

## SCRIPT MOD4S2B: HETEROSKEDASTIC REGRESSION VIA SIMULATION

INSTRUCTOR: KLAUS MOELTNER

### GENERATE DATA

```
R> n<-1000 #sample size
R> x1<-rep(1,n)
R> x2<-rnorm(n,-1.4,1)
R> x3<-rnorm(n,3,2)
R> #
R> bvec<-c(1.2,2,0.8)
R> #
R> X<-cbind(x1,x2,x3)
R> k<-ncol(X)
```

### EXAMPLE 1: NO HETEROSKEDASTICITY

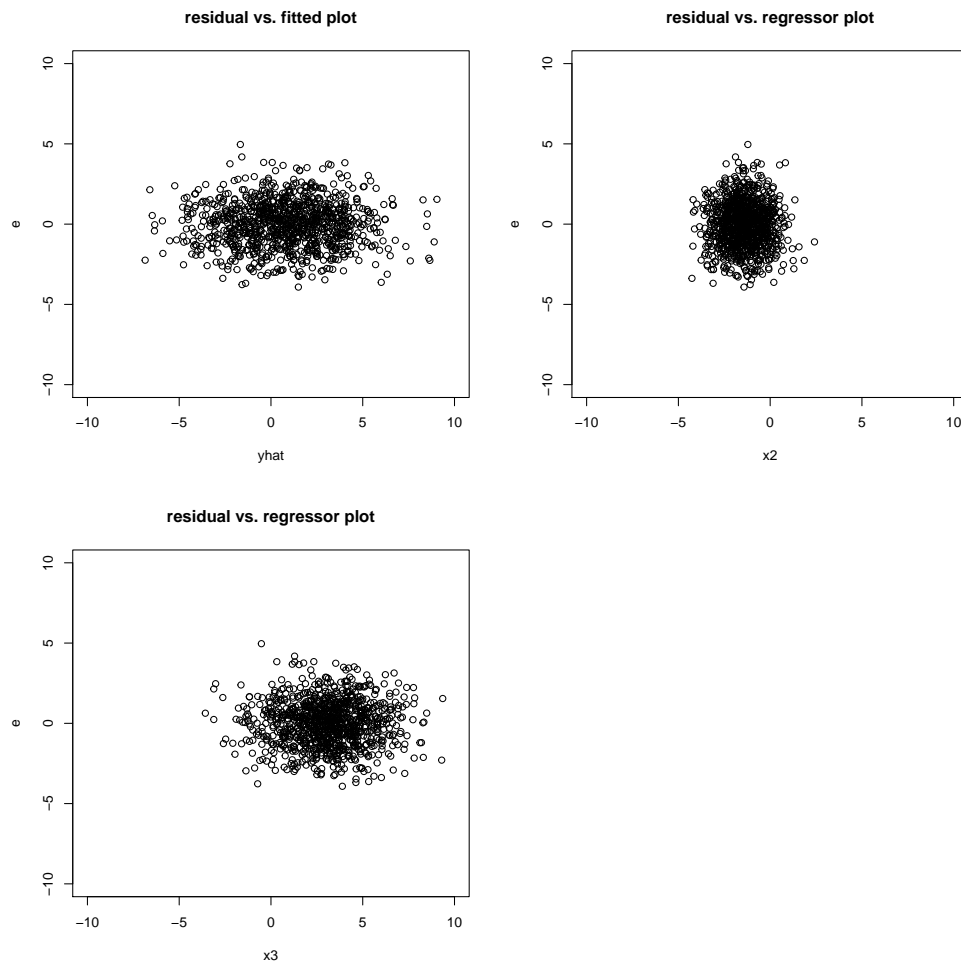
Simulate a well-behaved CLRM, show the output, and create residual plots

```
R> sig2true<-2 #"true" error variance
R> eps<-rnorm(n,0,sqrt(sig2true))
R> y<-X %*% bvec + eps
R> #
R> bols<-solve((t(X)) %*% X) %*% (t(X) %*% y)
R> e<-y-X%*%bols
R> yhat<-X%*%bols
R> SSR<-(t(e)%*%e)
R> s2<-(t(e)%*%e)/(n-k)
R> Vb<-s2[1,1]*solve((t(X))%*%X)
R> se=sqrt(diag(Vb))
R> tval=bols/se
R> #
R> ttols<-data.frame(col1=c("constant","x2","x3"),
                    col2=bols,
                    col3=se,
                    col4=tval)
R> colnames(ttols)<-c("variable","estimate","s.e.", "t")
```

