Test Exercise 2

Notes: • See website for how to submit your answers and how feedback is organized. • For parts (e) and (f), you need regression results discussed in Lectures 2.1 and 2.5. Goals and skills being used: • Use matrix methods in the econometric analysis of multiple regression. • Employ matrices and statical methods in multiple regression analysis. • Give numerical verification of mathematical results. Questions This test exercise is of a theoretical nature. In our discussion of the F-test, the total set of explanatory factors was split in two parts. The factors in X1 are always included in the model, whereas those in X2 are possibly removed. In questions (a), (b), and (c) you derive relations between the two OLS estimates of the effects of X1 on y, one in the large model and the other in the small model. In parts (d), (e), and (f), you check the relation of question (c) numerically for the wage data of our lectures. We use the notation of Lecture 2.4.2 and assume that the standard regression assumptions A1-A6 are satisfied for the unrestricted model. The restricted model is obtained by deleting the set of g explanatory factors collected in the last g columns X2 of X. We wrote the model with X = (X1 X2) and corresponding partitioning of the OLS estimator b in b1 and b2 as

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon = X_1 b_1 + X_2 b_2 + e$$

. We denote by bR the OLS estimator of β1 obtained by regressing y on X1, so that

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

Further, let

$$P = (X_1'X_1)^{-1}X_1X_2$$

(a) Prove that

$$E(b_R) = \beta 1 + P\beta 2$$

. (b) Prove that

$$var(b_R) = \sigma 2(X_1'X1)^{-1}$$

(c) Prove that

$$b_R = b_1 + Pb_2$$

Now consider the wage data of Lectures 2.1 and 2.5. Let y be log-wage (500×1 vector), and let X1 be the (500×2) matrix for the constant term and the variable ‘Female’. Further let X2 be the (500 × 3) matrix with observations of the variables ‘Age’, ‘Educ’ and ‘Parttime’. The values of bR were given in Lecture 2.1, and those of b in Lecture 2.5. (d) Argue that the columns of the (2 × 3) matrix P are obtained by regressing each of the variables ‘Age’, ‘Educ’, and ‘Parttime’ on a constant term and the variable ‘Female’. (e) Determine the values of P from the results in Lecture 2.1. (f) Check the numerical validity of the result in part (c). Note: This equation will not hold exactly because the coefficients have been rounded to two or three decimals; preciser results would have been obtained for higher precision coefficients.

(a) Prove that

$$E(b_R) = \beta 1 + P\beta 2$$

since

$$b_R = X_1' X_1^{-1} X_1' y$$
$$E(b_R) = E((X_1'X_1)^{-1} X_1' y)$$
$$= E((X_1'X_1)^{-1} X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon))$$
$$= E((X_1'X_1)^{-1} X_1' X_1 \beta_1 + (X_1'X_1)^{-1} X_1' X_2 \beta_2 + (X_1'X_1)^{-1} X_1' \varepsilon)) \ldots\ldots (1)$$
$$(X_1'X_1)^{-1} X_1' X_1 = I \ldots\ldots (2)$$

$$P = (X_1'X_1)^{-1} X_1 X_2 \ldots\ldots (3)$$
$$E(\varepsilon) = 0 \ldots\ldots (A3)$$

X are not random ....(A2)

since (2),(3),A(2), A（3）

$$(1) = E(\beta_1 + P\beta_2) + (X_1'X_1)^{-1} X_1' E(\varepsilon) = \beta 1 + P\beta 2$$

(b) Prove that

$$var(b_R) = \sigma 2(X_1' X1)^{-1}$$

first,

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

then

$$b_R - E(b_R) = (X_1'X_1)^{-1} X_1' \varepsilon$$
$$var(b_R) =$$
$$E((b_R - E(b_R))(b_R - E(b_R))') =$$
$$E((X_1'X_1)^{-1} X_1' \varepsilon)((X_1'X_1)^{-1} X_1' \varepsilon)') =$$
$$(X_1'X_1)^{-1} X_1' E(\varepsilon\varepsilon') X_1 (X_1'X_1)^{-1} \ldots\ldots (1)$$

since

$$E(\varepsilon\varepsilon') = \sigma^2 \ldots\ldots A(4)$$
$$X_1' X_1 (X_1'X_1)^{-1} = I \ldots\ldots (1)$$

since A4 and (1)

