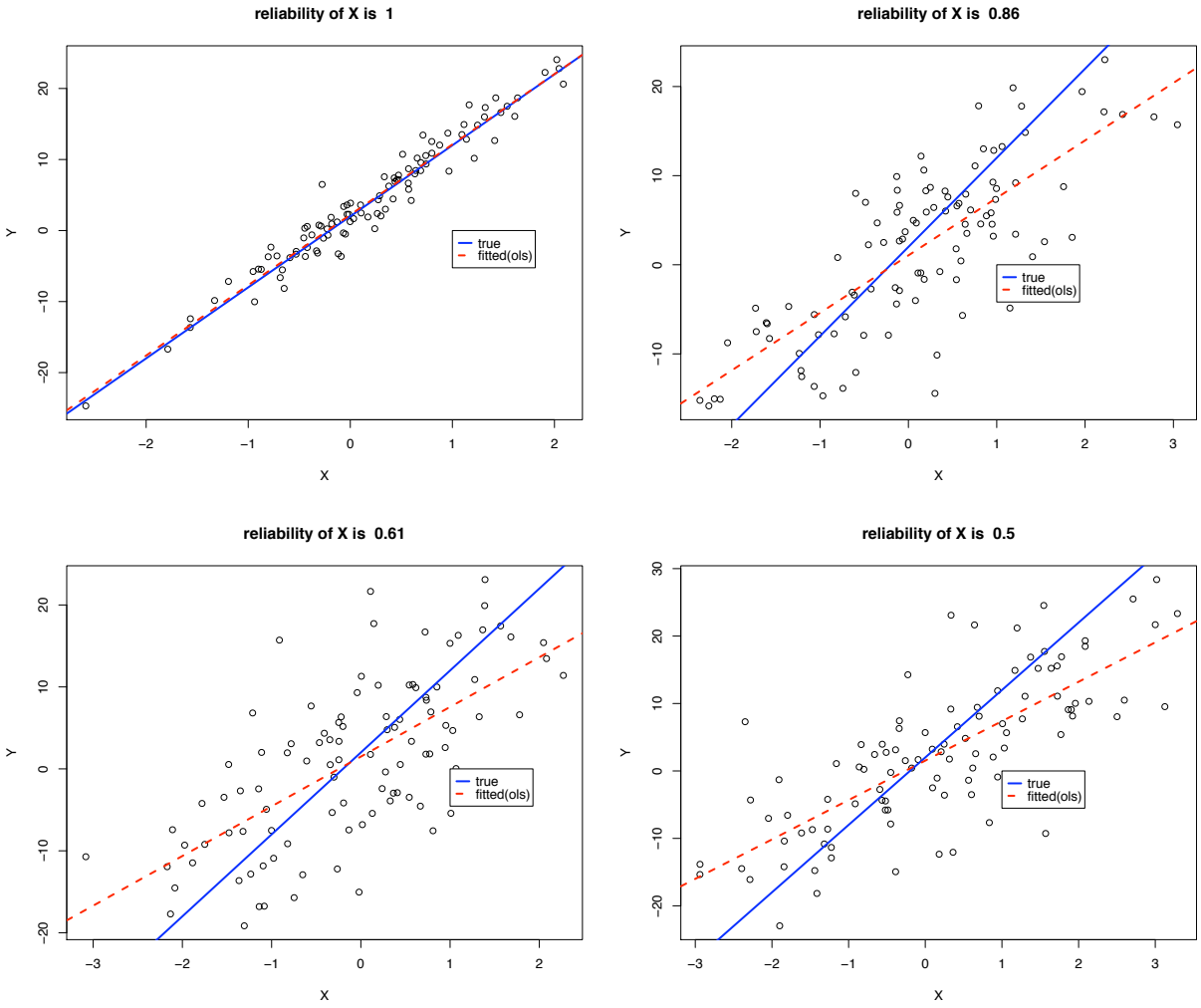
As examples of application of measurement error models prenatal exposure Consider the following illustration with simulated data: The model is

```
Y = 2 + 10x + e
X = x +  u
```

In the measurement error model X = x + u, with x uncorrelated with u, var(X) = var(x) + var(u). The reliabiliy coefficient is defined as rel = var(x)/var(X) = 1 - var(u)/var(X); i.e., it is a number between 0 and 1. In the simulation we consider sample size equal to 10, and different values for the reliability of the regressor.



where we see that there is an increasing attenuation with the decrease of the reliability.

Errors in variables regression

This is extracted from
ucla web One of the assumptions of regression is that the predictor variables are measured without error. The problem is that measurement error in predictor variables in OLS regression leads to under estimation of the regression coefficients. Errors-in-variables regression models are useful when one or more of the independent variables are measured with error. One can adjust for the biases if one knows the reliability of the variable,

```
r = 1 - (variance of measurement error)/(total variance)

The model we wish to estimate is

y = X* beta  + e

where X* are the true values and

X = X* + U
```

the X are the observed values. The estimates b of b are obtained by

b = A^-1 X'y, where

A = X'X - S

S is a diagonal matrix with elements N(1-ri)si2, where the ri are the reliability coefficients.

