Solution for HW3 (Mistakes corrected)

1. Type I error: Reject H_0 when it is true
   Type II error: Accept H_0 when it is false
   Suppose we use too high confident level. That means it is easier to accept H_0. Therefore, we create more type II error, vice versa.

2. a)To test whether the IBM is volatile than the market, we test the following hypothesis

$$H_0 : \beta_{IBM} \le 1$$
$$H_a : \beta_{IBM} > 1$$

To test the hypothesis, we use t-stat.

$$\hat{t} = \frac{\beta_{IBM} - 1}{s.e.(\beta_{IBM})} = \frac{1.0598 - 1}{0.0728} = 0.8214$$

Assuming the 95% confident level with the one-tail test, $t_{0.05,238} = 1.645$. Thus, we fail to reject H_0 at 0.05 level of significant. So, IBM is not more volatile than the market

b) To test whether the constant is significant, we test the following hypothesis

$$H_0 : \alpha_{IBM} = 0$$
$$H_a : \alpha_{IBM} \ne 0$$

We still use t-test to test this 2-tail hypothesis testing.

$$\hat{t} = \frac{\alpha_{IBM} - 0}{s.e.(\alpha_{IBM})} = \frac{0.7264 - 0}{0.3001} = 2.42$$

Assuming the 95% confident level with the two-tail test, $t_{\frac{0.05}{2},238} = 1.96$. Thus, we reject H_0 at 0.05 level of significant. So, the constant is significantly different than zero.

So, the test suggests that even though the market return is zero, the IBM still generate the positive return. It proves the existence of Jensen's alpha

3. a) If LFPR of women in 1986 increases by 1 unit, the LFPR in 1972 will be expected to increase by 0.6560. If the LFPR in 1986 is 0, the LFPR in 1972 is expected to be 0.2033.

b) $H_0 : \beta_2 = 1; H_a : \beta_2 > 1$. We use 1-tail t-test to test this hypothesis because it is the individual test.

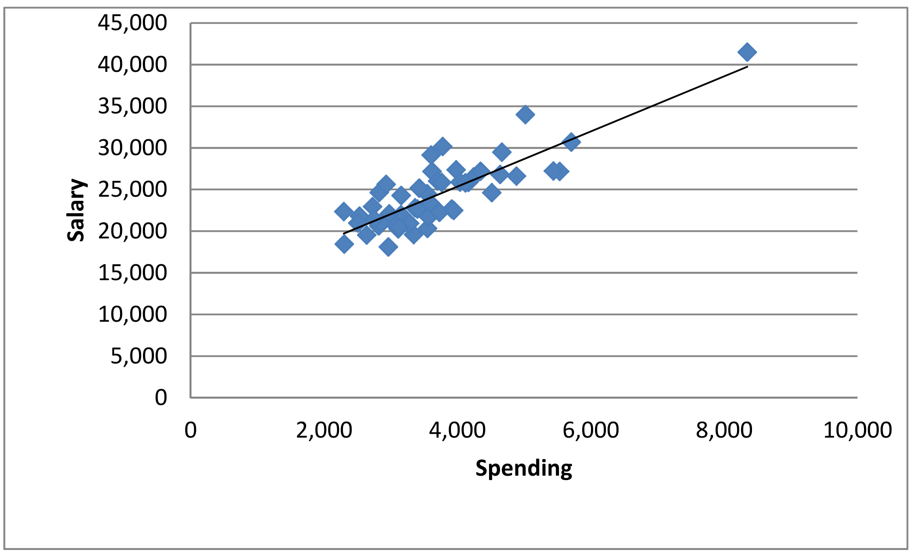
$$\hat{t} = \frac{\beta_2 - 1}{s.e.(\beta_2)} = \frac{0.6560 - 1}{0.1961} = -1.7542$$

Assuming 95% confident level, the critical t-value is $t_{0.05,17} = 1.74$. So, we reject H_0 at 95% level of confident and conclude that $\beta_2 > 1$.

c) $\hat{Y} = 0.2033 + 0.6560(0.58) = 0.5838$. We cannot construct 95% confident level because of insufficient information.

d) You can use Jarque-Bera test. However, we do not have the information on skewness and kurtosis. Thus, we cannot perform this test.

