Forecasting India's GDP over 2015-2019 to assess the performance of Modi Government

Forecasting Analytics: Group A7

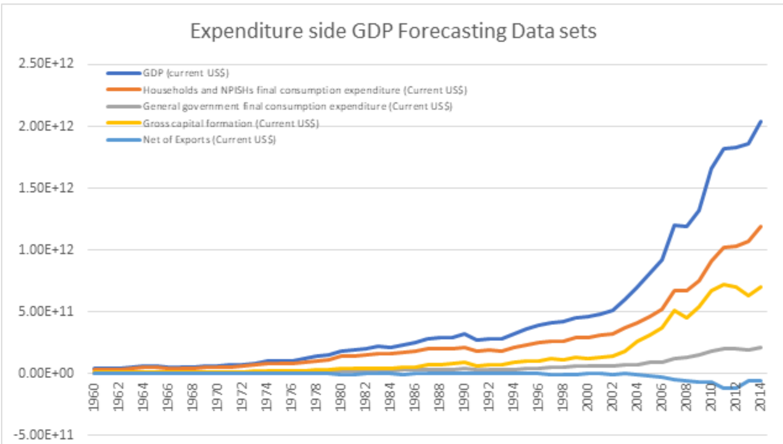
⇒ PG ID	⇒ **Student Name**
⇒ 61910640	⇒ Shashank Suryae
⇒ 61910066	⇒ Amol Kamath
⇒ 61910411	⇒ Soumya Joshi
⇒ 61910805	⇒ Mithilesh Parvataneni
⇒ 61911003	⇒ Gary Rohan Singh
⇒ 61910474	⇒ Rashi Choudhari

Executive Summary

Problem Description

The project aims to forecast India's GDP through 3 methods – employment, GDP trend and the sum of 4 expenditure components (household, government expenditure, investment and net exports) and to use the same to assess if the Modi Government has been able to change the India's GDP growth trajectory over its 5 years in power (2015-2019). The client is the Department of Expenditure, Ministry of Finance. Using the project, they can do a high-level assessment of impact of government policies and can also use it for pitching the success of the incumbent party, in case of favourable results. A favourable result for the client will show outperformance of actual GDP to the forecasted GDP, showing a change in trajectory, which could be attributed to the government policies. In terms of implication of forecasting errors, a model that under-forecasts the GDP implies that the Government was successful in making India better as compared to the usual growth pattern and paints the Government in a favorable light while a model that over-forecasts does the exact opposite.

Brief description of data



The data from Penn World Tables (original source) was only available until 2014 and the fact that the employment data had missing links due to ILO not reporting retrospective data was a problem. The current calculations have been carried out basis the data from World Bank for India from 1960 to 2017. The data has been cleansed for any erratic values, missing values, reporting error, and converted to US$ Bn. (2011 ER equivalent) to account for any irregular reporting or exchange rate effect in the entire data set.

Forecasting Methods

The project forecasts 6 series and uses the methods of Naive, Naive Seasonal, Moving Average, Exponential, Double Exponential, Holt-Winters, Linear Regression additive, Linear Regression Multiplicative and Linear Regression Quadratic. The data was partitioned to create a validation period was of 5 years from 2010 – 2014 as that is the length of the forecasting horizon. The balance observations were taken as training period. The forecast period is 2015 – 2019. The metrics to judge the same include MAPE and RMSE. Further, residual plots were used to check if trend and seasonality were captured. Across models, we see improvement from the benchmark of Naïve model. For every dataset, we compared the metrics and chose the model with lowest errors and appropriate residual plot. Using this model the forecasts were made.

