Imposing and Testing Equality Constraints in Models

Overview. We have previously discussed how to impose and test certain restrictions on models. In particular, we have examined how to (1) test whether a subset of coefficients equals zero (2) test whether all coefficients equal zero (3) test whether a single coefficient equals zero. In this section we will extend this discussion by explaining how to test whether (4) two or more coefficients within a model are equal; we'll also show how to test more complicated sorts of equality constraints. In later handouts we'll discuss how to impose and test equality constraints across populations, e.g. see whether income has the same effect on both men and women.

Test for equality of parameters within a model. Suppose you wish to test

H_0: $\beta_1 = \beta_2$
H_A: $\beta_1 \neq \beta_2$

It may be helpful to note that this is the same as testing

H_0: $\beta_1 - \beta_2 = 0$
H_A: $\beta_1 - \beta_2 \neq 0$

That is, you want to test whether two variables have equal effects. For example, in a model of family decision-making, you might hypothesize that wives have the same amount of influence as their husbands. Or, you might want to test whether time spent in one type of activity has the same effect as time spent in another activity.

There are at least 3 ways of doing this. Note that

- All three of these approaches are fine for OLS regression.
- HOWEVER, for maximum likelihood techniques, like logistic regression, a variation of Option 1 (using chi-square rather than F statistics) works best. While the other options can be used with MLE techniques and generally work ok, they can occasionally produce results that are inaccurate or sub-optimal.
- These approaches can generally be modified to test more complicated hypotheses, e.g. $\beta_1 = \beta_2 = \beta_3$, $\beta_1 = 3^* \beta_2 + 7$, etc. The trick is figuring out how to set the problem up. Stata often makes this easier than SPSS does.

Option 1: Incremental F Test. We've talked a lot about incremental F tests (see Appendix B if you need a refresher). Some key advantages of this approach are that (a) you can use it with most statistical software, and (b) even though this isn't always the easiest approach, it is important to understand it because the strategy used here is similar to the strategy that is optimal for other statistical techniques like logistic regression. You can proceed as follows:

- Compute a new variable that is equal to the sum of the two variables you hypothesize to have equal effects, e.g. in SPSS you might have COMPUTE SUM12 = X1 + X2, whereas in Stata you'd have `gen sum12 = x1 + x2`

- Regress Y on X1, X2, and any other IVs in the model. Retrieve the Unconstrained Error Sum of Squares (SSE_u). We refer to this as the unconstrained model, because the effects of X1 and X2 are not constrained to be equal. That is, you are estimating the model

$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \sum_{k=3}^{K} \beta_k X_k + \varepsilon$$

- Run a second regression in which you regress Y on SUM12 and any other IVs in the model. (Do NOT include X1 and X2 though.) Retrieve the constrained error sum of squares (SSE_c). We refer to this as the constrained model, because, by adding X1 and X2 together, you are forcing their estimated effects to be equal (i.e. only one beta is being estimated for both variables). That is, you are estimating the model

$$y = \alpha + \beta_1 (X_1 + X_2) + \sum_{k=3}^{K} \beta_k X_k + \varepsilon$$

In practice, we estimate this via

$$y = \alpha + \beta_1 \text{Sum12} + \sum_{k=3}^{K} \beta_k X_k + \varepsilon$$

- Do an incremental F test. In this case, J = 1, so we get

$$F_{1,N-K-1} = \frac{(SSE_c - SSE_u)/1}{SSE_u/(N-K-1)} = \frac{(SSE_c - SSE_u)*(N-K-1)}{SSE_u * 1} = \frac{(R_u^2 - R_c^2)*(N-K-1)}{(1-R_u^2)*1}$$

- This procedure can be easily modified to test similar hypotheses. For example, if you hypothesize that X1 and X2 have equal but opposite effects, compute a variable like DIFF12 = X1 - X2. If you think that the effect of X1 is twice that of X2, compute something like WSUM12 = 2X1 + X2. If you hypothesize that 3 variables have equal effects, compute SUM123 = X1 + X2 + X3. (Note that J =2 in this case.)

