Measrement_Error_-Apr8.R

danny

2020-04-08

```r
# Author: Gordon Burtch and Gautam Ray
# Course: MSBA 6440
# Session: Selection and Measurement Error
# Topic: Measurement Error

suppressWarnings(suppressPackageStartupMessages({
library(MASS)
library(AER)
library(stargazer)
}))


# X and Y have classical measurement error. The true value are Xt and Yt, but they are meas-
# ured with error.
# X is measured as true X (Xt) plus error (ex)
# Y is measured as true Y (Yt) plus error (ey)
# The mean of Yt, Xt, ey and ex is (10, 7, 0, 0)
# The standard deviation of Yt, Xt, ey, and ex is (4, 8, 3, 6)
# The correlation of Yt and Xt is 0.7; Yt and Xt are uncorrelated with ey and ex; and ey and
# ex are uncorrelated with each other.

set.seed(1234)


Yt_Xt_ey_ex <- (mvrnorm(10000, c(10, 7, 0, 0), matrix(c(16, 22.4, 0.0, 0.0, 22.4, 64,0, 0, 0,
                                                         0, 9,0, 0, 0, 0, 36), ncol = 4)))
Yt <- Yt_Xt_ey_ex[,1]
Xt <- Yt_Xt_ey_ex[,2]
ey <- Yt_Xt_ey_ex[,3]
ex <- Yt_Xt_ey_ex[,4]

Y <- Yt + ey
X <- Xt + ex


# Check everything worked as expected.

cov(Yt_Xt_ey_ex)
```

```
##              [,1]       [,2]       [,3]       [,4]
## [1,] 15.8500152 21.7983191 -0.1185045  0.3525733
## [2,] 21.7983191 62.4120430 -0.3538506  0.4489615
## [3,] -0.1185045 -0.3538506  8.9843536 -0.0209746
## [4,]  0.3525733  0.4489615 -0.0209746 36.4322839
```

```r
cor(Yt_Xt_ey_ex)
```

```
##              [,1]         [,2]         [,3]         [,4]
## [1,]  1.000000000  0.693064996 -0.009930629  0.014672070
## [2,]  0.693064996  1.000000000 -0.014943156  0.009415246
## [3,] -0.009930629 -0.014943156  1.000000000 -0.001159330
## [4,]  0.014672070  0.009415246 -0.001159330  1.000000000
```

```r
sd(ey)
```

```
## [1] 2.997391
```

```r
sd(ex)
```

```
## [1] 6.035916
```

```r
sd(Yt)
```

```
## [1] 3.981208
```

```r
sd(Xt)
```

```
## [1] 7.900129
```

```r
mean (Yt)
```

```
## [1] 10.04655
```

```r
mean(Xt)
```

```
## [1] 7.037755
```

```r
mean(ey)
```

```
## [1] 0.01225815
```

```r
mean(ex)
```

```
## [1] -0.04648214
```

```r
#1. Measurement error in X, underestimates the effect of X on Y. The reliability of mismea-
# surement is the magnitude of the mismeasurement.

summary(lm(Yt~Xt))
```

```
##
## Call:
## lm(formula = Yt ~ Xt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.0613  -1.9560   0.0189   1.9343  10.9935
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.588506   0.038439  197.42   <2e-16 ***
## Xt          0.349265   0.003633   96.13   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.87 on 9998 degrees of freedom
## Multiple R-squared:  0.4803, Adjusted R-squared:  0.4803
## F-statistic:  9241 on 1 and 9998 DF,  p-value: < 2.2e-16
```

```
summary(lm(Yt~X))
```

```
##
## Call:
## lm(formula = Yt ~ X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.0113  -2.2426  -0.0154   2.2387  13.9810
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.493914   0.040360  210.45   <2e-16 ***
## X           0.222081   0.003311   67.08   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.306 on 9998 degrees of freedom
## Multiple R-squared:  0.3104, Adjusted R-squared:  0.3103
## F-statistic:  4500 on 1 and 9998 DF,  p-value: < 2.2e-16
```

```
Reliability <- var(Xt)/var(X)

Reliability
```

```
## [1] 0.6257333
```

```
summary(lm(Yt~X))$coefficients[2,1]/ summary(lm(Yt~Xt))$coefficients[2,1]
```

```
## [1] 0.6358541
```

```r
#2. Measurement error in Y, does not influence the coefficient of X, but exaggerartes the
# standard error of the regression coefficient.

summary(lm(Yt~Xt))
```

```
##
## Call:
## lm(formula = Yt ~ Xt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.0613  -1.9560   0.0189   1.9343  10.9935
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.588506   0.038439  197.42   <2e-16 ***
## Xt          0.349265   0.003633   96.13   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.87 on 9998 degrees of freedom
## Multiple R-squared:  0.4803, Adjusted R-squared:  0.4803
## F-statistic:  9241 on 1 and 9998 DF,  p-value: < 2.2e-16
```

```r
summary(lm(Y~Xt))
```

```
##
## Call:
## lm(formula = Y ~ Xt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.4804  -2.7805   0.0196   2.7619  16.3189
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.640666   0.055594  137.44   <2e-16 ***
## Xt          0.343595   0.005255   65.39   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.151 on 9998 degrees of freedom
## Multiple R-squared:  0.2996, Adjusted R-squared:  0.2995
## F-statistic:  4276 on 1 and 9998 DF,  p-value: < 2.2e-16
```

```r
#3. Given that the measurement error in Y is more innocuous than the measurement error in