TABLE 1. OLS for Model w/o HSK

variable	estimate	s.e.	t
constant	1.261	0.101	12.538
x2	2.061	0.044	46.701
x3	0.815	0.022	36.936

FIGURE 1. HSK diagnostic plots, no HSK



EXAMPLE 2: HETEROSKEDASTICITY RELATED TO  $E(Y)$

```
R> #sig2true<-2*abs(X %>% bvec) #"true" error variance
R> sig2true<-0.1*(X %>% bvec)^2
R> eps<-rnorm(n,0,sqrt(sig2true))
R> y<-X %>% bvec + eps
R> #
R> bols<-solve((t(X)) %>% X) %>% (t(X) %>% y)
R> e<-y-X%>%bols
R> yhat<-X%>%bols
R> SSR<-(t(e)%>%e)
R> s2<-(t(e)%>%e)/(n-k)
R> Vb<-s2[1,1]*solve((t(X))%>%X)
R> se=sqrt(diag(Vb))
R> tval=bols/se
R> #
R> ttols<-data.frame(col1=c("constant","x2","x3"),
```

```

col2=bols,
col3=se,
col4=tval)
R> colnames(ttols)<-c("variable","estimate","s.e.,"t")

```

TABLE 2. OLS for Model with HSK related to y

variable	estimate	s.e.	t
constant	1.283	0.065	19.686
x2	2.007	0.029	70.183
x3	0.786	0.014	55.030