4. a)



b)

```
. reg salary spending

      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  1,    49) =  112.60
       Model |  608555015     1   608555015            Prob > F      =  0.0000
    Residual |  264825250    49  5404596.94            R-squared     =  0.6968
-------------+------------------------------           Adj R-squared =  0.6906
       Total |  873380265    50  17467605.3            Root MSE      =  2324.8

------------------------------------------------------------------------------
      salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spending |   3.307585   .3117043    10.61   0.000     2.681192    3.933978
       _cons |   12129.37   1197.351    10.13   0.000     9723.204    14535.54
------------------------------------------------------------------------------
```

c) According to $\beta_2 = 3.3076$, if the per pupil expenditure increases by 1 dollar, the expected salary will increase by 3.3076 dollars. In addition, if there is no spending on public schools per pupil, the salary will be 12129.37 dollars per year.

The slope of the model makes sense. A district with high spending per pupil implies that district concern more on the education of pupils. So, the teachers in high-spending-per-pupil district can be paid higher as well. On the other hand, the intercept term does not have much economic sense. If the school does not spend any money at all (spending on public schools per pupils = 0), then it means the school does not be able to pay the teachers.

d) The 95% confidence interval of β_2 is

$$\hat{\beta}_2 \pm t_{\frac{0.05}{2},49} se(\hat{\beta}_2) = \left[3.3076 - 2.010(0.3117), 3.3076 + 2.010(0.3117)\right] = [2.6811, 3.9341]$$

From the hypothesis,

$$H_0 : \beta_2 = 3$$
$$H_a : \beta_2 \neq 3$$

Since the 95% confidence interval of β_2 includes 3, therefore we do not to reject that the true slope coefficient is 3.

e) The mean an individual forecast value when per pupil spending is \$5,000 is:

$$Salary = 12129.37 + 3.3076(5000) = 28,667.37$$

Mean prediction:

$$\text{var}(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right] = 5404596.94 \left[\frac{1}{51} + \frac{(5000 - 3696.61)^2}{55,625,998.16} \right] = 271,029.61$$

So, the 95% confidence interval is

$$\hat{Y}_0 \pm t_{\frac{0.05}{2},49} se(\hat{Y}_0) = 28,667.37 \pm 2.01(\sqrt{271,029.61}) = [27620.58, 29713.42]$$

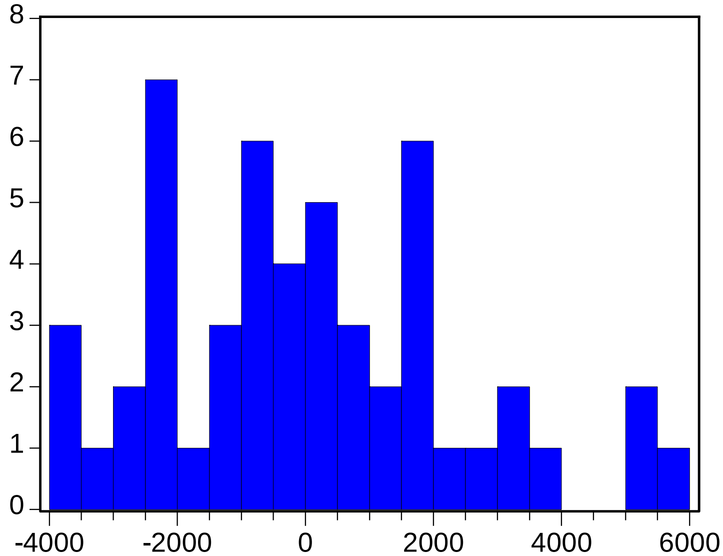
Individual prediction:

$$\text{var}(Y_0 - \hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right] = 5404596.94 \left[1 + \frac{1}{51} + \frac{(5000 - 3696.61)^2}{55,625,998.16} \right] = 5675626.55$$

So, the 95% confidence interval is

$$\hat{Y}_0 \pm t_{\frac{0.05}{2},49} se(Y_0 - \hat{Y}_0) = 28,667.37 \pm 2.01(\sqrt{5675626.55}) = [23878.83, 33455.91]$$

f) We will use JB-test.



Series: Residuals	
Sample 1 51	
Observations 51	
Mean	1.34e-12
Median	-217.5192
Maximum	5529.342
Minimum	-3847.976
Std. Dev.	2301.414
Skewness	0.499126
Kurtosis	2.807557
Jarque-Bera	2.196273
Probability	0.333492

The critical value is $\chi^2_{0.05,2} = 5.99$. So, JB-stat is less than the critical value. We accept the null hypothesis of normality at 95% level of confident.

