

46-926 Homework 1, Part 1

Jingyi Guo

1/22/2017

1. Categorical predictors: `id`, `bond_id`, `is_callable`, `trade_type`, `trade_type_last1`, `trade_type_last2`, ..., `trade_type_last10`

exclude categorical predictors and predictors from time steps 5 and prior:

```
bonddata = read.table("http://www.stat.cmu.edu/~cschafer/MSCF/bonddata.txt", sep=",", header=T)
newdata=subset(bonddata[1:31], select=-c(id, bond_id, is_callable, trade_type, trade_type_last1, trade_type_last2, ..., trade_type_last10))
```

2. Fit a model using all of the predictors that remain:

```
fitmodel = lm(trade_price ~ ., data = newdata)
```

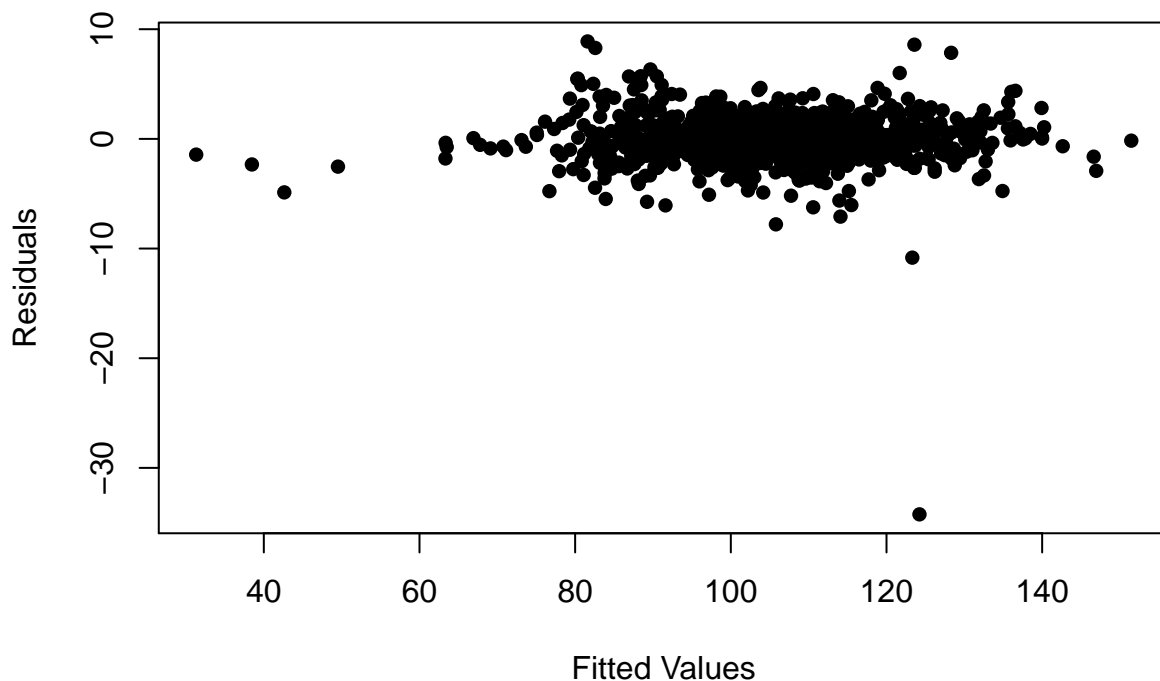
AIC for this model:

```
AIC(fitmodel)
```

```
## [1] 6563.772
```

