(80 marks)

4. You are conducting an econometric investigation into the hourly wage rates of unionized and non-unionized female employees. The sample data consist of observations for 3286 female employees on the following variables:

W_i = the hourly wage rate of the i-th employee, in dollars per hour;

ED_i = years of formal education completed by the i-th employee, in years;

EXP_i = years of work experience accumulated by the i-th employee, in years;

UN_i = an indicator variable defined such that $UN_i = 1$ if the i-th employee is unionized, and $UN_i = 0$ if the i-th employee is non-unionized.

The regression model you propose to use is the log-lin (semi-log) regression equation

$$\ln W_i = \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 ED_i^2 + \beta_5 EXP_i^2 + \beta_6 ED_i EXP_i + \beta_7 UN_i + \beta_8 UN_i ED_i + \beta_9 UN_i EXP_i + \beta_{10} UN_i ED_i^2 + \beta_{11} UN_i EXP_i^2 + \beta_{12} UN_i ED_i EXP_i + u_i \quad \textbf{... (1)}$$

where the β_j (j = 1, 2, …, 12) are regression coefficients, $\ln W_i$ denotes the natural logarithm of the variable W_i, and u_i is a random error term.

$$\ln W_i = \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 ED_i^2 + \beta_5 EXP_i^2 + \beta_6 ED_i EXP_i + \beta_7 UN_i + \beta_8 UN_i ED_i + \beta_9 UN_i EXP_i + \beta_{10} UN_i ED_i^2 + \beta_{11} UN_i EXP_i^2 + \beta_{12} UN_i ED_i EXP_i + u_i \quad \textbf{... (1)}$$

- Regression function for ***non-unionized*** **workers**: set $UN_i = 0$

$$E(\ln W_i \mid ED_i, EXP_i, UN_i = 0) = \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 ED_i^2 + \beta_5 EXP_i^2 + \beta_6 ED_i EXP_i$$

- Regression function for ***unionized*** **workers**: set $UN_i = 1$

$$E(\ln W_i \mid ED_i, EXP_i, UN_i = 1) = \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 ED_i^2 + \beta_5 EXP_i^2 + \beta_6 ED_i EXP_i + \beta_7 + \beta_8 ED_i + \beta_9 EXP_i + \beta_{10} ED_i^2 + \beta_{11} EXP_i^2 + \beta_{12} ED_i EXP_i$$

$$= (\beta_1 + \beta_7) + (\beta_2 + \beta_8) ED_i + (\beta_3 + \beta_9) EXP_i + (\beta_4 + \beta_{10}) ED_i^2 + (\beta_5 + \beta_{11}) EXP_i^2 + (\beta_6 + \beta_{12}) ED_i EXP_i$$

- **Mean log-wage *difference*** between ***unionized*** **and *non-unionized* workers**:

$$E(\ln W_i \mid ED_i, EXP_i, UN_i = 1) - E(\ln W_i \mid ED_i, EXP_i, UN_i = 0) = \beta_7 + \beta_8 ED_i + \beta_9 EXP_i + \beta_{10} ED_i^2 + \beta_{11} EXP_i^2 + \beta_{12} ED_i EXP_i$$

Marginal log-wage effects of ED_i:

- For ***non-unionized*** **workers**:

$$E(\ln W_i | ED_i, EXP_i, UN_i = 0) = \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 ED_i^2 + \beta_5 EXP_i^2 + \beta_6 ED_i EXP_i$$

$$\frac{\partial E(\ln W_i | ED_i, EXP_i, UN_i = 0)}{\partial ED_i} = \beta_2 + 2\beta_4 ED_i + \beta_6 EXP_i$$

- For ***unionized*** **workers**:

$$E(\ln W_i | ED_i, EXP_i, UN_i = 1) = (\beta_1 + \beta_7) + (\beta_2 + \beta_8) ED_i + (\beta_3 + \beta_9) EXP_i + (\beta_4 + \beta_{10}) ED_i^2 + (\beta_5 + \beta_{11}) EXP_i^2 + (\beta_6 + \beta_{12}) ED_i EXP_i$$

$$\frac{\partial E(\ln W_i | ED_i, EXP_i, UN_i = 1)}{\partial ED_i} = (\beta_2 + \beta_8) + 2(\beta_4 + \beta_{10}) ED_i + (\beta_6 + \beta_{12}) EXP_i$$