5. 5.1)

$$\tilde{\beta}_1 = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{\sum X_i(\beta_0 + \beta_1 X_1 + u_i)}{\sum X_i^2}$$

$$\tilde{\beta}_1 = \frac{\beta_0 \sum X_i + \beta_1 \sum X_i^2 + \sum X_i u_i}{\sum X_i^2}$$

$$E(\tilde{\beta}_1) = \beta_1 + \beta_0 E\left(\frac{\sum X_i}{\sum X_i^2}\right) + E\left(\frac{\sum X_i u_i}{\sum X_i^2}\right); \beta_0 = 0 \text{ and } \sum X_i u_i = 0 \text{ (Exogeneity assumption)}$$

$$\therefore E(\tilde{\beta}_1) = \beta_1$$

Another case is when $\sum X_i = 0$.

5.2) From $\tilde{\beta}_1 = \beta_1 + \beta_0 \frac{\sum X_i}{\sum X_i^2} + \frac{\sum X_i u_i}{\sum X_i^2}$

$\tilde{\beta}_1 = \beta_1 + \beta_0 \sum k_i + \sum k_i u_i$ where $k_i = \frac{X_i}{\sum X_i^2}$

$$\text{var}(\tilde{\beta}_1) = \beta_0^2 \text{ var}(\sum k_i) + \text{var}(\sum k_i u_i)$$

$\text{var}(\tilde{\beta}_1) = \left(\sum k_i\right)^2 \text{var}(u_i)$ as X is fixed (by OLS assumption), so does k

$$\therefore \text{var}(\tilde{\beta}_1) = \frac{\text{var}(u_i)}{\sum X_i^2} = \frac{\sigma^2}{\sum X_i^2}$$

5.3) $\text{var}(\tilde{\beta}_1) = \dfrac{\sigma^2}{\sum X_i^2}; \text{var}(\hat{\beta}_1) = \dfrac{\sigma^2}{\sum (X_i - \bar{X})^2}$

Since $\sum X_i^2 \geq \sum (X_i - \bar{X})^2$, the denominator of $\text{var}(\tilde{\beta}_1)$ is more than $\text{var}(\hat{\beta}_1)$, therefore $\text{var}(\tilde{\beta}_1) \leq \text{var}(\hat{\beta}_1)$

5.4) If you use $\tilde{\beta}_1$, you will have a lower variance. However, you may face the problem of unbiasedness. On the other hand, if you use $\hat{\beta}_1$, you would not have the problem of unbiasedness, but you will have higher variance.

6. $Y_i = \beta_1 + \beta_2 X_i + u_i$: If income increases by 1 unit, expenditure will increase by β_2

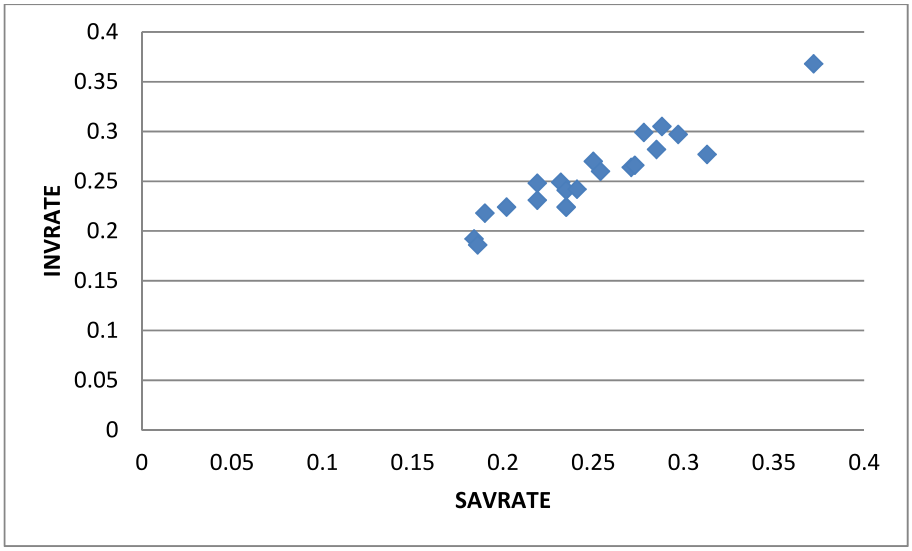
   $Y_i = \beta_1 + \beta_2 (1/X_i) + u_i : \dfrac{dY_i}{dX_i} = -\beta_2 \left(\dfrac{1}{X^2} \right)$ or a unit increase in income will reduce expenditure by $\beta_2 \left(\dfrac{1}{X^2} \right)$.

   $\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i : \dfrac{dY/Y}{dX/X} = \beta_2$ or 1% increase in income will increase expenditure by β_2 percent. Consequently, this one is the most appropriate one.