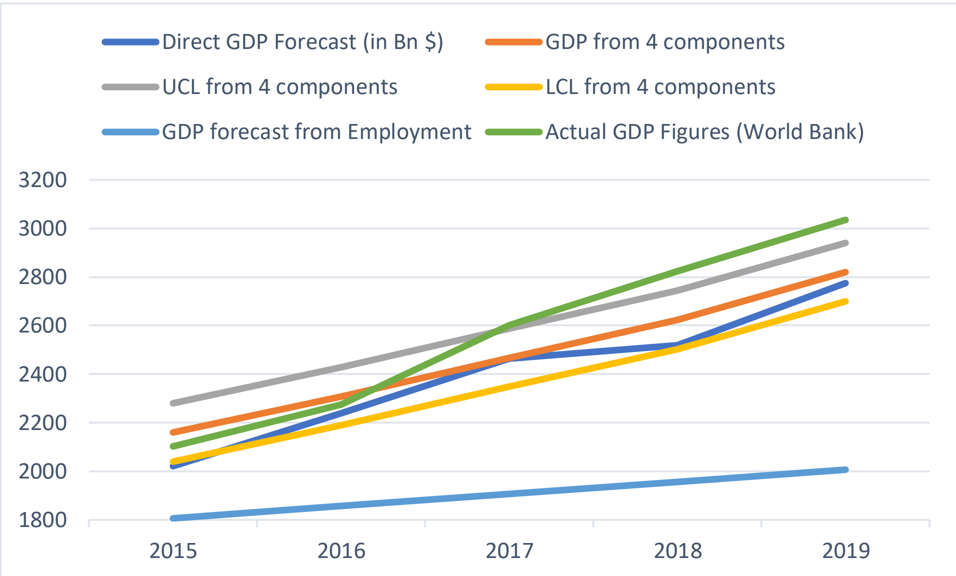
Direct GDP Forecast	Household Expenditure	Government Expenditure	Net Exports	Investments	Employment
Regression (Quadratic with seasonality)	ARIMA model with lag-1	Regression (Multiplicative Seasonality)	Exponential (optimized)	Holt-winters (additive)	Double Exponential

Year	Direct GDP Forecast (in Bn $)	Household Expenditure	Government Expenditure	Net Exports	Investments	Employment
2015	2020.8395	1285.6876	220.487	-60.8936	714.4857773	690003216.4
2016	2239.4323	1395.3661	238.619	-60.8936	735.300555	694336320.5
2017	2463.8382	1515.2553	257.124	-60.8936	756.1153326	698669424.5
2018	2518.4852	1646.3734	260.753	-60.8936	776.9301103	703002528.5
2019	2774.5451	1789.8465	293.107	-60.8936	797.7448879	707335632.6

Conclusions and Recommendations

Year	Direct GDP Forecast (in Bn $)	GDP from 4 components	UCL from 4 components	LCL from 4 components	GDP forecast from Employment	Actual GDP Figures (World Bank)
2015	2020.8395	2159.767	2279.96	2039.57	1805.946	2102.391
2016	2239.4323	2308.392	2428.58	2188.20	1856.053	2274.220
2017	2463.8382	2467.601	2587.79	2347.41	1906.16	2600.818
2018	2518.4852	2623.163	2743.35	2502.97	1956.267	2823 (P)
2019	2774.5451	2819.805	2940.00	2699.61	2006.374	3035 (P)

As per the data, the GDP has outperformed the mean prediction in all 3 cases. It is worth nothing that for the first 2 years, the actual GDP is between the UCL and LCL of the forecast, however the GDP exceeds the UCL in the last 3 years. Further, the GDP forecast from employment is significantly lower than the actual GDP. This leads to below conclusions:



- The Modi government has been able to change India's growth trajectory through its policies. Even its government expenditure has been higher than forecasted

- GDP forecasted through employment is significantly lower - indicating an increase in productivity amongst the employed workforce
- It is likely that it takes time (~2 years) for government policies to show impact in terms of metrics, explaining the trend seen in the first 2 years

The client could use the work to showcase achievements of the current govt. and the ruling party could use it for upcoming elections. This could also provide learnings for future policies.