EXAMPLE. Here is a modified version of the income/education/job experience example we have been using. I have reworked the data so that it is now a sample of 100 blacks and four hundred whites. We want to test whether a year of job experience (JOBEXP) has the same effect on income as a year of education (EDUC).

First, we estimate the unconstrained model. INCOME is regressed on EDUC and JOBEXP, yielding the following (I'm using SPSS but Stata would be very similar):

```
Variable(s) Entered on Step Number
   1..    JOBEXP
   2..    EDUC


Multiple R           .90347
R Square             .81626
Adjusted R Square    .81552
Standard Error      3.85420
```

```
Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           2         32798.40320      16399.20160
Residual           497          7382.84680         14.85482

F =    1103.96483       Signif F =  .0000


------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

EDUC           1.945120     .043700    .862872    44.511  .0000
JOBEXP          .708221     .034367    .399488    20.607  .0000
(Constant)    -7.382935     .802778               -9.197  .0000
```

Hence, $SSE_u = 7382.84680$, $R^2_u = .81626$, N = 500, K = 2.

Now, we estimate the constrained model. First, we compute JOBED = EDUC + JOBEXP. Then, we regress INCOME on JOBED. We get the following:

```
Variable(s) Entered on Step Number
   1..    JOBED


Multiple R           .77935
R Square             .60739
Adjusted R Square    .60660
Standard Error      5.62831

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           1         24405.69609      24405.69609
Residual           498         15775.55391         31.67782

F =     770.43486       Signif F =  .0000


------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

JOBED          1.160275     .041802    .779352    27.757  .0000
(Constant)    -3.166147    1.143318               -2.769  .0058
```

Thus, we get $SSE_c = 15775.55391$, $R^2_c = .60739$, J = 1.

The incremental F test is then

$$F_{1,N-K-1} = \frac{(SSE_c - SSE_u)*(N-K-1)}{SSE_u * 1} = \frac{(R_u^2 - R_c^2)*(N-K-1)}{(1-R_u^2)*1}$$

$$= \frac{(15775.55 - 7382.85)*497}{7382.85} = \frac{(.81626 - .60739)*497}{1 - .81626} = 565$$

This value is highly significant. Ergo, we reject the null hypothesis that education and job experience have equal effects.

Option 2. Wald Test. Wald tests are computed using the estimated coefficients and the variances/covariances of the estimates from the unconstrained model. A nice feature of Wald tests is that they only require the estimation of one model. This is the approach used by Stata's `test` command, where it is quite easy and simple to use. The main limitations are that (1) Not all programs make this as easy as Stata does, and (b) while you can do it, the use of a Wald test is NOT the optimal strategy with maximum likelihood techniques like logistic regression. A strategy more like Option 1 (using chi-square tests instead of Fs) is preferable.

Here is the rationale for this approach: Recall that testing $\beta_{Educ} = \beta_{Jobexp}$ is equivalent to testing $\beta_{Educ} - \beta_{Jobexp} = 0$. Also recall that

$$V(X \pm Y) = V(X) + V(Y) \pm 2\ COV(X,Y) = \sigma^2_{X \pm Y}$$

This implies that

$$V(\beta_{Educ} - \beta_{Jobexp}) = V(\beta_{Educ}) + V(\beta_{Jobexp}) - 2\ COV(\beta_{Educ}, \beta_{Jobexp})$$

Hence, an appropriate test statistic for this problem is

$$F_{1,N-K-1} = \left(\frac{(b_{Educ} - b_{Job\exp}) - (\beta_{Educ} - \beta_{Job\exp})_0}{\sqrt{s^2{}_{b_{Educ}} + s^2{}_{b_{Job\exp}} - 2s_{b_{Educ},b_{Job\exp}}}} \right)^2 = \left(\frac{(b_{Educ} - b_{Job\exp})}{\sqrt{s^2{}_{b_{Educ}} + s^2{}_{b_{Job\exp}} - 2s_{b_{Educ},b_{Job\exp}}}} \right)^2$$

We have seen this idea many times before: Observed Value – Value Predicted by the Null (the predicted difference in this case being zero) divided by the estimated standard error of the estimator. Since an F test is being reported, all of this is squared.