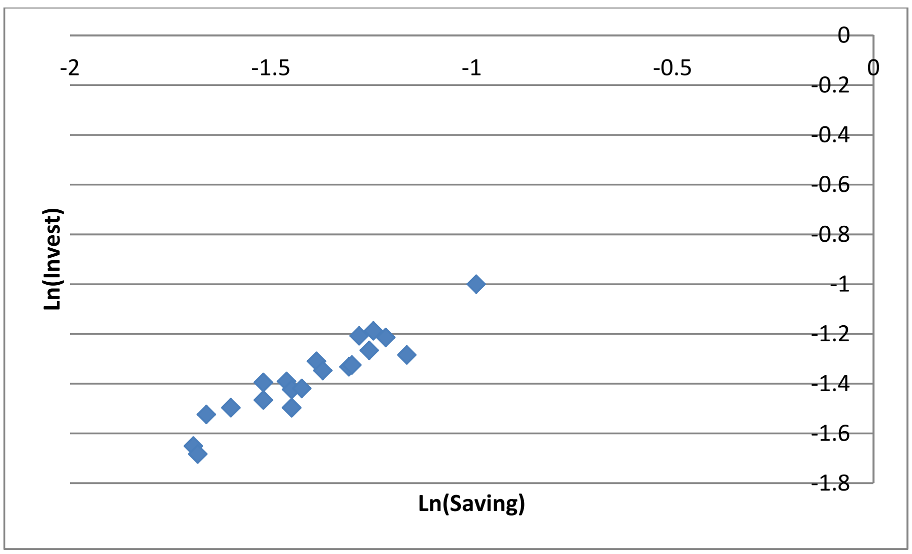
   $\ln Y_i = \ln \beta_1 + \beta_2 \ln(1/X_i) + u_i : \dfrac{dY/Y}{dX/X} = -\beta_2$ or 1% increase in income will decrease expenditure by β_2 percent.

   $Y_i = \beta_1 + \beta_2 \ln X_i + u_i : \dfrac{dY}{dX/X} = \beta_2$ or 1% increase in income will decrease expenditure by β_2 unit.

7. a)



b)



If you compare the scatter plots for both models, they are pretty much the same. Therefore, both models should fit the model equally well.

c) $Invrate = \beta_1 + \beta_2 Savrate + u_t$:

```
                delta:  1 unit
. reg invest saving

      Source |       SS       df       MS              Number of obs =      21
-------------+------------------------------           F(  1,    19) =  149.38
       Model |  .030726967     1  .030726967           Prob > F      =  0.0000
    Residual |  .003908174    19  .000205693           R-squared     =  0.8872
-------------+------------------------------           Adj R-squared =  0.8812
       Total |  .034635142    20  .001731757           Root MSE      =  .01434

------------------------------------------------------------------------------
      invest |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      saving |   .8467562   .0692801    12.22   0.000     .7017512    .9917612
```

```
       _cons |   .0435195   .0176297     2.47   0.023       .00662     .080419
------------------------------------------------------------------------------
```

$\ln Invrate = \alpha_1 + \alpha_2 Savrate + u_t$:

```
. reg l_inv l_sav

      Source |       SS       df       MS              Number of obs =      21
-------------+------------------------------           F(  1,    19) =  140.79
       Model |  .448303526     1  .448303526           Prob > F      =  0.0000
    Residual |  .060501114    19  .003184269           R-squared     =  0.8811
-------------+------------------------------           Adj R-squared =  0.8748
       Total |   .50880464    20  .025440232           Root MSE      =  .05643

------------------------------------------------------------------------------
       l_inv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       l_sav |   .8288074    .069851    11.87   0.000     .6826077    .9750072
       _cons |  -.2159075   .0985839    -2.19   0.041    -.4222461   -.009569
------------------------------------------------------------------------------
```

e) There will be a difference in the interpretation. For the first model, one percentage point increases in saving rate will increase investment rate by β_2 percentage point. For the second model, 1% increases in saving rate will increase investment rate by α_2 .

f) No. R^2 in the first model measures the variation in Invrate, but R^2 in the second model measures the variation in ln(Invrate).

g) Linear model: $Invrate = \beta_1 + \beta_2 Savrate + u_t$

$$\frac{dInvrate}{dSavrate} = \beta_2$$

$$\frac{dInvrate / Invrate}{dSavrate / Savrate} = \beta_2 \frac{Savrate}{Invrate}$$

Log-linear model: $\ln Invrate = \beta_1 + \beta_2 Savrate + u_t$

$$\frac{d \ln Invrate}{dSavrate} = \beta_2$$

$$\frac{dInvrate / Invrate}{dSavrate / Savrate} = \beta_2 Savrate$$

h) I would choose the first model because the test-statistics are more significant. In addition, the interpretation is more appealing.