Stata's eivreg command uses user-specified relibility coefficents to compute the S matrix which, in turn, takes measurement error into account when

Stata's eivreg command

```
  use http://www.ats.ucla.edu/stat/stata/webbooks/reg/hsb2
(highschool and beyond (200 cases))

. describe

Contains data from http://www.ats.ucla.edu/stat/stata/webbooks/reg/hsb2.dta
  obs:           200                          highschool and beyond (200
                                                cases)
 vars:            11                          20 Jun 2000 14:13
 size:         9,600 (99.1% of memory free)
------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
id              float  %9.0g
female          float  %9.0g       fl
race            float  %12.0g      rl
ses             float  %9.0g       sl
schtyp          float  %9.0g       scl        type of school
prog            float  %9.0g       sel        type of program
read            float  %9.0g                  reading score
write           float  %9.0g                  writing score
math            float  %9.0g                  math score
science         float  %9.0g                  science score
socst           float  %9.0g                  social studies score
------------------------------------------------------------------------------


use http://www.ats.ucla.edu/stat/stata/webbooks/reg/hsb2

regress write read female

  Source |       SS       df       MS                  Number of obs =     200
---------+------------------------------               F(  2,   197) =   77.21
   Model |  7856.32118     2  3928.16059               Prob > F      =  0.0000
Residual |  10022.5538   197  50.8759077               R-squared     =  0.4394
---------+------------------------------               Adj R-squared =  0.4337
   Total |   17878.875   199   89.843593               Root MSE      =  7.1327

------------------------------------------------------------------------------
   write |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    read |   .5658869   .0493849     11.459   0.000        .468496    .6632778
  female |   5.486894   1.014261      5.410   0.000        3.48669    7.487098
   _cons |   20.22837   2.713756      7.454   0.000       14.87663    25.58011
------------------------------------------------------------------------------
The predictor read is a standardized test score. Every test has measurement error. We don't know the exact reliability of read, but using .9 for the

eivreg write read female, r(read .9)

               assumed                          errors-in-variables regression
variable     reliability
------------------------                        Number of obs =     200
    read       0.9000                           F(  2,   197) =   83.41
       *       1.0000                           Prob > F      =  0.0000
                                                R-squared     =  0.4811
                                                Root MSE      = 6.86268

------------------------------------------------------------------------------
   write |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    read |   .6289607   .0528111     11.910   0.000        .524813    .7331085
  female |   5.555659   .9761838      5.691   0.000       3.630548     7.48077
   _cons |   16.89655   2.880972      5.865   0.000       11.21504    22.57805
Note that the F-ratio and the R2 increased along with the regression coefficient for read. Additionally, there is an increase in the standard error

Now, let's try a model with read, math and socst as predictors. First, we will run a standard OLS regression.

regress write read math socst female

  Source |       SS       df       MS                  Number of obs =     200
---------+------------------------------               F(  4,   195) =   64.37
   Model |  10173.7036     4  2543.42591               Prob > F      =  0.0000
Residual |  7705.17137   195  39.5136993               R-squared     =  0.5690
```

```
---------+------------------------------               Adj R-squared =  0.5602
   Total |   17878.875   199   89.843593               Root MSE      =   6.286

------------------------------------------------------------------------------
       write |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
        read |   .2065341   .0640006      3.227   0.001       .0803118    .3327563
        math |   .3322639   .0651838      5.097   0.000       .2037082    .4608195
       socst |   .2413236   .0547259      4.410   0.000        .133393    .3492542
      female |   5.006263   .8993625      5.566   0.000       3.232537     6.77999
       _cons |   9.120717   2.808367      3.248   0.001       3.582045    14.65939
------------------------------------------------------------------------------
```

Now, let's try to account for the measurement error by using the following reliabilities: read - .9, math - .9, socst - .8.

```
eivreg write read math socst female, r(read .9 math .9 socst .8)

              assumed                         errors-in-variables regression
variable     reliability
------------------------                      Number of obs =     200
        read       0.9000                     F(  4,   195) =   70.17
        math       0.9000                     Prob > F      =  0.0000
       socst       0.8000                     R-squared     =  0.6047
           *       1.0000                     Root MSE      = 6.02062

------------------------------------------------------------------------------
       write |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
        read |   .1506668   .0936571      1.609   0.109      -.0340441    .3353776
        math |    .350551   .0850704      4.121   0.000       .1827747    .5183273
       socst |   .3327103   .0876869      3.794   0.000        .159774    .5056467
      female |   4.852501   .8730646      5.558   0.000        3.13064    6.574363
       _cons |    6.37062   2.868021      2.221   0.027       .7142973    12.02694
------------------------------------------------------------------------------
```

Note that the overall F and R2 went up, but that the coefficient for read is no longer statistically significant.

Last modified: Wed Feb 7 16:24:16 CET 2007