4.2 Tests of Structural Changes

This section focuses on Chow Test and leaves general discussion on dummy variable models to other section.

4.2.1 Chow Test: Simple Example[1]

Chow Test examines whether parameters of one group of the data are equal to those of other groups. Simply put, the test checks whether the data can be pooled. If only intercepts are different across groups, this is a fixed effect model, which is simple to handle. Let us consider two groups.

$y = \alpha + \beta x + \varepsilon$ for all observations

$y = \alpha_1 + \beta_1 x + \varepsilon_1$ for n_1 observations (group 1)

$y = \alpha_2 + \beta_2 x + \varepsilon_2$ for n_2 observations (group 2)

The null hypothesis is $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. If the null hypothesis is rejected, two groups have different slopes and intercepts; data are not poolable.

$$F(J, n_1 + n_2 - 2K) = \frac{(e'e - e_1'e_1 - e_2'e_2)/J}{(e_1'e_1 + e_2'e_2)/(n_1 + n_2 - 2K)} = \frac{(SSE - SSE_1 - SSE_2)/J}{(SSE_1 + SSE_2)/(n_1 + n_2 - 2K)}$$ where

$e'e$ is the SSE of the pooled model and *J* is the number of restrictions (often equal to K—all parameters).[2]

In order to conduct the Chow test,

1. Run pooled OLS to get SSE_{pooled}
2. Run separate OLS to get SSE_1 and SSE_2
3. Apply the formula

In the following example, we assume that two groups have difference slopes of `cost` and different intercepts; there are two restrictions, *J*=2.

```
. use http://www.indiana.edu/~statmath/stat/all/panel/airline.dta, clear
(Cost of U.S. Airlines (Greene 2003))

. gen d1=(airline<=2)
. gen d0=(d1==0)
. gen d2=(airline>=2 & airline<=4)
. gen d3 =(airline>=5)

. gen cost0=cost*d0
. gen cost1=cost*d1
. gen cost2=cost*d2
. gen cost3=cost*d3
```

First, fit the pooled model.

---

[1] Greene 2003 (289-291), http://www.stata.com/support/faqs/stat/chow.html

[2] If we want to test the null hypothesis that only intercept is different, J will be K-1 (all the slopes are equal)

```
. regress output cost            // pooled model

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  1,    88) =  880.73
       Model |  107.123089     1  107.123089           Prob > F      =  0.0000
    Residual |  10.7034329    88   .12162992           R-squared     =  0.9092
-------------+------------------------------           Adj R-squared =  0.9081
       Total |  117.826522    89  1.32389351           Root MSE      =  .34875

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cost |   .9691952    .032658    29.68   0.000     .9042942    1.034096
       _cons |  -14.12819   .4380397   -32.25   0.000    -14.99871   -13.25768
------------------------------------------------------------------------------
```

And then fit separate models using two subsets of data.

```
. regress output cost if d1==0    // if airline >= 3

      Source |       SS       df       MS              Number of obs =      60
-------------+------------------------------           F(  1,    58) =  314.51
       Model |     30.7715     1     30.7715           Prob > F      =  0.0000
    Residual |   5.6746959    58  .097839584           R-squared     =  0.8443
-------------+------------------------------           Adj R-squared =  0.8416
       Total |  36.4461959    59  .617732134           Root MSE      =  .31279

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cost |   .8454175    .047671    17.73   0.000     .7499936    .9408413
       _cons |  -12.64286    .610876   -20.70   0.000    -13.86566   -11.42006
------------------------------------------------------------------------------

. regress output cost if d1==1    // if airline <= 2

      Source |       SS       df       MS              Number of obs =      30
-------------+------------------------------           F(  1,    28) =  166.32
       Model |  2.80505683     1  2.80505683           Prob > F      =  0.0000
    Residual |  .472224836    28  .016865173           R-squared     =  0.8559
-------------+------------------------------           Adj R-squared =  0.8508
       Total |  3.27728166    29  .113009712           Root MSE      =  .12987

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cost |   .5144647   .0398915    12.90   0.000     .4327507    .5961787
       _cons |  -7.328991   .5798709   -12.64   0.000    -8.516802   -6.141179
------------------------------------------------------------------------------
```

$$F = \frac{(e'e - e_1'e_1 - e_2'e_2)/J}{(e_1'e_1 + e_2'e_2)/(n_1 + n_2 - 2K)} = \frac{(10.7034 - 5.6747 - .4722)/2}{(5.6747 + .4722)/(60 + 30 - 2*2)} = 31.8745$$

```
. di ((10.7034329-.472224836-5.6746959)/2)/((.472224836+5.6746959)/(30+60-2*2))
31.8745

. di Ftail(2,86,31.8745)
4.393e-11
```

The large F 31.8745 (2, 68) rejects the null hypothesis of equal slope and intercept (p<.0000).

