The Empirical Human Capital Model

Simple Version

For the simplest version of the model

$$w_i = A_0 \cdot \exp(r_s \cdot S_i + v_i)$$

where $w_i = \text{wage rate}$, $S_i = \text{years of schooling}$, and v_i is a random disturbance. This model implies

(1) $$\ln w_i = \ln(A_0) + r_s \cdot S_i + v_i = \beta_0 + \beta_S \cdot S_i + v_i$$

The following results are run on just the 1979 male sample in CPSORG.

```
.   reg lnwage ed

      Source |       SS       df       MS              Number of obs =    3096
-------------+------------------------------           F(  1,  3094) =  350.61
       Model |  66.7595844     1  66.7595844           Prob > F      =  0.0000
    Residual |  589.123863  3094  .190408488           R-squared     =  0.1018
-------------+------------------------------           Adj R-squared =  0.1015
       Total |  655.883447  3095  .211917107           Root MSE      =  .43636

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .0499781   .0026691    18.72   0.000     .0447447    .0552115
       _cons |   2.202434   .0341515    64.49   0.000     2.135472    2.269396
------------------------------------------------------------------------------
```

More Realistic

We might think of allowing for other variables to enter the this human capital model. For instance, it is though that increases in labor market experience increase human capital and, thus, wages

(2) $$\ln w_i = \beta_0 + \beta_S \cdot S_i + \beta_{\text{ex}} \cdot \text{ex}_i + v_i$$

```
.   reg lnwage ed ex

      Source |       SS       df       MS              Number of obs =    3096
-------------+------------------------------           F(  2,  3093) =  411.49
       Model |   137.83928     2  68.9196402           Prob > F      =  0.0000
    Residual |  518.044167  3093  .167489223           R-squared     =  0.2102
-------------+------------------------------           Adj R-squared =  0.2096
       Total |  655.883447  3095  .211917107           Root MSE      =  .40925

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .0697895   .0026817    26.02   0.000     .0645315    .0750476
          ex |   .0121608   .0005903    20.60   0.000     .0110034    .0133183
       _cons |   1.719942   .0396798    43.35   0.000     1.642141    1.797744
------------------------------------------------------------------------------
```

One of the problems with the above model is that returns to labor market experience are constant when we think there are many reasons to believe they are not. The fix for this is to allow for the rate of return to a year of labor market experience to depend on how much experience you actually have. This can be done by including an experience squared term in the above regression equation

(3) $$\ln w_i = \beta_0 + \beta_s \cdot S_i + \beta_{ex} \cdot \text{ex}_i + \beta_{\text{ex}\,sq} \cdot \left(ex_i\right)^2 + v_i$$

```
      Source |       SS       df       MS              Number of obs =    3096
-------------+------------------------------           F(  3,  3092) =  339.40
       Model |  162.480273     3   54.160091           Prob > F      =  0.0000
    Residual |  493.403174  3092  .159574118           R-squared     =  0.2477
-------------+------------------------------           Adj R-squared =  0.2470
       Total |  655.883447  3095  .211917107           Root MSE      =  .39947

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .0629068   .0026755    23.51   0.000     .0576609    .0681528
          ex |    .036022   .0020048    17.97   0.000     .0320912    .0399528
        exsq |  -.0005437   .0000438   -12.43   0.000    -.0006295   -.0004579
       _cons |   1.644259   .0392069    41.94   0.000     1.567385    1.721133
------------------------------------------------------------------------------
```

One Step Closer To Reality

One problem with the latest version of the model is our estimates of the expected wage conditional on educational attainment and labor market experience is the same for blacks, whites, and Hispanics. We could relax this by allowing different intercepts for individuals from different racial or ethnic groups. These different intercepts could reflect different average levels of initial human capital $\left(A_0\right)$ or the effects of discrimination.

(4) $$\ln w_i = \beta_0 + \beta_{black} \cdot black_i + \beta_{hisp} \cdot Hisp_i + \beta_s \cdot S_i + \beta_{ex} \cdot \mathrm{ex}_i + \beta_{\mathrm{ex}\,sq} \cdot \left(ex_i\right)^2 + v_i$$

In this model the variables $black$ and $Hisp$ are know as **dummy variables** or indicator variables.

In model (4) there are effectively three intercept terms.