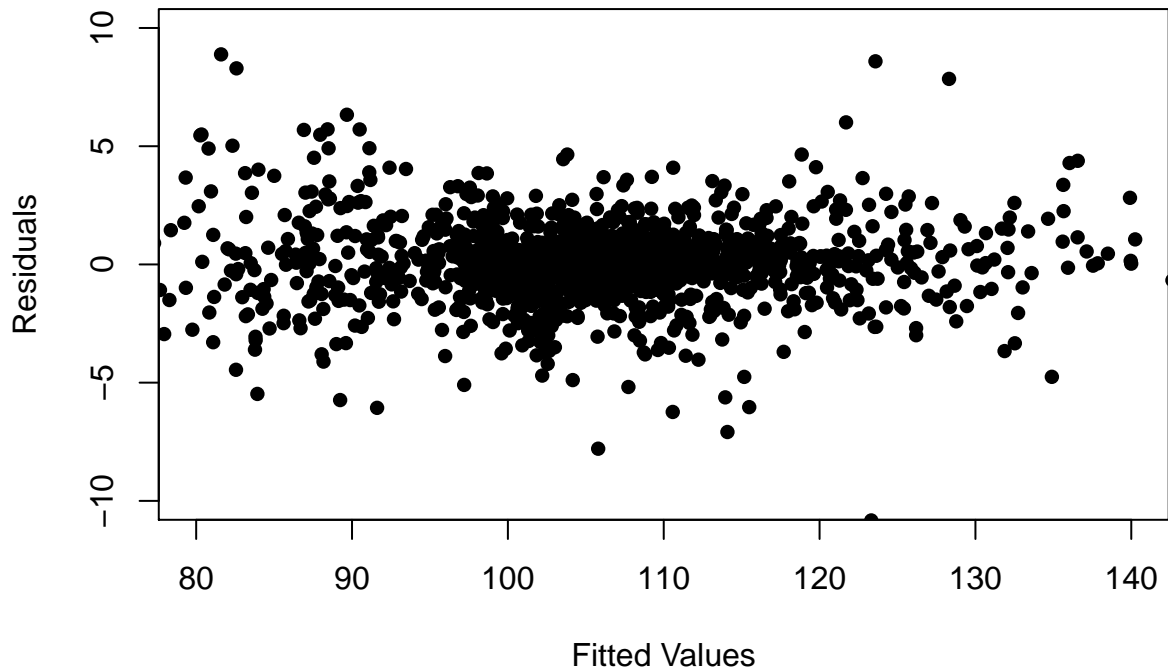
Plot of residuals versus fitted values:

```
plot(as.numeric(fitmodel$fit), as.numeric(fitmodel$resid), pch=16, xlab="Fitted Values", ylab="Residuals")
```



```
#changed the scales of axes
```

```
plot(as.numeric(fitmodel$fit),as.numeric(fitmodel$resid),pch=16,xlab="Fitted Values", ylab="Residuals",
```

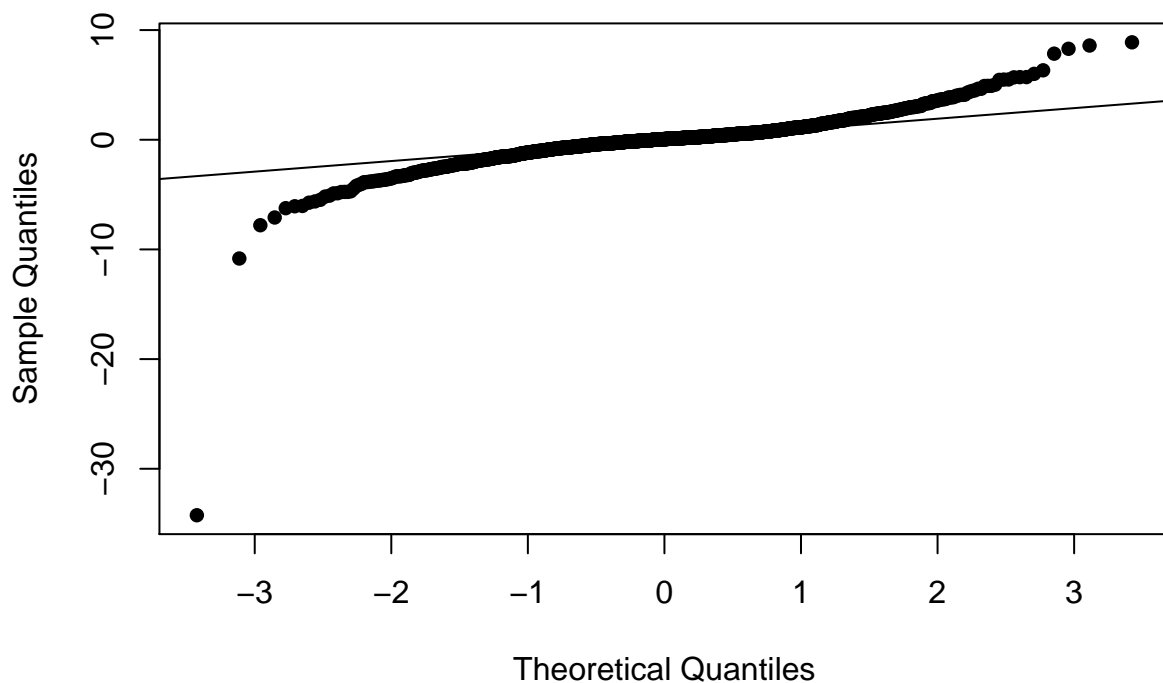


Comment: There is no significantly prevalent pattern in the plot of residual versus fitted values. However, there is one point for which the residual is quite extreme relative to others.

The normal probability plot of the residuals:

```
qqnorm(as.numeric(fitmodel$resid),pch=16,main="")
```

```
qqline(as.numeric(fitmodel$resid))
```



Comment: The normal probability plot of residuals suggests that errors have distribution with tails significantly heavier than normal distribution.