Homework 2 (CAPM)

(a) Prove that $E(\boldsymbol{b_R}) = \boldsymbol{\beta_1} + \boldsymbol{P\beta_2}$

By definition:

$$b_R = (X_1'X_1)^{-1}X_1'y$$

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

Substituting y into b_R:

$$b_R = (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon)$$

By expanding this expression:

$$b_R = (X_1'X_1)^{-1}X_1'X_1\beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon$$

where:

$$(X_1'X_1)^{-1}X_1'X_1 = I$$

$$P = (X_1'X_1)^{-1}X_1'X_2$$

Simplifying:

$$b_R = I\beta_1 + P\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon$$

$$b_R = \beta_1 + P\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon$$

Computing expected value:

$$E(b_R) = E[\beta_1 + P\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon]$$

By expanding this expression and using properties of expected value:

$$E(b_R) = E(\beta_1) + E(P\beta_2) + E[(X_1'X_1)^{-1}X_1'\varepsilon]$$

$$E(b_R) = \beta_1 + P\beta_2 + (X_1'X_1)^{-1}X_1'E[\varepsilon]$$

One assumption of the linear regression model indicates that the expected value of the error term is zero,

$$E[\varepsilon] = 0$$

Therefore,

$$E(b_R) = \beta_1 + P\beta_2$$

(b) Prove that $\boldsymbol{V(b_R) = \sigma^2(X_1'X_1)^{-1}}$

From part (a), I know that:

$$b_R = \beta_1 + P\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon$$

Equivalently,

$$b_R - \beta_1 - P\beta_2 = (X_1'X_1)^{-1}X_1'\varepsilon$$

The definition of variance is:

$$V(b_R) = E\left[\big(b_R - E(b_R)\big)\big(b_R - E(b_R)\big)'\right]$$

where:

$$E(b_R) = \beta_1 + P\beta_2$$

Substituting $E(b_R)$ into $V(b_R)$:

$$E\left[\big(b_R - E(b_R)\big)\big(b_R - E(b_R)\big)'\right] = E\{[(X_1'X_1)^{-1}X_1'\varepsilon][(X_1'X_1)^{-1}X_1'\varepsilon]'\}$$

where (by rules of transpose matrix):

$$[(X_1'X_1)^{-1}X_1'\varepsilon]' = \varepsilon'X_1(X_1'X_1)^{-1}$$

Computing expected value:

$$E[(b_R - \beta_1)(b_R - \beta_1)'] = E\{[(X_1'X_1)^{-1}X_1'\varepsilon][\varepsilon'X_1(X_1'X_1)^{-1}]\}$$

$$E[(b_R - \beta_1)(b_R - \beta_1)'] = (X_1'X_1)^{-1}X_1'E(\varepsilon\varepsilon')X_1(X_1'X_1)^{-1}$$

One assumption of the linear regression model indicates that the variance of the error term is constant and equal to σ^2. In matrix terms:

$$E(\varepsilon\varepsilon') = \sigma^2 I$$

$$E[(b_R - \beta_1)(b_R - \beta_1)'] = (X_1'X_1)^{-1}X_1'\sigma^2 I X_1(X_1'X_1)^{-1}$$

Simplifying:

$$V(b_R) = E[(b_R - \beta_1)(b_R - \beta_1)'] = \sigma^2(X_1'X_1)^{-1}$$

(c) Prove that $\boldsymbol{b_R = b_1 + Pb_2}$

By definition:

$$b_R = (X_1'X_1)^{-1}X_1'y$$

$$y = X_1b_1 + X_2b_2 + e$$

Substituting, expanding and simplifying:

$$b_R = (X_1'X_1)^{-1}X_1'(X_1b_1 + X_2b_2 + e)$$

$$b_R = (X_1'X_1)^{-1}X_1'X_1b_1 + (X_1'X_1)^{-1}X_1'X_2b_2 + (X_1'X_1)^{-1}X_1'e$$

$$b_R = Ib_1 + Pb_2 + (X_1'X_1)^{-1}X_1'e$$

Estimated errors are orthogonal to the data generating them.

$$X_1'e = 0$$

Thus,

$$b_R = Ib_1 + Pb_2 + (X_1'X_1)^{-1}0$$

$$b_R = b_1 + Pb_2$$

(d) Argue that the columns of the 2x3 columns matrix P are obtained by regressing each of the variables 'Age', 'Educ', 'Parttime' on a constant term and a variable 'Female'.

