	-0.006
x2	-0.0728	0.007	-10.787	0.000	-0.086	-0.060
x3	0.4384	0.007	63.572	0.000	0.425	0.452
x4	10.4730	0.054	192.608	0.000	10.366	10.580
x5	5.1968	0.074	70.033	0.000	5.051	5.342
x6	7.1298	0.104	68.551	0.000	6.926	7.334
x7	7.1674	0.112	64.095	0.000	6.948	7.387

Omnibus:	42539.300	**Durbin-Watson:**	1.937
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	421617.999
Skew:	-0.148	**Prob(JB):**	0.00
Kurtosis:	8.284	**Cond. No.**	1.33e+17

10. Y_i = "annual labour income" for all women sample

```
In [44]: est10.summary()
```

Out[44]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.062
Model:	OLS	**Adj. R-squared:**	0.062
Method:	Least Squares	**F-statistic:**	4736.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	03:06:39	**Log-Likelihood:**	-4.5779e+06
No. Observations:	428951	**AIC:**	9.156e+06
Df Residuals:	428944	**BIC:**	9.156e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-4648.7837	122.266	-38.022	0.000	-4888.421	-4409.147
x1	518.3108	4.915	105.456	0.000	508.678	527.944
x2	-514.6370	6.304	-81.634	0.000	-526.993	-502.281
x3	785.1222	7.402	106.064	0.000	770.614	799.631
x4	-3032.7934	46.397	-65.366	0.000	-3123.730	-2941.856
x5	300.6907	52.491	5.728	0.000	197.810	403.571
x6	-1496.1664	80.324	-18.627	0.000	-1653.599	-1338.734
x7	-420.5146	85.170	-4.937	0.000	-587.446	-253.583

Omnibus:	306878.777	**Durbin-Watson:**	1.936
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	11227870.026
Skew:	3.024	**Prob(JB):**	0.00
Kurtosis:	27.324	**Cond. No.**	2.47e+16

11. Y_i = "annual labour income" for married women sample

```
In [45]: est11.summary()
```

Out[45]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.064
Model:	OLS	**Adj. R-squared:**	0.064
Method:	Least Squares	**F-statistic:**	4712.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	03:07:48	**Log-Likelihood:**	-4.4394e+06
No. Observations:	415914	**AIC:**	8.879e+06
Df Residuals:	415907	**BIC:**	8.879e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	-3932.5618	125.834	-31.252	0.000	-4179.192	-3685.932
x1	503.0049	5.034	99.931	0.000	493.139	512.870
x2	-522.2820	6.401	-81.599	0.000	-534.827	-509.737
x3	779.3128	7.554	103.161	0.000	764.507	794.119
x4	-3058.8449	47.166	-64.853	0.000	-3151.289	-2966.401
x5	941.2500	56.672	16.609	0.000	830.175	1052.325
x6	-1423.8238	82.745	-17.207	0.000	-1586.002	-1261.646
x7	-391.1430	86.473	-4.523	0.000	-560.627	-221.659

Omnibus:	296518.835	**Durbin-Watson:**	1.939
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	10722113.648
Skew:	3.012	**Prob(JB):**	0.00
Kurtosis:	27.134	**Cond. No.**	3.86e+16

12. Y_i = "annual labour income" for husbands of married women sample

```
In [46]: est12.summary()
```

Out[46]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.157
Model:	OLS	**Adj. R-squared:**	0.157
Method:	Least Squares	**F-statistic:**	1.120e+04
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	03:08:34	**Log-Likelihood:**	-4.1339e+06
No. Observations:	361306	**AIC:**	8.268e+06
Df Residuals:	361299	**BIC:**	8.268e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.76e+04	235.870	-74.637	0.000	-1.81e+04	-1.71e+04
x1	767.5559	10.872	70.603	0.000	746.248	788.864
x2	-173.6417	12.263	-14.160	0.000	-197.677	-149.606
x3	2706.0122	12.525	216.045	0.000	2681.463	2730.561
x4	201.5779	98.766	2.041	0.041	7.999	395.157
x5	-9481.6107	134.786	-70.346	0.000	-9745.787	-9217.434
x6	-3436.0413	188.918	-18.188	0.000	-3806.315	-3065.768
x7	-4888.5288	203.116	-24.068	0.000	-5286.630	-4490.428

