SCRIPT `MOD4S2C`: HETEROSKEDASTIC REGRESSION USING HOTEL DATA

INSTRUCTOR: KLAUS MOELTNER

Load and Prepare Data

We will use data on the annual water consumption by 120 hotels / motels of small to moderate size (under 100,000 sft) in the Reno / Sparks area of Northern Nevada. Each establishments has data for up to 8 years (1993-2000), although some (newer) hotels are associated with fewer years. Thus, this is an example of an *unbalanced panel*. Courtesy of Truckee Meadows Water Authority (the local utility).

```
R> data<- read.table('c:/Klaus/AAEC5126/R/data/water.txt', sep="\t", header=FALSE)
R> #
R> #assign variable names
R> names(data)[1]<-"id"
R> names(data)[2]<-"year"
R> names(data)[3]<-"lncons"
R> names(data)[4]<-"age"
R> names(data)[5]<-"stories"
R> names(data)[6]<-"sf000"
```

Variable definitions:

```
% Contents of Data (columns)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% 1     id          hotel id (1:120)
% 2     year        1993 - 2000 (not complete for all hotels)
% 3     ln_cons     log of (annual consumption in gallons)
% 4     age         age of establishment
% 5     stories     number of stories
% 6     sf000       square footage in 1000 units
```

```
R> # Create year dummies
R> data$d94<-ifelse(data$year==1994,c(1),c(0))
R> data$d95<-ifelse(data$year==1995,c(1),c(0))
R> data$d96<-ifelse(data$year==1996,c(1),c(0))
R> data$d97<-ifelse(data$year==1997,c(1),c(0))
R> data$d98<-ifelse(data$year==1998,c(1),c(0))
R> data$d99<-ifelse(data$year==1999,c(1),c(0))
R> data$d00<-ifelse(data$year==2000,c(1),c(0))
R> data$age2<-data$age^2
R> #
R> save(data, file = "c:/Klaus/AAEC5126/R/data/water.rda")
R> attach(data)
R> #
```

```
R> y<-lncons
R> n<-length(y)
R> X<-cbind(rep(1,n),age,age2,stories,sf000,d94,d95,d96,d97,d98,d99,d00)
R> k<-ncol(X)
```

Generic OLS and Residual Plots

```
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","age","age2","stories",
        "sf000","d94","d95","d96","d97","d98","d99","d00"),
                col2=bols,
                col3=se,
                col4=tval)
R> colnames(ttols)<-c("variable","estimate","s.e.","t")
```

Table 1. OLS Results

variable	estimate	s.e.	t
constant	4.374	0.225	19.436
age	0.067	0.009	7.885
age2	-0.001	0.000	-5.617
stories	0.194	0.037	5.174
sf000	0.034	0.002	17.498
d94	0.054	0.120	0.447
d95	0.048	0.118	0.410
d96	0.023	0.117	0.194
d97	-0.007	0.117	-0.058
d98	-0.066	0.118	-0.562
d99	0.006	0.120	0.048
d00	-0.046	0.122	-0.381

White's general test

```
R> yaux<-e^2 #use squared OLS residuals as dep.var. in White test
R> #construct all permissible squared terms from the original X
R> Xsq<-cbind(stories^2, sf000^2) #note: we already have
R> #age2 in the original X
R> # we could / should add age2^2, but this produces a singularity error in this case
R> #construct all permissible interaction terms from the original X
R> #
```

