

## [Example8-4,5] Heteroskedasticity Test

Kei Sakamoto

```
load("~/計量経済学演習/R data sets for 5e/hprice1.RData")
hprice1<-data

(reg <- lm(price~lotsize+sqrft+bdrms, data=hprice1))

##
## Call:
## lm(formula = price ~ lotsize + sqrft + bdrms, data = hprice1)
##
## Coefficients:
## (Intercept)      lotsize      sqrft      bdrms
## -21.770308      0.002068      0.122778     13.852522
```

### Example 8-4 Breusch Pagan test (BP test)

$u$  は unobservable なので  $\hat{u}$ (residual)で置き換え、residual の 2 乗は( $E(\hat{u})=0$  から) $Var(\hat{u})$ なので、要するに regressor が  $\hat{u}$ の variance に対して説明力を持つかどうか test すればいい。具体的には auxiliary regression の F-test をする。

### Manual regression of squared residuals(auxiliary regression の summary で勝手に F-test までしてくれる。)

```
summary(lm( resid(reg)^2 ~ lotsize+sqrft+bdrms, data=hprice1))

##
## Call:
## lm(formula = resid(reg)^2 ~ lotsize + sqrft + bdrms, data = hprice1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9044  -2212  -1256    -97   42582
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.523e+03  3.259e+03  -1.694  0.09390 .
## lotsize      2.015e-01  7.101e-02   2.838  0.00569 **
## sqrft        1.691e+00  1.464e+00   1.155  0.25128
## bdrms        1.042e+03  9.964e+02   1.046  0.29877
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6617 on 84 degrees of freedom
## Multiple R-squared:  0.1601, Adjusted R-squared:  0.1301
## F-statistic: 5.339 on 3 and 84 DF,  p-value: 0.002048
```

## Automatic BP test

`library(lmtest)` *#lmtest 割と万能だけど、LM version なので若干結果違う*

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

bptest(reg)

##
## studentized Breusch-Pagan test
##
## data:  reg
## BP = 14.092, df = 3, p-value = 0.002782
```

manual でも automatic でも 1% significant なことは変わらない。つまり regressor が  $\hat{u}$  の variance に対して説明力を持つので heteroskedastic だと言えそう。

## Example8-5 White test

model を log transform してもう一回 BPtest して、今度は  $\hat{u}$  の variance は regressors に対して nonlinear を許している White test と比較する。df(自由度)下がって power(検出力)が下がるのを避けるために y の fitted を regressor として入れて model を定義する。nonlinear terms は y の fitted に含まれていることを期待。(cross product とか馬鹿正直に全部入れると自由度鬼下がってしまう。)

```
(reg <- lm(log(price)~log(lotsize)+log(sqrft)+bdrms, data=hprice1))

##
## Call:
## lm(formula = log(price) ~ log(lotsize) + log(sqrft) + bdrms,
```

```
##      data = hprice1)
##
## Coefficients:
## (Intercept)  log(lotsize)    log(sqrft)      bdrms
##      -1.29704      0.16797      0.70023      0.03696
```

## BP test

```
bptest(reg)
```

```
##
## studentized Breusch-Pagan test
##
## data:  reg
## BP = 4.2232, df = 3, p-value = 0.2383
```

log transform すると説明力下がっちゃう。(x が大きいと log transform が x の良い近似になっていない可能性。)

## White test

```
bptest(reg, ~ fitted(reg) + I(fitted(reg)^2) )
```

```
##
## studentized Breusch-Pagan test
##
## data:  reg
## BP = 3.4473, df = 2, p-value = 0.1784
```

同じく log transform のせいなのかもしれないが nonlinear terms 含めて coef all 0 説は significance level 10%でも reject できない。  
heteroskedastic だとは言い切れなくなってしまう。