Group 1 HW 3 - Q2

Siow Meng Low

17 December 2016

Forecasting Model

In the *Bass Model* article, the forecasting model is given as below:

$$F_t = (p + q[C_{t-1}/m])(m - C_{t-1})$$

$$F_t = pm + (q - p)C_{t-1} - \frac{q}{m}C_{t-1}^2$$

The paraemters used in the above equation is described below:

- F_t: Forecasted number of new adopters during time period t, denoted as $N(t)$ in the class lecture notes
- C_{t-1}: Cumulative number of people who have adopted the product up to time $t-1$
- p: Coefficient of innovation (a.k.a. Rate of spontatneous adoption in the class lecture notes)
- q: Coefficient of imitation (a.k.a. Rate of imitation in the class lecture notes)
- m: Market size, the number of people estimated to eventually adopt the new product

From the above equation, we can see that F_t is a function of C_{t-1} and C_{t-1}^2. In the article, the authors proposed to minimise the sum of squared forecast errors $\sum_{t=1}^{N} E_t^2$ where:

- S_t: the actual number of new adopters in period t, denoted as $S(t)$ in the class lecture notes
- $E_t = F_t - S_t$, E_t is the forecast error for period t

This is the same technique used in the Ordinary Least Square Estimation of Linear Regression, where we take F_t (the forecasted number of new adopters in period t) to be $\hat{S}_t$ (the estimated number of adopters in period t). The equation for linear regression can be written as:

$$F_t = \hat{p}\hat{m} + (\hat{q} - \hat{p})C_{t-1} - \frac{\hat{q}}{\hat{m}}C_{t-1}^2$$

$$\hat{S}_t = \hat{a} + \hat{b}C_{t-1} + \hat{c}C_{t-1}^2$$

where:

$$\hat{S}_t = F_t$$

$$\hat{a} = \hat{p}\hat{m}$$

$$\hat{b} = \hat{q} - \hat{p}$$

$$\hat{c} = -\frac{\hat{q}}{\hat{m}}$$

The system of equations involving $\hat{p}$, $\hat{q}$, and $\hat{m}$ can be rewritten as follows:

$$\hat{m} = \frac{-\hat{b} \pm \sqrt{\hat{b}^2 - 4\hat{a}\hat{c}}}{2\hat{c}}$$

$$\hat{p} = \frac{\hat{a}}{\hat{m}}$$

$$\hat{q} = -\hat{c}\hat{m}$$

Note that for $\hat{m}$, we are only interested in its non-negative real root, since the market size should always be greater than or equal to zero.

In the following section, we shall perform least square regression (dependent variable is S_t and independent variables are C_{t-1} and C_{t-1}^2) using R, in order to obtain the rolling horizon estimates of p, q, m, and F_t.

Rolling Horizon Estimate using Ordinary Least Square Regression

The table that contains the box office revenue data of "The Doctor" is included in the CSV file, *TheDoctorData.csv*. First, we shall read the revenue data from the CSV and calculate C_{t-1} for each period.

```
setwd("C:/Users/admin/Desktop/Imperial MSc/Network Analytics/Assignment 3")

doctorData <- read.csv("TheDoctorData.csv")
doctorData$"Ctlag" <- c(0, doctorData$Ct[1:(nrow(doctorData) - 1)])

pander(doctorData,
       caption = "Revenues and Cumulative Revenues in $ Millions for THE DOCTOR")
```

Table 1: Revenues and Cumulative Revenues in $ Millions for THE DOCTOR

Week	St	Ct	Ctlag
1	0.1	0.1	0
2	3	3.1	0.1
3	5.2	8.3	3.1
4	7	15.3	8.3
5	5.25	20.55	15.3
6	4.9	25.45	20.55
7	3	28.45	25.45
8	2.4	30.85	28.45
9	1.9	32.75	30.85
10	1.3	34.05	32.75
11	0.8	34.85	34.05
12	0.6	35.45	34.85

Next, starting at week 5 ($t = 5$), we shall use the observed cumulative revenues, C_{t-1} to obtain the rolloing-horizon estimates of the parameters.

In the following section, OLS regressions are performed (for $t = 5, 6, ..., 12$) and the eight sets of estimated parameters are tabulated below:

```
finalTable = data.frame()

# Start forecasting at t = 5
```

```r
tValues <- doctorData$Week[5 : nrow(doctorData)]
for (tValue in tValues) {
    modelDoctor <- lm(St ~ I(Ctlag) + I(Ctlag^2), data = doctorData[1 : (tValue - 1), ])
    a <- modelDoctor$coefficients[1]
    b <- modelDoctor$coefficients[2]
    c <- modelDoctor$coefficients[3]
    m <- c((-b + sqrt(b^2 - 4 * a * c)) / (2 * c), (-b - sqrt(b^2 - 4 * a * c)) / (2 * c))
    m <- m[!is.nan(m) & m >= 0] # Only get the positive root
    p <- a / m
    q <- -c * m
    Ft <- predict(modelDoctor, newdata = data.frame(Ctlag = doctorData$Ctlag[tValue]))
    finalTable <- rbind(finalTable, data.frame(week = tValue, p = p, q = q, m = m,
                                               Ft = Ft, St = doctorData$St[tValue],
                                               row.names = NULL))
}

panderOptions('round', 3)
pander(finalTable, caption = "Rolling Horizon Estimates of p, q, m, and Ft")
```

Table 2: Rolling Horizon Estimates of p, q, m, and Ft

week	p	q	m	Ft	St
5	0.094	1.668	15.3	-0.008	5.25
6	0.079	1.309	20.12	-0.611	4.9
7	0.074	0.928	25.82	0.361	3
8	0.072	0.774	28.76	0.262	2.4
9	0.072	0.674	30.84	-0.003	1.9
10	0.072	0.598	32.49	-0.175	1.3
11	0.073	0.547	33.62	-0.272	0.8
12	0.073	0.516	34.31	-0.321	0.6

From the above estimates, we can see that the estimated F_t is quite different from the actual S_t. The rolloing-horizon approach may not be suitable in this case.