International Research Journal of Social Sciences ISSN 2319–3565
Vol. **1(2),** 1-7, October **(2012)** *I Res. J. Social Sci.*

Modeling India's National Anthem: A Statistical Approach

Chakraborty Soubhik, Sarkar Saurabh, Tewari Swarima and Pal Mita
Department of Applied Mathematics, Birla Institute of Technology Mesra, Ranchi-835215, INDIA

Available online at: www.isca.in
Received 05th September 2012, revised 08th September 2012, accepted 10th September 2012

Abstract

A major strength of statistics lies in modeling. Modeling a musical structure or performance is both an interesting and challenging endeavour given that the true model is not only complex but unknown even to the composer. On the other hand, although statistical models are neither perfect nor unbiased, it should be understood that i. we can at least make the data objective or nearly so ii. the true model may have multiple parameters and we usually do not have explicit knowledge about them nor we know how or in what functional way they enter the model and iii. doing a stochastic realisation of this deterministic true model (the decision process of the composer is deterministic as any musical sequence of notes is planned and not random) is within the scope of statistics including controlling the errors in the model. The present work highlights our maiden attempt to model the national anthem of India using a simple exponential smoothing. The fit is found to be explaining the note progression well enough with a smoothing factor 0.716404. Should such a model work well for the national anthem of any other country, it is of interest to see how the smoothing factor varies. Experimenting with other sophisticated models like Kalman filter where the smoothing factor is not fixed but varying is also of interest.

Keywords: National anthem of India, simple exponential smoothing, information content, statistics

Introduction

India's national anthem: *Jana Gana Mana* (Bengali: জন গণ মন, Hindi: जन गण मन) is the national anthem of India. Written in highly Sanskritised (Tatsama) Bengali, it is the first of five stanzas of a Brahmo hymn composed and scored by Nobel laureate Rabindranath Tagore. It was first sung in Calcutta Session of the Indian National Congress on 27 December 1911. *Jana Gana Mana* was officially adopted by the constituent assembly as the Indian national anthem on 24 January 1950. 27 December 2011 marked the completion of 100 years of *Jana Gana Mana* since it was sung for the first time[1]. The time allotted for it is 52 seconds**.** The present note gives an interesting simple exponential model fitted to the national anthem of India. The fit is found to be explaining the note progression well enough with smoothing factor Alpha 0.716404.

Simple Exponential smoothing: A major strength of statistics lies in modeling. Modeling a musical structure or performance is both an interesting and challenging endeavour given that the true model is not only complex but unknown even to the composer. On the other hand, although statistical models are neither perfect nor unbiased, it should be understood that i. we can at least make the data objective or nearly so ii. the true model may have multiple parameters and we usually do not have explicit knowledge about them nor we know how or in what functional way they enter the model. Doing a *stochastic realisation* of this deterministic (the decision process of the composer is deterministic as any musical sequence of notes is planned and not random) true model is within the scope of statistics which includes controlling the errors in the model.

Simple Exponential smoothing: is used to statistically model time series data for smoothing purpose or for prediction. Although it was Holt[2] who proposed it first, it is Brown's simple exponential smoothing that is commonly used nowadays[3]. Simple exponential smoothing is achieved by the model

$F_{t+1} = \alpha Y_t + (1-\alpha)$ Ft, $0<\alpha<1$, to the data (t, Y_t) where F_t is the predicted against Y_t and initially $F_o = Y_o$.

Here α is the smoothing factor. This is the only parameter in the model that needs to be determined from the data. The smoothed statistic F_{t+1} is a simple weighted average of the previous observation Y_t and the previous smoothed statistic Ft. The term *smoothing factor* applied to α here is something of a misnomer, as larger values of α actually reduce the level of smoothing, and in the limiting case with α = 1 the output series is just the same as the original series (with lag of one time unit). Simple exponential smoothing is easily applied and it produces a smoothed statistic as soon as two observations are available. Values of α close to one have less of a smoothing effect and give greater weight to recent changes in the data, while values of α closer to zero have a greater smoothing effect and are less responsive to recent changes. There is no formally correct procedure for choosing α. Sometimes the statistician's judgment is used to choose an appropriate factor. Alternatively, a statistical technique may be used to *optimize* the value of α. For example, the method of least squares might be used to determine the value of α for which the sum of the quantities $(F_t - Y_t)^2$ is minimized [4].