Technical Summary

The raw data from World Bank gave the overall GDP and the % contribution of each component. Hence, the data for each component was extracted using the game. However, we faced 3 major issues with the data - Unavailability of reliable data for certain GDP components, availability of only annualized data resulting in small data sets and limiting of seasonality and discounting of external global factors and their impact on the Indian GDP

Employment

The data set is from 1991 to 2017. Its components include level, trend (upward trend) and noise. There is no seasonality and hence no seasonality models have been used. The benchmark used is the Naïve model. The below table lists the summary of models assessed:

	Naïve	Moving Average[1]	Exponential	**Double Expo**	Linear Reg	Quadratic Reg
RMSE (#)[2]	30707669.7	8021198.2	21222085	**42680979**	43431462.93	28222016.9
MAPE	4%	0.82%	3.01%	**6.30%**	6%	4%

The MA & D.Expo have decent Training & Validation RMSE & MAPE as compared to Naive. But, the MA method causes overfitting (Training RMSE is low but validation RMSE is higher) and therefore the prediction is not reliable. Therefore, going forward with the **Double Expo. model**. The forecast of employment was used to predict GDP though a linear regression.

Household Expenditure

The data covers the time span from 1961 to 2017. Like employment, there is an upward trend along with level and noise, with no seasonality. The benchmark used is Naïve model. The models have been summarized in the table. MA, Expo, D.Expo and LinReg models caused overfitting making the prediction unreliable. Hence, ARIMA model with lag-1 and t as the predictor and LN (Households consumption

[1] MA-Moving average, Exponential – Expo; Double exponential – D.Expo, Linear regression – Lin Reg and Quadratic Regression – Quad Reg
[2] Actual number of employed people

expenditure) was used as the model to forecast the values. This model has low RMSE and gives better results than other models. Using the linear regression coefficients the forecasted values on the right were calculated.

Forecasting Method	Metric	Values	
		Training	Validation
Benchmark - Naïve Model	RMSE		311.3628056
	MAPE		0.400458866
Moving Average	RMSE	39.3821475	335.1834027
	MAPE	10.69407	30.58698
Exponential	RMSE	83.00940872	514.4577257
	MAPE	21.77783	48.14565
Double Exponential	RMSE	55.09498567	323.0020823
	MAPE	17.60313	30.56914
Linear Regression	RMSE	52.35623168	378.5777641
ARIMA Model	RMSE	21.52	65.70

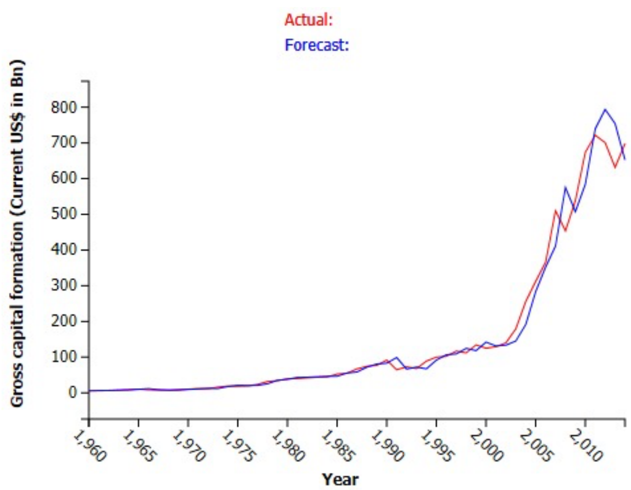
Net Exports

Net exports data has a level, a downward trend and noise, but no seasonality. All models have been assumed with no roll-forward prediction, as all predictions must be made before the start of the government tenure.[3]

	Naïve	Moving Average	Expo (0.2 alpha)	**Exponential (optimized)**	Double Expo	Linear Reg	Quadratic Reg
RMSE ($bn)	31.74	43.66	56.54	**31.75**	47.68	69.29	50.79

Expo. model with an optimized alpha has been used for forecasts. While both Naïve and chosen model have similar RMSE for validation data –the latter performs significantly better on the training period. Further, MAPE has not been used as the data has a large number of negative values. As we do not have actual data for the next 4 years, we will use the same forecast for the next 5 years which is USD Bn -60.88 (indicates imports > exports). The Forecast is given as = 0.99762*(Y_{2014}) + (1-0.99762)*(F_{2014}). A strong focus is given on last years value as the major driver of change in values is oil prices.

