liquid<-read.csv(file.choose(), header=TRUE)

attach(liquid)

model1 = lm(VOLUME ~ PRICE + SHARE + VALUE)

model2 = lm(VOLUME ~ PRICE)

extractAIC(model1)

[1] 4.0000 472.6637

extractAIC(model2)

[1] 2.0000 580.9976

#The smaller the better,so by the AIC critierion,it prefers model1

step(model1)

Start: AIC=472.66

VOLUME ~ PRICE + SHARE + VALUE

Df Sum of Sq RSS AIC

- PRICE 1 18.22 5395.5 471.08

<none> 5377.3 472.66

- VALUE 1 169.41 5546.7 474.48

- SHARE 1 1218.82 6596.1 495.79

Step: AIC=471.08

VOLUME ~ SHARE + VALUE

Df Sum of Sq RSS AIC

<none> 5395.5 471.08

- VALUE 1 164.96 5560.5 472.78

- SHARE 1 1606.73 7002.2 501.14

Call:

lm(formula = VOLUME ~ SHARE + VALUE)

Coefficients:

(Intercept) SHARE VALUE

7.03198 0.05638 0.25494

# in the 1st line there are 4 betas,3 coefficent and 1 intercept.

plot(residuals(model1) ~ fitted.values(model1), main="Residuals vs.Fitted Value")

lnVOLUME = log(VOLUME)

model3 = lm(lnVOLUME ~ PRICE + SHARE + VALUE)

plot(residuals(model3) ~ fitted.values(model3), main="LN Residuals vs.Fitted Value")

# This log transformation has made the amount of spread much worse and

# more unequal, it definitely has heterosedasticity in the log plot