Research Methodology

Time series is a series of observations in chronological order. Musical data can also be taken as a time series in which a note characterized by pitch Y_t is the *entry* corresponding to the *argument* time t which may mean time of clock in actual performance or just the instance at which the note is realized. In our case, since we are modelling only the structure of India's national anthem so the arguments will be simply the instances[1-3]. The desired note sequence is given in table-1. Western Art Music readers should refer to table 2 where corresponding western notations are provided. The tonic (Sa in Indian music) is taken at natural C. The present note gives an interesting Simple Exponential Modeling fit of the national anthem of India. The fit is found to be well enough with smoothing factor Alpha 0.716404. Analyzing the structure of a musical piece helps in giving an approximate model that captures the note progression in general without bringing the style of a particular artist into play. On the other hand, performance analysis gives additional features like note duration and the pitch movements between notes etc.[5]. In table-2, pitches of notes in three octaves are represented by corresponding integers, C of the middle octave being assigned the value 0.We are motivated by the works of Adiloglu, Noll and Obermayer[6]. The letters S, R, G, M, P, D and N stand for Sa, *Sudh* Re, *Sudh* Ga, *Sudh* Ma, Pa, *Sudh* Dha and *Sudh* Ni respectively. The letters r, g, m, d, n represent *Komal* Re, *Komal* Ga, *Tibra* Ma, *Komal* Dha and *Komal* Ni respectively. Normal type indicates the note belongs to middle octave; italics implies that the note belongs to the octave just lower than the middle octave while a bold type indicates it belongs to the octave just higher than the middle octave. The terms *Sudh*, *Komal* and *Tibra* imply, respectively, natural, flat and sharp.

Let us understand the national anthem of India with reference to the sequence of notes provided in table 1 before we move to the modeling part.

Translation in detail 1: Line 1: *Jana-Gana-mana-adhinayaka, jaya he Bharata-bhagya-vidhata* Translation: O lord of our destiny you are the captain of our souls and of the people of Bharat (India). This line corresponds to notes 1-21 in that sequence in table 1.

Line 2: *Punjaba-Sindhu-Gujarata-Maratha, Dravida-Utkala-Banga* Translation: Your name rouses the hearts of Punjab, Sindhu, Gujarat, Maratha, of the Dravida and Orissa and Bangla. This line corresponds to notes 22-44 in that sequence in table 1.

Line 3: *Vindhya-Himachala-Yamuna-Ganga Uchchala-Jaladhi-taranga Tava shubhaname jage Tava shubha asisa mage.* Translation: *Tava shubha name jage Tava shubha asisa mage* Your (the eternal charioteer of India) auspicious names echo in the hills of the Vindhyas and Himalyas, mingle in the music of Jamuna and Ganga and are chanted by the waves of the Indian Sea. This line corresponds to the notes 45-83 in that sequence in table 1.

Line 4: *Gahe tava jaya gatha* Translation: They pray for your blessing and sing your praise (unknowingly or knowingly they chant your names and are blessed.) Line 4 corresponds to notes 84-92 in that sequence of table 1.

Line 5: *Jana-gana-mangala-dayaka jaya he Bharata-bhagya-vidhata* Translation: The saving of all people waits in your hand, Oh You dispenser of India's destiny. Line 5 corresponds to notes 93-113 in that sequence of table 1.

Line 6: *Jaya he, jaya he, jaya he,Jaya jaya jaya, jaya he!* Translation: Victory, victory, victory to You Victory Victory Victory to You (our God of destiny, because you are ruling us, and we are your servants. So you always win) This last line corresponds to notes 114-131 in that sequence of table 1.

Next we provide the results of Single Exponential Smoothing obtained from Minitab 16.

Results and Discussion

Single Exponential Smoothing for C1 (Zero values of Yt exist; MAPE calculated only for non-zero Yt) Data C1 (C1 in Minitab represents pitch Yt), Length 131, Smoothing Constant, Alpha (α) = 0.716404

Accuracy Measures: MAPE (Mean Absolute Percent Error) = 35.3967, MAD (Mean Absolute Deviation) =1.2554, MSD (Mean Squared Deviation) = 3.6897, Table 3 gives the residuals.

