The Stata Journal (2011) 11, Number 2, pp. 1–3

Stata tip 97: Getting at ρs and σs

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

There are a number of models in Stata that estimate additional coefficients besides regular 'regression-like' coefficients. Often these have an interpretation as a standard deviation or a correlation of either error terms or random coefficients. Examples are [R] **xtmixed** or [R] **heckman**. Sometimes we want access to these coefficients to perform a test or impose a constraint. However, getting access to these parameters is not always straightforward. The problem is that Stata programs often do not estimate these coefficients directly but a transformed version of those parameters. In this Stata tip I will illustrate how to recover these parameters, test hypotheses, and impose constraints.

The reason for estimating the transformed parameters rather than the parameters directly is that that not all positive and negative numbers represent valid values for correlations or standard deviations, which can make it more difficult to estimate the model. The transformation is a way to work around this problem, as this transformation is chosen such that all numbers represent valid values for this parameter. For example, a standard deviation can only be larger than or equal to 0. If we model the logarithm of the standard deviation instead, then all positive and negative numbers represent valid values. Similarly, correlations are often transformed using a Fisher's z transformation (for a discussion of this transformation see: Cox 2008b). This transformation is represented by $z = \frac{1}{2}(\ln(1+\rho) - \ln(1-\rho))$. In Stata one would compute this as $z =$ `atanh(`ρ`)`. The inverses of these transformation are implemented in Stata as the `exp()` and `tanh()` functions ([D] **function**)[1].

The example below will use the `heckman` command as this will return both standard deviations and correlations on a transformed scale. Other commands where these strategies can be applied are for example [R] **heckprob**, [R] **treatreg**, [R] **intreg**, [XT] **xtmixed**, [XT] **xtlogit**[2], [XT] **xtprobit**[2], [XT] **xtmelogit**, [XT] **xtmepoisson**. Moreover, there are some programs that estimate additional parameters on a transformed scale that are not standard deviations or correlations but where this same strategy can also be applied, for example: [R] **nbreg** or [ST] **streg**.

If we want to recover the values of the parameters, we need to know how Stata is calling them. To find that out we can add the `coeflegend` option to our estimation command. We can than see that neither `rho` nor `sigma` have a legend attached to it. However, there are parameters called `athrho` and `lnsigma` that do have legends attached to it, and we can read in the manual (and deduce from their names) that these are the Fisher's z transformed correlation and the natural logarithm of the standard

1. The Mata equivalents of these functions have the same names and are documented in [M-5] **exp()** (for `exp()` and `ln()`) and [M-5] **sin()** (for `atanh()` and `tanh()`).
2. However, note that the `rho` in the output of these commands does not refer to a correlation.

 st0001

deviation. So we can recover these parameters as follows:

```
. webuse womenwk, clear
. heckman wage educ, select(married children educ) nolog coeflegend
Heckman selection model                         Number of obs      =      2000
(regression model with sample selection)        Censored obs       =       657
                                                Uncensored obs     =      1343
                                                Wald chi2(1)       =    403.39
Log likelihood = -5250.348                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        wage |      Coef.  Legend
-------------+----------------------------------------------------------------
wage         |
   education |   1.099506  _b[wage:education]
       _cons |   7.042147  _b[wage:_cons]
-------------+----------------------------------------------------------------
select       |
     married |   .5420304  _b[select:married]
    children |   .4409418  _b[select:children]
   education |   .0722993  _b[select:education]
       _cons |  -1.473038  _b[select:_cons]
-------------+----------------------------------------------------------------
     /athrho |   .8081049  _b[athrho:_cons]
    /lnsigma |   1.807547  _b[lnsigma:_cons]
-------------+----------------------------------------------------------------
         rho |   .6685435
       sigma |   6.095479
      lambda |   4.075093
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =    47.02   Prob > chi2 = 0.0000
------------------------------------------------------------------------------
. // standard deviation of the residual of the wage equation
. di exp([lnsigma]_b[_cons])
6.0954785
. //correlation between residuals of the wage and selection equation
. di tanh([athrho]_b[_cons])
.6685435
```

If we have specific hypotheses then one way of testing these is to rephrase these hypotheses in terms of the transformed metric. Here I compute the transformed standard deviations and correlations on the fly by using the trick to enter the computations as `` `= *exp* ' `` (Cox 2008a). This way *exp* is immediately evaluated, and `test` only sees the number that is the result of that computation. Notice that the transformations are not defined for standard deviations of 0 or correlations of -1 or 1. This is another way in which we can see that one needs to be careful when testing hypotheses on ‘the boundary of the parameter space’ (for example Gutierrez et al. 2001).

```
. test ( [lnsigma]_b[_cons]   = `= ln(6)'     ) ///
>      ( [athrho]_b[_cons]    = `= atanh(.7)' )
 ( 1)  [lnsigma]_cons = 1.791759
 ( 2)  [athrho]_cons = .8673005
           chi2(  2) =    2.29
         Prob > chi2 =    0.3177
```

Using similar tricks, we can also impose constraints on these transformed auxiliary parameters.

```
. constraint 1 [lnsigma]_cons = `= ln(6)´

. constraint 2 [athrho]_cons  = `= atanh(.7)´

. heckman wage educ,                   ///
>         select(married children educ) ///
>         constraint(1 2) nolog

Heckman selection model                         Number of obs      =      2000
(regression model with sample selection)        Censored obs       =       657
                                                Uncensored obs     =      1343

                                                Wald chi2(1)       =    465.26
Log likelihood = -5251.522                      Prob > chi2        =    0.0000

 ( 1)  [lnsigma]_cons = 1.791759
 ( 2)  [athrho]_cons = .8673005

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
wage         |
   education |   1.104284   .0511957    21.57   0.000     1.003943    1.204626
       _cons |   6.911303   .7037765     9.82   0.000     5.531927     8.29068
-------------+----------------------------------------------------------------
select       |
     married |   .5405253   .0639579     8.45   0.000     .4151702    .6658805
    children |   .4405534   .0258773    17.02   0.000     .3898347     .491272
   education |   .0725339    .010469     6.93   0.000     .0520151    .0930527
       _cons |  -1.467094   .1436783   -10.21   0.000    -1.748699    -1.18549
-------------+----------------------------------------------------------------
     /athrho |   .8673005          .        .       .            .           .
    /lnsigma |   1.791759          .        .       .            .           .
-------------+----------------------------------------------------------------
         rho |         .7          .                            -1           1
       sigma |          6          .                             .           .
      lambda |        4.2          .                             .           .
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =    44.67   Prob > chi2 = 0.0000
------------------------------------------------------------------------------
```

References

Cox, N. J. 2008a. Stata tip 59: Plotting on any transformed scale. *The Stata Journal* 8: 142–145.

———. 2008b. Speaking Stata: Correlation with confidence, or Fisher's z revisited. *The Stata Journal* 8(3): 413–439.

Gutierrez, R., S. Carter, and D. M. Drukker. 2001. On boundary-value likelihood-ratio tests. *Stata Technical Bulletin* 60: 15–18.