

# Measurement Error(Monte Carlo Sim)

Kei Sakamoto

Dependent variableの measurement errorの影響と、regressorsの measurement errorの影響を推定する。どちらもerrorが(dependent variableともregressorとも独立なのが重要。独立でないのはIVでしか対処不可能。)

Monte Carlo Simurationのmethodはいつもと同じ。(simple regression)

## ME in Dependent variable

```
set.seed(1234567)
b0<-1; b1<-0.5
b1hat <- numeric(10000)
b1hat.me <- numeric(10000)

# Draw a sample of x, fixed over replications(number of observation = 1000)
x <- rnorm(1000,4,1)
# repeat 10000 times
for(j in 1:10000) {
  u <- rnorm(1000)
  y_star <- b0 + b1*x + u #we wanna truely observe y_star(not y)
  bhat <- coef( lm(y_star~x) )
  b1hat[j] <- bhat["x"]

  # Measurement error and mismeasured y
  e0 <- rnorm(1000) #ここが救い。i.i.d errorでしかもmeanが0。
  y <- y_star + e0
  bhat.me <- coef(lm(y~x))
  b1hat.me[j] <- bhat.me["x"]
}
```

## compare the each Mean

```
c(mean(b1hat),mean(b1hat.me))
```

```
## [1] 0.5003774 0.5001819
```

meanには影響が出ない。 $E(e_0)=0$ だから。

## compare the each Variance

```
c(var(b1hat),var(b1hat.me))
```

```
## [1] 0.0009990556 0.0019991960
```

**variance**は大きくなる(**population**では $\text{cov}(y\_star, e_0)=0$ なはずなので**variance**はこの2つの**variance**の単純な和で表せるので**1+1**ぐらいになる。)ので**efficiency**の**loss**になるだけ。あと**intercept**がずれるだけで**x**は**exogenous**なのは保たれるので**consistency**は保たれるのでそんなに大きな問題ではない。

## ME in Explanatory variable(regressor)

```
set.seed(1234567)
b0<-1; b1<-0.5
b1hat <- numeric(10000)
b1hat.me <- numeric(10000)

# Draw a sample of x, fixed over replications(number of observation = 1000)
x_star <- rnorm(1000,4,1) #we wanna truely observe x_star(not x)
# repeat 10000 times
for(j in 1:10000) {
  u <- rnorm(1000)
  y <- b0 + b1*x + u
  bhat <- coef( lm(y~x_star) )
  b1hat[j] <- bhat["x_star"]

  # Measurement error and mismeasured x
  e1 <- rnorm(1000) #ここが救い。 i.i.d error
  x <- x_star + e1
  bhat.me <- coef(lm(y~x))
  b1hat.me[j] <- bhat.me["x"]
}
```

## compare the each Mean

```
c(mean(b1hat),mean(b1hat.me))
```

```
## [1] 0.5002785 0.2490650
```

今度は**mean**にも影響が出る。これが**attenuation bias**。  
 $\{\text{Var}(x\_star)\}/\{\text{Var}(a\_star)+\text{Var}(e_1)\}$ 分だけ縮む。**consistency**も保たれなくなるのでまずい。たとえ**error**が今回のように**iidNormal(0,1)**でも。

## compare the each Variance

```
c(var(b1hat),var(b1hat.me))
```

```
## [1] 0.0012542225 0.0006640531
```

**β1\_hat**に**attenuaton bias**がかかっているので、**variance**はむしろ小さくなる。