Alternatively, you may regress *y* on two dummies and two interaction terms with the intercept suppressed.[3] Parameter estimates are identical to those of the above, while standard errors are

[3] Therefore, R^2 and standard errors are not reliable.

different. This estimation is handy since parameter estimates are slopes and intercepts of individual groups; any further computation is not need at all.

```
. regress output cost0 d0 cost1 d1, noconstant

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  4,    86) =  824.72
       Model |  235.789788     4  58.9474471           Prob > F      =  0.0000
    Residual |  6.14692073    86  .071475822           R-squared     =  0.9746
-------------+------------------------------           Adj R-squared =  0.9734
       Total |  241.936709    90  2.68818566           Root MSE      =  .26735

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       cost0 |   .8454175   .0407452    20.75   0.000     .7644187    .9264162
          d0 |  -12.64286    .522126   -24.21   0.000    -13.68081   -11.60491
       cost1 |   .5144647   .0821229     6.26   0.000     .3512098    .6777197
          d1 |  -7.328991   1.193756    -6.14   0.000    -9.702098   -4.955883
------------------------------------------------------------------------------

. test _b[cost0]=_b[cost1]
 ( 1)  cost0 - cost1 = 0

       F(  1,    86) =   13.03
            Prob > F =    0.0005

. test _b[d0]=_b[d1], accum

 ( 1)  cost0 - cost1 = 0
 ( 2)  d0 - d1 = 0

       F(  2,    86) =   31.87
            Prob > F =    0.0000
```

More convenient way is to regress *y* on a regressor of interest, `cost`, an interaction term, and a dummy with the intercept included. The intercept is the intercept of the baseline group and the dummy coefficient is the deviation from the baseline intercept. The coefficient of the regressor is the slope of the baseline, while the coefficient of the interaction term is the deviation of slope from the baseline slope. That is, the intercept of compared group is 7.3290 (=5.3139-12.6429) and the slope is .5145 (=-.3310+.8454).

```
. regress output cost cost1 d1

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  3,    86) =  520.83
       Model |  111.679602     3  37.2265339           Prob > F      =  0.0000
    Residual |  6.14692073    86  .071475822           R-squared     =  0.9478
-------------+------------------------------           Adj R-squared =  0.9460
       Total |  117.826522    89  1.32389351           Root MSE      =  .26735

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cost |   .8454175   .0407452    20.75   0.000     .7644187    .9264162
       cost1 |  -.3309528   .0916752    -3.61   0.001     -.513197   -.1487085
          d1 |    5.31387   1.302946     4.08   0.000     2.723699    7.904041
       _cons |  -12.64286    .522126   -24.21   0.000    -13.68081   -11.60491
------------------------------------------------------------------------------

. test _b[cost1]=0
 ( 1)  cost1 = 0

       F(  1,    86) =   13.03
            Prob > F =    0.0005
```

```
. test _b[d1]=0, accum
 ( 1)  cost1 = 0
 ( 2)  d1 = 0

       F(  2,    86) =   31.87
            Prob > F =    0.0000
```

4.2.2 Chow Test: Comparing Three Groups

What if we want to compare three groups? Let us fit two remaining models for group 2 and 3. Restrictions here are 1) slop2 is equal to the baseline slope (slop1), 2) slope3 is equal to the baseline slope, 3) intercept2 is equal to the baseline intercept, and 4) intercept3 is equal to the baseline intercept. Degrees of freedom is 84=N-(group*K)=90-3*2.

```
. regress output cost if d2==1

      Source |       SS       df       MS              Number of obs =      30
-------------+------------------------------           F(  1,    28) =   43.01
       Model |  4.48338312     1  4.48338312           Prob > F      =  0.0000
    Residual |  2.91859218    28  .104235435           R-squared     =  0.6057
-------------+------------------------------           Adj R-squared =  0.5916
       Total |   7.4019753    29  .255240528           Root MSE      =  .32286

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cost |   .6205532   .0946202     6.56   0.000     .4267326    .8143739
       _cons |  -9.498564   1.255486    -7.57   0.000    -12.07031   -6.926819
------------------------------------------------------------------------------

. regress output cost if d3==1

      Source |       SS       df       MS              Number of obs =      30
-------------+------------------------------           F(  1,    28) =  817.74
       Model |  9.93465282     1  9.93465282           Prob > F      =  0.0000
    Residual |  .340168734    28  .012148883           R-squared     =  0.9669
-------------+------------------------------           Adj R-squared =  0.9657
       Total |  10.2748216    29  .354304191           Root MSE      =  .11022

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cost |   .7370425   .0257742    28.60   0.000     .6842466    .7898385
       _cons |  -11.47175   .3181417   -36.06   0.000    -12.12344   -10.82007
------------------------------------------------------------------------------
```