- For ***unionized-non-unionized*** **difference in marginal effect of ED**:

$$\frac{\partial E(\ln W_i | ED_i, EXP_i, UN_i = 1)}{\partial ED_i} - \frac{\partial E(\ln W_i | ED_i, EXP_i, UN_i = 0)}{\partial ED_i} = \beta_8 + 2\beta_{10} ED_i + \beta_{12} EXP_i$$

Marginal log-wage effects of EXP_i:

- For ***non-unionized*** **workers**:

$$E\left(\ln W_i \mid ED_i, EXP_i, UN_i = 0\right) = \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 ED_i^2 + \beta_5 EXP_i^2 + \beta_6 ED_i EXP_i$$

$$\frac{\partial E\left(\ln W_i \mid ED_i, EXP_i, UN_i = 0\right)}{\partial EXP_i} = \beta_3 + 2\beta_5 EXP_i + \beta_6 ED_i$$

- For ***unionized*** **workers**:

$$E\left(\ln W_i \mid ED_i, EXP_i, UN_i = 1\right) = (\beta_1 + \beta_7) + (\beta_2 + \beta_8) ED_i + (\beta_3 + \beta_9) EXP_i + (\beta_4 + \beta_{10}) ED_i^2 + (\beta_5 + \beta_{11}) EXP_i^2 + (\beta_6 + \beta_{12}) ED_i EXP_i$$

$$\frac{\partial E\left(\ln W_i \mid ED_i, EXP_i, UN_i = 1\right)}{\partial EXP_i} = (\beta_3 + \beta_9) + 2(\beta_5 + \beta_{11}) EXP_i + (\beta_6 + \beta_{12}) ED_i$$

- For ***unionized-non-unionized*** **difference in marginal effect of EXP**:

$$\frac{\partial E\left(\ln W_i \mid ED_i, EXP_i, UN_i = 1\right)}{\partial ED_i} - \frac{\partial E\left(\ln W_i \mid ED_i, EXP_i, UN_i = 0\right)}{\partial ED_i} = \beta_9 + 2\beta_{11} EXP_i + \beta_{12} ED_i$$

<u>Question 4</u>: OLS Sample Regression Equations for $\ln W_i$ (*standard errors* in parentheses)

Regressors		(1)	(2)	(3)	(4)
Intercept	$\hat{\beta}_1$	0.5175 (0.2475)	0.6914 (0.2299)	0.7055 (0.2295)	0.7412 (0.2313)
ED_i	$\hat{\beta}_2$	0.08449 (0.02894)	0.06910 (0.02666)	0.06516 (0.02655)	0.05574 (0.02674)
EXP_i	$\hat{\beta}_3$	0.040026 (0.008062)	0.03417 (0.007574)	0.03566 (0.007524)	0.03771 (0.007580)
ED_i^2	$\hat{\beta}_4$	0.001601 (0.0008865)	0.001837 (0.0008112)	0.001957 (0.0008091)	0.002445 (0.0008130)
EXP_i^2	$\hat{\beta}_5$	–0.0003937 (0.0000890)	–0.0003664 (0.0000879)	–0.0004177 (0.0000841)	–0.0004624 (0.0000846)
ED_iEXP_i	$\hat{\beta}_6$	–0.001447 (0.0004434)	–0.001096 (0.0004094)	–0.001049 (0.0004063)	–0.001020 (0.0004095)
UN_i	$\hat{\beta}_7$	1.762 (0.7416)	–0.007131 (0.1059)	0.1709 (0.02311)	----
UN_iED_i	$\hat{\beta}_8$	–0.1526 (0.08174)	----	----	----
UN_iEXP_i	$\hat{\beta}_9$	–0.03941 (0.02491)	0.02097 (0.01065)	----	----
$UN_iED_i^2$	$\hat{\beta}_{10}$	0.002564 (0.002455)	----	----	----
$UN_iEXP_i^2$	$\hat{\beta}_{11}$	–0.0001760 (0.0002791)	–0.0005027 (0.0002447)	----	----
$UN_iED_iEXP_i$	$\hat{\beta}_{12}$	0.0032978 (0.001229)	----	----	----