Investment or Gross Capital Formation



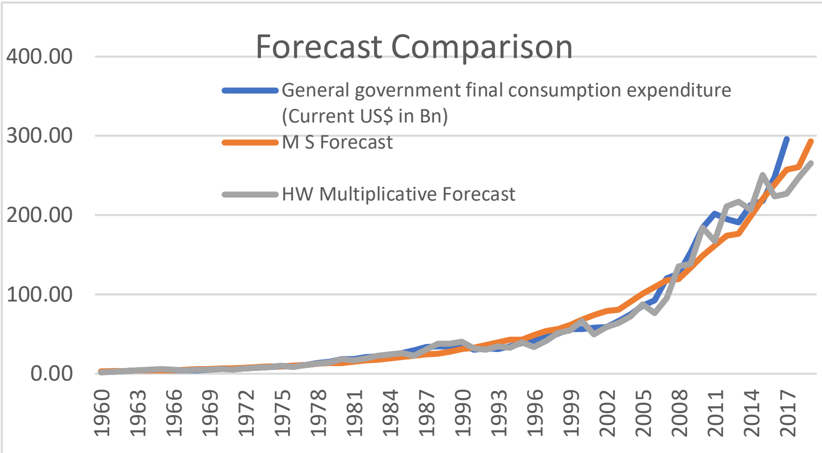
The time series components that appear in the pre-event plot are level, trend, noise and very minute seasonality. A smooth trend can be observed for the data set until 1991. Post liberalization and globalization, the trend seems to be exponential. On the seasonality front, there is no significant seasonality visible. It is understandable from the fact that the data is yearly and there is very low effect that creeps in due to the change of government. Different forecasting methods were applied and compared on the basis of MAPE (Appendix). Of these, the Holt Winters Additive Model with an optimize alpha of 0.75, beta of 0.39 and gamma of 0.56, was selected due to the lowest MAPE value. On the basis of this, a forecast

[3] Rolling forward Naïve performs worse than normal Naïve in this case

has been made for the future years. The model had a MAPE of 10.44% instead of the Naïve MAPE od 21.30%.

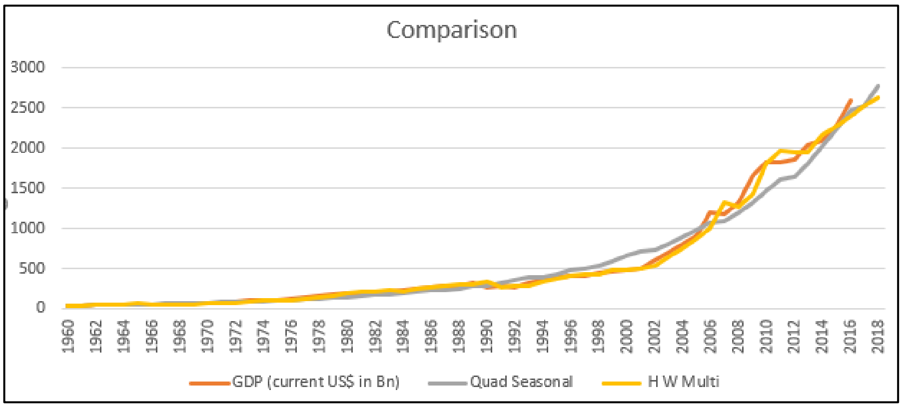
Government Expenditure

The time series components that appear in the pre-event plot include level, trend and noise, with seasonality as per election year[4]. The trend is upward exponential trend. The plots shows global seasonality and pattern within the years. The expenditure behaviour of the party's changes in the 4th year of their power. and the parties that are in power for 1 or 2 years, they tend to have correlated spending in the consecutive years of being in power. We can consider using a regression to account for the changes. Of the methods, Holt's Winters Multiplicative and Lin. Multiplicative regression gives us the lowest MAPE. We can observe that the Linear regression with seasonal dummies (M S Forecast) provides a better forecast without overfitting the original plot. As we can see by the blue line, the government has increased its expenditure to improve well-being.