Intercept for whites: β_0

Intercept for blacks: $\beta_0 + \beta_{black}$

Intercept for Hispanics: $\beta_0 + \beta_{Hisp}$

An equivalent model in terms of information content would be

(5) $$\begin{aligned} \ln w_i = {} & \alpha_{white} \cdot white_i + \alpha_{black} \cdot black_i + \alpha_{hisp} \cdot Hisp_i + \beta_s \cdot S_i \\ & + \beta_{ex} \cdot \mathrm{ex}_i + \beta_{\mathrm{ex}\,sq} \cdot \left(ex_i\right)^2 + v_i \end{aligned}$$

In this model the intercept for whites is

Intercept for whites: α_{white}

Intercept for blacks: α_{black}

Intercept for Hispanics: α_{Hisp}

The results from the model (4) estimation are

```
. reg lnwage black hisp ed ex exsq

      Source |       SS       df       MS              Number of obs =    3096
-------------+------------------------------           F(  5,  3090) =  210.53
       Model |  166.661253     5  33.3322506           Prob > F      =  0.0000
    Residual |  489.222194  3090  .158324335           R-squared     =  0.2541
-------------+------------------------------           Adj R-squared =  0.2529
       Total |  655.883447  3095  .211917107           Root MSE      =   .3979

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |  -.1330717   .0285048    -4.67   0.000    -.1889619   -.0771816
        hisp |  -.0836429   .0328886    -2.54   0.011    -.1481287   -.0191571
          ed |   .0602155   .0027439    21.95   0.000     .0548355    .0655955
          ex |   .0362958   .0019979    18.17   0.000     .0323783    .0402132
        exsq |  -.0005539   .0000437   -12.68   0.000    -.0006395   -.0004682
       _cons |   1.691663   .0405569    41.71   0.000     1.612142    1.771185
------------------------------------------------------------------------------
```

And the results from the model (5) specification

```
. reg lnwage black white hisp ed ex exsq, nocon

      Source |       SS       df       MS              Number of obs =    3096
-------------+------------------------------           F(  6,  3090) =26182.05
       Model |  24871.5301     6  4145.25502           Prob > F      =  0.0000
    Residual |  489.222194  3090  .158324335           R-squared     =  0.9807
-------------+------------------------------           Adj R-squared =  0.9807
       Total |  25360.7523  3096  8.19145747           Root MSE      =   .3979

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |   1.558592   .0460419    33.85   0.000     1.468316    1.648868
       white |   1.691663   .0405569    41.71   0.000     1.612142    1.771185
        hisp |   1.608021   .0457152    35.17   0.000     1.518385    1.697656
          ed |   .0602155   .0027439    21.95   0.000     .0548355    .0655955
          ex |   .0362958   .0019979    18.17   0.000     .0323783    .0402132
        exsq |  -.0005539   .0000437   -12.68   0.000    -.0006395   -.0004682
------------------------------------------------------------------------------
```

Note the relationship between the model (4) and model (5) coefficients.

(Hypothesis) Test Involving More than Just One Coefficient Value (The F-test)

Note that

(3) $\ln w_i = \beta_0 + \beta_s \cdot S_i + \beta_{ex} \cdot \mathrm{ex}_i + \beta_{\mathrm{ex}\,sq} \cdot \left(ex_i\right)^2 + v_i$

is just a restricted version of

(4) $\ln w_i = \beta_0 + \beta_{black} \cdot black_i + \beta_{hisp} \cdot Hisp_i + \beta_s \cdot S_i + \beta_{ex} \cdot \mathrm{ex}_i + \beta_{\mathrm{ex}\,sq} \cdot \left(ex_i\right)^2 + v_i$

Basically model (3) is equivalent to model (4) with the values of $\beta_{black} = \beta_{Hisp} = 0$.

The F-test provides a way of testing these and other linear restrictions. The F-test is based on the test statistic

$$F = \frac{\dfrac{SSE^* - SSE}{r}}{\dfrac{SSE}{n-(k+1)}} \sim F(r, n-(k+1))$$

where

r = the number of restrictions (the number of equal signs)

SSE^* = the error sum of squares from a restricted model

SSE = the error sum of squares from the unrestricted model

If we had the error sum of squares from the restricted and unrestricted models we could compute the value of the test statistic. It turns out that there is a much easier way to compute the value of the test statistic. If I divide both the numerator and denominator of the F-statistic by SST (Total Sum of Squares) we get

$$F = \frac{\dfrac{1}{r}\left[\dfrac{SSE^*}{SST} - \dfrac{SSE}{SST}\right]}{\dfrac{1}{n-(k+1)}\left[\dfrac{SSE}{SST}\right]} = \frac{\dfrac{1}{r}\left[1 - R_*^2 - \left(1 - R^2\right)\right]}{\left(\dfrac{1-R^2}{n-(k+1)}\right)} = \frac{\left(\dfrac{R^2 - R_*^2}{r}\right)}{\left(\dfrac{1-R^2}{n-(k+1)}\right)}$$

Basically all I need to do to conduct the test is estimate both the restricted and the unrestricted models, get the R-squared terms from these regressions, form the F-statistic and go to the tables.

It is now possible to test the restrictions implied by model (3) given only the output from the regression table provided in this handout.

$$F = \frac{\left(\frac{R^2 - R_*^2}{r}\right)}{\left(\frac{1-R^2}{n-(k+1)}\right)} = \frac{\left(\frac{0.2541-0.2477}{2}\right)}{\left(\frac{1-0.2541}{3090}\right)} \approx 13.26$$

The p-value associated with this test statistic is infinitesimal $\Rightarrow$ reject the restrictions.

It turns out there are is also a very easy way to compute this test statistic (and associated p-value) using STATA's post estimation commands.

