The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–3

Stata tip 108: On adding and constraining

Maarten L. Buis
Department of Sociology
Tübingen University
Tübingen, Germany
maarten.buis@uni-tuebingen.de

In many estimation commands one can use [R] **constraint** to impose linear constraints. The most common of these is the constraint that two or more regression coefficients are equal. A sometimes useful characteristic of models with that constraint is that they are equivalent to a model that includes the sum of the variables that are constrained. Consider the relevant part of a regression equation:

$$\beta_1 x_1 + \beta_2 x_2$$

If we constrain the effects of x_1 and x_2 to be equal than we can replace β_1 and β_2 with β:

$$\beta x_1 + \beta x_2 = \beta(x_1 + x_2)$$

One situation where this characteristic can be useful occurs when one has created a variable by adding several variables and one wonders whether that was a good idea. In the example below there are three variables on the degree of trust a respondent has in the executive, legislative, and judicial branch of the US federal government, `confed`, `conlegis`, and `conjudge` respectively. These can take the values 0 (hardly any trust), 1 (only some trust), or 2 (a great deal of trust). I think that these three variables say something about the trust in the federal government and I created a single variable that captures that, `congov`, which I use to predict whether or not a respondent voted for Barack Obama in the 2008 US presidential election. This results in model `sum1`. If I want to check whether adding these three confidence measures was a good idea, I can use the fact that adding variables is equivalent to constraining their effects to be equal. So one can operationalize the rather vague idea 'adding these variables is a good idea' to the testable statement 'the effects of these three variables are the same'. As a check I first estimated a model that constrains the effects to be equal. This is model `constr1`, and as expected the resulting coefficients, standard errors, log-likelihood are all exactly the same. I than estimate a model with the three confidence variables without constraint, `unconstr1`. The resulting coefficients are very different from one another: the effects do not even have the same sign[1]. A likelihood ratio test also rejects the hypothesis that these variables have the same effect on voting for Obama. So adding the sum of the three confidence measures was not a good idea in this case.

1. These are odds ratios, so the sign is determined by whether the ratio is larger or smaller than 1.

 st0001

```
. use gss10.dta, clear
(extract from the 2010 General Social Survey)

.
. gen byte congov = confed + conlegis + conjudge
(463 missing values generated)

.
. qui logit obama congov, or nolog

. est store sum1

.
. constraint 1 confed = conlegis

. constraint 2 confed = conjudge

.
. qui logit obama confed conlegis conjudge, or constraint(1 2) nolog

. est store constr1

.
. qui logit obama confed conlegis conjudge, or nolog

. est store unconstr1

.
. est tab sum1 constr1 unconstr1, stats(ll N) eform ///
>         b(%9.3g) se(%9.3g) stfmt(%9.4g)

----------------------------------------------------
    Variable |    sum1       constr1    unconstr1   
-------------+--------------------------------------
      congov |      1.62                            
             |       .11                            
      confed |                  1.62         3.47   
             |                   .11         .576   
    conlegis |                  1.62         1.69   
             |                   .11         .305   
    conjudge |                  1.62         .674   
             |                   .11         .107   
       _cons |      .461        .461         .689   
             |     .0833       .0833         .134   
-------------+--------------------------------------
          ll |    -347.8      -347.8       -324.9   
           N |       557         557          557   
----------------------------------------------------
                                        legend: b/se

.
. lrtest unconstr1 constr1

Likelihood-ratio test                                 LR chi2(2)  =     45.77
(Assumption: constr1 nested in unconstr1)             Prob > chi2 =    0.0000
```

Another situation where this characteristic can be useful occurs when one has two or more ordinal or categorical variables that one wants to combine. Consider the example below. In that example I want to treat education as a ordinal variable, and I want to see the effect of 'family educational background' on the educational attainment of the children. I think of 'family educational background' as some sort of sum of the father's and mother's education, but how do I create a sum of two ordinal variables? That is hard, but it is easy to consider the equivalent model that constrains the effects of father's education to be equal to the effects of mother's education. In this example the effects of mother's and father's education are fairly similar, and the test of the hypothesis that they are equal cannot be rejected (compare `unconstr2` with `constr2`). It also shows that constraining effects to be the same is equivalent to adding the sums of the indicator variables (compare `constr2` with `sum2`).

```
. qui ologit degree i.madeg i.padeg, or nolog

. est store unconstr2

.
. constraint 1 1.madeg = 1.padeg

. constraint 2 2.madeg = 2.padeg

.
. qui ologit degree i.madeg i.padeg, or constraint(1 2) nolog

. est store constr2

.
. gen byte p_hs   = 1.madeg + 1.padeg
(329 missing values generated)

. gen byte p_mths = 2.madeg + 2.padeg
(329 missing values generated)

.
. qui ologit degree p_hs p_mths , or nolog

. est store sum2

.
. est tab unconstr2 constr2 sum2, stats(ll N) eform ///
>        b(%9.3g) se(%9.3g) stfmt(%9.4g) keep(degree:)

------------------------------------------------
    Variable | unconstr2    constr2      sum2
-------------+----------------------------------
       madeg |
           1 |       2.5       2.17
             |      .449       .199
           2 |      4.89       5.51
             |      1.26       .691
             |
       padeg |
           1 |      1.88       2.17
             |      .322       .199
           2 |      6.08       5.51
             |      1.46       .691
             |
        p_hs |                            2.17
             |                            .199
      p_mths |                            5.51
             |                            .691
-------------+----------------------------------
          ll |    -799.1     -800.5     -800.5
           N |       972        972        972
------------------------------------------------
                                    legend: b/se

.
. lrtest unconstr2 constr2

Likelihood-ratio test                                 LR chi2(2)  =      2.79
(Assumption: constr2 nested in unconstr2)             Prob > chi2 =    0.2484
```