AAE 637 – Applied Econometrics II – Spring 2018

Assignment #1 – solutions

Answer Key by Eduardo Cenci, with edits by Charng-Jiun Yu and Adam Theising

1)

1.a)

The fact that the individual values of the x_i's are independent from each other imply that $\operatorname{var}(x_i) = \sigma^2, \forall i = 1, \ldots, N$. Therefore

$$\begin{aligned}
\operatorname{var}(\hat{\mu}) &= \operatorname{var}\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) \\
&= \operatorname{var}\left(\frac{1}{N}x_1 + \cdots + \frac{1}{N}x_N\right) \\
&= \frac{1}{N^2}\operatorname{var}(x_1) + \cdots + \frac{1}{N^2}\operatorname{var}(x_N) \\
&= \frac{1}{N^2}(N\operatorname{var}(x_i)) \\
&= \frac{1}{N}\sigma^2
\end{aligned}$$

1.b)

The sample mean of per capita electricity consumption in the data is $\hat{\mu} = 5{,}111$. To test whether this value is statically equal to $3{,}750$ we can use the following t-test:

$$t = \frac{\hat{\mu} - \mu_0}{s(\hat{\mu})}, \qquad s(\hat{\mu}) = \sqrt{\operatorname{var}(\hat{\mu})} = \frac{1}{\sqrt{N}}\sigma \therefore t = \sqrt{N}\left(\frac{\hat{\mu} - \mu_0}{\sigma}\right)$$

The standard deviation in the data is $\sigma = 3{,}919$. Therefore

$$t = \sqrt{12018}\left(\frac{5111 - 3750}{3919}\right) = 10.15$$

The value of the t-statistic is great enough to guarantee significance at the 5% level ($c = 1.96$), at the 1% level ($c = 2.576$), or less. Thus, we can reject the hypothesis that the sample mean is equal to $4{,}750$ KWH per capita.

Matlab output

```
The t-stat for a test that the mean of PC KWH = 3750 is 38.1857 and its p-value is 0.0000
```

1.c)

For the sub sample of households with less than $2{,}750$ sqft of heating/cooling we have

$$\hat{\mu}_1 = 4{,}856, \qquad \sigma_1 = 3{,}647, \qquad N_1 = 9{,}404,$$

For the sub sample of households with $2{,}750$ sqft of heating/cooling or more we have

$$\hat{\mu}_2 = 6{,}012, \qquad \sigma_2 = 4{,}641, \qquad N_2 = 2{,}679$$

Denote by $\widehat{D}$ the difference $\hat{\mu}_2 - \hat{\mu}_1$. We want to test $H_0: \widehat{D} = 0$. The variance of the difference is given by

$$\operatorname{var}(\widehat{D}) = \operatorname{var}(\hat{\mu}_2 - \hat{\mu}_1) = \operatorname{var}(\hat{\mu}_2) + \operatorname{var}(\hat{\mu}_1) - 2\operatorname{cov}(\hat{\mu}_2, \hat{\mu}_1) = N_2^{-1}\sigma_2^2 + N_1^{-1}\sigma_1^2$$

because independence implies $\operatorname{cov}(\hat{\mu}_2, \hat{\mu}_1) = 0$. The t-statistic will be

$$t = \frac{\widehat{D}}{\sqrt{\operatorname{var}(\widehat{D})}} = 11.8959$$

Therefore we can reject H_0, i.e., that the two means are equal.

Matlab output:

```
The t-stat for testing the equality of the subgroup means is   11.8959 (p-value = 0.0000)
```

2)

Regression equation

$$\begin{aligned}
KWH_t = \beta_0 + \beta_1 \ln(HDD65_t) + \beta_2 \ln(CDD65_t) + \beta_3 \ln(HouseAge_t) + \beta_4 HHSize_t + \beta_5 \ln(Tot_SqFt_H/C_t) \\
+ \beta_6 \ln(HH_Inc_t) + \beta_7 Elec_Pr_t + \varepsilon_t \qquad (2.1)
\end{aligned}$$

2.a)

I dropped all observations that would produce a null value for the log of the independent variable, and I used this sub-sample in all the following exercises

Little more than 33% of the total variation in electric energy consumption about the mean is explained by the estimated regression according to obtained R^2 and adjusted R^2.