RSS =	698.894	700.839	701.744	713.442
TSS =	905.589	905.589	905.589	905.589
N =	3286	3286	3286	3286

Note: The symbol "----" means that the corresponding regressor was omitted from the estimated sample regression equation. The figures in parentheses below the coefficient estimates are the estimated *standard errors*. RSS is the Residual Sum-of-Squares, TSS is the Total Sum-of-Squares, and N is the number of sample observations.

4. (continued)

(10 marks)

(a) Compare the goodness-of-fit to the sample data of the four sample regression equations (1), (2), (3) and (4) in the table. Calculate the value of an appropriate goodness-of-fit measure for each of the sample regression equations (1), (2), (3) and (4) in the table. Which of the four sample regression equations provides the best fit to the sample data? Which of the four sample regression equations provides the worst fit to the sample data?

Must use ***adjusted*** **R-squared**, because the regression equations (1) to (4) have **different numbers of regressors** – different values of K – but the ***same*** **regressand**.

$$\bar{R}^2 = 1 - \frac{RSS/(N-K)}{TSS/(N-1)} \qquad TSS/(N-1) = 905.589/3285 = \mathbf{0.275674}$$

Eq (1): $\bar{R}^2 = 1 - \frac{RSS/(N-K)}{TSS/(N-1)} = 1 - \frac{698.894/3274}{0.275674} = 1 - \frac{0.213468}{0.275674} = 0.22565 = \mathbf{0.2257}$ **best**

Eq (2): $\bar{R}^2 = 1 - \frac{RSS/(N-K)}{TSS/(N-1)} = 1 - \frac{700.839/3277}{0.275674} = 1 - \frac{0.213866}{0.275674} = 0.22421 = \mathbf{0.2242}$

Eq (3): $\bar{R}^2 = 1 - \frac{RSS/(N-K)}{TSS/(N-1)} = 1 - \frac{701.744/3279}{0.275674} = 1 - \frac{0.214012}{0.275674} = 0.22368 = \mathbf{0.2237}$

Eq (4): $\bar{R}^2 = 1 - \frac{RSS/(N-K)}{TSS/(N-1)} = 1 - \frac{713.442/3280}{0.275674} = 1 - \frac{0.217513}{0.275674} = 0.21098 = \mathbf{0.2110}$ **worst**

(10 marks)

(b) Use the estimation results for regression equation (3) in the above table to perform a test of the proposition that unionized employees of any given education and experience have higher average log-wages than non-unionized employees of the same education and experience. Perform the test at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). State the null hypothesis H_0 and the alternative hypothesis H_1. Show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

Equation (3) is: $\ln W_i = \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 ED_i^2 + \beta_5 EXP_i^2 + \beta_6 ED_i EXP_i + \beta_7 UN_i + u_i$

Mean log-wage differential between ***unionized*** **and** ***non-unionized*** **workers** in equation (3) is:

$$E(\ln W_i | ED_i, EXP_i, UN_i = 1) - E(\ln W_i | ED_i, EXP_i, UN_i = 0) = \beta_7$$

Test: H_0: $\beta_7 = 0$ versus H_1: $\beta_7 > 0$ a ***right-tail*** **t-test**

Compute ***sample value*** **of t-statistic**:

$$t_0(\hat{\beta}_7) = \frac{\hat{\beta}_7 - 0}{\hat{se}(\hat{\beta}_7)} = \frac{\hat{\beta}_7}{\hat{se}(\hat{\beta}_7)} = \frac{0.1709}{0.02311} = \mathbf{7.395}$$

Null distribution of t-statistic $t(\hat{\beta}_7)$ is $t[N - K] = t[3286 - 7] = t[3279] = t[\infty]$

4 (b):

Decision rule for a *right-tail* t-test:

1. If $t_0 \le t_\alpha[N-K] = t_\alpha[\infty]$, ***retain* (do not reject) H_0** at the 100α % significance level.
2. If $t_0 > t_\alpha[N-K] = t_\alpha[\infty]$, ***reject* H_0** at the 100α % significance level.

