



*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

*No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.*

## Stata tip 97: Getting at $\rho$ 's and $\sigma$ 's

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

There are several models in Stata that estimate coefficients other than 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 `xtmixed` or `heckman` (see [XT] `xtmixed` or [R] `heckman`). Sometimes, we want access to these coefficients to perform a test or to 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 instead estimate 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 estimating the parameters directly is that not all positive and negative numbers represent valid values for correlations or standard deviations, which can make it more difficult to fit the model. The transformation is a way to work around this problem because 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 [2008a]). This transformation is represented by  $z = 1/2\{\ln(1 + \rho) - \ln(1 - \rho)\}$ . In Stata, you would compute this as  $z = \text{atanh}(\rho)$ . The inverses of these transformations are implemented in Stata as the `exp()` and `tanh()` functions (see [D] **functions**).<sup>1</sup>

The example below will use the `heckman` command because `heckman` will return both standard deviations and correlations on a transformed scale. Other commands where these strategies can be applied are, for example, `heckprob`, `treatreg`, `intreg`, `xtmixed`, `xtlogit`,<sup>2</sup> `xtprobit`,<sup>2</sup> `xtmelogit`, and `xtmepoisson`. Moreover, there are some programs that estimate additional parameters on a transformed scale that are not standard deviations or correlations but to which this same strategy can also be applied, for example, `nbreg` or `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 then see that neither `rho` nor `sigma` has a legend attached to it. However, there are parameters called `athrho` and `lnsigma` that do have legends attached to them, and we can read in the manual (and deduce from their names) that these are the

---

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, the `rho` in the output of these commands does not refer to a correlation.

Fisher's  $z$  transformed correlation and the natural logarithm of the standard deviation, respectively. So we can recover these parameters as follows:

```
. http://www.stata-press.com/data/r11/womenwk
. 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 2008b). 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, see Gutierrez, Carter, and Drukker [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. Speaking Stata: Correlation with confidence, or Fisher's z revisited. *Stata Journal* 8: 413–439.
- . 2008b. Stata tip 59: Plotting on any transformed scale. *Stata Journal* 8: 142–145.
- Gutierrez, R. G., S. Carter, and D. M. Drukker. 2001. sg160: On boundary-value likelihood-ratio tests. *Stata Technical Bulletin* 60: 15–18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 269–273. College Station, TX: Stata Press.