

# Network Analytics

## Homework 3 - Group Part

Group 1

18 December 2016

### 2.

#### 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 parameters 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 spontaneous 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.

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 rolling-horizon estimates of the parameters.

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

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 rolling-horizon approach may not be suitable in this case.

### 3. a.

#### Overview

The paper models the spread of information by looking at the take-up of microfinance in rural villages, with no exposure to microfinance institutions, in the south of India, to answer the following question: does the set of “leaders” initially informed have an influence on the long-run participation in microfinance? First, we will replicate some of the data used in the paper and construct some of the regression models used to estimate the importance of the leaders. Then, we will draw some conclusions from the results and finally we will describe the dynamic model and the structural estimation procedure that was used in the paper.

#### Data Replication and Regression Models

The data we were able to replicate for the households of the villages and get the exact same results were the following:

- Microfinance take-up rate of non-leader households
- Average degree of leaders
- Average eigenvector centrality for leaders
- Number of households
- Fraction of leaders
- Fraction of taking leaders
- Average eigenvector centrality of taking leaders

Besides these characteristics we reproduced the average betweenness centrality and closeness centrality of the leaders but we got different results. With these data available we were able to reconstruct regression models 1, 2, 3 and 5.

In the first regression model, we assessed whether eigenvector centrality of the leaders and number of households have any effect on the take-up rate. We can see that eigenvector centrality of the leaders is significant but the number of households is not.

In the second regression model, we looked to see if take-up rate is correlated between the number of households and degree of the leaders. Number of households appears to be significant in determining the uptake of microfinance.

In the third regression model, we looked at eigenvector centrality of the leaders, number of households and degree of leaders. Again, eigenvector centrality is significant in determining uptake of microfinance.

In the last regression model, we regressed four different variables against the take-up rate of microfinance. Fraction of taking leaders appears to be significant along with eigenvector centrality of leaders.

#### Results Interpretation

The summaries of the four regression models are shown in table 3. From these as well as from the summaries of the models we didn’t manage to reconstruct, we can come to a conclusion that the injection points are crucial in determining the uptake of microfinance within a village. Looking at the results from the regression models, we can see that eigenvector centrality of leaders, which is the indicative of the influence of individuals is particularly important, given its high coefficients, as well as fraction of taking leaders to a lesser degree. Other relevant measures such as degree of the leaders and eigenvector centrality of taking leaders do not appear to be as important. On the other hand, social network characteristics, such as number of households and savings, are not as important in determining the percentage of uptake by a village, given their small coefficients and insignificance in the regression models. This is interesting because one would theoretically expect that the amount of savings each household has should influence whether this household participates in microfinance or not.

These findings reinforce the importance and influence of leaders, who act as key conduits for disseminating information within a village and determining the outcome of rolling out a scheme. In addition, the findings from the panel analysis

Table 3: Leaders/Injection points

	<i>Dependent variable:</i>			
	Take-up Rate			
	(1)	(2)	(3)	(4)
Eigenvector Centrality of Leaders	1.634233* (0.880280)		1.934449** (0.928651)	1.253972 (0.913598)
Number of Households	-0.000382 (0.000270)	-0.000704*** (0.000213)	-0.000270 (0.000292)	-0.000305 (0.000233)
Degree of Leaders		-0.001109 (0.003154)	-0.003237 (0.003198)	
Fraction of Taking Leaders				0.322830*** (0.081432)
Eigenvector Centrality of Taking Leaders				-0.175042 (0.421896)
Constant	0.150208 (0.114858)	0.362092*** (0.066447)	0.161842 (0.115396)	0.092399 (0.100286)
Observations	43	43	43	43
R <sup>2</sup>	0.293095	0.234552	0.311190	0.501782
Adjusted R <sup>2</sup>	0.257750	0.196280	0.258205	0.449338
<i>Note:</i>			*p<0.1; **p<0.05; ***p<0.01	

also corroborate the individual regression models we generated and the fact that eigenvector centrality of leaders matters most over time in determining uptake of microfinance, as opposed to just the later stages in the roll-out of the scheme.

### 3. b.

#### Model

The data was collected through a questionnaire that focused on village, household and individual characteristics, including network data. This was supported by and matched with regular administrative data on who joined the program. The discrete time models that were estimated had the following structure:

1. The set of initial leaders (exogenously chosen) were informed and chose whether or not to participate.
2. In each time period, households passed information on with a certain probability that varied according to whether the household had chosen to participate ( $q^P$ ) or not ( $q^N$ ).
3. In each time period, the previously informed households decide whether to participate or not. In the baseline “information” model, the probability that an individual takes up microfinance, given that they are informed,  $p_i(\alpha, \beta)$  is given by:

$$p_i = P(\text{participation} | X_i) = \Delta(\alpha + X_i\beta)$$

Where X contains attributes of the individual and its neighbours (but does not include “endorsement”: whether the neighbours have taken up or not).

#### Estimation Procedure

The final model was fit using a Method of Simulated Moments approach. The dynamic simulation was run 75 times for each possible combination of parameters,  $\theta = (q^N, q^P, p_i(\alpha, \beta))$ , the moments were averaged across those simulations,

and then the set of parameters that minimised the square of the average difference between the simulated moments and the empirically collected moments is identified as the set of estimates for those parameters.

The moments used to estimate the model include

- One non-network moment: the share of leaders who take up microfinance
- Three proportion-based network moments: The share of households who take up microfinance with none of their neighbours having taken up, the share of households that take up microfinance - and are in the neighbourhood of a leader that takes up microfinance, and the share of households that take up microfinance and are in the neighbourhood of a leader who doesn't take up microfinance.
- Two covariance-based network moments: The covariance of the proportion of households taking up compared to the share of their neighbours who take up microfinance, and the covariance of the proportion of households taking up compared to the proportion of second-degree neighbours that take up

To estimate the standard errors for the parameter estimates, a grid-based Bayesian bootstrap algorithm is used, using 1000 samples (resampling with replacement), finding the optimal parameters for each sample (weighted due to the random sampling method), and using this to estimate the sample distribution.

## Robustness Checks

The following were undertaken to check and justify the use of the “information model”:

- A model including endorsement effects was also run, and those effects not found to be significant.
- The model was estimated with a different set of moments, and found that the parameters were comparable, and still significantly different.
- Microfinance participation was replaced with a “placebo” outcome, “type of roof”, which gave very high parameters, which was expected for cases where information transmission is not highly relevant.
- The model was compared to a model only using distance to the leaders.

## Conclusions Drawn

With the chosen model (without endorsement), the paper estimates  $q^N = 0.1$  and  $q^P = 0.5$ , both being individually statistically significant, and significantly different from each other. This implies that the probability of a neighbour taking up microfinance from someone who is informed is higher if that person has also taken up microfinance.

The robustness assessment implies the adoption of microfinance depends on more than just the closeness to the leaders, although there is an identified limitation that the dynamic model is more accurate for later time periods.