GDP

At the overall level, GDP has a level, upward trend which becomes exponential in the latter stages, some seasonality and noise. Holt's Winters Multiplicative gives us the lowest MAPE (appendix). On comparison of the different trends for linear regression techniques and their outcomes, we get the best forecast with the lowest residuals for the linear regression with quadratic trend and multiplicative seasonality with seasonal dummies for the election years.



[4] See Appendix

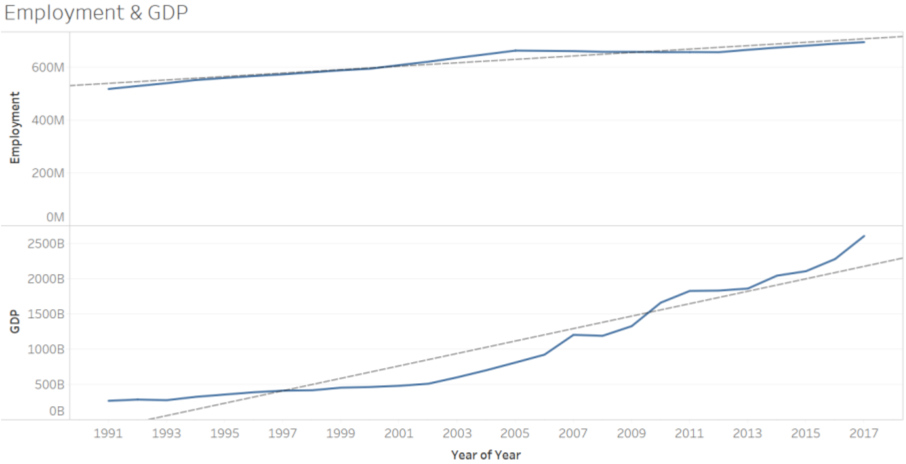
Appendix

Data Description

The World Bank - https://data.worldbank.org/country/india was used as the data source for this forecasting exercise. All Data components were Annualized and included the below:

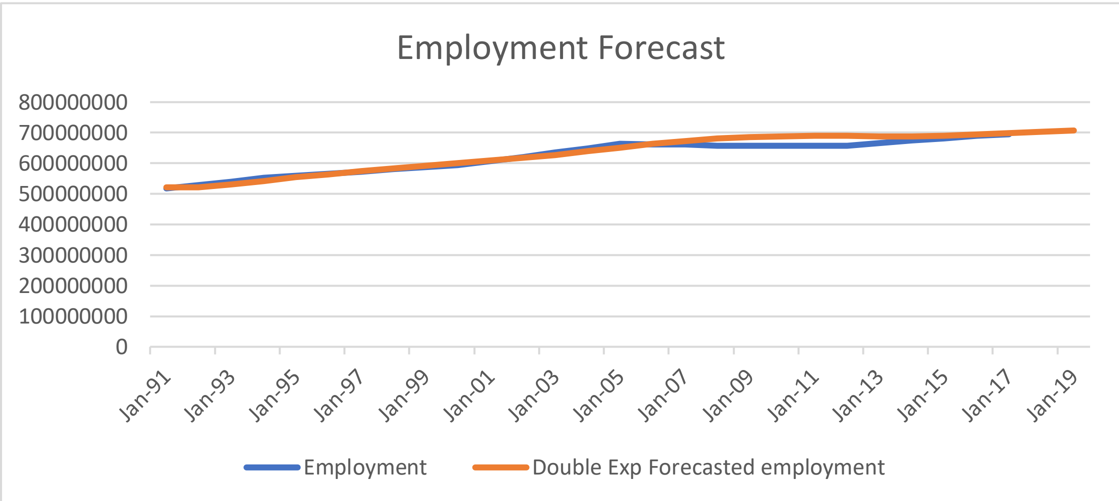
- GDP (current US$) – 1960 - 2017
- Households and NPISHs final consumption expenditure (Current US$) – 1960 - 2017
- General government final consumption expenditure (Current US$) – 1960 - 2017
- Gross capital formation (Current US$) – 1960 - 2017
- Net of Exports (Current US$) – 1960 – 2017
- Employment to population ratio – 1991 – 2017

Employment



The trends of sum of Employment and sum of GDP for Year Year.