In general, OLS estimation indicates that:

$$y = X\beta + \varepsilon \Rightarrow b = \left(X^{'}X\right)^{-1}X^{'}y$$

Now, consider that the dependent variables can be any of the following three possibilities: 'Age' (x_3), 'Educ' (x_4), 'Parttime' (x_5), where they are included in matrix X_2. Finally, variable 'Female' is the independent variable. Then,

$$X_2 = X_1P + \varepsilon \Rightarrow P = (X_1'X_1)^{-1}X_1'X_2$$

where P contains the estimated parameters of the regression between the variables 'Age', 'Educ', 'Parttime' (included in X_2) on a constant term and a variable 'Female' and:

$$X_2 = \begin{pmatrix} x_{31} & x_{41} & x_{51} \\ x_{32} & x_{42} & x_{52} \\ & \cdots & \\ x_{3n} & x_{4n} & x_{5n} \end{pmatrix}$$

(e) Determine the values of P from the results in Lecture 2.1

By definition:

$$P = (X_1'X_1)^{-1}X_1'X_2$$

where n=500 and the sums are:

$$(X_1'X_1)^{-1} = \begin{pmatrix} n & \Sigma x_2 \\ \Sigma x_2 & \Sigma x_2^2 \end{pmatrix}^{-1} = \begin{pmatrix} 500 & 184 \\ 184 & 184 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{46}{14536} & -\frac{46}{14536} \\ -\frac{46}{14536} & \frac{125}{14536} \end{pmatrix}$$

$$X_1'X_2 = \begin{pmatrix} \Sigma x_3 & \Sigma x_4 & \Sigma x_5 \\ \Sigma x_2 x_3 & \Sigma x_2 x_4 & \Sigma x_2 x_5 \end{pmatrix} = \begin{pmatrix} 20005 & 1039 & 144 \\ 7349 & 325 & 82 \end{pmatrix}$$

Substituting:

$$P = \frac{1}{14536}\begin{pmatrix} 582176 & 32844 & 2852 \\ -1605 & -7169 & 3626 \end{pmatrix} = \begin{pmatrix} 40.050 & 2.259 & 0.196 \\ -0.110 & -0.493 & 0.249 \end{pmatrix}$$

Which are equal the coefficients obtained from Lecture 2.1:

$$Age = 40.05 - 0.11Female + e$$

$$Educ = 2.26 - 0.49Female + e$$

$$Parttime = 0.196 - 0.249Female + e$$

(f) Check the numerical validity of the result in part (c).

The OLS estimation of the unrestricted model is:

Dependent Variable: LOGWAGEINDEX
Method: Least Squares
Date: 11/15/15 Time: 10:46
Sample: 1 500
Included observations: 500

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.052685	0.055334	55.16839	0.0000
FEMALE	-0.041101	0.024711	-1.663296	0.0969
AGE	0.030606	0.001273	24.04101	0.0000
EDUC	0.233178	0.010660	21.87382	0.0000
PARTTIME	-0.365449	0.031571	-11.57563	0.0000

R-squared	0.703976	Mean dependent var	4.641408
Adjusted R-squared	0.701584	S.D. dependent var	0.448926
S.E. of regression	0.245237	Akaike info criterion	0.036765
Sum squared resid	29.76986	Schwarz criterion	0.078911
Log likelihood	-4.191347	Hannan-Quinn criter.	0.053303
F-statistic	294.2901	Durbin-Watson stat	1.873691
Prob(F-statistic)	0.000000		

where:

$$b_1 = \begin{pmatrix} 3.052 \\ -0.041 \end{pmatrix}$$

$$b_2 = \begin{pmatrix} 0.030 \\ 0.233 \\ -0.365 \end{pmatrix}$$

The OLS estimation of the restricted model is:

Dependent Variable: LOGWAGEINDEX
Method: Least Squares
Date: 11/15/15 Time: 10:51
Sample: 1 500
Included observations: 500

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.733644	0.024344	194.4490	0.0000
FEMALE	-0.250643	0.040130	-6.245809	0.0000

R-squared	0.072643	Mean dependent var	4.641408
Adjusted R-squared	0.070781	S.D. dependent var	0.448926
S.E. of regression	0.432746	Akaike info criterion	1.166662
Sum squared resid	93.26019	Schwarz criterion	1.183521
Log likelihood	-289.6656	Hannan-Quinn criter.	1.173278
F-statistic	39.01014	Durbin-Watson stat	1.383433
Prob(F-statistic)	0.000000		

where:

$$b_R = \begin{pmatrix} 4.733 \\ -0.250 \end{pmatrix}$$

Checking the numerical validity:

$$b_R = b_1 + Pb_2$$

$$\begin{pmatrix} 4.733 \\ -0.250 \end{pmatrix} \approx \begin{pmatrix} 3.052 \\ -0.041 \end{pmatrix} + \begin{pmatrix} 40.050 & 2.259 & 0.196 \\ -0.110 & -0.493 & 0.249 \end{pmatrix} \begin{pmatrix} 0.030 \\ 0.233 \\ -0.365 \end{pmatrix} = \begin{pmatrix} 4.721 \\ -0.250 \end{pmatrix}$$