Matlab output:

```
Nr. of observations used =    12018
F-statistic = 856.7742
R^2 =  0.3331
adjusted R^2 =  0.3327
Estimate of the error variance = 39029851.5401

Table of Results
-----------------------------------------------------------------
     Variables      Value    Std.Err    T-Value    P-Value
-----------------------------------------------------------------
     intercept  -37118.06    1627.48   -22.8070     0.0000
      ln_HDD65     213.81      98.04     2.1809     0.0292
      ln_CDD65    2353.56      97.17    24.2201     0.0000
  ln_House_Age    -563.21      64.93    -8.6739     0.0000
        HHSize     919.97      38.78    23.7227     0.0000
ln_Tot_SqFt_HC    4156.01     103.49    40.1587     0.0000
     ln_HH_Inc     377.59      65.73     5.7450     0.0000
       Elec_Pr  -39030.99    1288.47   -30.2926     0.0000
-----------------------------------------------------------------

Parameter covariance matrix =
   1.0e+06 *

    2.6487   -0.1212   -0.1288   -0.0268    0.0028   -0.0380   -0.0347   -0.3556
   -0.1212    0.0096    0.0069   -0.0000    0.0001   -0.0016    0.0006    0.0128
   -0.1288    0.0069    0.0094    0.0007   -0.0001   -0.0003    0.0006    0.0126
   -0.0268   -0.0000    0.0007    0.0042    0.0000    0.0007    0.0004   -0.0117
    0.0028    0.0001   -0.0001    0.0000    0.0015   -0.0007   -0.0002   -0.0008
   -0.0380   -0.0016   -0.0003    0.0007   -0.0007    0.0107   -0.0026    0.0067
   -0.0347    0.0006    0.0006    0.0004   -0.0002   -0.0026    0.0043   -0.0053
   -0.3556    0.0128    0.0126   -0.0117   -0.0008    0.0067   -0.0053    1.6601
```

2.b)

The coefficients make sense: both increase electricity usage, but cooling degree days have a much higher effect (almost an order of magnitude). To test if the difference between the two coefficients is statistically significant, we perform a t-test similar to the previous one.

Specifically, let $\hat{\mu}_H$ denote the mean of $HDD65_t$ and $\hat{\mu}_C$ denote the mean of $CDD65_t$. The marginal effects are

$$\left.\frac{\partial KWH_t}{\partial HDD65_t}\right|_{HDD65_t=\hat{\mu}_H} = \frac{\beta_1}{\hat{\mu}_H}, \qquad \left.\frac{\partial KWH_t}{\partial CDD65_t}\right|_{CDD65_t=\hat{\mu}_C} = \frac{\beta_2}{\hat{\mu}_C}$$

Denote by θ their difference. We want to test $H_0\colon \theta = 0$.

The variance of the difference is given by

$$\operatorname{var}(\hat{\theta}) = \operatorname{var}\left(\frac{\hat{\beta}_1}{\hat{\mu}_H} - \frac{\hat{\beta}_2}{\hat{\mu}_C}\right) = \operatorname{var}\left(\frac{\hat{\beta}_1}{\hat{\mu}_H}\right) + \operatorname{var}\left(\frac{\hat{\beta}_2}{\hat{\mu}_C}\right) - 2\operatorname{cov}\left(\frac{\hat{\beta}_1}{\hat{\mu}_H}, \frac{\hat{\beta}_2}{\hat{\mu}_C}\right) = \left[\hat{\mu}_H^{-2}\operatorname{var}(\hat{\beta}_1) + \hat{\mu}_C^{-2}\operatorname{var}(\hat{\beta}_2) - 2\hat{\mu}_H^{-1}\hat{\mu}_C^{-1}\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2)\right]$$

and the t-statistic will be

$$t = \hat{\theta}\Big/\sqrt{\operatorname{var}(\hat{\theta})} = -29.8132$$

Therefore we can reject H_0, i.e., that the marginal effect of HDD's equals the marginal effect of CDD's.