Table-1
Note sequence of India's National Anthem

Instance of note realization t	Musical Note	Pitch Y_t
1.	S	0
2.	R	2
3.	G	4
4.	G	4
5.	G	4
6.	G	4
7.	G	4
8.	G	4
9.	G	4
10.	R	2
11.	G	4
12.	M	5
13.	G	4
14.	G	4
15.	G	4
16.	R	2
17.	R	2
18.	R	2
19.	*N*	-1
20.	R	2
21.	S	0

22.	S	0
23.	S	0
24.	P	7
25.	P	7
26.	P	7
27.	P	7
28.	P	7
29.	P	7
30.	P	7
31.	P	7
32.	P	7
33.	m	6
34.	D	9
35.	P	7
36.	M	5
37.	M	5
38.	M	5
39.	G	4
40.	G	4
41.	G	4
42.	R	2
43.	M	5
44.	G	4
45.	G	4
46.	G	4
47.	G	4
48.	G	4
49.	G	4
50.	R	2
51.	P	7
52.	P	7
53.	P	7
54.	M	5
55.	M	5
56.	M	5
57.	G	4
58.	G	4
59.	G	4
60.	R	2
61.	R	2
62.	R	2
63.	*N*	-1
64.	R	2
65.	S	0
66.	G	4
67.	G	4
68.	G	4
69.	G	4
70.	G	4
71.	G	4
72.	R	2
73.	G	4
74.	M	5
75.	G	4
76.	M	5
77.	P	7

78.	P	7
79.	M	5
80.	G	4
81.	R	2
82.	M	5
83.	G	4
84.	G	4
85.	G	4
86.	G	4
87.	R	2
88.	R	2
89.	R	2
90.	*N*	-1
91.	R	2
92.	S	0
93.	P	7
94.	P	7
95.	P	7
96.	P	7
97.	P	7
98.	P	7
99.	P	7
100.	P	7
101.	P	7
102.	m	6
103.	D	9
104.	P	7
105.	M	5
106.	M	5
107.	M	5
108.	G	4
109.	G	4
110.	G	4
111.	R	2
112.	M	5
113.	G	4
114.	N	11
115.	N	11
116.	**S**	12
117.	D	9
118.	D	9
119.	N	11
120.	P	7
121.	P	7
122.	D	9
123.	S	0
124.	S	0
125.	R	2
126.	R	2
127.	G	4
128.	G	4
129.	R	2
130.	G	4
131.	M	5

Table-2
Numbers representing pitches of musical notes in three octaves

C	Db	D	Eb	E	F	F#	G	Ab	A	Bb	B	Western notation
S	*r*	*R*	*g*	*G*	*M*	*m*	*P*	*d*	*D*	*n*	*N*	Notes (lower octave)
-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	Numbers for Pitch
S	r	R	g	G	M	m	P	d	D	n	N	Notes (middle octave)
0	1	2	3	4	5	6	7	8	9	10	11	Numbers for Pitch
S	**r**	**R**	**g**	**G**	**M**	**m**	**P**	**d**	**D**	**n**	**N**	Notes (higher octave)
12	13	14	15	16	17	18	19	20	21	22	23	Numbers for Pitch