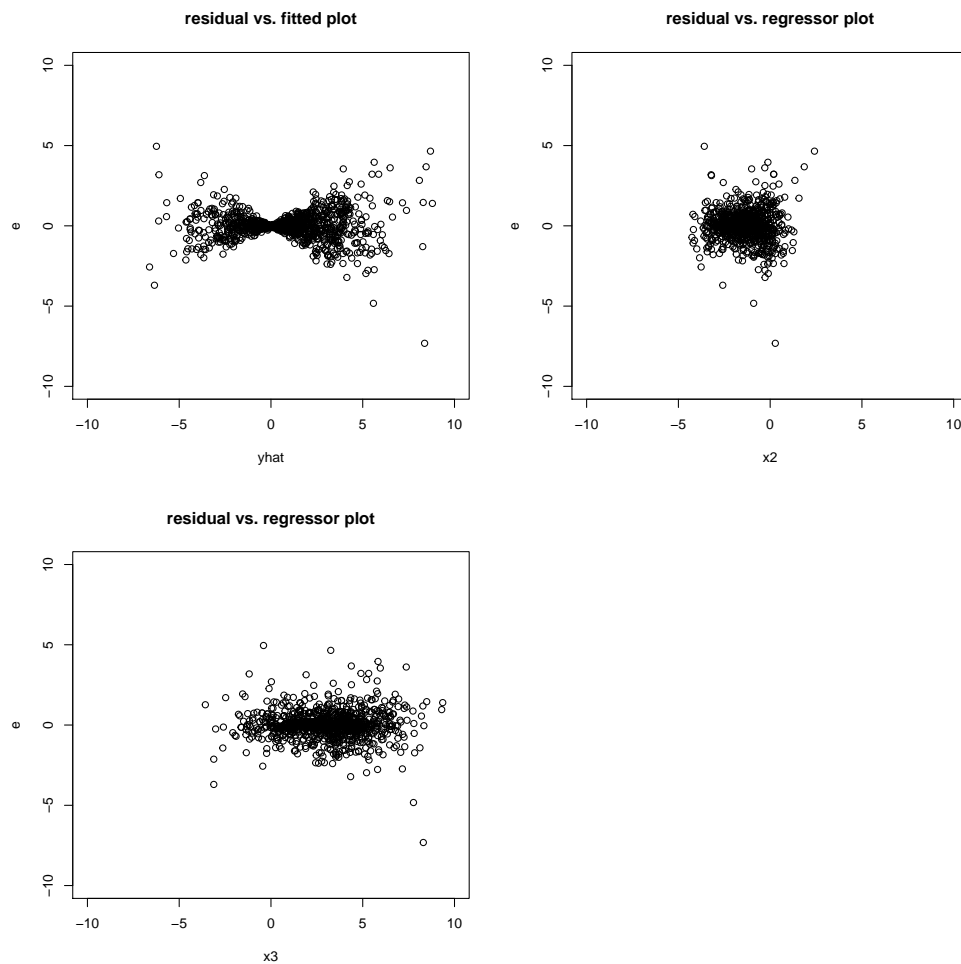


FIGURE 2. HSK diagnostic plots,  $HSK = f(E(y))$

### EXAMPLE 3: HETEROSKEDASTICITY RELATED TO X

```
R> sig2true<-2*abs(x2) #"true" error variance
R> eps<-rnorm(n,0,sqrt(sig2true))
R> y<-X %*% bvec + eps
R> #
R> bols<-solve((t(X)) %*% X) %*% (t(X) %*% y)
R> e<-y-X%*%bols
R> yhat<-X%*%bols
R> SSR<-(t(e)%*%e)
R> s2<-(t(e)%*%e)/(n-k)
R> Vb<-s2[1,1]*solve((t(X))%*%X)
R> se=sqrt(diag(Vb))
R> tval=bols/se
R> #
R> ttols<-data.frame(col1=c("constant","x2","x3"),
                    col2=bols,
                    col3=se,
                    col4=tval)
R> colnames(ttols)<-c("variable","estimate","s.e.,""t")
```

TABLE 3. OLS for Model with HSK related to x2

variable	estimate	s.e.	t
constant	1.115	0.121	9.200
x2	1.941	0.053	36.496
x3	0.813	0.027	30.595

### EXAMPLE 4: RANDOM GROUPWISE-HETEROSKEDASTICITY

(This one is difficult to detect from graphs alone)

```
R> sig2true<-c(rep(1,n/4),rep(2,n/4),rep(4,n/4),rep(9,n/4)) #"true" error variance
R> eps<-rnorm(n,0,sqrt(sig2true))
R> y<-X %*% bvec + eps
R> #
R> bols<-solve((t(X)) %*% X) %*% (t(X) %*% y)
R> e<-y-X%*%bols
R> yhat<-X%*%bols
R> SSR<-(t(e)%*%e)
R> s2<-(t(e)%*%e)/(n-k)
R> Vb<-s2[1,1]*solve((t(X))%*%X)
R> se=sqrt(diag(Vb))
R> tval=bols/se
R> #
R> ttols<-data.frame(col1=c("constant","x2","x3"),
                    col2=bols,
                    col3=se,
                    col4=tval)
R> colnames(ttols)<-c("variable","estimate","s.e.,""t")
```