Matlab output:

```
The marginal effect of HDD at the mean of the data is   0.0516
The marginal effect of CDD at the mean of the data is   1.6677
```

```
The t-stat for testing the equality of marginal effects is -29.8132 (p-value = 0.0000)
```

2.c)

The elasticity of HDD is given by

$$\varepsilon_H = \frac{\partial KWH_t}{\partial HDD65_t} \cdot \frac{HDD65_t}{KWH_t} \therefore \varepsilon_H|_{HDD65_t=\hat{\mu}_H} = \frac{\beta_1}{\hat{\mu}_H} \cdot \frac{\hat{\mu}_H}{KWH_t} = \frac{\beta_1}{KWH_t}$$

Therefore, both elasticities evaluated at the mean of the data ($\hat{\mu}_K$) will be

$$\varepsilon_H|_{KWH_{t_t}=\hat{\mu}_K} = \frac{\beta_1}{\hat{\mu}_K}, \qquad \varepsilon_C|_{KWH_{t_t}=\hat{\mu}_K} = \frac{\beta_2}{\hat{\mu}_K}$$

Again, denote by θ their difference. We want to test H_0:

$$\theta = 0 \Leftrightarrow \frac{1}{\hat{\mu}_K}(\beta_1 - \beta_2) = 0 \Leftrightarrow \beta_1 - \beta_2 = 0$$

so we can drop the constant $\hat{\mu}_K^{-1}$ and simplify our calculations. Call δ this simplified difference. Its variance is

$$\operatorname{var}(\hat{\delta}) = \operatorname{var}(\hat{\beta}_1 - \hat{\beta}_2) = [\operatorname{var}(\hat{\beta}_1) + \operatorname{var}(\hat{\beta}_2) - 2\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2)]$$

and the t-statistic will be

$$t = \hat{\delta}/\sqrt{\operatorname{var}(\hat{\delta})} = -29.5662$$

Therefore we can reject H_0, i.e., that the elasticity of HDD's equals the elasticity of CDD's.

Matlab output:

```
The elasticity of HDD at the mean of the data is   0.0189
The elasticity of CDD at the mean of the data is   0.2084
The t-stat for testing the equality of elasticities is -29.5662 (p-value = 0.0000)
```

Note that running a new “log-log” regression will note yield the same elasticities from above or from the next question, because a log-log model implicitly assumes constant elasticity over all values of the data (here we are evaluating it at the mean).

2.d)

The price elasticity evaluated at the mean of the data is

$$\varepsilon_P = \frac{\partial KWH_t}{\partial Elec\_Pr_t} \cdot \frac{Elec\_Pr_t}{KWH_t} \therefore \varepsilon_P|_{KWH=\hat{\mu}_K, Elec\_Pr_t=\hat{\mu}_P} = \beta_7 \frac{\hat{\mu}_P}{\hat{\mu}_K}$$

We want to test H_0: $\varepsilon_P = 0$ and H_0: $\varepsilon_P - (-1) < 0$. The variance of the elasticity is

$$\operatorname{var}(\hat{\varepsilon}) = \operatorname{var}\left(\frac{\hat{\mu}_P}{\hat{\mu}_K}\hat{\beta}_7\right) = \left(\frac{\hat{\mu}_P}{\hat{\mu}_K}\right)^2 \operatorname{var}(\hat{\beta}_7)$$

and the t-statistics will be

$$t = \frac{\hat{\varepsilon}}{\sqrt{\operatorname{var}(\hat{\varepsilon})}} = -30.2926, \qquad t = \frac{\hat{\varepsilon}+1}{\sqrt{\operatorname{var}(\hat{\varepsilon})}} = 38.1926$$

Therefore we can reject H_0 in both cases, i.e., that the price elasticity is equal to 0 or less than -1.