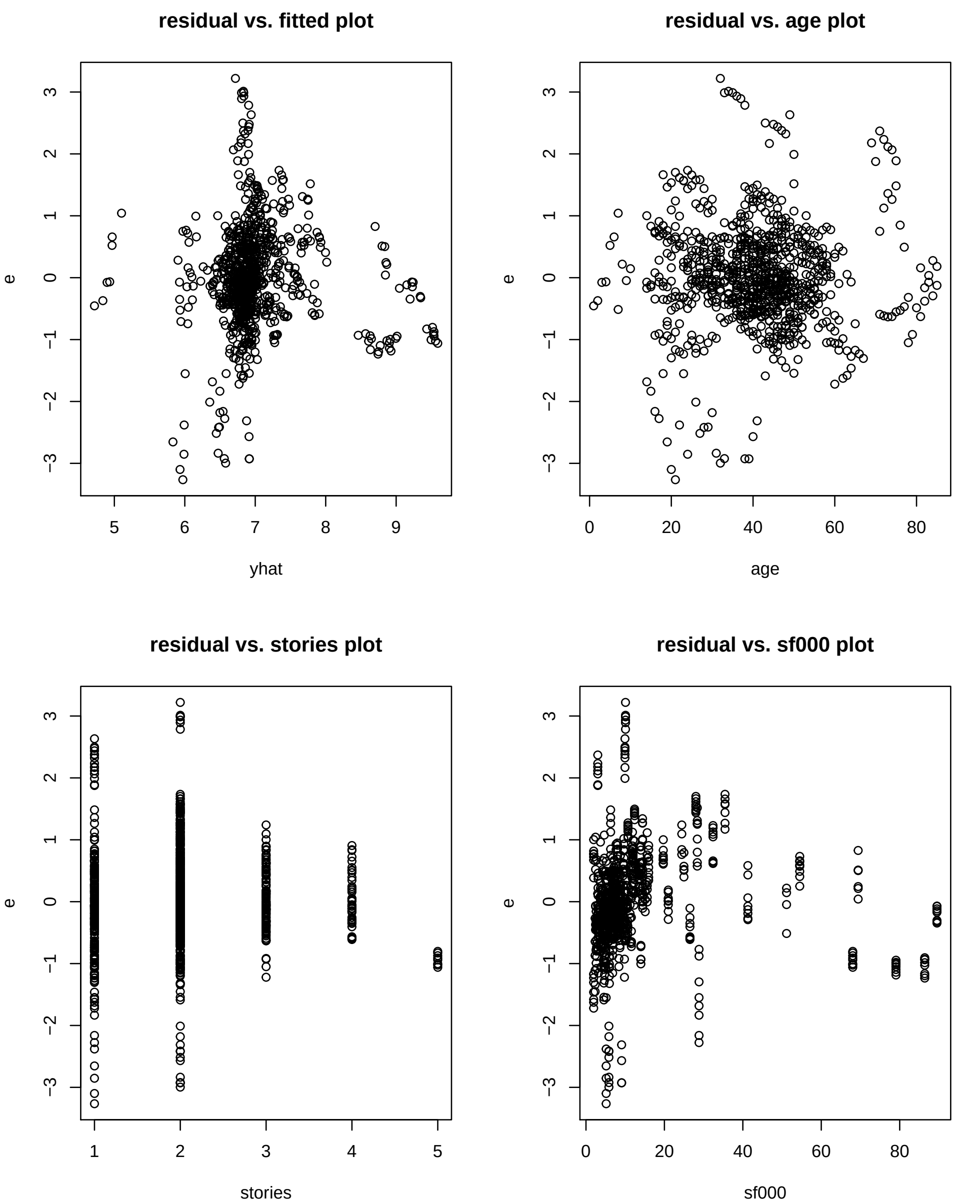

Figure 1. HSK diagnostic plots

```
R> # first for all continuous variables
R> Xc1<-age*age2
R> Xc2<-age*stories
R> Xc3<-age*sf000
R> Xc4<-age2*stories
R> Xc5<-age2*sf000
R> Xc6<-sf000*stories
R> #
```

```
R> #next for the continuous with indicators
R> dmat<-cbind(d94, d95, d96, d97, d98, d99, d00)
R> Xciage<-matrix(rep(age,7),nrow=n)*dmat
R> Xciage2<-matrix(rep(age2,7),nrow=n)*dmat
R> Xcistories<-matrix(rep(stories,7),nrow=n)*dmat
R> Xcisf000<-matrix(rep(sf000,7),nrow=n)*dmat
R> #
R> #interactions across year indicators don't make any sense
R> # (would lead to all-zero colums), so we're done.
R> #
R> #Next: Run auxiliary regression and capture R^2
R> Xaux<-cbind(X,Xsq,Xc1,Xc2,Xc3,Xc4,Xc5,Xc6,Xciage,Xciage2,Xcistories,Xcisf000)
R> kaux<-ncol(Xaux)
R> baux<-solve((t(Xaux)) %*% Xaux) %*% (t(Xaux) %*% yaux)
R> eaux<-yaux-Xaux%*%baux
R> I<-diag(n)
R> i<-rep(1,n)
R> Mo<-I-i %*% solve(t(i) %*% i) %*% t(i)
R> SSE<-t(eaux) %*% eaux
R> SST<-t(yaux) %*% Mo %*% yaux
R> R2<-1-SSE/SST
R> Wh<-n*R2
R> pval=1-pchisq(Wh,kaux-1)
```

The White-statistic for this test is 105.804. The degrees of freedom for the test are 47. The corresponding p-value is 0.

Breusch-Pagan test

```
R> int<-(t(e)%*% e)/n
R> g<-(e^2/(int[1,1]))-1
R> #capture variables you think may be related to HSK
R> Z<-cbind(rep(1,n), age, stories, sf000)
R> kz<-ncol(Z)
R> LM<-(1/2)*(t(g) %*% Z %*% solve(t(Z) %*% Z) %*% t(Z) %*% g)
R> pval=1-pchisq(LM,kz-1)
```

The BP-statistic for this test is 56.628. The degrees of freedom for the test are 3. The corresponding p-value is 0.

Robust Breusch-Pagan test (Koenker, 1981)

```
R> int<-(t(e)%*% e)/n
R> V<-mean((e^2-(int[1,1]))^2)
R> LM<-(1/V)*
 (t(e^2-int[1,1]) %*% Z %*% solve(t(Z) %*% Z) %*% t(Z) %*% (e^2-int[1,1]))
R> pval=1-pchisq(LM,kz-1)
```

The robust BP-statistic for this test is 25.016. The degrees of freedom for the test are 3. The corresponding p-value is 0.

Robust OLS

```
R> bols<-solve((t(X)) %*% X) %*% (t(X) %*% y)
R> e<-as.vector(y-X%*%bols)
R> # The "as.vector" coercion is apparently needed for the following diag command
R> # to work properly
R> S<-diag(e^2)
R> Vb<-solve((t(X))%*%X) %*% t(X) %*% S %*% X %*% solve((t(X))%*%X)
R> se=sqrt(diag(Vb))
R> tval=bols/se
R> #
R> ttols<-data.frame(col1=c("constant","age","age2","stories",
        "sf000","d94","d95","d96","d97","d98","d99","d00"),
               col2=bols,
               col3=se,
               col4=tval)
R> colnames(ttols)<-c("variable","estimate","s.e.","t")
```

Table 2. Robust OLS Results

variable	estimate	s.e.	t
constant	4.374	0.254	17.199
age	0.067	0.008	8.076
age2	-0.001	0.000	-5.828
stories	0.194	0.035	5.491
sf000	0.034	0.002	17.609
d94	0.054	0.123	0.438
d95	0.048	0.125	0.385
d96	0.023	0.121	0.186
d97	-0.007	0.117	-0.058
d98	-0.066	0.120	-0.552
d99	0.006	0.121	0.047
d00	-0.046	0.117	-0.397

```
R> proc.time()-tic
   user  system elapsed
   0.26    0.06    0.33
```