Table 3
Residual (Error) Analysis

Time t	C1	Smooth	Predict	Error(C1-Predict)
1	0	0.2065	0.7280	-0.72803
2	2	1.4914	0.2065	1.79353
3	4	3.2886	1.4914	2.50864
4	4	3.7982	3.2886	0.71144
5	4	3.9428	3.7982	0.20176
6	4	3.9838	3.9428	0.05722
7	4	3.9954	3.9838	0.01623
8	4	3.9987	3.9954	0.00460
9	4	3.9996	3.9987	0.00131
10	2	2.5671	3.9996	-1.99963
11	4	3.5936	2.5671	1.43291
12	5	4.6012	3.5936	1.40637
13	4	4.1705	4.6012	-0.60116
14	4	4.0483	4.1705	-0.17049
15	4	4.0137	4.0483	-0.04835
16	2	2.5711	4.0137	-2.01371
17	2	2.1620	2.5711	-0.57108
18	2	2.0459	2.1620	-0.16196
19	-1	-0.1362	2.0459	-3.04593
20	2	1.3942	-0.1362	2.13619
21	0	0.3954	1.3942	-1.39419
22	0	0.1121	0.3954	-0.39539
23	0	0.0318	0.1121	-0.11213
24	7	5.0238	0.0318	6.96820
25	7	6.4396	5.0238	1.97615
26	7	6.8411	6.4396	0.56043
27	7	6.9549	6.8411	0.15894
28	7	6.9872	6.9549	0.04507
29	7	6.9964	6.9872	0.01278
30	7	6.9990	6.9964	0.00363
31	7	6.9997	6.9990	0.00103
32	7	6.9999	6.9997	0.00029
33	6	6.2836	6.9999	-0.99992
34	9	8.2296	6.2836	2.71643
35	7	7.3487	8.2296	-1.22963
36	5	5.6661	7.3487	-2.34872
37	5	5.1889	5.6661	-0.66609
38	5	5.0536	5.1889	-0.18890
39	4	4.2988	5.0536	-1.05357
40	4	4.0847	4.2988	-0.29879
41	4	4.0240	4.0847	-0.08474
42	2	2.5740	4.0240	-2.02403
43	5	4.3120	2.5740	2.42599
44	4	4.0885	4.3120	-0.31200
45	4	4.0251	4.0885	-0.08848
46	4	4.0071	4.0251	-0.02509
47	4	4.0020	4.0071	-0.00712
48	4	4.0006	4.0020	-0.00202
49	4	4.0002	4.0006	-0.00057
50	2	2.5672	4.0002	-2.00016
51	7	5.7429	2.5672	4.43276
52	7	6.6435	5.7429	1.25711
53	7	6.8989	6.6435	0.35651
54	5	5.5385	6.8989	-1.89889
55	5	5.1527	5.5385	-0.53852
56	5	5.0433	5.1527	-0.15272
57	4	4.2959	5.0433	-1.04331
58	4	4.0839	4.2959	-0.29588
59	4	4.0238	4.0839	-0.08391
60	2	2.5739	4.0238	-2.02380
61	2	2.1628	2.5739	-0.57394
62	2	2.0462	2.1628	-0.16277
63	-1	-0.1361	2.0462	-3.04616
64	2	1.3942	-0.1361	2.13612
65	0	0.3954	1.3942	-1.39420
66	4	2.9777	0.3954	3.60461
67	4	3.7101	2.9777	1.02225
68	4	3.9178	3.7101	0.28991
69	4	3.9767	3.9178	0.08222
70	4	3.9934	3.9767	0.02332
71	4	3.9981	3.9934	0.00661
72	2	2.5667	3.9981	-1.99812
73	4	3.5935	2.5667	1.43334
74	5	4.6011	3.5935	1.40649
75	4	4.1705	4.6011	-0.60113
76	5	4.7648	4.1705	0.82952
77	7	6.3661	4.7648	2.23525
78	7	6.8202	6.3661	0.63391
79	5	5.5162	6.8202	-1.82023
80	4	4.4300	5.5162	-1.51621
81	2	2.6891	4.4300	-2.42999
82	5	4.3446	2.6891	2.31086
83	4	4.0977	4.3446	-0.34465

84	4	4.0277	4.0977	-0.09774
85	4	4.0079	4.0277	-0.02772
86	4	4.0022	4.0079	-0.00786
87	2	2.5678	4.0022	-2.00223
88	2	2.1610	2.5678	-0.56782
89	2	2.0457	2.1610	-0.16103
90	-1	-0.1363	2.0457	-3.04567
91	2	1.3942	-0.1363	2.13626
92	0	0.3954	1.3942	-1.39417
93	7	5.1270	0.3954	6.60462
94	7	6.4688	5.1270	1.87304
95	7	6.8494	6.4688	0.53119
96	7	6.9573	6.8494	0.15064
97	7	6.9879	6.9573	0.04272
98	7	6.9966	6.9879	0.01212
99	7	6.9990	6.9966	0.00344
100	7	6.9997	6.9990	0.00097
101	7	6.9999	6.9997	0.00028
102	6	6.2836	6.9999	-0.99992
103	9	8.2296	6.2836	2.71643
104	7	7.3487	8.2296	-1.22963
105	5	5.6661	7.3487	-2.34872
106	5	5.1889	5.6661	-0.66609
107	5	5.0536	5.1889	-0.18890