$$F = \frac{(e'e - e_1'e_1 - e_2'e_2 - e_3'e_3)/J}{(e_1'e_1 + e_2'e_2 + e_3'e_3)/(n_1 + n_2 + n_3 - gK)} = \frac{(10.703 - .472 - 2.919 - .340)/4}{(.472 + 2.919 + .340)/(30 + 30 + 30 - 3*2)} = 39.245$$

```
. di ((10.7034329-.472224836-2.91859218-.340168734)/4)/ ///
((.472224836+2.91859218+.340168734)/(30+30+30-3*2))
39.244693

. di Ftail(4,84,39.244693)
1.696e-18
```

Now, include all interaction terms and dummies without the regressor of interest. The model report all parameter estimates, which are identical to the above.

```
. regress output cost1 d1 cost2 d2 cost3 d3, noconstant

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  6,    84) =  893.83
       Model |  238.205723     6  39.7009539           Prob > F      =  0.0000
```

```
    Residual |  3.73098575    84  .044416497           R-squared     =  0.9846
-------------+------------------------------           Adj R-squared =  0.9835
       Total |  241.936709    90  2.68818566           Root MSE      =  .21075

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       cost1 |   .5144647   .0647376     7.95   0.000     .3857268    .6432026
          d1 |  -7.328991   .9410399    -7.79   0.000    -9.200352   -5.457629
       cost2 |   .6205532   .0617658    10.05   0.000     .4977251    .7433813
          d2 |  -9.498564   .8195514   -11.59   0.000    -11.12833   -7.868796
       cost3 |   .7370425    .049282    14.96   0.000     .6390399    .8350452
          d3 |  -11.47175   .6083095   -18.86   0.000    -12.68144   -10.26206
------------------------------------------------------------------------------

. test _b[cost1]=_b[cost2]=_b[cost3], notest
. test _b[d1]=_b[d2]=_b[d3], accum
 ( 1)  cost1 - cost2 = 0
 ( 2)  cost1 - cost3 = 0
 ( 3)  d1 - d2 = 0
 ( 4)  d1 - d3 = 0

       F(  4,    84) =   39.24
            Prob > F =    0.0000
```

Finally, include the regresor, two interaction terms, and two dummies excluding baseline interaction and dummy. The coefficient of the regressor is the baseline slope and the intercept is the baseline intercept. Coefficients of interaction terms are deviations from the baseline slope. As a result, the slope of group 1 is .5144647 (=-.2225778+.7370425) and the intercept is -7.328991 (=4.142762-11.47175).

```
. regress output cost cost1 d1 cost2 d2

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  5,    84) =  513.75
       Model |  114.095537     5  22.8191073           Prob > F      =  0.0000
    Residual |  3.73098575    84  .044416497           R-squared     =  0.9683
-------------+------------------------------           Adj R-squared =  0.9665
       Total |  117.826522    89  1.32389351           Root MSE      =  .21075

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cost |   .7370425    .049282    14.96   0.000     .6390399    .8350452
       cost1 |  -.2225778   .0813614    -2.74   0.008    -.3843739   -.0607817
          d1 |   4.142762   1.120534     3.70   0.000     1.914458    6.371067
       cost2 |  -.1164893   .0790173    -1.47   0.144    -.2736239    .0406453
          d2 |   1.973189   1.020639     1.93   0.057    -.0564649    4.002842
       _cons |  -11.47175   .6083095   -18.86   0.000    -12.68144   -10.26206
------------------------------------------------------------------------------

. test _b[cost1]=0, notest
. test _b[cost2]=0, accum notest
. test _b[d1]=0, accum notest
. test _b[d2]=0, accum
 ( 1)  cost1 = 0
 ( 2)  cost2 = 0
 ( 3)  d1 = 0
 ( 4)  d2 = 0

       F(  4,    84) =   39.24
            Prob > F =    0.0000
```

Hypothesis testing is identical to the above.

