Bass Model

1. The first step of creating the data model is to load the dataset and calculate cumulative sales for each week. We shall also calculate the square transformation for the cumulative sales to include in as the dependent variable in the regression equation $S_t = a + b * N_{t-1} + c* N_{t-1} \wedge 2$

```
data sales;
      input week sales;
      datalines;
      1     160
      2     390
      3     800
      4     995
      5     1250
      6     1630
      7     1750
      8     2000
      9     2250
      10    2500
      ;
run;

data sales_cumulative;
      set sales; by week;
      if week = 1 then cum_sum = 0;
      cum_sum + lag(sales);
      cum_sum_sq = cum_sum ** 2;
run;

proc reg
      data = sales_cumulative outest = mylib.est;
      model sales = cum_sum cum_sum_sq / stb ;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: sales

Number of Observations Read	10
Number of Observations Used	10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5372770	2686385	98.05	<.0001
Error	7	191793	27399		
Corrected Total	9	5564563			

Root MSE	165.52823	R-Square	0.9655
Dependent Mean	1372.50000	Adj R-Sq	0.9557
Coeff Var	12.06020		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	1	417.46282	93.92583	4.44	0.0030	0
cum_sum	1	0.35647	0.04963	7.20	0.0002	1.78713
cum_sum_sq	1	-0.00001601	0.00000456	-3.51	0.0099	-0.87138

We see that the model is significant with extremely small p-value. The R^2 value of 0.9655 suggest that the model is able to explain 96.55% of the variation in the sales during the period of 10 weeks. To calculate the estimates for the p, q and M we need the intercept, and co-efficient for cumulative sales and transformed variable (square of cumulative sales).

```
data estimates;
      set mylib.est;

      a = Intercept;
      b = cum_sum;
      c = cum_sum_sq;

      M = ((-1 * b) - sqrt(b*b - 4 * a * c))/(2 * c);
      p = a / M;
      q = p + b;

      peak_time = log(q / p) * 1 / (p + q);
      peak_sales = M * ((p + q) ** 2) / (4 * q);

      call symput('p_coeff', p);
      call symput('q_coeff', q);
      call symput('M_coeff', M);

run;

proc print
      data = estimates;
run;
```

Model Estimates

Obs	_MODEL_	_TYPE_	_DEPVAR_	_RMSE_	Intercept	cum_sum	cum_sum_sq	sales	a	b	c	M	p	q	peak_time	peak_sales
1	MODEL1	PARMS	sales	165.526	417.463	0.35647	-.000016006	-1	417.463	0.35647	-.000016006	23386.22	0.017851	0.37432	7.75951	2402.20

From the output we have p = **0.0179**, q = **0.3743** and M = **23386.22**. Also, the peak sales, predicted is 2402.20 units of sales / currency. The peak time predicted is 7.76 i.e. around second half between 7th and 8th week of sales.

Predicting sales in each period using model parameters with sales at time period 0 = 0.

```
data predicted_sales;
      set sales_cumulative;
      if _N_ = 1 then predicted_sales =
symget('p_coeff')*symget('M_coeff');
      if _N_ = 1 then nt1 = 0;

      if _N_ > 1 then nt1 = nt1 + predicted_sales;
      if _N_ > 1 then
            predicted_sales = ((symget('p_coeff') +(symget('q_coeff') /
            symget('M_coeff')) * nt1) * (symget('M_coeff') - nt1));

      retain nt1;
      retain predicted_sales;
run;

proc print data = predicted_sales; run;
```

Predicted Sales

Obs	week	sales	cum_sum	cum_sum_sq	predicted_sales	nt1
1	1	160	0	0	417.46	0.00
2	2	390	160	25600	563.49	417.46
3	3	800	550	302500	751.74	980.95
4	4	995	1350	1822500	987.06	1732.69
5	5	1250	2345	5499025	1268.58	2719.75
6	6	1630	3595	12924025	1584.58	3988.33
7	7	1750	5225	27300625	1906.94	5572.92
8	8	2000	6975	48650625	2188.30	7479.86
9	9	2250	8975	80550625	2367.74	9668.16
10	10	2500	11225	126000625	2389.23	12035.90

Week	Observed Sales	Predicted Sales
1	160	417.46
2	390	563.49
3	800	751.74
4	995	987.06
5	1250	1268.58
6	1630	1584.58
7	1750	1906.94
8	2000	2188.3
9	2250	2367.74
10	2500	2389.23

Plotting a graph of actual versus predicted sales

```
proc sgplot
 data = predicted_sales;
 series X = week Y = sales /  legendlabel = 'Actual Sales' markers;
 series X = week Y = predicted_sales /  legendlabel = 'Actual Sales'
markers;
 title 'Actual v/s Predicted Sales';
 yaxis label = 'Sales in units';
 xaxis type = DISCRETE;
run;
```