# 回归分析 期中试题

#### 王立敏 2017E8018661153

1. 为了找出影响石油价格波动的主要因素，我们使用逐步回归法来剔除一些对于应变量即价格影响很小的自变量剔除出我们的模型。

我们分别把WTI Price Field Production of Crude Oil (Thousand Barrels) 等自变量的名称改为x1，x2……最后的突发事件需要用到哑变量，哑变量只需要2个即可，我们将其作为X49，X50，X51,三个参数并将它们的值”正影响”，”无影响”，”负影响”分别改为-1,0,1.。

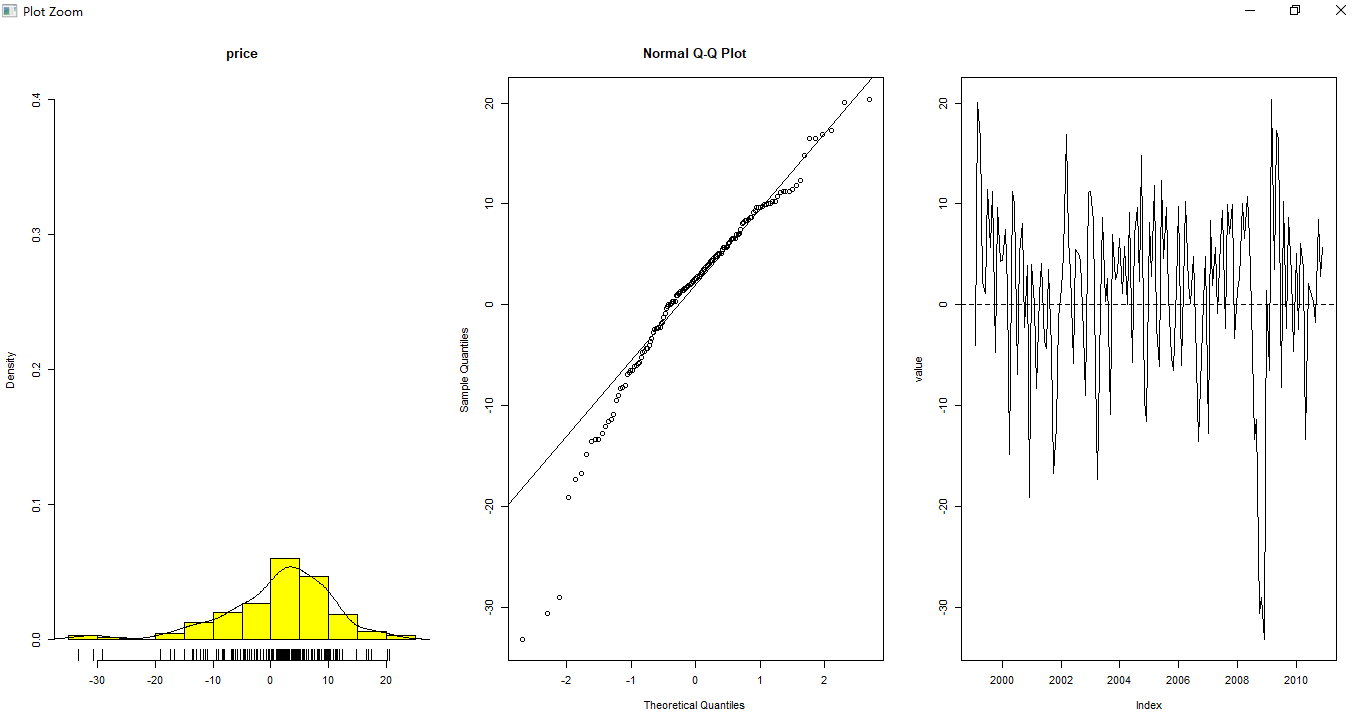
经过R语言处理以后我们得到模型

Y~x1 + x2 + x4 + x5 + x7 + x13 + x14 + x15 + x16 + x17 + x18 + x20 + x21 + x23 + x34 + x25 + x26 + x29 + x30 + x33 + x35 + x36 + x37 + x39 + x40 + x42 + x44 + x46 + x47 + x48 + x49 + x50

由此可得，影响较小的已经被剔除了。

1. 我们通过使用Garch模型来预测波动性，

先检验数据正态性，可以计算得出数据分布函数，QQ图，对数收益率序列折现图



> shapiro.test(rlogdiffdata)