4.2.3 Chow Test: Including Covariates

Now, suppose we need to include some covariates for control, `load` and `fuel`. We may regress the dependent variable on all interactions, dummies, and covariates with the intercept suppressed. Again R^2 is not reliable here. A coefficient of an interaction term is the slope of cost of the group, just as the dummy coefficient is the intercept of the group.

```
. regress output cost1 d1 cost2 d2 cost3 d3 load fuel, noconstant

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  8,    82) = 1526.59
       Model |  240.323102     8  30.0403877           Prob > F      =  0.0000
    Residual |  1.61360768    82  .019678142           R-squared     =  0.9933
-------------+------------------------------           Adj R-squared =  0.9927
       Total |  241.936709    90  2.68818566           Root MSE      =  .14028

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       cost1 |   1.017854   .0884737    11.50   0.000     .8418518    1.193856
          d1 |  -9.777424   .7217463   -13.55   0.000    -11.21321    -8.34164
       cost2 |   1.070322   .0839545    12.75   0.000     .9033093    1.237334
          d2 |  -10.56474   .5832504   -18.11   0.000    -11.72501   -9.404465
       cost3 |    1.10826   .0672781    16.47   0.000     .9744229    1.242098
          d3 |  -11.10194   .4070026   -27.28   0.000     -11.9116   -10.29229
        load |   2.041787    .398699     5.12   0.000     1.248648    2.834927
        fuel |  -.4733269   .0556113    -8.51   0.000    -.5839555   -.3626983
------------------------------------------------------------------------------
```

This model is written as,

```
output= -9.7774+1.0179*cost+2.0418*load-.4733*fuel (group 1)
output=-10.5647+1.0703*cost+2.0418*load-.4733*fuel (group 2)
output=-11.1019+1.1083*cost+2.0418*load-.4733*fuel (group 3)
```

Let us conduct the hypothesis test for the four restrictions.

```
. test _b[cost1]=_b[cost2]=_b[cost3], notest
. test _b[d1]=_b[d2]=_b[d3], accum
 ( 1)  cost1 - cost2 = 0
 ( 2)  cost1 - cost3 = 0
 ( 3)  d1 - d2 = 0
 ( 4)  d1 - d3 = 0

       F(  4,    82) =    0.86
            Prob > F =    0.4923
```

We may also fit the same model including the regressor of interest and excluding baseline interaction term and its intercept. Covariates remain unchanged. Note that slope2 is 1.0703 (=-.0379+1.1083) and the intercept of group 2 is -10.5647 (=.5372079-11.10194).

```
. regress output cost cost1 d1 cost2 d2 load fuel

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  7,    82) =  843.67
       Model |  116.212915     7  16.6018449           Prob > F      =  0.0000
    Residual |  1.61360768    82  .019678142           R-squared     =  0.9863
-------------+------------------------------           Adj R-squared =  0.9851
       Total |  117.826522    89  1.32389351           Root MSE      =  .14028

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cost |    1.10826   .0672781    16.47   0.000     .9744229    1.242098
       cost1 |  -.0904063   .0572606    -1.58   0.118    -.2043159    .0235034
          d1 |    1.32452   .8454028     1.57   0.121    -.3572557    3.006295
```

```
       cost2 |  -.0379388   .0545639    -0.70   0.489    -.1464838    .0706061
          d2 |   .5372079   .7215669     0.74   0.459    -.8982187    1.972634
        load |   2.041787    .398699     5.12   0.000     1.248648    2.834927
        fuel |  -.4733269   .0556113    -8.51   0.000    -.5839555   -.3626983
       _cons |  -11.10194   .4070026   -27.28   0.000     -11.9116   -10.29229
------------------------------------------------------------------------------

. test _b[cost1]=0, notest
. test _b[cost2]=0, accum notest
. test _b[d1]=0, accum notest
. test _b[d2]=0, accum

 ( 1)  cost1 = 0
 ( 2)  cost2 = 0
 ( 3)  d1 = 0
 ( 4)  d2 = 0

       F(  4,    82) =    0.86
            Prob > F =    0.4923
```