108	4	4.2988	5.0536	-1.05357
109	4	4.0847	4.2988	-0.29879
110	4	4.0240	4.0847	-0.08474
111	2	2.5740	4.0240	-2.02403
112	5	4.3120	2.5740	2.42599
113	4	4.0885	4.3120	-0.31200
114	11	9.0399	4.0885	6.91152
115	11	10.4441	9.0399	1.96008
116	12	11.5588	10.4441	1.55587
117	9	9.7257	11.5588	-2.55876
118	9	9.2058	9.7257	-0.72565
119	11	10.4912	9.2058	1.79421
120	7	7.9901	10.4912	-3.49117
121	7	7.2808	7.9901	-0.99008
122	9	8.5124	7.2808	1.71922
123	0	2.4141	8.5124	-8.51244
124	0	0.6846	2.4141	-2.41409
125	2	1.6270	0.6846	1.31537
126	2	1.8942	1.6270	0.37303
127	4	3.4028	1.8942	2.10579
128	4	3.8306	3.4028	0.59719
129	2	2.5192	3.8306	-1.83064
130	4	3.5800	2.5192	1.48084
131	5	4.5973	3.5800	1.41996

Fig. 1 and Fig. 2 next give a graphical summary of the results of statistical analysis.

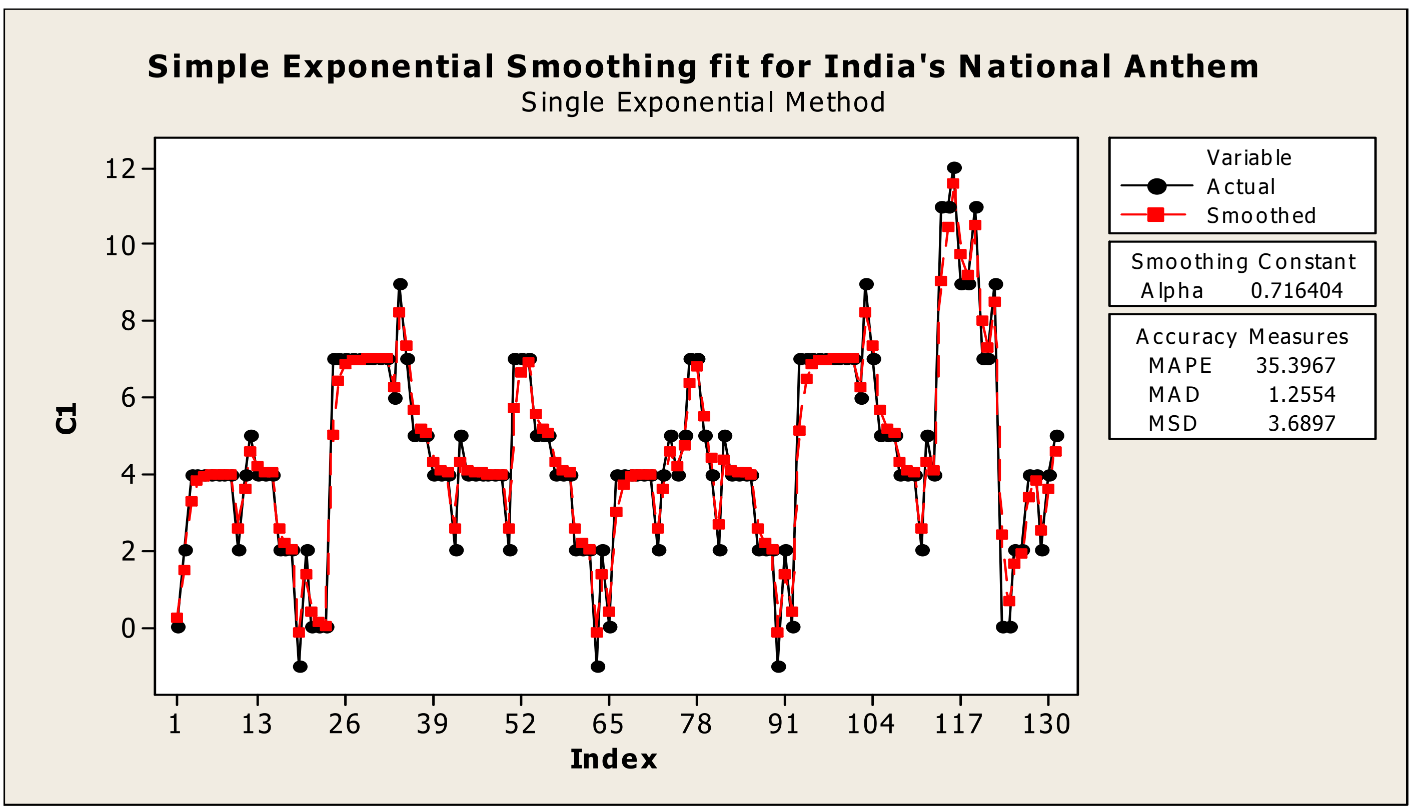

Figure-1
Single Exponential Smoothing Plot for C1

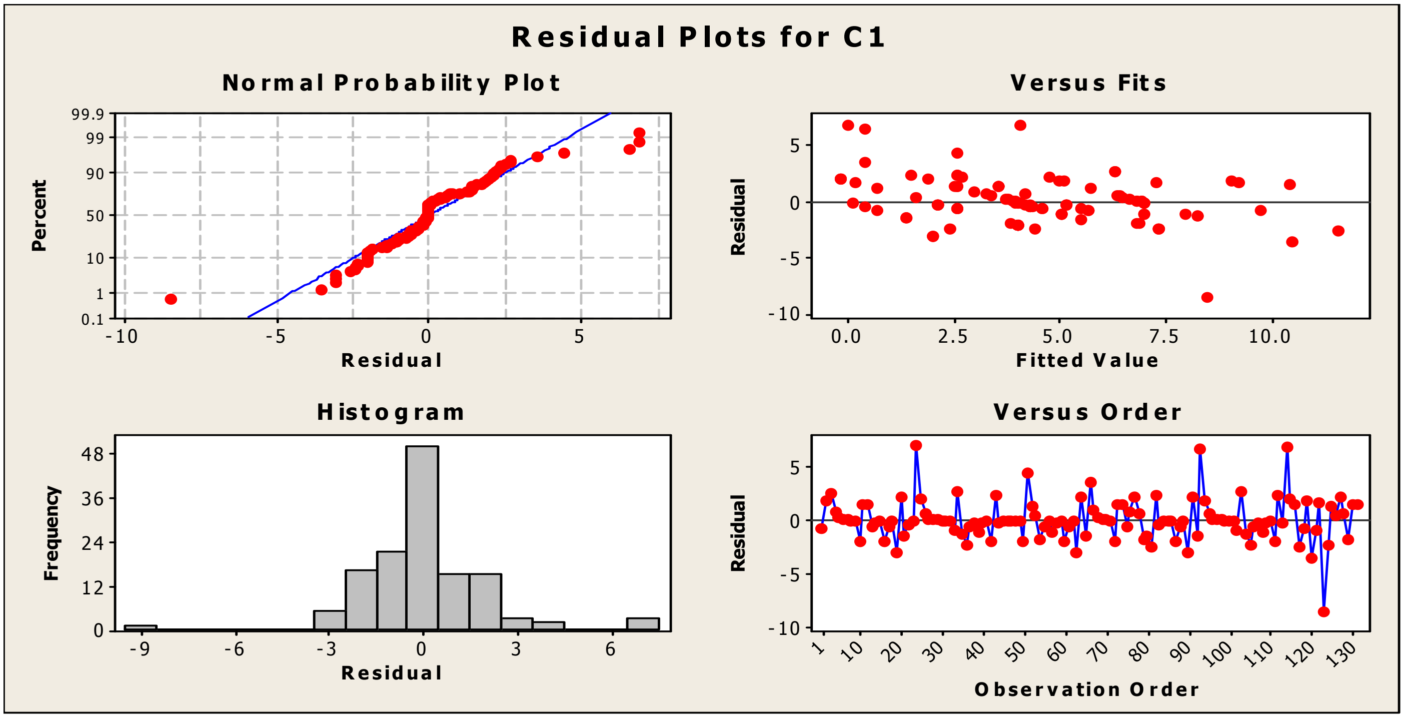

Figure-2
Residual Plots for C1

Interpretations from figures 1 and 2: The random pattern of the residuals (figure-2) together with the closeness of smoothed data with the observed one (figure-1) justifies the Simple Exponential Smoothing. A detailed discussion of the findings is given next.

MAPE (Mean Absolute Percent Error) - measures the accuracy of fitted time series values. It expresses accuracy as a percentage. MAD (Mean Absolute Deviation) - measures the accuracy of fitted time series values. It expresses accuracy in the same units as the data, which helps conceptualize the amount of error. MSD (Mean Squared Deviation) - measures the accuracy of fitted time series values. MSD is always computed using the same denominator (the number of forecasts) regardless of the model, so one can compare MSD values across models and hence compare the accuracy of two different models. For all three measures, smaller values generally indicate a better fitting model. In case we fit other models to the same data, it is of interest to compare the corresponding MAPE, MAD and MSD values. This is reserved as a rewarding future work.