At significance level **α = 0.05**, **critical value** is $t_{0.05}[\infty] = \mathbf{1.645}$

<u>**Inference**</u>**:** Since $\mathbf{t_0 = 7.395 > 1.645 = t_{0.05}[\infty]}$, ***reject*** H_0 at the **5 % significance level**.

(10 marks)

(c) State the coefficient restrictions that regression equation (2) in the table imposes on regression equation (1). Explain in words what the restrictions mean. Use the estimation results given in the table to perform a test of these coefficient restrictions at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). State the null and alternative hypotheses, show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test. Would your inference be the same at the 1 percent significance level (i.e., for significance level $\alpha = 0.01$)? Based on the outcome of the test, which of the two regression equations would you choose, equation (1) or equation (2)?

Equation (2) imposes on equation (1) the exclusion restrictions $\boldsymbol{\beta_8 = 0}$ **and** $\boldsymbol{\beta_{10} = 0}$ **and** $\boldsymbol{\beta_{12} = 0}$

Test: H_0: $\beta_8 = 0$ and $\beta_{10} = 0$ and $\beta_{12} = 0$ versus H_1: $\beta_8 \neq 0$ and/or $\beta_{10} \neq 0$ and/or $\beta_{12} \neq 0$

Interpretation of H_0: The marginal log-wage effect of ED_i is equal (identical) for unionized and non-unionized workers.

Compute *sample value* of general F-statistic:

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1}$$

where:

$RSS_0 = 700.839$ with $df_0 = N - K_0 = 3286 - 9 = 3277$
$RSS_1 = 698.894$ with $df_1 = N - K = 3286 - 12 = 3274$

$RSS_1/df_1 = 698.894/3274$ **= 0.213468**

4 (c):

$RSS_0 - RSS_1 = 700.839 - 698.894 = 1.945$ and $df_0 - df_1 = 3$

$$F_0 = \frac{1.945/3}{698.894/3274} = \frac{0.648333}{0.213468} = 3.03715 = \mathbf{3.037}$$

Null distribution of F_0: is $F[df_0 - df_1, df_1] = F[K - K_0, N - K] =$ **F[3, 3274]**

Decision Rule: at significance level α

1. If $F_0 > F_\alpha[3, 3274]$, ***reject*** $\mathbf{H_0}$ at the 100α percent significance level.
2. If $F_0 \le F_\alpha[3, 3274]$, ***retain*** **(do not reject)** $\mathbf{H_0}$ at the 100α percent significance level.

Critical value of F[3, 3274] at 5% significance level ($\alpha = 0.05$) is $\mathbf{F_{0.05}[3, \infty] = 2.60}$
Critical value of F[3, 3274] at 1% significance level ($\alpha = 0.01$) is $\mathbf{F_{0.01}[3, \infty] = 3.78}$

Inference: is different at the 5% and 1% significance levels.

- Since $\mathbf{F_0 = 3.037 > 2.60 = F_{0.05}[5, \infty]}$, ***reject*** H_0 **at 5% significance level.**
- Since $\mathbf{F_0 = 3.037 < 3.78 = F_{0.01}[5, \infty]}$, ***retain (do not reject)*** H_0 **at 1% significance level.**

Choose equation (1): restrictions incorporated in equation (2) are rejected at a sufficiently low significance level.