Simply estimate the model

```
.   reg lnwage black hisp ed ex exsq

      Source |       SS       df       MS              Number of obs =    3096
-------------+------------------------------           F(  5,  3090) =  210.53
       Model |  166.661253     5  33.3322506           Prob > F      =  0.0000
    Residual |  489.222194  3090  .158324335           R-squared     =  0.2541
-------------+------------------------------           Adj R-squared =  0.2529
       Total |  655.883447  3095  .211917107           Root MSE      =   .3979

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |  -.1330717   .0285048    -4.67   0.000    -.1889619   -.0771816
        hisp |  -.0836429   .0328886    -2.54   0.011    -.1481287   -.0191571
          ed |   .0602155   .0027439    21.95   0.000     .0548355    .0655955
          ex |   .0362958   .0019979    18.17   0.000     .0323783    .0402132
        exsq |  -.0005539   .0000437   -12.68   0.000    -.0006395   -.0004682
       _cons |   1.691663   .0405569    41.71   0.000     1.612142    1.771185
------------------------------------------------------------------------------
```

After estimating the model you can use the test command to get a variety of test statistics

```
. test black=0

 ( 1)  black = 0.0

       F(  1,  3090) =   21.79
            Prob > F =    0.0000
```

This is the test statistic associated with the hypothesis that the coefficient on *black* is equal to zero.

```
. test hisp=0, accum

 ( 1)  black = 0.0
 ( 2)  hisp = 0.0

       F(  2,  3090) =   13.20
            Prob > F =    0.0000
```

This is a test statistic associated with testing the restrictions implied by model (3) (i.e. $\beta_{black} = \beta_{Hisp} = 0$). Note the use of the accum option.

Another Example

Suppose we thought that returns to school vary by race. We could estimate a model that allowed returns to school to vary by race by including race-schooling **interaction variables** into the model (4) specification.

$$(6) \quad \begin{aligned} \ln w_i = \beta_0 + \beta_{black} \cdot black_i + \beta_{hisp} \cdot Hisp_i + \beta_{black \cdot S} \cdot \left(black_i \cdot S_i\right) + \beta_{Hisp \cdot S} \cdot \left(Hisp_i \cdot S_i\right) \\ + \beta_s \cdot S_i + \beta_{ex} \cdot \mathrm{ex}_i + \beta_{\mathrm{ex}\,sq} \cdot \left(ex_i\right)^2 + v_i \end{aligned}$$

In this model the rate of return to a year of school for whites is β_S. For blacks and Hispanics the rates of return to a year of schooling $\beta_s + \beta_{black \cdot S}$ and $\beta_S + \beta_{Hisp \cdot S}$ respectively.

Note that model (4) is just a restricted version of this model. The restrictions implied by (4) are $\beta_{black \cdot S} = \beta_{Hisp \cdot S} = 0$

We can test these restrictions using the F-Statistic. As with the previous example there are two approaches

1. Estimate the restricted and unrestricted models, obtain the R-squared terms, and construct the test statistic. (You will need to be able to do this for the exam).
2. Estimate the unrestricted model and use STATA's test command.

I will use the second approach

```
. gen black_ed=black*ed

. gen hisp_ed=hisp*ed

.   reg lnwage black hisp black_ed hisp_ed ed ex exsq

      Source |       SS       df       MS              Number of obs =    3096
-------------+------------------------------           F(  7,  3088) =  150.53
       Model |  166.862028     7  23.8374325           Prob > F      =  0.0000
    Residual |   489.02142  3088  .158361859           R-squared     =  0.2544
-------------+------------------------------           Adj R-squared =  0.2527
       Total |  655.883447  3095  .211917107           Root MSE      =  .39795

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |  -.2529284   .1115613    -2.27   0.023    -.4716703   -.0341866
        hisp |  -.1117459   .0983421    -1.14   0.256    -.3045685    .0810766
    black_ed |   .0104171   .0093896     1.11   0.267    -.0079934    .0288275
     hisp_ed |   .0025325   .0088625     0.29   0.775    -.0148444    .0199094
          ed |   .0592706   .0029542    20.06   0.000     .0534781     .065063
          ex |   .0362238    .002003    18.09   0.000     .0322966    .0401511
        exsq |  -.0005515   .0000438   -12.58   0.000    -.0006375   -.0004656
       _cons |   1.703744   .0427701    39.83   0.000     1.619884    1.787605
------------------------------------------------------------------------------

. test black_ed=0

 ( 1)  black_ed = 0.0

       F(  1,  3088) =    1.23
            Prob > F =    0.2673


. test hisp_ed=0, accum

 ( 1)  black_ed = 0.0
 ( 2)  hisp_ed = 0.0

       F(  2,  3088) =    0.63
            Prob > F =    0.5306
```

On the basis of this test statistic we cannot reject restrictions implied by model (4).

As an additional exercise you should compute the value of the test statistic manually. If it is not approximately equal to 0.63 you did something wrong.