Employment forecast



The validation metrics for the model are given below

Metric	SSE	RMSE	MAPE	MAD	CFE	MFE	TSE
Value	5.72E+15	15434258.9	1.83248242	11659005	-148367258	-6181969.1	-12.73

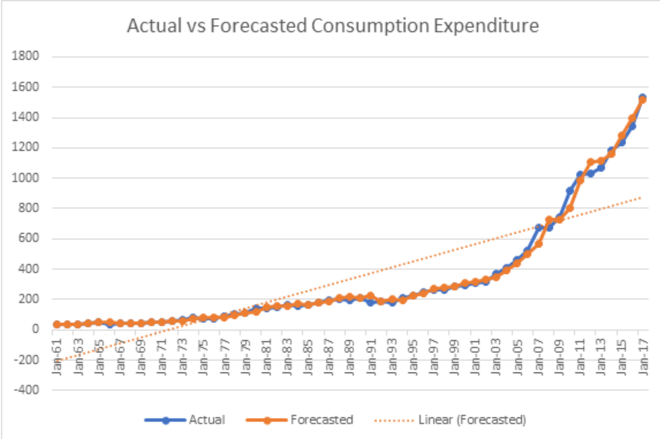
Consumption Expenditure

Coefficients for the model

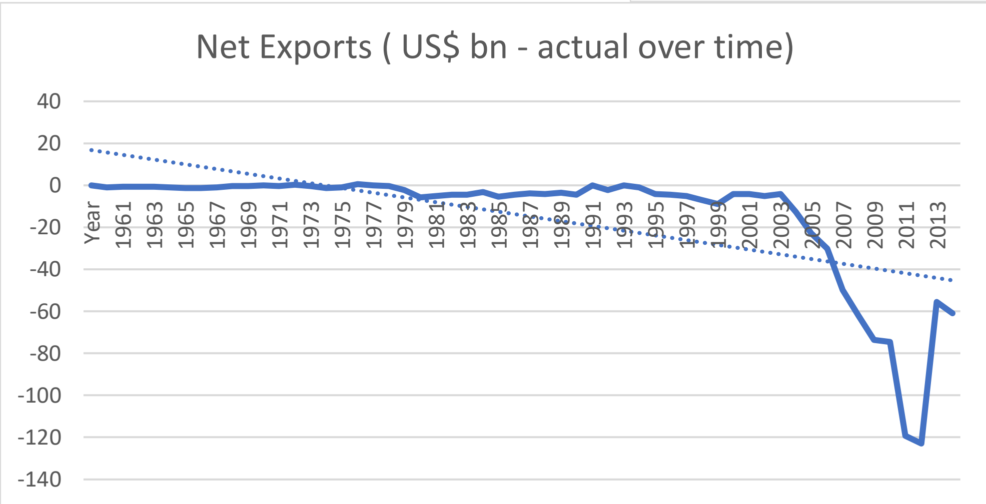
Predictor	Estimate	Confidence Interval: Lower	Confidence Interval: Upper	Standard Error	T-Statistic	P-Value
Intercept	0.04972	-0.002569082	0.102009992	0.025992188	1.912899951	0.062
t	0.00056	-0.001202792	0.002330607	0.000878195	0.64212138	0.52

Standard Errors for the ARIMA model used for forecasting consumption expenditure

	Training	Validation
SSE	0.35522661	0.02427704
MSE	0.00724952	0.00485541
RMSE	0.08514413	0.06968075
MAD	0.06040636	0.05938521
R2	0.00869647	-0.04973334