Matlab output:

```
The price elasticity at the mean of the data is  -0.4423
The t-stat for testing price elasticity equal to zero is -30.2926 (p-value = 0.0000)
The t-stat for testing price elasticity less than -1 is  38.1926 (p-value = 0.0000)
```

2.e)

In this question, we compare two models: the complete model and the restricted model, where CDD and HDD are excluded as explanatory variables. We want to test if their coefficients are jointly equal to zero. To test this we carry the following F-test.

$$F = \frac{(SSE_R - SSE_C)/q}{SSE_C/(n-k)} = 542.8102$$

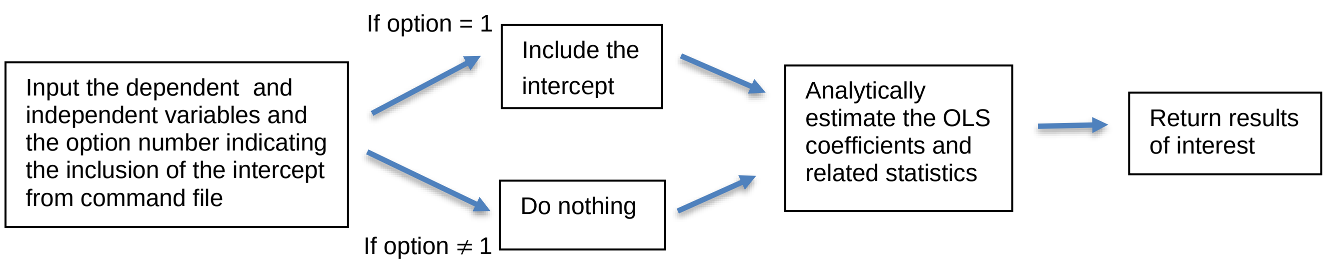
where k is the number of covariates in the complete model, q is the number of restrictions (covariates excluded in the restricted model), and SSE are the sum of the squared errors in each model.

We compare the results to the desired critical value in the F table is F_{v_1,v_2} where $v_1 = q$ and $v_2 = n - k$. In this case, we can reject H_0 i.e., that the coefficients for CDD and HDD are zero making the restricted model, the true model.

```
Matlab output:
The F-stat to test if the restricted model is "true" is 542.8102
```

3)
Develop your own function to answer the question #2.

3.a)



3.b)
The results are identical to those from question #2a table.

4)
I scaled the new variable by 1,000 to produce coefficients that are similar in magnitude and to avoid precision errors

$$HHINC \times Price = (HH\_Inc \times Elec\_\mathrm{Pr})/1000$$

The new estimated equation, therefore, is

$$KWH_t = \beta_0 + \beta_1 \ln(HDD65_t) + \beta_2 \ln(CDD65_t) + \beta_3 \ln(HouseAge_t) + \beta_4 HHSize_t + \beta_5 \ln(Tot\_SqFt\_H/C_t) + \beta_6 \ln(HH\_Inc_t) + \beta_7 Elec\_Pr_t + \beta_8 (HH\_Inc_t \times Elec\_Pr_t)/1000 + \varepsilon_t \qquad (3.1)$$

4.a)
The estimated marginal effect of a change in the price of electricity is

$$\frac{\partial KWH_t}{\partial Elec\_Pr_t} = \hat{\beta}_7 + \hat{\beta}_8 \frac{HH\_Inc}{1000}$$