Omnibus:	179164.278	**Durbin-Watson:**	1.789
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	2478899.964
Skew:	2.051	**Prob(JB):**	0.00
Kurtosis:	15.159	**Cond. No.**	1.33e+17

13. Y_i = "log of family income" for all women sample

In [47]:
```
est13.summary()
```

Out[47]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.077
Model:	OLS	**Adj. R-squared:**	0.077
Method:	Least Squares	**F-statistic:**	6004.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	03:09:39	**Log-Likelihood:**	-7.2526e+05
No. Observations:	428951	**AIC:**	1.451e+06
Df Residuals:	428944	**BIC:**	1.451e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.1652	0.015	401.122	0.000	6.135	6.195
x1	0.0404	0.001	65.390	0.000	0.039	0.042
x2	0.0070	0.001	8.864	0.000	0.005	0.009
x3	0.0875	0.001	94.014	0.000	0.086	0.089
x4	1.8106	0.006	310.440	0.000	1.799	1.822
x5	1.2002	0.007	181.892	0.000	1.187	1.213
x6	1.5246	0.010	150.986	0.000	1.505	1.544
x7	1.6297	0.011	152.218	0.000	1.609	1.651

Omnibus:	451190.371	**Durbin-Watson:**	1.925
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	28102283.354
Skew:	-5.421	**Prob(JB):**	0.00
Kurtosis:	41.141	**Cond. No.**	2.47e+16

14. Y_i = "log of family income" for married women sample

In [48]:
```
est14.summary()
```

Out[48]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.060
Model:	OLS	**Adj. R-squared:**	0.060
Method:	Least Squares	**F-statistic:**	4399.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	03:10:26	**Log-Likelihood:**	-6.8992e+05
No. Observations:	415914	**AIC:**	1.380e+06
Df Residuals:	415907	**BIC:**	1.380e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.3599	0.015	415.733	0.000	6.330	6.390
x1	0.0356	0.001	58.213	0.000	0.034	0.037
x2	0.0078	0.001	9.997	0.000	0.006	0.009
x3	0.0832	0.001	90.590	0.000	0.081	0.085
x4	1.8036	0.006	314.535	0.000	1.792	1.815
x5	1.3547	0.007	196.629	0.000	1.341	1.368
x6	1.5679	0.010	155.862	0.000	1.548	1.588
x7	1.6337	0.011	155.399	0.000	1.613	1.654

Omnibus:	451157.873	**Durbin-Watson:**	1.932
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	32222198.898
Skew:	-5.660	**Prob(JB):**	0.00
Kurtosis:	44.608	**Cond. No.**	3.86e+16

15. Y_i = "log of non-wife income" for married women sample

```
In [49]: est15.summary()
```

Out[49]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.054
Model:	OLS	**Adj. R-squared:**	0.054
Method:	Least Squares	**F-statistic:**	3942.
Date:	Thu, 25 Apr 2019	**Prob (F-statistic):**	0.00
Time:	03:11:15	**Log-Likelihood:**	-9.3014e+05
No. Observations:	415914	**AIC:**	1.860e+06
Df Residuals:	415907	**BIC:**	1.860e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	5.7699	0.027	211.691	0.000	5.717	5.823
x1	0.0047	0.001	4.331	0.000	0.003	0.007
x2	0.0686	0.001	49.447	0.000	0.066	0.071
x3	0.0523	0.002	31.962	0.000	0.049	0.056
x4	1.9585	0.010	191.695	0.000	1.938	1.978
x5	0.6295	0.012	51.283	0.000	0.605	0.654
x6	1.5329	0.018	85.527	0.000	1.498	1.568
x7	1.6491	0.019	88.041	0.000	1.612	1.686

Omnibus:	288515.662	**Durbin-Watson:**	1.945
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3127479.221
Skew:	-3.399	**Prob(JB):**	0.00
Kurtosis:	14.587	**Cond. No.**	3.86e+16