Net exports



Investment or Gross Capital Formation

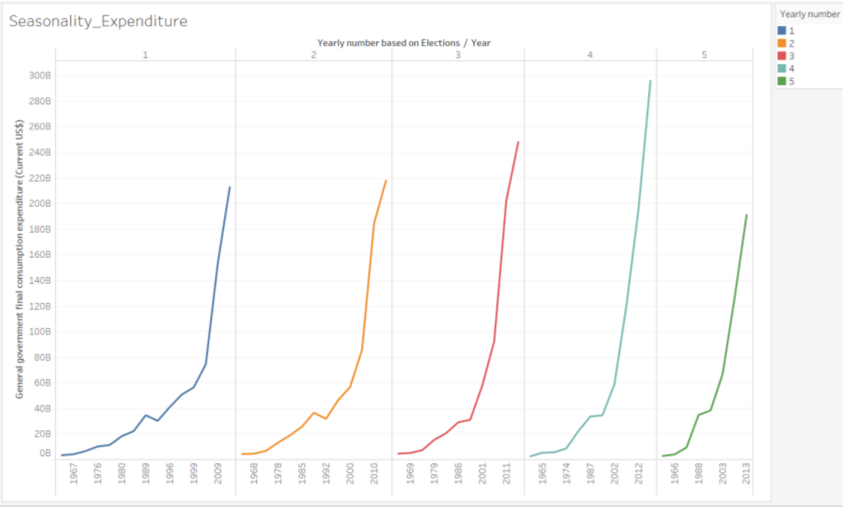
Capital formation refers to additions of capital stock, such as equipment, tools, transportation assets and electricity, used for production of goods and services.

Technique/Model	Notes	MAPE (%)
Naïve Forecast		21.30
Seasonal Naïve Forecast		36.26
Moving average	W=5	31.92
Exponential Smoothing	α=0.2	47.47
Exponential (Optimized) Smoothing	α=0.9976	21.33
Double Exponential Smoothing	α=0.2, ß=0.15	23.85
Double Exponential (Optimized) Smoothing	α=0.6762, ß=0.4441	11.16
Holt Winters Multiplicative Smoothing	α=0.2, ß=0.15, Ɣ=0.05	19.70
Holt Winters Multiplicative (Optimized) Smoothing	α=0.7966, ß=0.5711, Ɣ=0.8686	10.47

Holt Winters Additive Smoothing	α=0.2, ß=0.15, Ɣ=0.05	23.10
Holt Winters Additive (Optimized) Smoothing	α=0.7453, ß=0.3874, Ɣ=0.5566	10.44
Holt Winters No Trend Smoothing	α=0.2, Ɣ=0.05	45.99
Holt Winters No Trend (Optimized) Smoothing	α=0.9347, Ɣ=0.9892	21.39
Linear Regr. of Ln(Investment) w/o Seasonality		18.65
Linear Regr. of Ln(Investment) w/ Seasonality		18.44