The normal probability graph plots the residuals versus their expected values when the distribution is normal. The residuals from the analysis should be normally distributed. In practice, for data with a large number of observations, moderate departures from normality do not seriously affect the results. The normal probability plot of the residuals should roughly follow a straight line. One can use this plot to look for the following:

This pattern...	Indicates...
Not a straight line	Nonnormality
Curve in the tails	Skewness
A point far away from the line	An outlier
Changing slope	An unidentified variable

As is clear from figure-2, the plot roughly follows a straight line with one outlier: The next graph plots the residuals versus the fitted values. The residuals should be scattered randomly about zero. One can use this plot to look for the following:

This pattern...	Indicates...
Fanning or uneven spreading of residuals across fitted values	Non-constant variance
Curvilinear	A missing higher-order term
A point far away from zero	An outlier

As is clear from figure-2, the points are randomly scattered: A histogram of the residuals shows the distribution of the residuals for all observations. One can use the histogram as an exploratory tool to learn about the following characteristics of the data: i. Typical values, spread or variation, and shape, ii. Unusual values in the data

This pattern...	Indicates...
Long tails	Skewness
A bar far away from the other bars	An outlier

The histogram of the residuals should be bell-shaped. One can use this plot to look for the following:

Because the appearance of the histogram can change depending on the number of intervals used to group the data, use the normal probability plot and goodness-of-fit tests to assess whether the residuals are normal. We have already given the normal probability plot for residuals.

The graph residuals versus order plots the residuals in the order of the corresponding observations. The plot is useful when the order of the observations may influence the results, which can occur when data are collected in a time sequence (as in our case) or in some other sequence, such as geographic area. This plot can be particularly helpful in a designed experiment in which the runs are not randomized.

The residuals in the plot should fluctuate in a random pattern around the center line as in figure-2. One can examine the plot to see if any correlation exists between error terms that are near each other. Correlation among residuals may be signified by: i. An ascending or descending trend in the residuals, ii. Rapid changes in signs of adjacent residuals

Remark: Simple exponential smoothing is useful when i. there is no trend ii. there is no seasonal variation iii. there is no missing value iv. we want short term forecast. For the sake of completeness we show figure-3 which clearly indicates an absence of linear trend. There was no missing value in table-1. The melody phrase {G, G, G, R, R, R, *N*, R S} of length 9 (as there are nine notes in it) comes thrice which is maximum for any melody phrase in the entire sequence and this leads to a melody significance[7] of 9x3=27.

Table 4 gives the information content (IC) of the notes. This measures the surprise a note carries when it appears. Higher the probability, lower the information content and hence lower the surprise. The formula is

IC (note) = $-\log_2\{\text{Probability(note)}\}$.