(10 marks)

(d) State the coefficient restrictions that regression equation (3) in the table imposes on regression equation (1). Explain in words what the restrictions mean. Use the estimation results given in the table to perform a test of these coefficient restrictions at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). State the null and alternative hypotheses, show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test. Would your inference be the same at the 1 percent significance level (i.e., for significance level $\alpha = 0.01$)? Based on the outcome of the test, which of the two regression equations would you choose, equation (1) or equation (3)?

Equation (3) imposes on equation (1) the restrictions $\boldsymbol{\beta_8 = 0}$ **and** $\boldsymbol{\beta_9 = 0}$ **and** $\boldsymbol{\beta_{10} = 0}$ **and** $\boldsymbol{\beta_{11} = 0}$ **and** $\boldsymbol{\beta_{12} = 0}$

Test: H_0: $\beta_8 = 0$ and $\beta_9 = 0$ and $\beta_{10} = 0$ and $\beta_{11} = 0$ and $\beta_{12} = 0$ or $\beta_j = 0$ for all j = 8, 9, …, 12

versus

H_1: $\beta_8 \neq 0$ and/or $\beta_9 \neq 0$ and/or $\beta_{10} \neq 0$ and/or $\beta_{11} \neq 0$ and/or $\beta_{12} \neq 0$

Interpretation of H_0: The marginal log-wage effects of both ED_i and EXP_i are equal (identical) for unionized and non-unionized workers; the union-nonunion mean log-wage difference is a constant that does not vary with ED_i or EXP_i.

4 (d):

Compute *sample value* of general F-statistic:

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1}$$

where:

$RSS_0 = 701.744$ with $df_0 = N - K_0 = 3286 - 7 = 3279$
$RSS_1 = 698.894$ with $df_1 = N - K = 3286 - 12 = 3274$

$RSS_1/df_1 = 698.894/3274$ **= 0.213468**

$RSS_0 - RSS_1 = 701.744 - 698.894 = 2.850$ and $df_0 - df_1 = 5$

$$F_0 = \frac{2.850/5}{698.894/3274} = \frac{0.570000}{0.213468} = 2.67019 = \mathbf{2.670}$$

Null distribution of F_0: is $F[df_0 - df_1, df_1] = F[K - K_0, N - K] =$ **F[5, 3274]**

Decision Rule: at significance level α

1. If $F_0 > F_\alpha[5, 3274]$, ***reject*** $\mathbf{H_0}$ at the 100α percent significance level.
2. If $F_0 \le F_\alpha[5, 3274]$, ***retain*** **(do not reject)** $\mathbf{H_0}$ at the 100α percent significance level.

4 (d):

Critical value of F[5, 3274] at 5% significance level ($\alpha = 0.05$) is $\mathbf{F_{0.05}[5, \infty] = 2.21}$
Critical value of F[5, 3274] at 1% significance level ($\alpha = 0.01$) is $\mathbf{F_{0.01}[5, \infty] = 3.02}$

Inference: is different at the 5% and 1% significance levels.

- Since $\mathbf{F_0 = 2.670 > 2.21 = F_{0.05}[5, \infty]}$, ***reject*** $\mathbf{H_0}$ **at 5% significance level.**
- Since $\mathbf{F_0 = 2.670 < 3.02 = F_{0.01}[5, \infty]}$, ***retain (do not reject)*** $\mathbf{H_0}$ **at 1% significance level.**

Choose equation (1): restrictions incorporated in equation (3) are rejected at a sufficiently low significance level.

4. (continued)

(10 marks)

(e) State the coefficient restrictions that regression equation (4) in the table imposes on regression equation (1). Explain in words what the restrictions mean. Use the estimation results given in the table to perform a test of these coefficient restrictions at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). State the null and alternative hypotheses, show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test. Would your inference be the same at the 1 percent significance level (i.e., for significance level $\alpha = 0.01$)? Based on the outcome of the test, which of the two regression equations would you choose, equation (1) or equation (4)?