# X, we might run the reverse regression.
# The coefficient of the regular regression and the inverse of the coefficient of the rev-
# erse regression, bracket the true coefficient.

regular_reg <- (lm(Yt~X))
b <- summary(regular_reg)$coefficients[2,1]
```

```r
reverse_reg <- (lm(X~Yt))
reverse_reg_coeff <- summary(reverse_reg)$coefficients[2,1]

g <- 1/(summary(reverse_reg)$coefficients[2,1])

b/g
```

```
## [1] 0.3103656
```

```r
summary((lm(Yt~X)))
```

```
##
## Call:
## lm(formula = Yt ~ X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.0113  -2.2426  -0.0154   2.2387  13.9810
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.493914   0.040360  210.45   <2e-16 ***
## X           0.222081   0.003311   67.08   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.306 on 9998 degrees of freedom
## Multiple R-squared:  0.3104, Adjusted R-squared:  0.3103
## F-statistic:  4500 on 1 and 9998 DF,  p-value: < 2.2e-16
```

```r
#4. If there is a good instrument for Xt, then the true estimate can be recovered.
# Lets say that Z is a good instrument for Xt. Like Xt, Z has a mean of 7 and a sd of 8. Z
# has a correlation of 0.5 with Xt, and Z is uncorrelated with ey and ex.
# Z has a correlation of 0.35 with Yt which is the product of 0.7 and 0.5 i.e, the correl-
# ation between Yt and Xt and the correlation between Xt and Z.


Yt_Xt_ey_ex_Z <- (mvrnorm(10000, c(10, 7, 0, 0, 7), matrix(c(16, 22.4, 0.0, 0.0, 11.2, 22.4,
                    64,0, 0, 32, 0, 0, 9,0, 0, 0, 0, 0, 36, 0, 11.2, 32, 0, 0, 64), ncol = 5)))

Yt <- Yt_Xt_ey_ex_Z[,1]
Xt <- Yt_Xt_ey_ex_Z[,2]
ey <- Yt_Xt_ey_ex_Z[,3]
ex <- Yt_Xt_ey_ex_Z[,4]
Z <- Yt_Xt_ey_ex_Z[,5]

Y <- Yt + ey
X <- Xt + ex

cov(Yt_Xt_ey_ex_Z)
```

```
##             [,1]        [,2]        [,3]        [,4]       [,5]
## [1,] 16.0047135 22.63816128 -0.00282960 -0.37864749 11.2970827
## [2,] 22.6381613 64.48127066  0.02019033 -1.05799144 32.3030631
## [3,] -0.0028296  0.02019033  8.92867464 -0.08615143 -0.1550035
## [4,] -0.3786475 -1.05799144 -0.08615143 35.60204905 -0.9907565
## [5,] 11.2970827 32.30306312 -0.15500346 -0.99075646 64.8426342
```

```r
cor(Xt, Z)
```

```
## [1] 0.4995703
```

```r
cor(X, Z)
```

```
## [1] 0.3928657
```

```r
cor(Yt, Z)
```

```
## [1] 0.3506808
```

```r
cor(Y, Z)
```

```
## [1] 0.277137
```

```r
ols_true <- lm((Yt~Xt))

ivreg1 <- ivreg(formula=Yt ~ X | Z)

ivreg2 <- ivreg(formula=Y ~ X | Z)

stargazer(ols_true,ivreg1, ivreg2,type="text",title="True vs.Instrumented",
          column.labels = c("True","IV1", "IV2"))
```

```
##
## True vs.Instrumented
## =====================================================================================
##                                                 Dependent variable:
##                                     -------------------------------------------------
##                                              Yt ~ Xt              Yt           Y
##                                                OLS          instrumental instrumental
##                                                                variable     variable
##                                               True               IV1          IV2
##                                                (1)               (2)          (3)
## -------------------------------------------------------------------------------------
## Xt                                          0.351***
##                                             (0.004)
##
## X                                                              0.361***     0.356***
##                                                                (0.009)      (0.012)
##
## Constant                                    7.567***           7.516***     7.548***
```

```
##                                             (0.038)             (0.073)      (0.095)
##
## ---------------------------------------------------------------------------------------
## Observations                                10,000              10,000       10,000
## R2                                           0.497               0.207        0.136
## Adjusted R2                                  0.497               0.207        0.136
## Residual Std. Error (df = 9998)              2.839               3.563        4.641
## F Statistic                        9,862.679*** (df = 1; 9998)
## =======================================================================================
## Note:                                                   *p<0.1; **p<0.05; ***p<0.01
```