ORF 505: Statistical Analysis of Financial Data

Tuesday Precept 4: Least Square Regression

Oct 1, 2024

Lecturer: Qishuo Yin *Scribe: Qishuo Yin*

1 Review: Least Square (LS) Simple Linear Regression

- Univariate response, Univariate explanatory variable

$$(x_1, y_1), (x_2, y_2), \cdots\cdots, (x_n, y_n)$$

- Search for a Regression function φ (for now) linear or affine

$$\varphi(x) = \beta_0 + \beta_1 x$$

- Least squares estimation

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg \min_{(\beta_0, \beta_1)} \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i|^2$$

- Least Absolute Deviations (LAD or L1) estimation

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg \min_{(\beta_0, \beta_1)} \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i|$$

2 Play with Simple Least Square R Codes

1. We revisit the utility data, we first perform the least squares regression of the utility index daily log-return against the ENRON daily log-return.

```r
# play with R function lm - example 1
UeL2 <- lsfit(EnronLRet, UtilLRet) # lsfit: perform statistical inference
Ue <- lm(UtilLRet ~ EnronLRet) # lm: perform least squares regression
```

Listing 1: lm example 1

2. We can also conduct simple least squares linear regression of `UtilLRet` against `DukeLRet`.

```r
# play with R function lm - example 2
UdL2 <- lsfit(DukeLRet, UtilLRet)
Ud <- lm(UtilLRet ~ DukeLRet)
```

Listing 2: lm example 2

3. To make the model even more complex, we perform the multiple least squares linear regression of UtilLRet against the variables `EnronLRet` and `DukeLRet` together.

```
# multiple least squares linear regression
Ued <- lm(UtilLRet  ~ EnronLRet + DukeLRet)
```

Listing 3: lm multiple least squares linear regression example

If you have more interest in the simple or multiple regression, please see section 4.4, especially section 4.4.1 of [1] for further reference.

3 Extension: Polynomial Regression

We use the data set `FRWRD` containing the values in US $ of the 36 Henry Hub natural gas forward contracts traded on March 23rd 1998. We use those prices as observations of our response variable, and we use the index of the delivery month of the forward contract as the explanatory variable. Since natural gas forward curves have a strong seasonal component which distinguishes them from most of the other commodity forward curves, we conduct several polynomial regressions of the forward prices against the month indexes.

```
plot(1:36, FRWRD, main="Polynomial Gas Forward Curves")
lines(1:36, fitted(lm(FRWRD ~ poly(1:36, 3))), lty=1)
lines(1:36, fitted(lm(FRWRD ~ poly(1:36, 6))), lty=3)
lines(1:36, fitted(lm(FRWRD ~ poly(1:36, 8))), lty=6)
```

Listing 4: polynomial regression example

References

[1] R. Carmona. *Statistical analysis of financial data in R*, volume 2. Springer, 2014.