Equation (4) imposes on equation (1) the restrictions $\boldsymbol{\beta_j = 0}$ **for all** $\boldsymbol{j = 7, 8, \ldots, 12}$

Test: H_0: $\beta_7 = 0$ and $\beta_8 = 0$ and $\beta_9 = 0$ and $\beta_{10} = 0$ and $\beta_{11} = 0$ and $\beta_{12} = 0$
or $\beta_j = 0$ for all $j = 7, 8, \ldots, 12$

versus

H_1: $\beta_7 \neq 0$ and/or $\beta_8 \neq 0$ and/or $\beta_9 \neq 0$ and/or $\beta_{10} \neq 0$ and/or $\beta_{11} \neq 0$ and/or $\beta_{12} \neq 0$
or $\beta_j \neq 0$, $j = 7, 8, \ldots, 12$

Interpretation of H_0: The union-nonunion mean log-wage difference is zero for all values of ED_i and EXP_i.

4 (e):

Compute *sample value* of general F-statistic:

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1}$$

where:

$RSS_0 = 713.442$ with $df_0 = N - K_0 = 3286 - 6 = 3280$
$RSS_1 = 698.894$ with $df_1 = N - K = 3286 - 12 = 3274$

$RSS_1/df_1 = 698.894/3274$ **= 0.213468**

$RSS_0 - RSS_1 = 713.442 - 698.894 = 14.548$ and $df_0 - df_1 = 6$

$$F_0 = \frac{14.548/6}{698.894/3274} = \frac{2.424667}{0.213468} = 11.3585 = \mathbf{11.36}$$

Null distribution of F_0: is $F[df_0 - df_1, df_1] = F[K - K_0, N - K] =$ **F[6, 3274]**

Decision Rule: at significance level α

1. If $F_0 > F_\alpha[6, 3274]$, ***reject*** $\mathbf{H_0}$ at the 100α percent significance level.
2. If $F_0 \le F_\alpha[6, 3274]$, ***retain*** **(do not reject)** $\mathbf{H_0}$ at the 100α percent significance level.

4 (e):

Critical value of F[6, 3274] at 5% significance level ($\alpha = 0.05$) is $\mathbf{F_{0.05}[6, \infty] = 2.10}$
Critical value of F[6, 3274] at 1% significance level ($\alpha = 0.01$) is $\mathbf{F_{0.01}[6, \infty] = 2.80}$

Inference: is the same at both 5% and 1% significance levels.

- Since $\mathbf{F_0 = 11.36 > 2.10 = F_{0.05}[6, \infty]}$, ***reject*** H_0 **at 5% significance level.**
- Since $\mathbf{F_0 = 11.36 > 2.80 = F_{0.01}[6, \infty]}$, ***reject*** H_0 **at 1% significance level.**

Choose equation (1): restrictions incorporated in equation (4) are rejected.

(15 marks)

(f) Write the expression (or formula) for the marginal effect of ED_i on $\ln W_i$ for *non-unionized* employees implied by regression equation (1). Use regression equation (1) to compute a test of the proposition that the **marginal effect of ED_i on $\ln W_i$ for *non-unionized* employees is equal to zero for non-unionized employees** with any given values of ED_i and EXP_i. State the coefficient restrictions on regression equation (1) implied by this proposition; that is, state the null hypothesis H_0 and the alternative hypothesis H_1. Write the *restricted* regression equation implied by the null hypothesis H_0. OLS estimation of this *restricted* regression equation yields a Residual Sum-of-Squares value of **RSS = 836.832**. Use this information, together with the results from OLS estimation of equation (1), to calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test. **Choose an appropriate significance level for the test.**

For ***non-unionized* workers**, the marginal effect of ED is:

$$\frac{\partial E(\ln W_i \mid ED_i, EXP_i, UN_i = 0)}{\partial ED_i} = \beta_2 + 2\beta_4 ED_i + \beta_6 EXP_i$$

Test: H_0: $\beta_2 = 0$ and $\beta_4 = 0$ and $\beta_6 = 0$ or $\beta_j = 0$ for all j = 2, 4, 6

versus

H_1: $\beta_2 \neq 0$ and/or $\beta_4 \neq 0$ and/or $\beta_6 \neq 0$ or $\beta_j \neq 0$, j = 2, 4, 6