In SPSS, you include the BCOV parameter on the regression command to get the variance/covariance matrix of the estimators:

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS BCOV R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT income
  /METHOD=ENTER educ jobexp  .
```

The output includes

Coefficient Correlations[a]

Model			JOBEXP	EDUC
1	Correlations	JOBEXP	1.0000000	.1274659
		EDUC	.1274659	1.0000000
	Covariances	JOBEXP	.0011811	.0001914
		EDUC	.0001914	.0019097

a. Dependent Variable: INCOME

Hence,

$$F_{1,N-K-1} = \left(\frac{(b_{Educ} - b_{Job\exp})}{\sqrt{s^2_{b_{Educ}} + s^2_{b_{Job\exp}} - 2s_{b_{Educ},b_{Job\exp}}}} \right)^2 = \left(\frac{(1.94512 - .7082212)}{\sqrt{.0019097 + .0011811 - 2 * .0001914}} \right)^2$$

$$= \left(\frac{1.236973312}{.052038447} \right)^2 = 23.77037317^2 = 565.03$$

Wald tests can be used to test more complicated hypotheses than the above; but it is a lot easier to do so if you have a program like Stata figuring out how to do it, rather than you going through and figuring out exactly how to set the problem up.

Stata. To do a Wald test, use Stata's `test` command after `regress`:

```
. use http://www.nd.edu/~rwilliam/stats2/statafiles/blwh.dta, clear

. reg  income educ jobexp

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  2,   497) = 1103.96
       Model |  32798.4018     2  16399.2009           Prob > F      =  0.0000
    Residual |  7382.84742   497  14.8548238           R-squared     =  0.8163
-------------+------------------------------           Adj R-squared =  0.8155
       Total |  40181.2493   499  80.5235456           Root MSE      =  3.8542

------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    1.94512   .0436998    44.51   0.000     1.859261     2.03098
      jobexp |   .7082212   .0343672    20.61   0.000     .6406983     .775744
       _cons |  -7.382935   .8027781    -9.20   0.000    -8.960192   -5.805678
------------------------------------------------------------------------------

. test educ = jobexp

 ( 1)  educ - jobexp = 0

       F(  1,   497) =  564.98
            Prob > F =    0.0000
```

If for some reason you prefer to do the calculation yourself , you can use the `vce` command after running `regress` to get the variances and covariances (`estat vce` also works)

```
. vce

             |     educ   jobexp    _cons
-------------+---------------------------
        educ |   .00191
      jobexp |  .000191  .001181
       _cons | -.027719 -.018488  .644453
```

Incidentally, in Stata, if you want to see what the constrained parameter estimates look like, add the `coef` parameter to the test command, e.g.

```
. test educ = jobexp, coef

 ( 1)  educ - jobexp = 0

       F(  1,   497) =  564.98
            Prob > F =    0.0000

Constrained coefficients

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.160275   .0286252    40.53   0.000     1.104171     1.21638
      jobexp |   1.160275   .0286252    40.53   0.000     1.104171     1.21638
       _cons |  -3.166146   .7829306    -4.04   0.000    -4.700662    -1.63163
------------------------------------------------------------------------------
```

Especially if you have tried to test a fairly complicated hypothesis, it is good to look at the constrained coefficients to make sure you specified things correctly, e.g. if for some reason you hypothesized that $\beta_{Educ} = 3 * \beta_{Jobexp} - 1$,

```
. test educ = 3*jobexp - 1, coef

 ( 1)  educ - 3 jobexp = -1

       F(  1,   497) =   59.09
            Prob > F =    0.0000


Constrained coefficients

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.848938   .0418704    44.16   0.000     1.766873    1.931002
      jobexp |   .9496459   .0139568    68.04   0.000     .9222911    .9770007
       _cons |  -9.381235   .7595261   -12.35   0.000    -10.86988   -7.892591
------------------------------------------------------------------------------
```

`test` also makes it easy to test simpler hypotheses, e.g.