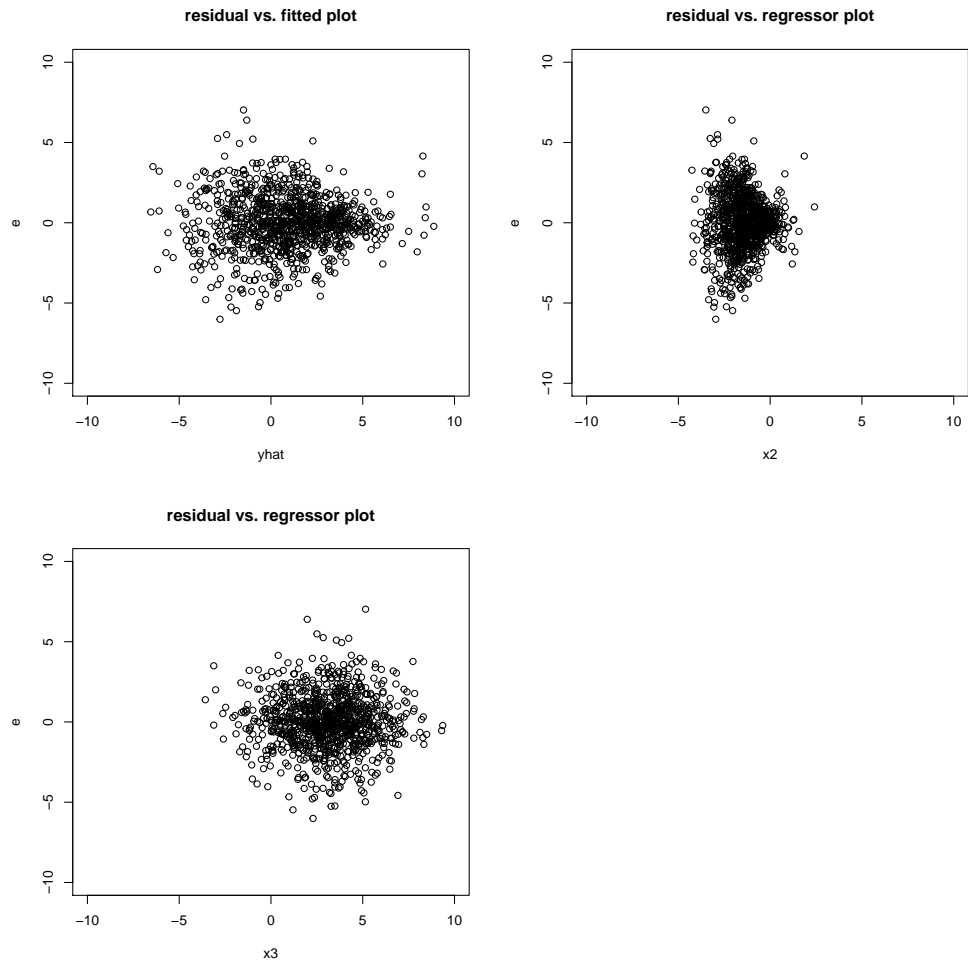


FIGURE 3. HSK diagnostic plots,  $HSK=f(x_2)$

TABLE 4. OLS for Model with group-wise HSK

variable	estimate	s.e.	t
constant	1.222	0.149	8.218
x2	1.999	0.065	30.642
x3	0.783	0.033	24.027

```
R> proc.time()-tic
  user system elapsed
 0.84  0.05  0.89
```

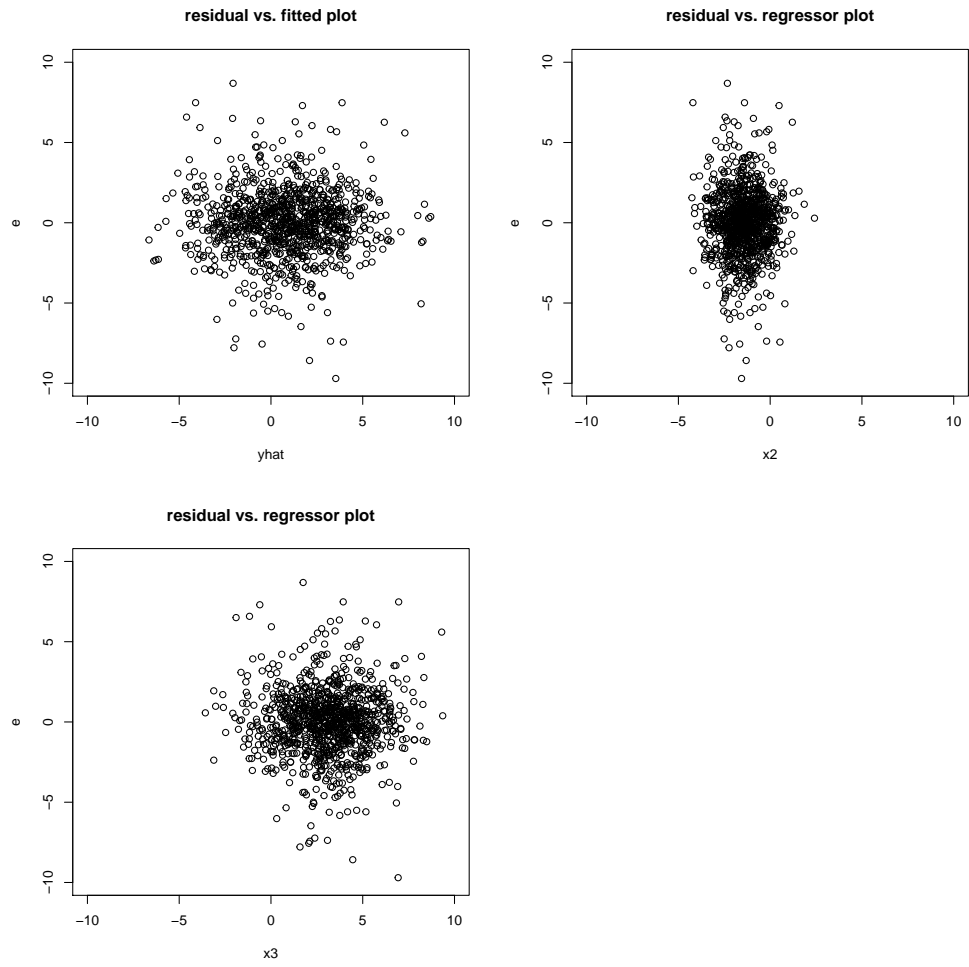


FIGURE 4. HSK diagnostic plots, group-wise HSK