4 (f):

To get restricted regression model implied by H_0, set $\beta_2 = 0$ and $\beta_4 = 0$ and $\beta_6 = 0$ in equation (1):

$$\begin{aligned}\ln W_i = \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 ED_i^2 + \beta_5 EXP_i^2 + \beta_6 ED_i EXP_i + \beta_7 UN_i + \beta_8 UN_i ED_i \\ + \beta_9 UN_i EXP_i + \beta_{10} UN_i ED_i^2 + \beta_{11} UN_i EXP_i^2 + \beta_{12} UN_i ED_i EXP_i + u_i \qquad \textbf{... (1)}\end{aligned}$$

Restricted model is therefore:

$$\begin{aligned}\ln W_i = \beta_1 + \beta_3 EXP_i + \beta_5 EXP_i^2 + \beta_7 UN_i + \beta_8 UN_i ED_i \\ + \beta_9 UN_i EXP_i + \beta_{10} UN_i ED_i^2 + \beta_{11} UN_i EXP_i^2 + \beta_{12} UN_i ED_i EXP_i + u_i \qquad \textbf{... (2)}\end{aligned}$$

Compute *sample value* of general F-statistic:

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1}$$

where:

$RSS_0 = 836.832$ with $df_0 = N - K_0 = 3286 - 9 = 3277$
$RSS_1 = 698.894$ with $df_1 = N - K = 3286 - 12 = 3274$

$RSS_1/df_1 = 698.894/3274$ **= 0.213468**

$RSS_0 - RSS_1 = 836.832 - 698.894 = 137.938$ and $df_0 - df_1 = 3$

4 (f):

$$F_0 = \frac{137.938/3}{698.894/3274} = \frac{45.9793}{0.213468} = 215.392 = \mathbf{215.39}$$

Null distribution of F_0: is $F[df_0 - df_1, df_1] = F[K - K_0, N - K] = \mathbf{F[3, 3274]}$

Decision Rule: at significance level α

1. If $F_0 > F_\alpha[3, 3274]$, ***reject* H_0** at the 100α percent significance level.
2. If $F_0 \le F_\alpha[3, 3274]$, ***retain* (do not reject) H_0** at the 100α percent significance level.

Critical value of F[3, 3274] at 5% significance level ($\alpha = 0.05$) is $\mathbf{F_{0.05}[3, \infty] = 2.60}$
Critical value of F[3, 3274] at 1% significance level ($\alpha = 0.01$) is $\mathbf{F_{0.01}[3, \infty] = 3.78}$

Inference: is the same at both 5% and 1% significance levels.

- Since $\mathbf{F_0 = 215.39 > 2.60 = F_{0.05}[3, \infty]}$, ***reject* H_0 at 5% significance level.**
- Since $\mathbf{F_0 = 215.39 > 3.78 = F_{0.01}[3, \infty]}$, ***reject* H_0 at 1% significance level.**

(15 marks)

(g) Write the expression (or formula) for the marginal effect of ED_i on $\ln W_i$ for *unionized* employees implied by regression equation (1). Use regression equation (1) to compute a test of the null hypothesis that **the marginal effect of ED_i on $\ln W_i$ for *unionized* employees is equal to zero** for unionized employees with any given values of ED_i and EXP_i. State the coefficient restrictions on regression equation (1) implied by this proposition; that is, state the null hypothesis H_0 and the alternative hypothesis H_1. Write the *restricted* regression equation implied by the null hypothesis H_0. OLS estimation of this *restricted* regression equation yields a Residual Sum-of-Squares value of **RSS = 722.988**. Use this information, together with the results from OLS estimation of equation (1), to calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test. **Choose an appropriate significance level for the test.**

For ***unionized* workers**, the marginal effect of ED is:

$$\frac{\partial E(\ln W_i \mid ED_i, EXP_i, UN_i = 1)}{\partial ED_i} = (\beta_2 + \beta_8) + 2(\beta_4 + \beta_{10})ED_i + (\beta_6 + \beta_{12})EXP_i$$