```
. test educ = 2

 ( 1)  educ = 2

       F(  1,   497) =    1.58
            Prob > F =    0.2098
```

Conversely, Stata's `testnl` command lets you test complicated nonlinear sorts of relationships among coefficients, e.g. if for some reason you hypothesize Sqrt(β_{Educ}) = β_{Jobexp},

```
. testnl sqrt(_b[educ]) = _b[jobexp]

  (1)  sqrt(_b[educ]) = _b[jobexp]

            F(1, 497) =      365.49
             Prob > F =        0.0000
```

Option 3. T-Test. This is less widely used so I have moved it to Appendix A.

Conclusion. If you are doing OLS regression and you are using Stata, Option 2 (Wald tests) is probably the easiest way to go (although Stata can handle the other options too). If using SPSS for OLS regression, Option 1 (incremental F test) is probably best. Option 3 (T-Test) is also good, but interpretation of the results is a little trickier, and more complicated tests may be hard to set up this way.

HOWEVER, if you are using a maximum likelihood technique like logistic regression, a modified version of Option 1 (using a chi-square statistic instead of F) tends to be optimal. We will discuss this more later.

Also, while we have primarily talked about testing the equality of 2 coefficients, e.g. $\beta_1 = \beta_2$, we have also seen that much more complicated sorts of tests are possible. You can also do simpler tests, like $\beta_1 = 3$.

Of course, any test you do should have a rationale behind it. You don't do tests just because they are possible, you do them because there are substantive reasons that motivate them.

Appendix A. Option III: T-Test

[NOTE: In the interests of time, I may skip over this in class, but you should read it on your own.] This approach has the advantage of letting the computer compute the test statistic for you. However, it is a little harder to interpret the results; with more complicated hypotheses it can be difficult or impossible to set up; and again it is not the optimal strategy for testing similar hypotheses when using MLE techniques.

With this approach, as an alternative to running two regressions, you could regress Y on X1, SUM12, and all other IVs (but NOT X2). If the effects of X1 and X2 do not significantly differ, the T value for X1 will not be statistically significant. Further, it will be the case that Incremental F = T^2_{X1}. (If you prefer, you can use X2 rather than X1; conclusions will be the same). Logically, if X1 and X2 have equal effects, the coefficient for SUM12 will capture this; adding X1 to the model will provide no additional useful information.

In the current example, we can regress INCOME on JOBED and EDUC. We then get

```
Variable(s) Entered on Step Number
   1..    JOBED
   2..    EDUC


Multiple R           .90347
R Square             .81626
Adjusted R Square    .81552
Standard Error      3.85420

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           2         32798.40320      16399.20160
Residual           497          7382.84680         14.85482

F =    1103.96483       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

EDUC           1.236899     .052038    .548699    23.769  .0000
JOBED           .708221     .034367    .475709    20.607  .0000
(Constant)    -7.382935     .802778               -9.197  .0000
```

Note that the T value for EDUC, 23.769, is highly significant. Further, note that $23.769^2 = 565$, the same value we got for the incremental F test. The fact that the coefficient for EDUC is positive means that EDUC has a stronger effect than JOBED does.

Incidentally, if you prefer, you can regress INCOME on JOBEXP and JOBED. If so, you get

```
Variable(s) Entered on Step Number
   1..    JOBED
   2..    JOBEXP


Multiple R           .90347
R Square             .81626
Adjusted R Square    .81552
Standard Error      3.85420

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           2         32798.40320      16399.20160
Residual           497          7382.84680         14.85482

F =    1103.96483       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

JOBEXP        -1.236899     .052038   -.697701   -23.769  .0000
JOBED          1.945120     .043700   1.306530    44.511  .0000
(Constant)    -7.382935     .802778               -9.197  .0000
```

Note that the T value for JOBEXP is simply the negative of the T value we earlier got for EDUC. The fact that the coefficient for JOBEXP is negative means that JOBEXP has a weaker effect on income than does EDUC.

Here is why this works: In the unconstrained model, you are estimating

$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

Note that this can be rewritten as Y

$= \alpha + \beta_1 X_1 + \beta_1 X_2 + \beta_2 X_2 - \beta_1 X_2 + \varepsilon$	Add and subtract $\beta_1 X_2$
$= \alpha + \beta_1 (X_1 + X_2) + (\beta_2 - \beta_1) X_2 + \varepsilon$	Rearrange terms
$= \alpha + \beta_1 (SUM12) + \beta_2^* X_2 + \varepsilon$	Make appropriate substitutions

Note that

- The final equality is the model for regressing Y on SUM12 and X2. Hence, the original unconstrained model which uses X1 and X2, and the alternative model, which uses Sum12 and X2, are equivalent to each other. If you know one model, you can determine the coefficients for the other.
- In both models, the value of the intercept is the same. In the alternative model, the coefficient for SUM12 corresponds to the coefficient for X1 in the original model. β_2* in the alternative model = β_2 -β_1 from the original model.
- If $\beta_1 = \beta_2$, as the null hypothesis claims, then $\beta_2^* = 0$. Hence, the T value for b_2^* in the alternative model is a test of whether or not the betas are equal.

To confirm: Results from the original unconstrained model were

