Measurement Error Example (Supplemental)

Here we give a hypothetical example that illustrates the properties shown in the measurement handout. We create a data set where the true measures (Yt and Xt) have a correlation of .7 with each other – but the observed measures (Y and X) both have some degree of random measurement error, and the reliability of both is .64. The way I am constructing the data set, using the `corr2data` command, there will be no sampling variability, i.e. we can act as though we have the entire population.

```
. matrix input corr = (1,.7,0,0\.7 ,1,0,0\0,0,1,0\0,0,0,1)

. matrix input sd = (4,8,3,6)

. matrix input mean = (10,7,0,0)

. corr2data Yt Xt ey ex, corr(corr) sd(sd) mean(mean) n(500)

(obs 500)

. * Create flawed measures with random measurement error
. gen Y = Yt + ey

. gen X = Xt + ex
```

A & B. We see that the flawed, observed measures have the same means as the true measures – but their variances & standard deviations are larger:

```
. sum  Yt Y Xt X

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          Yt |       500          10           4  -2.639851   22.83863
           Y |       500          10           5  -3.706503   26.55569
          Xt |       500           7           8  -16.16331   28.80884
           X |       500           7          10  -23.81675   38.49127
```

C. Random measurement error in X, however, does NOT affect the covariance:

```
. corr Yt Xt X, cov

(obs=500)

             |       Yt       Xt        X
-------------+---------------------------
          Yt |       16
          Xt |     22.4       64
           X |     22.4       64      100
```

D. Nonetheless, the correlation is attenuated by random measurement error:

```
. corr Yt Xt X
(obs=500)

             |       Yt       Xt        X
-------------+---------------------------
          Yt |   1.0000
          Xt |   0.7000   1.0000
           X |   0.5600   0.8000   1.0000
```

Note that the correlation between X and Xt is .8 – and that the correlation between X and Yt (.56) is only .8 times as large as the correlation between Xt and Yt (.7). Also, the .8 correlation between X and Xt means that the reliability of X is .64.

E. In a bivariate regression, random measurement error in X causes the slope coefficient to be attenuated, i.e. smaller in magnitude. First we run the regression between the true measures, and then we run the regression of Yt with the flawed measure X:

```
. reg Yt Xt

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =  478.47
       Model |  3912.16007     1  3912.16007           Prob > F      =  0.0000
    Residual |  4071.84001   498  8.17638555           R-squared     =  0.4900
-------------+------------------------------           Adj R-squared =  0.4890
       Total |  7984.00008   499  16.0000002           Root MSE      =  2.8594

------------------------------------------------------------------------------
          Yt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Xt |        .35   .0160008    21.87   0.000     .3185627    .3814373
       _cons |       7.55    .169994    44.41   0.000     7.216006    7.883994
------------------------------------------------------------------------------

. reg Yt X

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =  227.52
       Model |  2503.78247     1  2503.78247           Prob > F      =  0.0000
    Residual |  5480.21761   498   11.004453           R-squared     =  0.3136
-------------+------------------------------           Adj R-squared =  0.3122
       Total |  7984.00008   499  16.0000002           Root MSE      =  3.3173

------------------------------------------------------------------------------
          Yt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           X |       .224   .0148503    15.08   0.000     .1948231    .2531769
       _cons |      8.432   .1811488    46.55   0.000      8.07609     8.78791
------------------------------------------------------------------------------
```

Note that X has a reliability of .64 – and the slope coefficient using the flawed X (.224) is only .64 times as large as the slope coefficient using the perfectly measured Xt (.35).

F. Random measurement error in Y, however, does not cause the slope coefficient to be biased – but it does cause the standard error for the slope coefficient to be larger and the t value smaller. Again we run the true regression followed by the regression of Y with Xt.

```
. reg Yt Xt

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =  478.47
       Model |  3912.16007     1  3912.16007           Prob > F      =  0.0000
    Residual |  4071.84001   498  8.17638555           R-squared     =  0.4900
-------------+------------------------------           Adj R-squared =  0.4890
       Total |  7984.00008   499  16.0000002           Root MSE      =  2.8594

------------------------------------------------------------------------------
          Yt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Xt |        .35   .0160008    21.87   0.000     .3185627    .3814373
       _cons |       7.55    .169994    44.41   0.000     7.216006    7.883994
------------------------------------------------------------------------------

. reg Y Xt

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  1,   498) =  227.52
       Model |  3912.16001     1  3912.16001           Prob > F      =  0.0000
    Residual |  8562.84011   498   17.194458           R-squared     =  0.3136
-------------+------------------------------           Adj R-squared =  0.3122
       Total |  12475.0001   499  25.0000002           Root MSE      =  4.1466

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Xt |        .35   .0232035    15.08   0.000     .3044111    .3955889
       _cons |       7.55   .2465171    30.63   0.000     7.065658    8.034342
------------------------------------------------------------------------------
```