Test: H_0: $\beta_2 + \beta_8 = 0$ and $\beta_4 + \beta_{10} = 0$ and $\beta_6 + \beta_{12} = 0$

versus

H_1: $\beta_2 + \beta_8 \neq 0$ and/or $\beta_4 + \beta_{10} \neq 0$ and/or $\beta_6 + \beta_{12} \neq 0$

4 (g):

To get restricted regression model implied by H_0, set $\beta_2 + \beta_8 = 0$ and $\beta_4 + \beta_{10} = 0$ and $\beta_6 + \beta_{12} = 0$ in equation (1):

$$\begin{aligned} \ln W_i = {} & \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 ED_i^2 + \beta_5 EXP_i^2 + \beta_6 ED_i EXP_i + \beta_7 UN_i + \beta_8 UN_i ED_i \\ & + \beta_9 UN_i EXP_i + \beta_{10} UN_i ED_i^2 + \beta_{11} UN_i EXP_i^2 + \beta_{12} UN_i ED_i EXP_i + u_i \end{aligned} \quad \textbf{... (1)}$$

Set $\beta_8 = -\beta_2$ and $\beta_{10} = -\beta_4$ and $\beta_{12} = -\beta_6$ in equation (1); **restricted model** is therefore equation (3):

$$\begin{aligned} \ln W_i = {} & \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 ED_i^2 + \beta_5 EXP_i^2 + \beta_6 ED_i EXP_i + \beta_7 UN_i - \beta_2 UN_i ED_i \\ & + \beta_9 UN_i EXP_i - \beta_4 UN_i ED_i^2 + \beta_{11} UN_i EXP_i^2 - \beta_6 UN_i ED_i EXP_i + u_i \end{aligned}$$

$$\begin{aligned} \ln W_i = {} & \beta_1 + \beta_2 (1 - UN_i) ED_i + \beta_3 EXP_i + \beta_4 (1 - UN_i) ED_i^2 + \beta_5 EXP_i^2 + \beta_6 (1 - UN_i) ED_i EXP_i + \beta_7 UN_i \\ & + \beta_9 UN_i EXP_i + \beta_{11} UN_i EXP_i^2 + u_i \end{aligned} \quad \textbf{... (3)}$$

Compute *sample value* of general F-statistic:

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1}$$

where:

$RSS_0 = 722.988$ with $df_0 = N - K_0 = 3286 - 9 = 3277$
$RSS_1 = 698.894$ with $df_1 = N - K = 3286 - 12 = 3274$

$RSS_1/df_1 = 698.894/3274$ **= 0.213468**

4 (g):

$RSS_0 - RSS_1 = 722.988 - 698.894 = 24.094$ and $df_0 - df_1 = 3$

$$F_0 = \frac{24.094/3}{698.894/3274} = \frac{8.03133}{0.213468} = 37.6231 = \mathbf{37.62}$$

Null distribution of F_0: is $F[df_0 - df_1, df_1] = F[K - K_0, N - K] =$ **F[3, 3274]**

Decision Rule: at significance level α

1. If $F_0 > F_\alpha[3, 3274]$, ***reject*** **H_0** at the 100α percent significance level.
2. If $F_0 \le F_\alpha[3, 3274]$, ***retain*** **(do not reject) H_0** at the 100α percent significance level.

Critical value of F[3, 3274] at 5% significance level ($\alpha = 0.05$) is **$F_{0.05}[3, \infty] = 2.60$**
Critical value of F[3, 3274] at 1% significance level ($\alpha = 0.01$) is **$F_{0.01}[3, \infty] = 3.78$**

Inference: is the same at both 5% and 1% significance levels.

- Since **$F_0 = 37.62 > 2.60 = F_{0.05}[3, \infty]$,** ***reject*** H_0 **at 5% significance level.**
- Since **$F_0 = 37.62 > 3.78 = F_{0.01}[3, \infty]$,** ***reject*** H_0 **at 1% significance level.**