```
------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta          T  Sig T

EDUC           1.945120     .043700    .862872    44.511  .0000
JOBEXP          .708221     .034367    .399488    20.607  .0000
(Constant)    -7.382935     .802778               -9.197  .0000
```

Results from the alternative specification of the unconstrained model were

```
------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta          T  Sig T

EDUC           1.236899     .052038    .548699    23.769  .0000
JOBED           .708221     .034367    .475709    20.607  .0000
(Constant)    -7.382935     .802778               -9.197  .0000
```

Note that

- The intercepts are the same
- The coefficient for Jobed (Jobexp + Educ) in the alternative model is the same as the coefficient for Jobexp in the original model
- The coefficient for Educ in the alternative model = b_{educ} - b_{jobexp} from the original model

Specifying alternative but equivalent models is often a useful way for testing hypotheses.

Appendix B: Review of Incremental F Tests

[Go over this if you need to; we probably won't go over this in class.] Suppose we wish to test hypotheses concerning a subset of the variables in a model. For example, suppose a model includes 3 demographic variables (X1, X2, and X3) and 2 personality measures (X4 and X5). We may want to determine whether the personality measures actually add anything to the model, i.e. we want to test

H_0: $\beta_4 = \beta_5 = 0$
H_A: β_4 and/or $\beta_5 \neq 0$

One procedure for testing this is as follows.

1. Estimate the model with all 5 IVs included. This is known as the *unconstrained model*. Retrieve the values for SSE and/or R^2 (hereafter referred to as SSE_u and R^2_u.)
2. Estimate the model using only the 3 demographic variables. We refer to this as the *constrained model*, because the coefficients for the excluded variables are, in effect, constrained to be 0. Retrieve the values for SSE and/or R^2 (hereafter referred to as SSE_c and R^2_c).
3. Compute the following:

$$F_{J,N-K-1} = \frac{(SSE_c - SSE_u)/J}{SSE_u/(N-K-1)} = \frac{(SSE_c - SSE_u)*(N-K-1)}{SSE_u * J} = \frac{(R_u^2 - R_c^2)*(N-K-1)}{(1-R_u^2)*J}$$

where J = the number of constraints imposed (in this case, 2) and K = the number of variables in the *unconstrained* model (in this case, 5). Put another way, J = the error d.f. for the constrained model minus the error d.f. for the unconstrained model. This is sometimes known as an incremental F test (because you are testing whether additions to the model produce significant improvements in fit.)

We have previously discussed two special cases of the above:

- All betas are hypothesized to equal zero. In that case, SSE_c = SST (Error sums of squares and total sums of squares are the same, since none of the variability in Y is explained), SSE_c - SSE_u = SST - SSE_u = SSR_u, J = K (number of constraints = number of variables in the unconstrained model), $R_c^2 = 0$ (no variance is explained). Hence, the incremental F test becomes the same as the global F test for whether or not any betas differ from zero, i.e.

$$F_{K,N-K-1} = \frac{SSR/K}{SSE/(N-K-1)} = \frac{MSR}{MSE} = \frac{R^2*(N-K-1)}{(1-R^2)*K}$$

- The other special case is when J = 1, i.e. we are testing whether one particular coefficient = 0. In this case, the incremental F value produced by the procedure above will equal T^2. Further, in a bivariate regression, incremental F = global F = T^2.