Government Expenditure

Seasonality plots

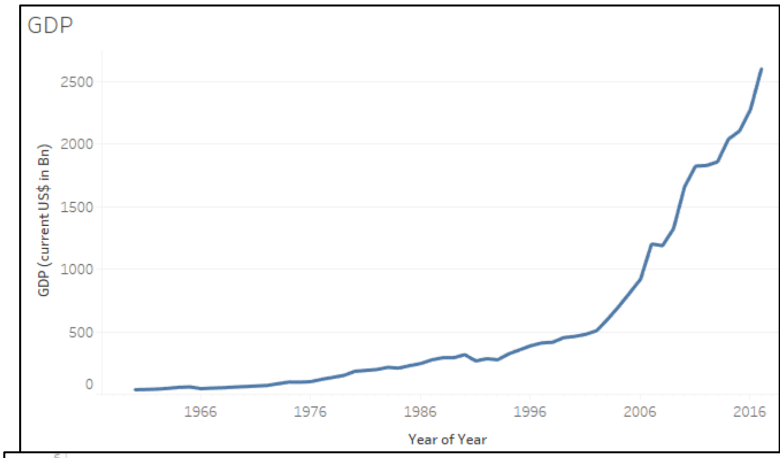


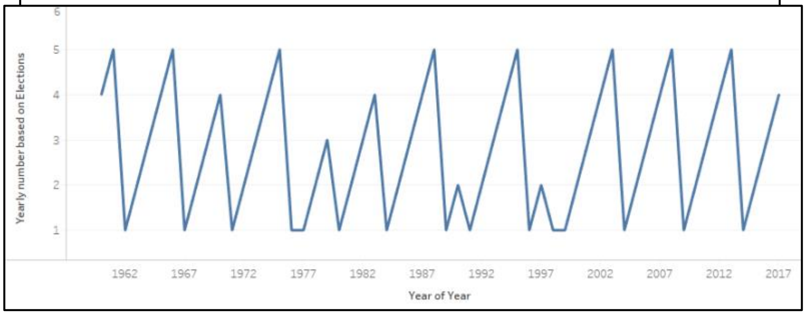
Model Summary

Technique/model	Notes	MAPE in %
Naïve Forecast		21.99
Seasonal Naïve Forecast		41.61
Moving average	W=5, W=2	36.72, 26.94
Exponential	α=0.2	48.31
Exponential	α=0.33	39.05
Exponential	optimized (α=0.9976)	22.02
Double Exponential (Holt's)	α=0.2, ß=0.15	29.78
Double Exponential (Holt's)	α=0.33, ß=0.15	20.59
Double Exponential (Holt's)	α=0.9976, ß=0.16	6.78
Double Exponential (Holt's)	Optimized (α=0.62, ß=0.99)	12.54
Holt's Winter Multiplicative	α=0.2, ß=0.15, Ɣ=0.05	28.1
Holt's Winter Multiplicative	α=0.33, ß=0.15, Ɣ=0.05	18.9
Holt's Winter Multiplicative	α=0.9976, ß=0.15, Ɣ=0.05	6.47
Holt's Winter Multiplicative	Optimized (α=0.5, ß=0.87, Ɣ=0.5)	12.03
Holt's Winter Additive	α=0.2, ß=0.15, Ɣ=0.05	29.2729
Linear Additive regression		91.87
Linear Additive regression (seasonal)		87.57
Linear Multiplicative regression	Seems to be overfitting	1.02
Linear Multiplicative regression (seasonal)	Best suited based on the visual analysis	1.09

GDP

The first exhibit shows the GDP trend over time. The second exhibit shows the mapping of years to elections – which has been used for calculating seasonality. Note that the yearly number is to be read as 1=Year of election, and following numbers as the years in which the elected party was in power. There have been instances where the elected party is in power less than the normal tenure of 5 years (shown in subplot). The table on the right shows the performance of the various models.





Technique/model	Notes	MAPE in %
Naïve Forecast		28
Seasonal Naïve Forecast		41
Moving average	W=5	36.88
Moving average	W=2	30.44
Exponential	α=0.2	48.94
Exponential	Optimized	27.77
Double Exponential (Holt's)	α=0.2, ß=0.15	29.896
Double Exponential (Holt's)	α=0.33, ß=0.15	21.087
Double Exponential (Holt's)	Optimized	11.56
Holt's Winter Multiplicative	α=0.2, ß=0.15, Ɣ=0.05	28.72
Holt's Winter Multiplicative	α=0.33, ß=0.15, Ɣ=0.05	20.5
Holt's Winter Multiplicative	Optimized (α=0.5, ß=0.87, Ɣ=0.5)	9.5
Holt's Winter Additive	α=0.2, ß=0.15, Ɣ=0.05	28.53

UCL and LCL for prediction

	S.E	Variance
Household Expenditure	29.14	849.14
Government Expenditure	0.17	0.03
Net Exports	19.88	395.26
Investments	50.16	2516.03
Total Variance		3760.45
SE of GDP forecast		61.32