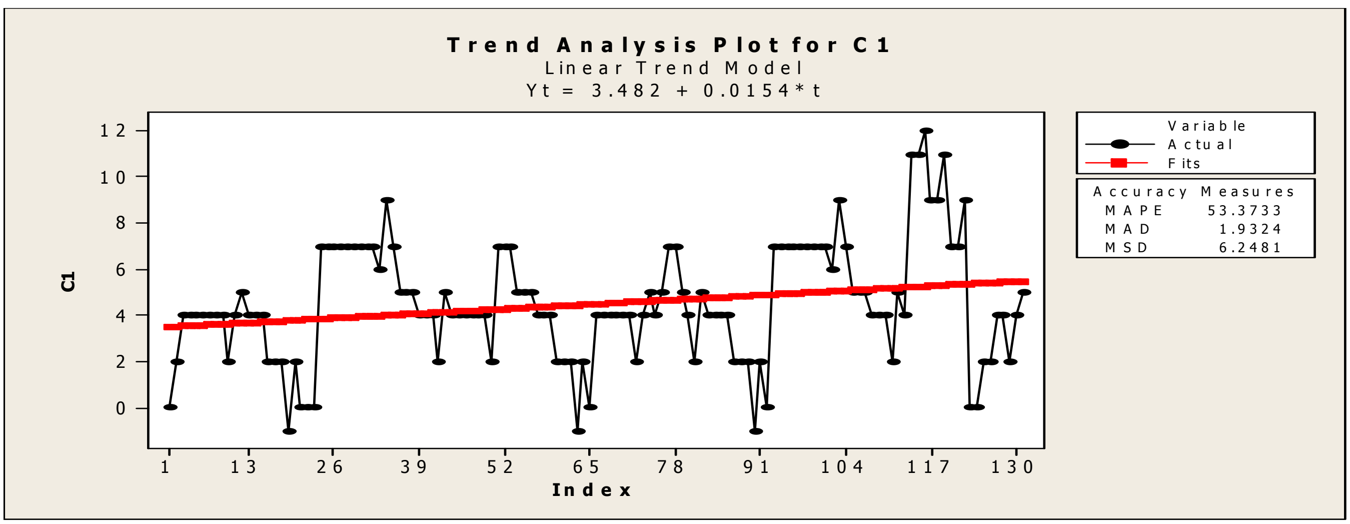

Figure-3
No linear trend in India's national anthem

Table-4
Probability of the notes and their information content

Sl. No.	Note	No. of Occurrence=f	Probability=f/131	Information Content
1	S	9	0.0687	3.8635
2	R	22	0.1679	2.5740
3	G	43	0.3282	1.6072
4	M	17	0.1298	2.9460
5	m	2	0.0153	6.0334
6	P	27	0.2061	2.2785
7	D	5	0.0382	4.7115
8	N	6	0.0458	4.4485

Conclusion

Simple Exponential Model fitted to the national anthem of India is found to be explaining the note progression well enough with smoothing factor 0.716404. Possibly this is the first attempt to model India's national anthem statistically. One advantage with simple exponential modeling is that the inner mechanism is fairly straightforward and as such can be easily understood by a general audience. Should such a model work well for the national anthem of any other country, it is of interest to see how the smoothing factor varies. Experimenting with more sophisticated smoothing techniques like Kalman filter where the smoothing factor dynamically changes is also within the scope of our research and is reserved as a rewarding future work, given that the three measures MAPE, MAD and MSD given by the Minitab package are not very informative by themselves but they are helpful in comparing the fits obtained by using different methods.

Remark: According to the first author (who, apart from being a statistician, is also a harmonium player) India's national anthem is based on the raga *Yaman Kalyan*. But unlike a raga rendition where notes are rendered with *meend* (glide), in the national anthem the notes are rendered flat, i.e., in straight transition. Therefore the raga mood does not build, neither it is necessary. In fact, because of the flat rendition of the notes, India's national anthem sounds very melodious if rendered properly by an orchestra.

Acknowledgement

This research is a part of an ongoing UGC Major Research Project in our department. We thank the University Grants Commission (UGC) for the sponsorship (Sanction Letter No.F.N.5-412/2010 (HRP); draft no.392889 dated 28.02.2011).

References

1. http://en.wikipedia.org/wiki/Jana_Gana_Mana **(2012)**
2. Holt C. C., Forecasting Trends and Seasonal by Exponentially Weighted Averages". Office of Naval Research Memorandum **52**, (1957) reprinted in Holt, C. C. (January–March 2004). Forecasting Trends and Seasonal Exponentially Weighted Averages, International Journal of Forecasting **20 (1),** 5–10 **(2004)**
3. Brown, R. G., Smoothing Forecasting and Prediction of Discrete Time Series. Englewood Cliffs, NJ: Prentice-Hall **(1963)**
4. http://en.wikipedia.org/wiki/Exponential_smoothing **(2012)**
5. Chakraborty, S., Ranganayakulu, R., Chauhan, S., Solanki, S. S. and Mahto, K. A Statistical Analysis of Raga Ahir Bhairav. Journal of Music and Meaning, Vol. 8, sec. 4, http://www.musicandmeaning.net/issues/show Article.php?artID=8.4 **(2009)**
6. Adiloglu K., Noll T. and Obermayer K., A Paradigmatic Approach to Extract the melodic Structure of a Musical Piece, *Journal of New Music Research,* **35(3),** 221-236 **(2006)**
7. Chakraborty, S., Krishnapryia, K., Loveleen, Chauhan, S., Solanki, S. S. and Mahto, K., Melody, Revisited: Tips from Indian Music Theory, *International Journal of Computational Cognition,* **8(3),** 26-32 **(2010)**