Therefore, it changes with the level of income, unless $\beta_8 = 0$ which is not the case as we can tell by the p-value of this coefficient in the results table below ($\text{t-stat} = 5.7506$, $\text{p-value} = 0.0000$).

```
Matlab output:
Nr. of observations used =    12018
F-statistic = 755.8128
R^2 =  0.3349
adjusted R^2 =  0.3344
Estimate of the error variance = 38925911.9178
-----------------------------------------------------------------
-----------------------------------------------------------------
 Variables      Value     Std.Err     T-Value     P-Value
-----------------------------------------------------------------
 intercept -31822.49769 1868.06550  -17.03500    0.00000
  ln_HDD65   242.34051   98.03228    2.47205    0.01345
  ln_CDD65  2378.49646   97.14131   24.48491    0.00000
ln_House_Age -547.92265   64.89928   -8.44266    0.00000
    HHSize   911.29535   38.75797   23.51246    0.00000
```

```
ln_Tot_SqFt_HC 4046.04387  105.10574   38.49498    0.00000
 ln_HH_Inc   -82.72533  103.51789   -0.79914    0.42422
   Elec_Pr -45750.82962 1738.17077  -26.32125    0.00000
Income x Price  115.46136   20.07826    5.75056    0.00000

-----------------------------------------------------------------

Parameter covariance matrix =
   1.0e+06 *

    3.4897   -0.1163   -0.1245   -0.0243    0.0014   -0.0555   -0.1083   -1.4307    0.0185
   -0.1163    0.0096    0.0069    0.0000    0.0001   -0.0017    0.0002    0.0070    0.0001
   -0.1245    0.0069    0.0094    0.0008   -0.0001   -0.0004    0.0002    0.0075    0.0001
   -0.0243    0.0000    0.0008    0.0042    0.0000    0.0006    0.0002   -0.0148    0.0001
    0.0014    0.0001   -0.0001    0.0000    0.0015   -0.0006   -0.0001    0.0010   -0.0000
   -0.0555   -0.0017   -0.0004    0.0006   -0.0006    0.0110   -0.0011    0.0290   -0.0004
   -0.1083    0.0002    0.0002    0.0002   -0.0001   -0.0011    0.0107    0.0883   -0.0016
   -1.4307    0.0070    0.0075   -0.0148    0.0010    0.0290    0.0883    3.0212   -0.0235
    0.0185    0.0001    0.0001    0.0001   -0.0000   -0.0004   -0.0016   -0.0235    0.0004
```

4.b)

The price elasticity evaluated at the mean of the data is

$$\varepsilon_P = \frac{\partial KWH_t}{\partial Elec\_Pr_t} \cdot \frac{Elec\_Pr_t}{KWH_t} \therefore \hat{\varepsilon}_P|_{KWH=\hat{\mu}_K, Elec\_Pr_t=\hat{\mu}_P, HH\_Inc_t=\hat{\mu}_I} = \left(\hat{\beta}_7 + \hat{\beta}_8 \frac{\hat{\mu}_I}{1000}\right)\frac{\hat{\mu}_P}{\hat{\mu}_K}$$

We want to test $H_0\colon \varepsilon_P - (-1) = \varepsilon_P + 1 = 0$.

Let by $\widehat{m}_I \equiv \hat{\mu}_I/1000$. The variance of the elasticity is

$$\mathrm{var}(\hat{\varepsilon}) = \mathrm{var}\left[\left(\hat{\beta}_7 + \hat{\beta}_8 \widehat{m}_I\right)\frac{\hat{\mu}_P}{\hat{\mu}_K}\right] = \left(\frac{\hat{\mu}_P}{\hat{\mu}_K}\right)^2 \left[\mathrm{var}\left(\hat{\beta}_7 + \hat{\beta}_8 \widehat{m}_I\right)\right] = \left(\frac{\hat{\mu}_P}{\hat{\mu}_K}\right)^2 \left[\mathrm{var}\left(\hat{\beta}_7\right) + \widehat{m}_I^2 \,\mathrm{var}\left(\hat{\beta}_8\right) + 2m_I \,\mathrm{cov}\left(\hat{\beta}_7, \hat{\beta}_8\right)\right]$$

and the t-statistic will be

$$t = \frac{\hat{\varepsilon} + 1}{\sqrt{\mathrm{var}(\hat{\varepsilon})}} = 37.3850$$

Therefore we cannot reject H_0, i.e., that the price elasticity is equal to -1.