$$var(b_R) = \sigma 2(X_1' X1)^{-1}$$

(c) Prove that

$$b_R = b_1 + Pb_2$$
$$b_R = X_1'X_1^{-1}X_1'y$$
$$= X_1'X_1^{-1}X_1'(X_1b_1 + X_2b_2 + e) =$$
$$X_1'X_1^{-1}X_1'X_1\beta_1 + X_1'X_1^{-1}X_1'X_2\beta_2 + X_1'X_1^{-1}X_1'e =$$
$$b_1 + Pb_2 + X_1'X_1^{-1}X_1'e.\dots\dots\dots(1)$$

since least squares residuals are orthogonal to all regressors

$$X_1'e = 0$$
$$b_R = b_1 + Pb_2$$

Now consider the wage data of Lectures 2.1 and 2.5. Let y be log-wage (500×1 vector), and let X1 be the (500×2) matrix for the constant term and the variable 'Female'. Further let X2 be the (500 × 3) matrix with observations of the variables 'Age', 'Educ' and 'Parttime'. The values of bR were given in Lecture 2.1, and those of b in Lecture 2.5. (d) Argue that the columns of the (2 × 3) matrix P are obtained by regressing each of the variables 'Age', 'Educ',and 'Parttime' on a constant term and the variable 'Female'.

ans: Equation

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

is proved if we replace b_R to P and y into X2, the above equation will have the format as

$$P = (X_1'X_1)^{-1}X_1X_2$$

, which similar to regress each of the variables 'Age', 'Educ',and 'Parttime' on a constant term and the variable 'Female.

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

In [21]:

```
import sys
sys.path.append('/Users/CJ/Documents/bitbucket/xforex_v1/xforex_v3')
%matplotlib inline
from numpy import matrix
from numpy import linalg
import pandas as pd
import xforex.BackTesting.econometrics_tools
from xforex.BackTesting.econometrics_tools import Econometrics_Tool


dat = pd.read_csv('/Users/CJ/Documents/bitbucket/xforex_v1/xforex_v3/training/ec
onometrics/week2-multiple-linear-regression/Dataset2.txt',
                      sep='\t')

X = dat[['Age','Educ', 'Parttime','Female']]
y = dat['LogWage']
model = Econometrics_Tool().linear_fit(X, y)

X1 = matrix([dat['Female'],model.resid]).T

X2 =  matrix([dat['Age'], dat['Educ'], dat['Parttime']]).T
print 'P ='
P = (X1.T*X1).I*X1.T*X2
print P
```

```
                            OLS Regression Results
==================================================================
==========
Dep. Variable:                LogWage   R-squared:
     0.704
Model:                            OLS   Adj. R-squared:
     0.702
Method:                 Least Squares   F-statistic:
     294.3
Date:                Fri, 09 Sep 2016   Prob (F-statistic):
 2.51e-129
Time:                        15:06:26   Log-Likelihood:
   -4.1790
No. Observations:                 500   AIC:
     18.36
Df Residuals:                     495   BIC:
     39.43
Df Model:                           4

Covariance Type:            nonrobust

==================================================================
==========
                 coef    std err          t      P>|t|      [95.0% C
onf. Int.]
------------------------------------------------------------------
----------
const          3.0527      0.055     55.170      0.000       2.944
     3.161
Age            0.0306      0.001     24.041      0.000       0.028
     0.033
Educ           0.2332      0.011     21.873      0.000       0.212
     0.254
Parttime      -0.3654      0.032    -11.575      0.000      -0.427
    -0.303
Female        -0.0411      0.025     -1.662      0.097      -0.090
     0.007
==================================================================
==========
Omnibus:                        0.968   Durbin-Watson:
     1.874
Prob(Omnibus):                  0.616   Jarque-Bera (JB):
     0.779
Skew:                           0.001   Prob(JB):
     0.677
Kurtosis:                       3.193   Cond. No.
      223.
==================================================================
==========

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors
 is correctly specified.
P =
[[  3.99402174e+01   1.76630435e+00   4.45652174e-01]
 [  1.10818021e-11   5.69766456e-13   7.70841724e-14]]
```

(f) Check the numerical validity of the result in part (c). Note: This equation will not hold exactly because the coefficients have been rounded to two or three decimals; preciser results would have been obtained for higher precision coefficients.

In [25]:

```
b_R = (X1.T*X1).I*X1.T*matrix(y).T
print b_R

b1 = matrix([3.0527, -0.0411]).T
b2= matrix([0.0306,0.2332,-0.3654]).T

print b1+P*b2
```

```
[[ 4.48302717]
 [ 1.        ]]
[[ 4.52393152]
 [-0.0411    ]]
```