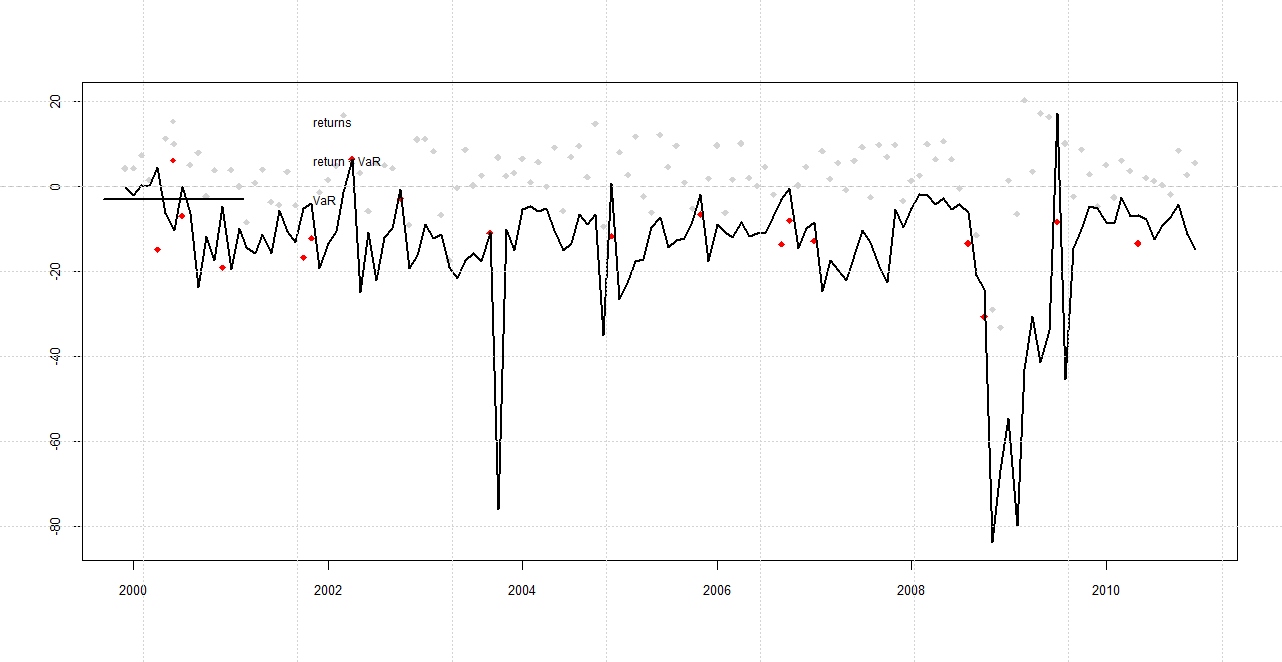
Shapiro-Wilk normality test

data: rlogdiffdata

W = 0.94315, p-value = 1.458e-05

由QQ图以及p值可见，数据大致上符合正态分布。

最后用Var曲线来进行预警剧烈波动。



红色的点就是剧烈波动前的预警点。

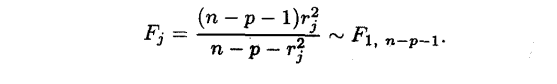
1. 我们可以通过使用cook统计量来寻找强影响点，因此我们用R语言的influence.measures()函数来进行影响分析。



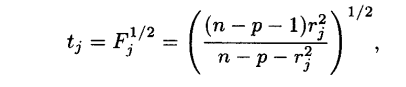
右侧带有\*号的表示强影响点。（不止图中的两个点，点太多，未放全）

1. 我们通过学生化残差来构造F检验，最终得到t检验，以此来检测异常点。

通过stdres<-rstudent(lm.sol)来得到学生化残差，然后通过公式

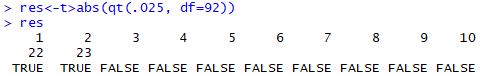


来计算Fj，并且最终转化为tj，t=sqrt((144-51-1)\*stdres^2/(144-51-stdres^2))



最后我们可以检查，如果则它为异常点。

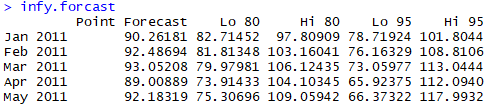
R语言执行res<-t>abs(qt(.025, df=92))可以直接得到大于对应t值的布尔值。



值为True的则可能为异常点。（不止有这些，值太多，未放全）

1. 预测

我们使用了HoltWinters来进行预测我们的油价区间

真实值基本都在预测的范围内，但是想要净赚预测还是比较困难的。