Matlab output:

```
The price elasticity at the mean of the data is  -0.4467
The t-stat for testing price elasticity equal to -1 is  37.8876 (p-value = 0.0000)
```

4.c)

The income elasticity evaluated at the mean of the data is

$$\eta = \frac{\partial KWH_t}{\partial HH\_Inc} \cdot \frac{HH\_Inc}{KWH_t} \therefore \hat{\eta}|_{KWH=\hat{\mu}_K, Elec\_Pr_t=\hat{\mu}_P, HH\_Inc_t=\hat{\mu}_I} = \left(\hat{\beta}_6 + \hat{\beta}_8 \frac{\hat{\mu}_P \hat{\mu}_I}{1000}\right)\frac{1}{\hat{\mu}_K} = \text{唷 ´ `}$$

We want to test $H_0\colon \eta - 0.5 = 0$.

Let by $\widehat{m}_{PI} \equiv \hat{\mu}_P \hat{\mu}_I/1000$. The variance of the elasticity is

$$\mathrm{var}(\hat{\eta}) = \mathrm{var}\left(\left(\hat{\beta}_6 + \hat{\beta}_8 \widehat{m}_{PI}\right)\frac{1}{\hat{\mu}_K}\right) = \hat{\mu}_K{}^{-2}\,\mathrm{var}\left(\hat{\beta}_6 + \hat{\beta}_8 m_{PI}\right) = \hat{\mu}_K{}^{-2}\left[\mathrm{var}\left(\hat{\beta}_6\right) + \widehat{m}_{PI}^2 \,\mathrm{var}\left(\hat{\beta}_8\right) + 2\widehat{m}_{PI} \,\mathrm{cov}\left(\hat{\beta}_6, \hat{\beta}_8\right)\right]$$

and the t-statistic will be

$$t = \frac{\hat{\eta} - 0.5}{\sqrt{\mathrm{var}(\hat{\eta})}} = -54.9734$$

Therefore we can reject H_0, i.e., that the income elasticity is equal to 0.5.

Matlab output:

```
The income elasticity at the mean of the data is   0.0644
The t-stat for testing income elasticity equal to 0.5 is -54.9734 (p-value = 0.0000)
```

4.d)

Let $\hat{\mu}_P^{75}$ denote the value of 75% of the sample mean for price. Define $\hat{\mu}_P^{125}$ in a similar fashion. The correspondent income elasticities will be

$$\eta_{75}|_{\hat{\mu}_K,\hat{\mu}_I,\hat{\mu}_P^{75}} = \left(\beta_6 + \beta_8 \frac{\hat{\mu}_I \hat{\mu}_P^{75}}{1000}\right)\frac{1}{\hat{\mu}_K}, \qquad \eta_{125}|_{\hat{\mu}_K,\hat{\mu}_I,\hat{\mu}_P^{125}} = \left(\beta_6 + \beta_8 \frac{\hat{\mu}_I \hat{\mu}_P^{125}}{1000}\right)\frac{1}{\hat{\mu}_K}$$

Denote by θ their difference. We want to test

$$H_0: \theta = 0 \Leftrightarrow \beta_8\left(\hat{\mu}_P^{125} - \hat{\mu}_P^{75}\right)\frac{1}{1000}.\frac{\hat{\mu}_I}{\hat{\mu}_K} = 0 \Leftrightarrow \beta_8 = 0 \text{ or } \hat{\mu}_P^{125} = \hat{\mu}_P^{75}$$

Therefore, the income elasticities will be the same if and only if $\beta_8 = 0$, which is not the case, as seen in 3.a. (The other possible case is when $\hat{\mu}_P^{125} = \hat{\mu}_P^{75}$, which is not an interesting one.)

Alternatively, we can use the usual procedure to calculate the t-statistic:

$$t = \frac{\hat{\theta} - 0}{\sqrt{\mathrm{var}(\hat{\theta})}} = \frac{\hat{\theta} - 0}{\sqrt{\left(\frac{0.75\hat{\mu}_P\hat{\mu}_I}{1000\hat{\mu}_K}\right)^2 \mathrm{var}(\widehat{\beta_8})}} = 5.7506$$

Matlab output:

```
The t-stat for a test of whether income elasticity varies at 75 versus 150 percent of mean income
is 5.7506 (p-value = 0.0000)
```