4.2.4 Chow Test: Different Slopes of Multiple Regressors[4]

If we assume that more than one regressor have different slopes across group, model will be complicated. But the underlying logic remains same. Let us create a set of interaction terms for the regressor `load`.

```
. gen load1=load*d1
. gen load2=load*d2
. gen load3=load*d3
```

First, include all interactions, dummies, and covariate `fuel`. Do not forget the suppress the intercept.

```
. regress output cost1 load1 d1 cost2 load2 d2 cost3 load3 d3 fuel, noconstant

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F( 10,    80) = 1207.56
       Model |  240.344451    10  24.0344451           Prob > F      =  0.0000
    Residual |  1.59225866    80  .019903233           R-squared     =  0.9934
-------------+------------------------------           Adj R-squared =  0.9926
       Total |  241.936709    90  2.68818566           Root MSE      =  .14108

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       cost1 |   1.014764   .0975898    10.40   0.000     .8205541    1.208974
       load1 |   1.817979   .6811907     2.67   0.009     .4623659    3.173591
          d1 |  -9.754334   .7455903   -13.08   0.000    -11.23811   -8.270562
       cost2 |   1.019311   .0981769    10.38   0.000     .8239324    1.214689
       load2 |   2.661021   .7210198     3.69   0.000     1.226146    4.095896
          d2 |  -10.38952   .6108743   -17.01   0.000     -11.6052   -9.173845
       cost3 |   1.111908   .0687766    16.17   0.000     .9750386    1.248778
       load3 |   1.703393   .7009252     2.43   0.017     .3085071    3.098278
          d3 |  -11.11345   .4096207   -27.13   0.000    -11.92862   -10.29827
        fuel |  -.4615669   .0578651    -7.98   0.000    -.5767222   -.3464117
------------------------------------------------------------------------------

. test _b[cost1]=_b[cost2]=_b[cost3], notest
. test _b[load1]=_b[load2]=_b[load3], accum notest
. test _b[d1]=_b[d2]=_b[d3], accum

 ( 1)  cost1 - cost2 = 0
 ( 2)  cost1 - cost3 = 0
 ( 3)  load1 - load2 = 0
```

[4] http://www.stata.com/support/faqs/stat/chow3.html

```
 ( 4)  load1 - load3 = 0
 ( 5)  d1 - d2 = 0
 ( 6)  d1 - d3 = 0

       F(  6,    80) =    0.74
            Prob > F =    0.6152
```

Now, include two regressors of interest, two sets of interactions, two dummies, and a covariate, excluding baseline interaction terms and its dummy. Make sure you include two regresors of interest, `cost` and `load`.

```
. regress output cost load cost1 load1 d1 cost2 load2 d2 fuel

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  9,    80) =  648.89
       Model |  116.234264     9  12.9149182           Prob > F      =  0.0000
    Residual |  1.59225866    80  .019903233           R-squared     =  0.9865
-------------+------------------------------           Adj R-squared =  0.9850
       Total |  117.826522    89  1.32389351           Root MSE      =  .14108

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cost |   1.111908   .0687766    16.17   0.000     .9750386    1.248778
        load |   1.703393   .7009252     2.43   0.017     .3085071    3.098278
       cost1 |  -.0971444   .0777234    -1.25   0.215    -.2518189    .0575301
       load1 |   .1145858   .9857632     0.12   0.908    -1.847146    2.076317
          d1 |   1.359111   .8696813     1.56   0.122    -.3716095    3.089832
       cost2 |  -.0925978   .0789714    -1.17   0.244    -.2497558    .0645603
       load2 |   .9576281   1.020969     0.94   0.351    -1.074165    2.989421
          d2 |    .723922   .7484569     0.97   0.336    -.7655546    2.213399
        fuel |  -.4615669   .0578651    -7.98   0.000    -.5767222   -.3464117
       _cons |  -11.11345   .4096207   -27.13   0.000    -11.92862   -10.29827
------------------------------------------------------------------------------

. test _b[cost1]=0, notest
. test _b[cost2]=0, accum notest
. test _b[load1]=0, accum notest
. test _b[load2]=0, accum notest
. test _b[d1]=0, accum notest
. test _b[d2]=0, accum

 ( 1)  cost1 = 0
 ( 2)  cost2 = 0
 ( 3)  load1 = 0
 ( 4)  load2 = 0
 ( 5)  d1 = 0
 ( 6)  d2 = 0

       F(  6,    80) =    0.74
            Prob > F =    0.6152
```